

Teachers' Perspectives on the Standards of Learning School Reform in Virginia

Melanie A. Bolt

Submitted to the Faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

in

Educational Research and Evaluation

April 2003

Blacksburg, Virginia

Committee chair: Kusum Singh

Committee members: Lawrence Cross, Jim C. Fortune, Jim Garrison, and Sue Magliaro

Keywords: high-stakes accountability, testing, standards,
classroom effects, curricula, instructional strategies, student learning,
teacher professional autonomy, teacher tension, school quality

Copyright 2003, Melanie A. Bolt

Abstract

Teachers' Perspectives on the SOL School Reform in Virginia

Melanie A. Bolt

School of Human Sciences and Education

Virginia Polytechnic Institute and State University

Dissertation Committee Chairperson: Kusum Singh

This study discussed the need for a broader public discourse on high-stakes accountability-based school reform that underscores teachers' perspectives. Also, the study discussed the need for fuller disclosure of the possible undesirable classroom effects of the reform. To address these needs, this study described teachers' perspectives on the Standards of Learning (SOL) school reform in Virginia, focusing upon teachers' views on the reform's classroom effects. The domains of interest were (1) the adequacy of curriculum and the diversity of teachers' instructional strategies, (2) the quality of student learning, (3) teachers' sense of professional autonomy and level of teacher tension, and (4) school quality. The study examined whether there are differences in teachers' views based on the income level of the school locale where the teachers teach (low-, middle-, or high-income), school type (elementary, middle, or high school), and teachers' status on whether they teach a SOL-tested subject (yes/no).

The participants of the study included 360 randomly selected teachers who were listed as members of Virginia Education Association (VEA). A survey research design was employed. The instrument included 80 Likert-type items, eight demographic items, and three open-ended questions. Inferential and descriptive statistics were reported for eight scales of the survey, as were thematic trends in the qualitative data. The study's results suggested that the SOL program contributes to a hurried, high-pressure classroom culture that depletes the potentiality of the very ends of education the program is intended to achieve.

Teachers tended to report (1) an inadequacy of the SOL content standards, (2) a reduction of teachers' use of diverse instructional strategies and an inability of the SOL program to meet diverse student needs, (3) arbitrary SOL test cut-scores, (4) an inadequacy of the SOL pass rates to represent school quality, (5) a lack of diagnostic usefulness of SOL test scores, (6) an inadequacy of SOL testing and SOL test scores to hold schools accountable, (7) teachers' sense of diminished professional autonomy, and (8) teachers' mounting tension in the classroom. These results were juxtaposed to the views of policymakers and business leaders, the public at large, parents, and scholars in the field of education concerning the issue of high-stakes accountability-based school reform. Finally, the study underscored a conflict related to the purpose of public schooling between a prevailing narrative of many policymakers and business leaders and what have been the marginalized views of classroom teachers.

Acknowledgments

As a doctoral student of educational research and evaluation at Virginia Tech, my role as an educational researcher has been extended and transformed by the expertise and generosity of others. First I would like to thank the professors at Virginia Tech whose excellent variety of coursework and expertise have been brought to bear upon the present study. I would like to thank Dr. Susan Magliaro for her interactive teaching style as well as her comprehensive and instructive feedback, particularly at the time this research was proposed. Also, I would like to thank Dr. Jim Garrison for his contemplative courses, particularly his course including Deweyan pragmatism that was recommended by a fine Deweyan scholar, Dr. Douglas Simpson, prior to my pursuing the doctorate. Dr. Garrison has been instrumental in enriching my thinking about public schooling and in suggesting a dialogic framework for this study's conclusions.

I would like to thank Dr. Jimmie Fortune for his collaborative approach to teaching evaluation and qualitative research methods. Having taken his courses, I have gained practical experience in conducting research with classmates, particularly Launcelot Brown, a colleague I have greatly esteemed. Further I am indebted to Dr. Fortune as he has provided me the opportunity to help teach his graduate level qualitative research methods course, which has afforded me invaluable teaching experience.

As my committee chairperson, I am grateful for the approachable teaching style of Dr. Kusum Singh, including her practical applications in teaching quantitative research. Also I am thankful for the opportunity Dr. Singh provided me to help teach a graduate level course in behavioral research methods alongside my respected colleague, Cecile Cachaper. This

opportunity has marked a great learning experience during my time in the educational research program. Finally I am appreciative to Dr. Singh for her conceptual and technical insight that has assisted me throughout this study.

I would like to thank Dr. Lawrence Cross for his witty and candid approach to teaching what was my first research course as a student at Virginia Tech in the educational research program. Additionally I am indebted to Dr. Cross as he has generously given of his time to discuss the focal points of this study and to offer copies of relevant journal articles and newspaper editorials. More importantly, the development of the study's survey instrument would not have been possible without his expert critiques and recommendations for improvement.

I am most appreciative to the Virginia Education Association (VEA) and Dr. Ralph Shotwell, Director of Research, for providing the sampling frame for the present study. I deeply appreciate the assistance of VEA as well as the numerous classroom teachers who participated in this research.

Finally I am indebted to my family who has helped care for my three-year-old daughter Erica Grace so that I could have generous blocks of time to conduct this study. Erica Grace reaped a wealth of affection from my parents Walt and Anne Bates, mother-in-law Ann Bolt, grandmother Emma Lee Weiss, and my sister-in-law Tammy Martin.

I would also like to note sincere appreciation for my husband Karl who has been a consistent encouragement throughout this research, and for my dear father-in-law, the late Ben D. Bolt, whose vision for others and whose delight in others' accomplishments helped me imagine possibilities for my life that otherwise I would have never considered.

Indeed it was the support of my professors, colleagues, and family who helped create the conditions for my intellectual growth and ultimately helped make this study possible. I am indebted to these co-partners whose expertise and generosity I will remember with deep gratitude.

Foreword

During the mid 1990s I was a teacher of sixth grade children of working class families and of rich ethnic diversity near the Dallas/Fort Worth Airport. While aiming to create conditions for students' learning that might engage their interests and meet their diverse learning needs that were related to the subjects of English and social studies, I was expected to meet the requirements of Texas' new public school reform effort. The Texas system was based upon high-stakes accountability for students and schools in an effort to improve academic achievement and equity. School reform programs such as this are common in most states today.

Although I enjoyed the support of a talented school administration, I became frustrated about the constraints that I perceived the State reform placed upon my pedagogical practice. There seemed to be a lack of time for teachers to communicate with one another concerning subject matter as well as a lack of time for children to deeply study important topics of their interest. In addition to these, expanding curriculum requirements contributed to my frustration. I did not desire our classroom activities to be utterly driven by the specified curriculum. Nor did I desire my pedagogical freedom to be impinged upon. The tension that I sensed led me to more thoughtfully consider my role and responsibilities as a classroom teacher.

While teaching in Texas I was a master's student at Texas Christian University studying educational research. The program emphasized a qualitative approach to research that was immersed in theory. Among the philosophical sworks that the program emphasized, the writings of John Dewey were particularly useful in helping me wrap words around my

efforts to teach and meet the requirements of the Texas high-stakes accountability program. Accordingly, my master's thesis resulted in a philosophical work that explored a Deweyan perspective of teachers' roles and responsibilities against the backdrop of my teaching experiences.

Out of this work grew a deeper interest in educational research and in how high-stakes accountability-based school reform affects the classroom as viewed by teachers. Deepening interest in these areas led me to pursue the doctorate in educational research and evaluation at Virginia Tech, where the program tended to focus upon quantitative research and statistics. This training provided me with another set of approaches for studying school improvement and high-stakes accountability systems as they affect the classroom. The questions and design of this study then became clear.

Given this background, I have attempted to be intellectually honest and have acknowledged that this study is an extension of my interest, an outgrowth of classroom teaching and almost ten years of reflection on high-stakes accountability. As such, I have been deeply concerned with the work of teachers and the quality of student learning. Also the participation of teachers in the public discourse on high-stakes accountability has been of vital interest to me, as classroom teachers are located at the intersection of state level policy, curriculum, pedagogy, and student learning.

The classroom is a complex social space that is transformative to all who enter in. To what end high-stakes accountability programs transform teaching and learning have been vociferously debated by policymakers and businesspersons, scholars, parents, and the public at large. However, classroom teachers' perspectives on this point have been too sparse to have a consequential bearing on the public debate. The importance of this research relates to

this point. This study is important because it enlarges the public discourse on high-stakes accountability by describing teachers' views, that is, views from the classroom that are tantamount to a more informed public discourse on the desirability of the contemporary reform effort.

Melanie A. Bolt

Teachers' Perspectives on the Standards of Learning School Reform in Virginia

Table of Contents

	Page
Chapter One.....	1
Introduction.....	3
Problem of Inquiry.....	10
Purpose of Inquiry.....	12
Questions for Research.....	12
Organization of the Study.....	14
Chapter Two.....	15
Review of Literature.....	15
Method of Review.....	15
Historical Context of Contemporary Accountability in the United States.....	16
Contemporary Accountability in Virginia.....	22
Teachers' Perspectives on Contemporary Accountability.....	27
Chapter Three.....	40
Method of Research.....	40
Research Design.....	40
Survey Development.....	41
Procedures.....	44
Validity and Reliability.....	46
Data Analyses.....	51
Chapter Four.....	54
Results.....	54
Participants.....	54
Item-Level Analysis.....	57
Scale-Level Analysis.....	66
Adequacy of the Standards of Learning.....	68
Ability of the SOL Program to Meet Diverse Student Needs.....	77
Appropriateness of SOL Cut-Scores to Signify Student Subject Mastery.....	89

Adequacy of the SOL Pass Rates to Signify School Quality.....	94
Diagnostic Value of SOL Test Scores.....	100
Adequacy of SOL Testing and SOL Test Scores to Hold Schools Accountable.....	108
Teacher Professional Autonomy.....	125
Level of Teacher Tension Within the SOL Reform.....	137
Chapter Five.....	158
Teachers' Perspectives on the SOL School Reform.....	158
Summary of Findings.....	158
Conclusions: Juxtapositions of Perspectives.....	163
Theoretical and Practical Considerations.....	181
Limitations of the Study and Suggestions for Future Research.....	186
Bibliography.....	189
Appendix A.....	204
Appendix B.....	205
Curriculum Vitae.....	219

Teachers' Perspectives on the Standards of Learning School Reform in Virginia

List of Tables and Figures

	Page
Table 1: Frequency of School Type by Teachers' Status on Teaching a SOL-Tested Subject.....	56
Figure 1: Response Frequency for Usefulness of SOL Test Scores for Pinpointing Students' Learning Difficulties by Income Level of School Locality.....	62
Figure 2: Response Frequency for Usefulness of SOL Test Scores for Determining Students' Course Placement by Teachers' Status on Teaching a SOL-Tested Subject.....	64
Figure 3: Response Frequency for Usefulness of SOL Test Scores for Determining Students' Course Placement by School Type.....	65
Table 2: Item Frequencies and Descriptive Statistics for Adequacy of the Standards of Learning.....	69
Figure 4: Teachers' Response Frequency for ADSOL.....	71
Table 3: Item Frequencies and Descriptive Statistics for Ability of the SOL Program to Meet Diverse Student Needs.....	79
Figure 5: Teachers' Response Frequency for DIVERSE.....	81
Table 4: Item Frequencies and Descriptive Statistics for the Appropriateness of SOL Cut-Scores to Signify Student Subject Mastery.....	90
Figure 6: Teachers' Response Frequency for CUTSCR.....	92
Table 5: Item Frequencies and Descriptive Statistics For Adequacy of SOL Pass Rates to Signify School Quality.....	96
Figure 7: Teachers' Response Frequency for PASSQ.....	98
Table 6: Item Frequencies and Descriptive Statistics for Diagnostic Value of SOL Test Scores.....	101
Figure 8: Teachers' Response Frequency for DIAGV.....	103
Figure 9: Comparison of ADTEST Means for School Type.....	109
Table 7: Item Frequencies and Descriptive Statistics for Adequacy of SOL Testing and SOL Test Scores to Hold Schools Accountable.....	111

Figure 10: Teachers' Response Frequency for ADTEST.....	113
Table 8: Descriptive Statistics for Support of Teacher Professional Autonomy Within the SOL Reform.....	127
Figure 11: Teachers' Response Frequency for TPROF.....	129
Table 9: Item Frequencies and Descriptive Statistics for Level of Teacher Tension Within the SOL Reform.....	138
Figure 12: Teachers' Response Frequency for TTEN.....	140

CHAPTER ONE

Over the past decade in the United States the landscape of public schooling has been influenced by a heightened political and public interest in results-oriented social policy of education, which has prompted most states to initiate high-stakes accountability programs. The notion of accountability has been supported by “the theory that measuring performance and coupling it to rewards and sanctions will cause schools and the individuals who work in them to perform at higher levels” (Elmore and Fuhrman, 2001, p. 67). The current accountability programs have been designed to improve academic achievement and equity for all students, with particular attention to children of color and to low-income children (McQuillan, 2001; Paris, Lawton, Turner, & Roth, 1991). There are several central features of the contemporary accountability programs that have been elucidated by Robert L. Linn (2000, p. 8) as follows:

(a) the emphasis on the development and use of ambitious content standards as the basis of assessment and accountability, (b) the dual emphasis on setting demanding performance standards and on the inclusion of all students, and (c) the attachment of high-stakes accountability mechanisms for schools, teachers, and, sometimes, students (Linn, 2000, p. 8).

The public school system in Virginia provides an example of the implementation of this contemporary effort. In 1995 Virginia Governor George Allen commenced the Standards of Learning (SOL) program that has been continued under the Gilmore administration and

the present Warner administration (Turner, 2001). The SOL program is a sophisticated statewide initiative comprised of various tools that are attached to the metaphorical tool belt of accountability. The tools of the program include (1) a set of high, uniform standards for all students; (2) the alignment of curricula to the state's academic standards, the Standards of Learning; (3) the testing of students at specified grade levels using tests that correspond to the Standards; and (4) the assessment of school quality based upon students' test scores (Virginia Department of Education, [http://www .pen.k12.va.us](http://www.pen.k12.va.us)). These tools can be described as comprising a technology of accountability that presents new possibilities as well as potentially producing "profound effects beyond the designed intent" (Madaus, 1994, p. 78). Madaus elaborates this point:

Technology gives people the power and will to create new possibilities, allowing for increased choices, opportunities, and freedom. However, although the benefits of technology are enormous, technology simultaneously creates problems, opens ways to make big mistakes, alters institutions in unanticipated negative ways, and impacts negatively on certain populations (1994, pp. 77-78).

Designed Intent of the SOL Program

The Virginia SOL accountability program is intended to produce the following outcomes: (1) a better specification of curriculum, (2) a clearer focus for teachers on subject matter, (3) access to test-score data that may help teachers diagnose students' learning difficulties, and (4) the presence of rewards and sanctions to help motivate students and

schools toward improved academic achievement (<http://www.pen.k12.va.us>). As such, the SOL technology is intended to help provide an equitable education and academic excellence for all students. As such, a member of the Virginia Board of Education, Mark Christie, describes the SOL program as one designed so “that all our schoolchildren, not just a lucky few, will be prepared to compete successfully in the global economy of the 21st century, whether they go to college or enter the workforce right out of high school, and will be informed, responsible citizens of a democracy” (Christie, 1999, p. 32).

Uses of Accountability Programs

The intent of current accountability systems may be better understood in terms of the specific uses of the tools comprising the systems. According to Heubert and Hauser (1999), generally the state-initiated programs involve the use of students’ test scores to make *high-stakes* decisions affecting students’ course placement, grade promotion, and high school graduation. For example in Virginia, beginning the 2004-2005 school year the State will link high school diplomas to whether students pass the SOL examinations. Further, beginning in 2007, SOL test results will determine school accreditation and students’ test scores will be used to represent school quality. Hence, the schools are to be evaluated based upon children’s scores on state-mandated tests, and by implication (or perhaps directly) teachers are to be judged based upon their students’ test performance. Provided these high-stakes decisions, state initiatives such as Virginia’s seem to possess “social and political potency” (Madaus, 1994, p. 78) to alter student learning, teacher morale, and teachers’ instructional strategies. The power of high-stakes to alter these features of public schooling has drawn much attention to the contemporary reform.

High-Stakes Accountability and the Discourse of Controversy

The technology of contemporary accountability has generated a widespread discourse of controversy. The public discourse has centered upon the technical tools of the SOL program and the possible undesirable classroom effects of the current reform. The tools of the program include the adequacy of content standards; the psychometric properties of the test instruments that relate to the content standards; and the appropriateness of the performance standards, the significance of student and school test performance, and consequences related to test performance. The possible unintended undesirable classroom effects include the outcomes of the SOL program that affect the adequacy of the classroom curriculum and teachers' instructional strategies, student learning, teachers' professional autonomy and level of teacher tension, as well as school quality. Various parent groups, teacher associations, policymakers and businesspersons, scholars, and the public at large have engaged in the accountability debate.

Parent concerns. In Virginia, grass-roots organizations such as Parents Across Virginia United to Reform SOLs (PAVURSOL) question whether the Virginia Standards are adequately specific, if their array and content are good, and whether the Standards indicate what students should know and be able to do (Wermers, 2001, <http://www.timesdispatch.com/editorials/index.shtml>). Further, PAVURSOL argues against the use of test score data to make important student decisions such as whether students graduate from high school.

Test specialists on measurement issues. Other concerns have hinged on the adequacy of the tests' psychometric properties, particularly in terms of test length and the ability of the tests to fairly measure the achievement of children of color as well as those living in lower-income communities (Sluss, 2001, B1). Measurement expert Lawrence H. Cross (2000)

states that the SOL tests are simply too brief to adequately assess the knowledge and skills that are represented by the broad content of the Standards. Due to the brevity of the tests Cross asserts, the tests' validity is inadequate and the important uses of the test scores are indefensible. Contrary to Cross' assertion, Mark Christie states that "what is more important about the SOL tests is not...whether they have more items or fewer, but the fact that *they count*, they have real, not sham, consequences (1999, p. 36). Cross contends Christie's (1999) statement and posits "the more important the decisions to be made from tests, the more important it is to have confidence in the tests (2000, p. 24).

Scholarly research on classroom effects. Although Virginia has reported an improvement in SOL test scores (Howell, 2002), which the Virginia Department of Education believes to provide evidence that student academic achievement is improving (<http://www.pen.k12.va.us>), there is some evidence to suggest that unintended, undesirable consequences of high-stakes accountability may exist at the classroom level. This might nullify the justification for the reform as a whole (Linn, 2000). Relevant literature suggests that teachers working within the contemporary reform tend to teach to the test which inflates the conception of achievement (Linn, 2000, p.7) and coincidentally appears to narrow the curriculum (Barksdale-Ladd & Thomas, 2000; Jones et al., 1999; McNeil, 2000; Perrault, 2000). With the narrowing of the curriculum, some teachers tend to use less of their creative pedagogical prowess in the classroom and lose their joy of teaching (McQuillan, 1999, p. 106). Further, Kubow and DeBard (2000), Shepard and Dougherty (1991), and Smith (1991) report that some teachers become deskilled and deprofessionalized within high-stakes accountability systems.

Aside from the reported unintended negative effects on teachers, Wideen et al. (1997), McNeil (2000) and Wassermann (2001, p. 32) note that high-stakes accountability systems appear to adversely affect students. As classroom curricula is narrowed and teachers teach to the tests, the quality and depth of student learning tends to diminish. Against a backdrop of these reported program effects and escalating teacher frustration, some experienced teachers are leaving the classroom. This trend seems to result in part due to criticism by the news media and by individuals who do not work within the teachers' schools (McQuillan, 1999, p. 106).

Civil rights advocates make similar claims about the presence of undesirable classroom consequences of the current reform. On this point, Hadderman (2000, p. 5) summarizes Orfield and Wald's (2000) observations:

Civil-rights advocates claim that most high-stakes testing policies, particularly those linking single standardized assessment scores to promotion and graduation, discriminate against minority youth, hamstring teachers, reduce complex learning opportunities, and punish victims, not perpetrators, of educational opportunities.

Orfield and Wald's (2000) observations are similar to those made by organizations including the Center for Research on Evaluation, Standards, and Student Testing (CRESST), the National Center for Fair and Open Testing (FairTest), and the Center for the Study of Testing, Evaluation, and Educational Policy (CSTEPP).

Corporate concerns. To the contrast, organizations such as the Thomas B. Fordham Foundation have criticized the views of testing experts who are associated with the abovementioned organizations. The Foundation lists Walter Haney of CSTEEP, the late Richard Jaeger of the American Educational Research Association and professor at the University of North Carolina, George Madaus of Boston College and CSTEEP, and Lorrie Shepard of the University of Colorado and CRESST. Writing for the Fordham Foundation, Richard P. Phelps asserts that the worldview of many testing experts and other educators is dominated by the educational philosophy of constructivism and is responsible for most of the objections with high-stakes testing. According to Phelps, constructivism is “the view that every student and teacher constructs his or her own meanings from classroom activities, books, etc. Hence, no construction is wrong or bad,” even “misplacing a decimal point” (<http://www.edexcellence.net/library/phelps.htm>). Phelps states, “Constructivists oppose school practices that they think ‘fix’ behavior. They see standardizing curricula and instructional practice as restricting teacher behavior and multiple-choice standardized tests as shackling student responses to problems” (<http://www.achieve.org>).

Similar to the Fordham Foundation, Achieve (<http://www.achieve.org>) maintains a prominent voice in the public dialogue on contemporary accountability. These organizations and sizable business groups across the United States are committed to high-stakes accountability programs in large part because they believe that such programs will provide a better- educated workforce, which will make the United states more globally competitive and contribute to American economic improvement (Christie, 1999). Accordingly, Achieve was formed by the U.S. governors and corporate leaders at the 1996 National Educational Summit and since that time has “help[ed] states address legitimate concerns and to provide

states with support for staying the reform course in the face of criticism” (<http://www.achieve.org>). Achieve asserts that the states are “contending with criticism from a vocal minority opposed to standards and accountability...In some states they have launched aggressive campaigns to discredit the testing programs and repeal accountability laws” (<http://www.achieve.org>). Achieve adds “We cannot improve the education all students receive without reliable achievement data and incentives for schools to get better” (<http://www.achieve.org>).

Teacher organizations. Also, teacher organizations such as the Virginia Education Association, including approximately 56, 000 teachers and education support personnel, articulates concern over Virginia’s use of SOL test-score data to determine whether students can graduate from high school. The Association seeks to develop “better measures of what students know and can do” as well as to provide “greater teacher contribution to decision making” (<http://www.veaweteach.org/detail.html>). Further, the American Federation of Teachers (AFT) has added to the discourse on contemporary accountability. Concerning the Virginia SOL system, the AFT released a report on November 2, 2001 that deemed the State’s SOL program as one of the best of its kind in the nation (Wermers, 2001, <http://www.timesdispatch.com/editorials/index.shtml>). Kirk T. Schroder, president of the Virginia Board of Education, responded to the report by asserting that it validates the SOL. Moreover, Schroder stated, “It is gratifying to know that Virginia is doing better than most other states in implementing this type of program . . . This will not stop the board from continuing to listen to reasonable concerns to improve the SOLs. I think we have made tremendous progress, but now is not the time to rest on our laurels” (Wermers, 2001, <http://www.timesdispatch.com/editorials/index.shtml>). The author of the AFT report, Heidi Glidden, showed

tempered optimism about the SOL program. Glidden stated “At this point, there is still too much focus on accountability and testing . . . The tests are seen as punitive when they should be diagnostic” (Wermers, 2001, <http://www.timesdispatch.com/editorials/index.shtml>).

Federal and state level policy. Shortly following the AFT report on December 18, 2001, the United States Congress passed legislation that will require all states to test students annually in the subjects of reading and mathematics. Thus, the trend of standards and test-based accountability received unprecedented ubiquitous approval. The federal legislation requires states to have “annual testing programs in place by the 2005-06 school year. The bill provides \$490 million to states to develop these programs . . . Schools that fail four consecutive years would be subject to restaffing. After the fifth year, these schools would be subject to state takeover, conversion to charter schools or being run by a private contractor” (Wermers, 2001, <http://www.timesdispatch.com/editorials/index.shtml>). In response to the passage of the bill, Schroder stated:

Anyone who thought SOLs or school accountability was going to go away has got to wake up and realize this is now a federal issue . . . The good news for Virginia is, we’re way ahead of the curve. Many states are only starting to build an accountability program” (Wermers, Times Dispatch on the Web).

Problem of Inquiry

In Virginia, provided the ostensible thrust and mainstay of the contemporary accountability effort there has been a need for current and full examination of the classroom effects of the effort. In the fall of 1998, the National Research Council “warned Congress that schools should refrain from basing important decisions like who gets promoted or graduates solely on test scores, and called for more exploration of the unintended consequences of high-stakes exams” (McGinn, 1999, p. 48). In addition, the American Educational Research Association (AERA) recommends that one of the primary conditions to be placed upon the use of standards and high-stakes testing is that there should be full disclosure of likely negative consequences attributable to such programs (Lewis, 2000, p. 263). Questions over the possible undesirable classroom effects of Virginia’s SOL program emerged in the political arena some time ago.

Some Virginia legislators called for a detailed examination of the State’s SOL program (Turner, 2001, B7). A resolution in the State’s House of Delegates was introduced that called for “the Joint Legislative Audit and Review Commission [JLARC] to study the SOLs’ effects on students and teachers” in an effort to determine “whether the SOL tests are valid measures of students and schools” (2001, B7). In addition to this motion, Kirk Schroder asserted that the House Education Committee must, at some point, “wrestle with the role of a test” (2001, B7).

Virginia has begun to take up the work of wrestling with the role of a test and disclosing the classroom consequences of the SOL reform. The SOL Technical Advisory Committee (TAC) has recommended that the SOL program be examined in terms of its possible negative unintended consequences, including the possible negative effects on

teacher morale, student learning, and instructional strategies. Full disclosure of the program's possible negative effects, however, will not be available for nearly 10 years (The Southwest Times, 2001, A2).

Related to the need for full disclosure of the classroom effects of the SOL program, there has been a need for a broader public discourse on high-stakes accountability. In considering teachers' first hand experience working amid the SOL reform, their views seem to be indispensable among the different public perspectives on the reform's effects. John Goodlad, co-director of the Center for Educational Renewal at the University of Washington, underscores this point as he suggests that the public and political discourse on school improvement has virtually excluded many educators as well as any perspectives that are not aligned with high-stakes accountability (2000).

Goodlad, among other scholars, contend that educators need greater opportunities for participation in the discourse, rather than to receive token representation and to have their views often marginalized (Goodlad, 2000, p. 89; McNeil, 2000; Rapp, 2002, p. 218). Hence, the problem of the present inquiry called for a fuller description of teachers' perspectives on the SOL school reform, including the reform's effects at the classroom level. Preliminary research findings informed the problem of the study.

Preliminary Research on Teachers' Views of the SOL Program

A small-scale survey including a sample ($N = 48$) of Southwest Virginia teachers in Floyd County was conducted, including kindergarten through twelfth grade teachers. The study queried teachers' views on the State's SOL program, with a focus on the SOL testing component. Preliminary results suggested that teachers tended to believe the SOL reform

exerts a considerable power over the classroom but that the SOL tests and corresponding test-score data benefit children least when compared to other forms of student assessment such as teacher observations, unit tests, and projects (Bolt, 1999). Overall, teachers tended to feel overly stressed given the current reform. The preliminary research found that teachers tended to believe the SOL tests (1) possess a great deal of power over their classroom pedagogical practices, (2) impact some students quite negatively, (3) create a stressful school environment, (4) lack usefulness as student assessment tools, and (5) procure unacceptably harsh consequences for those who do not pass the tests (Bolt, 1999). These practitioners' perspectives provide a view from the classroom that should not be overlooked, adding depth and understanding to an abiding question over the unanticipated classroom effects of Virginia's high-stakes accountability program.

Purpose of Inquiry

The purpose of this study was to explore and describe VEA teacher-members' perspectives on the SOL program, including their views on the classroom effects of the SOL system. To enlarge and inform the public dialogue of controversy on high-stakes accountability, the teachers' views were discussed in juxtaposition to the multiple perspectives discussed earlier in the chapter.

Questions for Research

In keeping with Creswell's (1994) recommendation for conducting exploratory research, this study was organized in terms of research questions rather than stated hypotheses that are suitable for conducting confirmatory research. The first four questions of

the study addressed the most salient domains of public schooling that appear to be affected by high-stakes accountability programs, as suggested by relevant literature. The fifth and sixth questions queried teachers' beliefs about the possibilities and potentials of the present program as well as queried their beliefs about their roles and responsibilities as they work within the SOL system.

Additionally, because there is some evidence to suggest that schools located in various income-levels are affected differently by the current accountability programs, the questions for research were considered in terms of whether teachers taught in a low-, middle-, or high-income level school locale (Hoffman et al., 2001; Kubow & DeBard, 2000; Paris, Lawton, Turner, & Roth, 1991). Similarly, some evidence suggests that teachers in different school types (elementary, middle, and high school) have varied perceptions of the classroom effects of high-stakes accountability (Kubow & DeBard, 2000); therefore, school types were considered in relation to the research questions. Finally, the perceptions of teachers who teach an SOL-tested subject and those who do not were compared. The questions for research were as follows:

1. What are the views of VEA teacher-members on how the SOL program influences
 - the adequacy of curriculum and the diversity of instructional strategies?
 - student learning?
 - teachers' sense of professional autonomy and level of teacher tension?
 - school quality?

2. Are there differences in VEA teacher-members' views on the SOL program based upon

- whether teachers teach in a low-, middle-, or high-income school locale
- whether teachers teach in an elementary, middle, or high school
- whether teachers teach a SOL-tested subject

Organization of the Study

This study was organized into five chapters. Chapter one has provided a brief account of the public discourse on high-stakes accountability and argued for wider debate that underscores the views of classroom teachers. Furthermore, the chapter has stated the need for full disclosure of the possible undesirable effects of the contemporary reform, with particular attention to the effects at the classroom level. Chapter two provides a review of the literature that is germane to the historical context of high-stakes accountability, the technological features of the SOL program, and the prior research on teachers' views on the current reform.

Chapter three details the study's research methods, which were informed by consideration of the technological features of the SOL program and by prior research involving teachers' perspectives on high-stakes accountability. Chapter four presents the results of this study. Finally, chapter five summarizes the findings, discusses the conclusions and limitations of the study, as well as provides ideas for future research.

CHAPTER TWO

Review of Literature

Method of Review

To explore the questions for research, scholarly literature was canvassed, with three electronic social science-related databases searched with no limits placed on the years and manual cross-referencing performed. First, Education Abstracts yielded 108 records when the keyword combination “high-stakes testing” was searched. Each of the 108 abstracts was reviewed to determine its relevance to the present inquiry, and as a result, 46 records were selected. A second search of Education Abstracts using the keyword combination “high-stakes testing and school reform,” yielded 12 records that contained significant overlap with the prior search; only four additional records were available. Second, the same sequence and keyword combinations as stated above were utilized in searching ERIC via FirstSearch. Initially, 128 records were yielded, and upon the second search 17 records were found, providing only six additional, relevant records.

Finally, a search of PsychInfo was not amenable to the present topic of inquiry, yielding irrelevant, overly narrow psychological records. Of the papers that were collected, their references were cross-checked and additional articles and books were retrieved and reviewed. Furthermore, several experts in the field of education were consulted whose recommendations and provisions of pertinent publications were reviewed. A number of relevant websites were inspected as well.

The literature and the information from pertinent websites were divided into three sections. The first section addressed the contemporary reform in its historical context

whereas the second section included information concerning Virginia's SOL program. The third section included research on teachers' views on high-stakes accountability and how the reform effort affects the classroom. Accordingly, this chapter is organized in terms of three sections that coincide with the three bodies of information, that is, the historical context, SOL policy, and the research on teacher perspectives about the contemporary reform.

Historical Context of Contemporary Accountability in the United States

On October 10, 2001, governors, businesspersons, and educational leaders held the third National Educational Summit "to take collective action to make the nation's public schools truly world class," which was described in a statement of principles offered at the Summit:

We meet again . . . to reaffirm our commitment to the twin goals of excellence and equity in our schools. We must raise achievement for all students while closing the achievement gap separating the educational "haves" from the "have-nots." These goals are an irreducible educational minimum for the United States. Nothing less than their full attainment will serve the nation's social, democratic, and economic interests (<http://www.achieve.org>).

Further, the statement expressed that standards- and test-based reform is the modus operandi through which these goals can be met and is "the very best hope for creating a universally well-educated citizenry and a more equal society" (<http://www.achieve.org>).

The sentiments shared at the Summit could well have been articulated in 1848 by Horace Mann who envisioned the U.S. public schools as “the great equalizer of the conditions of men” (Perkinson, 1976, p. 86). Mann, in 1850 under the direction of the Massachusetts secretary of the Board of Education, oversaw the design of a thirty-item State test (Sacks, 1999) that was used to measure student achievement over the span of one school year and by implication measured the quality of teaching (Greene, 1980, p. 42). As such, the Massachusetts effort appeared to set the tone for the future direction of public schooling and public school reform within the United States.

Technological Advances and New Waves of Immigrants

The beginning of contemporary assessment and high-stakes accountability is contended among scholars in the field. However, ample evidence suggests that early renditions of the reform can be noted during the ten decades between the 1830s and the 1930s when the U.S. experienced extraordinary economic growth and a burgeoning competitive ethic in American life (Cohen & Haney, 1980, p. 7). “Vast technological advances of those times permitted the conversion of raw materials into finished products on a massive scale, transforming regionally based enterprises into companies of national and international scope” (Sacks, 1999, p. 72). The confluence of developing industries and new waves of immigrants entering the U.S. seemed to stimulate a sociopolitical outlook of the day that those within the lower echelons of society could be brought up to a decent, or minimum, level of education, health, and economic status through the medium of the public schools (Cohen & Haney, 1980; Greene, 1980, p. 46; Perkinson, 1976).

Furthermore, as school “enrollments more than doubled as waves of immigrants created a newly diverse population,” there came a “demand for more efficient school management, including ‘the objective and efficient classification, or grading of pupils’” (Tyack, 1974, p. 44). With the demand to categorize the influx of students, tests became widely used. Test use was preferred for its perceived fairness and efficiency in contrast to the period’s system that used essay examinations that received various marks from different teachers (Haney, 1984; Heubert & Hauser, 1999).

Sorting, Classifying, and Tracking Students

By the mid 1900s tests were commonly used to sort, classify, and assign students to various programs of study. In reviewing five decades of assessment-based school reform, Robert L. Linn (2000) cites Cremin (1989) in noting that the rationale for these test uses was provided by James B. Conant in 1953. Conant argued for “universal elementary education, comprehensive secondary education, and highly selective meritocratic higher education” (Cremin, 1989, p. 22). Linn (2000, p. 5) describes Conant’s conceptualization of the educational system as one “for the purposes of selecting students for higher education and for identifying students for gifted programs within comprehensive high schools. In short, Conant believed that a primary problem for American education, as Linn (2000) cites Cremin (1989), included preserving the “quality of education for the academically talented in comprehensive high schools.”

Later during this decade, questions over school quality drew focused attention, especially following the Soviet’s launch of Sputnik in 1957. Pilling (1999, p. 23) cites Rickover (1959, p. 157) in describing that the “powerful thrust of Sputnik’s launching device

did more than penetrate outer space”; the event elicited blame and public scrutiny of U.S. schools and contributed to public dissatisfaction with American education. Subsequent to the Soviet success, American interest was rallied for the improvement of public education, particularly in the areas of mathematics and science (Pilling, 1999) in what seemed to be an effort to restore “faith in America’s present and future technological supremacy” (Rickover, 1959, p. 157).

Proliferation of Research and Evaluation

The Sputnik era gave way to a plethora of research and evaluation particularly between the years of 1966-1968 (Cohen & Haney, 1980, p. 10) where each study “weighed the impact of programs in terms of school output, not of service delivery. And all used one sort of output—students’ scores on standardized tests of achievement or ability” (Cohen & Haney, 1980, p. 10). From this research, Cooper and Leiter observe, “gross interpretations” of test score data were presented in the mass media that extended a deficiency model of American education and generated politics of blame (1980, p. 35). The longstanding supposition was challenged that a strong relationship between resources and academic results existed (Cohen & Haney, 1980, p. 10-11) and a push for accountability followed. Lewis (1999, p. 179) summarizes the cultural milieu of the day by stating that the seed of accountability

was planted in the 1960s when Congress passed the Elementary and Secondary Education Act, which included Title I. Title I required some accountability for student progress, and it combined with the state-level

minimum competency movement that was developing. Together, they created a much more favorable climate for external testing (Lewis, 1999, p. 179).

Moreover, the intense focus upon achievement test scores seemed to suggest to the public that the schools had somehow declined. No longer did the public view that the schools effectively educated the young, nor did they seem to believe as they once did that investing in the schools, by itself, was a wise practice. Rather, the pervasive public opinion called for schools to be held accountable for their actions. That is, the schools must somehow demonstrate that the public's monetary inputs were commensurate with school outputs. This sentiment signaled "a turning point in social policy—a shift away from aspirations to distribute resources more fairly, toward attempts to change the distribution of results" and, consequently, at the time made it "difficult to imagine a change of greater moment for social policy in education" (Cohen & Haney, 1980, p. 11).

Demand for Minimum Competency and Instructive Student Feedback

The turning point in social policy in education led to actions to ensure that minimum results were achieved in the public schools, as measured by minimum-competency tests. Also there grew a demand for test results to provide instructive, that is, remedial, student feedback (Cypress, 1980, p. 32). Thus tests were designed so they might "obtain a description of the specific knowledge and skills each student can demonstrate" (Linn & Gronlund, 1995, p. 17) and, accordingly, assist teachers in addressing specific academic weaknesses of students.

Linn (2000, p. 6) describes this era, the period of the 1970s and early 1980s, as one in which “minimum-competency testing (MCT) reforms swiftly spread from state to state. In a single decade (1973-1983) the number of states with some form of minimum-competency testing requirement went from 2 to 34.” The focus of the reform effort centered upon minimal basic skills as reasonable criteria that students must meet in order to graduate from high school (Linn, 2000, p. 6). Linn writes that the new requirements were of great importance for some students but had little relevance for most students. Gains in student achievement were observed but they occurred mostly at the low end of the distribution” (2000, p. 6).

Popular Portrayal of Public School Mediocrity and the Push for Accountability

The minimum-competency movement seemed to be reinforced by the release of the federal report *A Nation at Risk* in 1983, which portrayed the U.S. public schools as sites of mediocrity (Pilling, 1999). Subsequently, however, the notion of student and school accountability gained heightened political attention that stimulated the development of sophisticated technological tools for reforming American schools. The “[s]tate systemic reform movements during the [1980s and 1990s shifted] from instituting minimum competencies to an emphasis on quality in core academic courses” (Pilling, 1999, p. 95) and expanded the use of test results for accountability purposes (Linn, 2000, p. 7).

According to Linn (2000, p. 7), “Accountability programs took a variety of forms, but shared the common characteristic that they increased real or perceived stakes of results for teachers and educational administrators.” Although initially most states and school districts observed increased student achievement for the first several years of accountability testing

programs, Linn (1990) reports that the test results displayed an inflated impression of student achievement, also known as the Lake Wobegon effect (Koretz, 1988). Among the reported reasons for why this occurred include “the use of old norms, the repeated use of the same test form year after year, the exclusion of students from participation in accountability testing programs at a higher rate than they are excluded from norming studies, and the narrow focusing of instruction on the skills and question types used on the test (Koretz, 1988; Linn, Graue, & Sanders, 1990; Linn, 1990; Shepard, 1990).

With the increasing state-level attention on the quality of core academic courses and the use of test results for accountability purposes, interest was generated to set rigorous national standards in disciplines such as mathematics, science, English, and history/social studies. In effect, national organizations such as the National Council of Teachers of Mathematics and the National Academy of Science assembled commissions to develop national content standards for what students should know and be able to accomplish (Pilling, 1999, p. 82). These actions appeared to set the backdrop for widespread state-level development of content and performance standards that would be coupled with corresponding high-stakes tests for the purposes of student and school accountability.

Contemporary Accountability in Virginia

Virginia Standards of Learning (SOL) Program

From 1990-1996 national standards setting became the focus of school reform in the U.S. With the development of national standards across disciplines, there ensued a fervent political focus on results-oriented education, and school and student accountability became the mandate of the States. With it, many states developed content and performance standards that linked students’ test scores to course placement, high school graduation, and school

accreditation (Heubert & Hauser, 1999). In light of this trend, in 1995 the state of Virginia developed the SOL program that was designed to produce higher academic achievement and school quality. The State program, in effect, was intended to help prepare students for the challenges of the twenty-first century. As such, by 2004, SOL test scores will determine whether students can graduate from high school, and by 2007 should 70% of a school's student population not pass the tests, the school will lose its state accreditation. Schools are accredited by their reaching benchmarks, or passing rates on tests in English, math, science, and history/social studies administered at grades three, five, eight, and high school. Moreover, as reported in the July 2000 Revised Regulations Establishing Standards for Accrediting Public Schools in Virginia (<http://www.pen.k12.va.us>), "the standards propose a system of rewards for high-performing schools, school divisions, and students. Also included are consequences, actions, and assistance for low-performing schools."

In terms of consequences and actions, the State has planned a phase-in period extending to 2005. Through this period, schools may not be denied accreditation, but will be accredited with warning if they have not met the benchmarks established for the SOL tests. For schools that are accredited with warning, Virginia has delineated a plan of assistance: (1) A school academic review process and monitoring plan will be funded to assist schools; (2) Schools warned in regard to English or mathematics are "expected to adopt a recommended instructional method that has a proven track record of success at raising student achievement"; (3) The superintendent and principal will verify to the Board of Education that a successful instructional method has been implemented; and (4) "A three-year School Improvement Plan must be developed and implemented, based on the results of an academic review" of the warned school (<http://www.pen.k12.va.us>). In sum, SOL test

scores will become the ultimate criterion used by the State to determine school accreditation (Virginia Department of Education, 1997). Added to this, high-stakes decisions such as whether students may graduate from high school will be tied to students' SOL test results.

As discussed in chapter one, Virginia's accountability efforts have been underpinned by the views of national leaders, including governors, business leaders, and educational leaders, some of whom who have articulated that standards- and test-based accountability is the preferred means by which to promote excellence and equity in the public schools (<http://www.achieve.org>). McQuillan (2001, p. 102) observes that "proponents see these accountability systems as offering a mechanism to promote greater equity for low-income students of color, to help teachers and administrators identify curricular priorities, and to aid parents in identifying the best schools for their children." However, Madaus (1994, p. 79) cautions that the use of testing during the past 150 years has been "an ill-disguised attempt to coerce the actions of teachers or students or both, and to alter the instructional delivery system by linking results to high stakes." "Though there is no doubt that such pressure works," Madaus continues, "it does so at the price of corrupting instruction and the very tests themselves."

Further, when test results are linked to high-stakes decisions such as whether students are able to graduate from high school, additional questions have been raised. Richard R. Valencia and colleagues contend, "The logic of having an exit examination linked to high school graduation makes sense in a perfect world in which equal educational opportunity exists. Of course, this is not the case" (Valencia, Valenzuela, Sloan, & Foley, 2001, p. 318). At the same time that Valencia and colleagues' article was published in December of 2001, which emphasized the need for equal educational opportunity for students of color and low-

income students, Congress passed a bill that requires states to test students annually in the subjects of reading and mathematics. However, prior to Congress passing the new federal legislation, most of the states already implemented some type of accountability system, involving standards and test results that are tied to rewards and sanctions (Harvard Civil Rights Project, 2001; Linn, 2000; Madaus, 1993; McQuillan, 2001). On this point, McQuillan writes

only one state does not require regular standardized exams for students in its schools (L. Shepard, personal communication, May, 2001). Those who endorse such accountability measures believe they offer a variety of benefits. Most fundamentally, they represent a means to enhance student achievement because now the standards are explicit, testing procedures are in place, and students, teachers, and administrators understand the consequence of failing to meet the standards (2001, p. 102).

Questions Over the Adequacy and Appropriateness of the SOL Program

In Virginia, the high-stakes nature of the SOL system has generated perhaps the most debate. In February of 2001 questions over this feature of the system seemed to prompt some policymakers to endorse a bill (HB 2163) “that would force the state board to reduce its emphasis on Standards of Learning tests and consider other criteria in its formula for accrediting public schools” (Sluss, 2001). Additionally, in response to questions over the influence SOLs have upon student learning and their ability to graduate from high school,

The House of Delegates passed a bill (HB 2394) that would establish a “sliding scale” for struggling students, allowing high schools to blend SOL scores and classroom grades in awarding diplomas . . . The bill would apply only to students now in grades six through nine, giving the state board time to assess the new law’s impact (Sluss, 2001).

As noted in chapter one, other concerns have hinged on the adequacy of the tests’ psychometric properties, particularly in light of test length and the ability of the tests to fairly measure the achievement of children of color as well as those living in lower-income communities (Sluss, 2001). Even so, proponents of the test-based accountability program maintain that the SOL tests demonstrate good reliability and validity. To support the dimension of testing in the SOL program, the Virginia Department of Education (VDOE) has addressed the issues of reliability and validity of the State’s SOL tests and, according to a review of statistical analyses by the Department’s independent body of testing experts, the tests are psychometrically sound. The State believes the SOL test scores to be trustworthy indicators of students’ academic performance and school quality.

However, others question the technical integrity of the tests. The Virginia Parent Teacher Association states “There is absolutely no way a 40- or 50-item multiple-choice test is going to fully assess all of the Standards in a curriculum” (The Roanoke Times, 2000, p. B8). Cross (2000), a measurement expert, concurs with the parents’ view: “And test length does matter! Not only is the test able to sample more adequately the skills and knowledge to be tested, but a fundamental fact about test scores is that reliability increases with test length.” Notwithstanding this debate over the psychometric soundness of the SOL tests, the

question remains as to what are teachers' views on how the SOL program influences the classroom. To inform this question, teachers' perspectives on high-stakes accountability programs were examined.

Teachers' Perspectives on Contemporary Accountability

The previous two sections of the chapter have emphasized the longstanding American tradition of standards and testing, which has developed into a sophisticated technology of student and school accountability, particularly flourishing over the past decade with the increased participation of governors, business leaders, and federal legislators in support of the reform effort. The political power that appears to support the contemporary reform seems to be bolstered by reports of a deficiency model of American schools as well as political and public interest in improving excellence and equity in schools for all children, particularly for children of color and for low-income children. In effect, the reports have helped generate results-oriented social policy of education, leading individual states to legislate school reform efforts that are built upon assessment and high-stakes accountability.

In addition to the historical accounts of standards and testing, there has been an enormous volume of conceptual papers produced on teachers' perspectives on high-stakes accountability mechanisms and the outcomes of these mechanisms. These include opinion pieces, theoretical and philosophical papers, as well as views that are based upon informal observations or conversations with educators (Airasian, 1988; Bostingl, 2001; Cochran-Smith, 2000; Cook, Cunningham, & Tashlik, 2000; Coleman, 2000; Darling-Hammond, 1994; Gehring, 2000; Gitlin & Margonis, 1995; Haertel, 1999; Hanson, 1993; Hargreaves & Goodson, 1996; Hess & Brigham, 2000; Houston, 2000; Kanpol, 1995; Kohn, 2000; Labaree,

2000; Lewis, 1999; Linn, 2000; Madaus, 1988, 1991; Neill & Medina, 1989; Miller, 2001; Pearson, Vyas, Sensale, & Kim, 2001; Piphoo, 1995; Popham, 1991, 1999; Ramirez, 1999; Schmoker, 2000; Shepard, 1991; Soder, 1986; Smith & Fey, 2000; Stake, 1999; Suarez & Gottovi, 1992; Taylor, 1994; Tyack & Cuban, 1995; and so on). The overwhelming majority of these papers and books discuss undesirable outcomes of high-stakes accountability systems that are related to students and teachers.

Although there are many conceptual works that address the contemporary reform effort, there is a dearth of research that queries teachers' perspectives on the contemporary reform and its classroom effects, and that which does seem to lack evidentiary warrant that supports the validity and reliability of its measures. Prior to the present study, seventeen studies were identified in the empirical literature, seven of which were predominantly qualitative in nature and the remaining studies included survey research. Of the 10 studies that utilized surveys, only one reported a check on the construct validity, or meaningfulness, of the scales of the questionnaire (Shepard & Dougherty, 1991), and just one study reported the reliability coefficients for the survey scales (Urduan & Paris, 1994).

Of the surveys conducted, two reported random sampling of teachers or educators (Haney, 2000; and Hoffman, Assaf, & Paris, 2001, respectively), one reported a three-level stratified random sampling process (Jones, Jones, Hardin, Chapman, Yarbrough, & Davis, 1999), and one reported a random sample of schools along with an intentional selection of teachers based upon the grade level taught (Shepard & Dougherty, 1991). The sample sizes of the 10 surveys varied from $N=58$ to $N=360$ with one exception, a study conducted by Nolen, Haladyna, and Haas (1989) including over 2,000 teachers and administrators.

Furthermore, the surveys employed 4-, 5-, or 6-point Likert scales that typically queried to what extent the participants agreed with statements about how the contemporary reform influences the classroom. The surveys tended to address issues related to the classroom curriculum, instructional strategies, student learning, and issues of teacher professionalism, autonomy, and stress. The overwhelming majority of the surveys found unfavorable teacher views on contemporary accountability programs as well as undesirable classroom effects of such programs.

The seven qualitative studies that pertained to teachers' perspectives on contemporary accountability involved multiple modes of inquiry, including in-depth interviews within case studies, focus group interviews, observations, and analysis of personal teacher essays. Barksdale-Ladd and Thomas (2000) conducted 59 in-depth interviews of teachers of grades one through eight also who were graduate students in two large states. Eighteen of the 59 were interviewed in groups of six in focus group discussions whereas the remaining 41 teachers were interviewed individually. Also, Perrealt (2000) employed focus group interviews, including eight groups comprised of 7-9 members.

Among the other qualitative studies in the body of empirical literature included the observational study of Smith, Edelsky, Draper, Rottenberg, & Cherland, (1989). Smith et al. (1989) conducted the study in two elementary schools in a metropolitan Phoenix district in order to develop a grounded theory of testing effects. Added to the Smith et al. (1989) study is the ethnographic study reported by McNeil (2000), which involved daily observations in classrooms over the course of a semester in two magnet schools in Texas. Further, the study conducted interviews with teachers as well as engaged in historical research. Interviews and classroom observations were also employed in the research of Grogan (2001). Grogan (2001)

performed nine interviews of Virginia teachers as well as observed their classrooms in order to describe the teachers' views on the State's SOL system. Finally, unlike the other qualitative studies reviewed, Glatthorn's (1999) research examined personal essays of classroom teachers.

Concerning the qualitative studies' methods of data analysis, most discussed the analyses in terms of identifying themes, or patterns, in the qualitative data; however most did not specify particular methods that have been recommended by scholars in the field, for example, analytic induction methods (Erickson, 1986) or constant-comparative (Strauss, 1987). As an exception to this trend, the study conducted by Smith et al. (1989), which also was reported in the *Educational Researcher* (Smith, 1991), described the methods of analysis in sufficient detail. Additionally, the research of Barksdale-Ladd and Thomas (2000) sufficiently explained their method of double or triple coding units of meaning as well as employing representative member-checks for the purposes of improved clarity and accuracy. Similar to the findings of the survey research that was reviewed, most of the qualitative studies reported undesirable outcomes of contemporary accountability. Taken together, the results of this body of research are reviewed in the following sections of the chapter.

Adequacy of Curriculum and the Diversity of Instructional Strategies

An overarching aim of the contemporary reform is for students to master a state-specified body of knowledge in order to best "serve the nation's social, democratic, and economic interests" (Achieve on the Web), which typically requires teachers to align the classroom curriculum to the curriculum objectives specified by their state (Achieve on-line, 2001). Such alignment underscores the influence of the current reform on the classroom

curriculum and the approaches that teachers use to communicate subject matter (McNeil, 2000). Allan A. Glatthorn (1999, p. 27) has examined personal essays of teachers on how they respond to the contemporary accountability effort and describes teachers' views on curriculum alignment. Glatthorn reports that teachers believe the current program of standards and high-stakes testing will not lose its public and political appeal anytime soon; therefore, the "issue is not 'Should we align?' Rather, it is 'How can we align so that alignment is teacher-directed and teacher-friendly?'" (1999, p. 27). Further, according to Glatthorn, most teachers "cope creatively" (1999, p. 27) in their effort to align the classroom curriculum to state objectives. He describes curriculum alignment as a tool, "one that can be used foolishly or wisely":

Used foolishly, curriculum alignment diminishes the art of teaching, sterilizes the curriculum, and makes the classroom a boring place. Used wisely, it offers teachers a practical method by which they may ensure that their students are well prepared for the mandated test (1999, p. 27).

Grogan's (2001) research, based upon nine interviews of Virginia teachers, found similar results. Grogan (2001) added that provided Virginia's SOL reform, the teachers interviewed had a deep knowledge of the content and planned their instruction carefully. While Grogan's (2001) and Glatthorn's (1999) research suggest that teachers cope creatively with the necessity of curriculum alignment, there is a substantial body of research that describes teachers' views on the current reform's influence upon curriculum and instruction in less desirable terms. This body of research suggests that teachers tend to feel compelled, or

forced, to emphasize basic skills instruction and student preparation for tests at the expense of teaching content that is not tested (Hoffman et al., 2001; Jones et al., 1999; Kubow & DeBard, 2000; Shepard & Dougherty, 1991). Thus, the classroom curriculum is reportedly narrowed, or constricted, to that which will be tested on high-stakes examinations as a consequence of the contemporary reform. Further, according to some elementary and junior high teachers who teach within the reform effort, their students' opportunities for broad school experiences including subjects not tested by high-stakes tests, recess, and lunchtime are sometimes absorbed by test preparatory work (Gordon & Reese, 1997; Haney, 2000; Hoffman et al., 2001; Smith et al., 1989; and Smith, 1991). Smith (1991) adds that the impetus to increase test scores seems to prevent classroom teachers from experimenting with promising programs.

There is some evidence to suggest that the narrowing of classroom curriculum, also described as curriculum constriction (Kubow & DeBard, 2000), affects elementary, junior high, and high school teachers differently. Kubow and DeBard (2000) report that standards and high-stakes testing might impede the teacher-generalist's curriculum plans and instructional strategies whereas for high school teachers, typically subject matter specialists, curriculum constriction is not sensed with the same degree of intensity. In sum, curriculum constriction appears to be sensed more by elementary and junior high teachers than by high school practitioners.

Additionally, there is evidence to suggest that curriculum constriction is experienced more often by youth who attend schools in low-income locales as compared to those who attend schools in more affluent areas. Schools in low-income localities, which tend to display lower high-stakes test scores than schools in higher-income areas, sometimes view the low

test results as a need for more test preparation and basic skills instruction at the expense of activities that might cultivate students' critical thinking (Kubow & DeBard, 2000; McNeil, 2000).

Along with narrowing the classroom curriculum (Barksdale-Ladd & Thomas, 2000; Jones et al., 1999; McNeil, 2000; Perrault, 2000), the literature also suggests that teachers reduce the variety of instructional approaches they employ in order to teach state specified curriculum standards (Wideen et al., 1997, p. 428). The reduction of diverse instructional strategies is reportedly connected to an over-emphasis upon teaching basic skills and fact-based information that teachers believe will appear on students' high-stakes tests. Thus teachers perceive the rewards and sanctions associated with high-stakes tests, seek to prepare students for those tests, and in so doing, narrow the classroom curriculum to the specified standards. This process generally reduces teachers' variety of instructional strategies (Nolen et al., 1989; Smith et al., 1989, Smith, 1991; Wideen et al., 1997).

Quality of Student Learning

Wideen et al. (1997) report that as a consequence of the current reform, the classroom curriculum is narrowed and teachers reduce their use of a variety of instructional strategies. Consequently, this tends to diminish the quality of student learning (Gordon & Reese, 1997; Haney, 2000; Hoffman et al., 2001; McNeil, 2000; Wideen et al., 1997). An over-emphasis upon basic skills and information likely to be tested on high-stakes examinations tends to supplant students' opportunities to interact cooperatively in small groups with the subject matter. Furthermore, teachers report that they and their students have little opportunity for whole class discussion of the subject matter. High stakes testing, in effect, may discourage

“teachers from using strategies which [promote] enquiry and active student learning,” thus impoverish classroom discourse (Wideen et al., 1997, p. 428). McNeil’s ethnographic research in Houston, Texas public schools supports this claim (2000).

Teachers tend to report that the diminished classroom discourse associated with high-stakes accountability programs bears upon students’ opportunities to think critically about the subject content. Provided fewer opportunities to think critically about a subject, teachers indicate that students’ intellectual needs are not sufficiently met and that students are not making meaningful connections within the subject matter (Wideen et al., 1997). Although the research does not focus on teachers’ perspectives on the contemporary reform, the relevant research of Fordham and Ogbu (1986), Romo and Falbo (1996), and Valenzuela (1999) underscore the undesirable effects of diminished classroom discourse and opportunities for children to think critically. Moreover, teachers tend to report that provided a high-stakes accountability system, students are prepared for high-stakes tests yet do not tend to develop the intellectual depth of knowledge and application of that knowledge that teachers believe contributes to high quality learning in the classroom (Gordon & Reese, 1997; Haney, 2000; Hoffman et al., 2001; and McNeil, 2000). Without the presence of high-stakes systems, the 59 teachers of Barnes, Clarke, and Stephens’ (2000) survey indicated that “they would tend to do more interest-type activities...and the children might enjoy it more” (p. 642).

In addition to teachers reporting that students do not have ample opportunities to think critically about the subject matter provided high-stakes accountability systems, there is some evidence to suggest that students experience considerable stress and sometimes have physical illness as a consequence of high-stakes testing conditions. Kubow and DeBard (2000) report that teachers observe increased anxiety and stress in students as well as noting

that students with learning difficulties tend to feel overwhelmed by the contemporary reform. Further, Hoffman and colleagues (2001, p. 487) surveyed 200 Texas teachers regarding the State's high-stakes testing program and found that "According to teachers, many students experience headaches and stomachaches while taking the TAAS [high-stakes examination]. A surprising number are anxious, irritable, or aggressive." These negatives seem to be of greater consequence for "at-risk," low-performing, and minority students (Hoffman et al., 2001). The research of Urdan and Paris (1994) suggest similar consequences for students.

Teachers' Professional Autonomy and Level of Tension Experienced in the Classroom

Practitioners tend to report a diminished sense of teacher professionalism and mounting tension in the classroom as a result of high-stakes accountability. Kubow and DeBard (2000, p. 17) describe the wide attention to the subject of teacher professionalism (Cochran-Smith, 2000; Darling Hammond, 2000; van den Berg & Ros, 1999; Ryan & Cooper, 1998; Soder, 1986) and explain its meaning in terms of a "teacher professional" as follows:

A teacher professional is understood as one who possesses particular knowledge and expertise that are rendered to benefit others. In turn, a teacher is granted a fair degree of autonomy to employ his or her knowledge with the expectation that such knowledge and expertise will enrich students and that such service will be rendered in an ethical manner.

From their survey of 193 Ohio teachers ($n = 102$ high school and junior high teachers, $n = 91$ elementary school teachers) representing one suburban school district, Kubow and DeBard (2000) report that more than 96% of the participants believed that [high-stakes] proficiency testing had been “imposed upon” the school district. Since teachers believed they were not consulted in an on-going fashion about the program’s classroom impact and what is best for students, the participants felt their teacher professionalism was disregarded. As a result, mistrust developed between the Ohio teachers and policymakers (2000, p. 18).

The findings of Kubow and DeBard (2000, p. 22), similar to those of Wideen and colleagues (1997), suggest that a majority (64%) of the Ohio participants believed high-stakes testing minimized their creative teaching activities and hindered them from making decisions based upon their professional expertise. These consequences underscore a point in the empirical literature that suggests that the diminished opportunities for teachers to creatively employ their craft may produce a deskilling of the teaching profession (McNeil, 2000; Smith, 1991). In turn, this may encourage less critical thought and action on the part of classroom practitioners (McNeil, 2000). Teachers worry that “teaching to the test” may replace the broader curricula and their best knowledge and expertise about what their students’ should learn (Gordon & Reese, 1997; Haney, 2000; Hoffman et al., 2001; McNeil, 2000; Smith, 1991; Wilson & Corbett, 1991).

The interviews of Barksdale-Ladd and Thomas (2000) and Perreault (2000) add that there is considerable teacher frustration toward policies with a high-stakes testing centerpiece, that teachers’ decisions as to what is best for students is dismissed, and that teachers feel disempowered as a result of their diminished professional autonomy. In sum, teachers desire to contribute meaningfully to young people’s learning in the classroom and

do not equate this desire with the purpose of preparing students for high-stakes tests (Hoffman et al., 2001; Wideen et al., 1997).

When teachers' professional autonomy is not supported, practitioners tend to report that they sense a dramatic shift in the purpose of their teaching (Hoffman et al., 2001). Rather than experiencing a full opportunity to employ their expertise and creativity in selecting the classroom curriculum and designing approaches to instruction, some classroom teachers report a constraint upon the use of their expertise and creativity as they work within a high-stakes accountability program: "We are 'required' to teach to the [test]. I became a teacher to teach children," stated a teacher in Hoffman and colleagues' research (2001, p. 489).

Further, some teachers may seek to teach a grade level where the purpose of teaching does not appear to be shifted dramatically, or, similarly, they may seek to teach in a different school where they believe they might have a fuller opportunity to pursue their purpose for teaching. Still, some teachers leave the field of teaching altogether. The salient research of Hoffman et al. (2001) and Kubow and DeBard (2000) document this trend. McQuillan's (2001, p. 108) work, although not focusing upon teachers' views of high-stakes accountability, posits that this trend would seem to impact student achievement in low-income schools where teacher turnover tends to be more pronounced. In citing the Harvard Civil Rights Project (2001, p. 2), McQuillan writes that a "student's performance...is significantly tied to the level of their teacher's experience. Minority and low-income students tend to have teachers with the lowest amounts of experience and are therefore likely to perform less well...than their white counterparts—and to be unfairly hurt by the test's consequences."

Along with teachers' general sense of diminished professional autonomy, Shepard and Dougherty (1991) report that some practitioners feel excessive pressure to improve test scores by their district administrators as well as the news media that tend to report gross comparisons between schools and districts. When publicized school test scores are reported in the news media, which are intended to reflect school quality, some teachers experience "shame, embarrassment, guilt, and anger...and the determination to avoid such feelings in the future" (Smith, 1991, p. 9). The pressure to improve test scores, as it is associated with the possibility of school sanctions, tends to permeate the work of teachers (Kubow & DeBard, 2000) and adversely influences teachers' morale (Kubow & DeBard, 2000, p. 20). Some teachers have reported that they have engaged in test polluting practices, or cheating, during test time to improve students' test scores on high-stakes examinations.

Jones et al. (1999) report that much of teachers' feelings of pressure and anxiety are related to their feelings that they cannot directly control the characteristics of students assigned to them. Teachers believe that high-stakes test scores are related to students' socio-economic status (1999). Classroom teachers' level of tension and anxiety do seem to differ according to grade level taught, with some evidence to suggest that elementary school teachers experience more pressure than their secondary counterparts in relation to high-stakes programs (Kubow & DeBard, 2000, p. 20).

In conclusion, this review of the empirical literature on teachers' views on the contemporary accountability effort was based upon 17 studies, with the majority of the studies not reporting evidence to support the validity or reliability of the measures employed in the research. Additionally, of the 10 surveys, it appeared that half did not use random sampling, which limited the generalizability of the survey results. Of the qualitative studies

that might have been most informative to the present investigation was the study conducted by Grogan (2001) in which nine Virginia teachers were interviewed and classroom observations were made. This study included the smallest sample of teachers of any of the qualitative studies reviewed, and questions were raised regarding its accuracy particularly since its results differed markedly from all but one of the qualitative studies as well as the survey research.

Thus, the scope of research that addresses teachers' perspectives on high-stakes accountability mechanisms is limited, and much of the research that is available seems to lack evidentiary warrant for its validity and reliability. The present research, however, adds an important layer to this literature by employing comprehensive survey research, drawing a random sample of teachers, and providing sufficient evidence to support the validity and reliability of the measures used. Moreover, this study addresses teachers' views on the contemporary accountability reform in the state of Virginia, where little research on teachers' perspectives has yet to be reported.

CHAPTER THREE

Method of Research

Research Design

Survey Research

As presented in chapter one, the questions for research were intended to explore teachers' beliefs related to how various domains of the SOL program influence teaching and learning in the classroom. According to Rea and Parker (1997, p. 4), when a compilation of information is to be collected that is descriptive, behavioral, and preferential in nature, a survey research design is amenable for use. Thus, because the present study aimed to collect descriptive information on teachers' beliefs related to a range of topics concerning the SOL program, a survey research design seemed most appropriate.

Population and Sample

The population of the study included 36, 000 classroom teachers who are members of Virginia Education Association (VEA). The total membership of the Association is comprised of approximately 46, 000 administrators, teachers, and school personnel, with roughly 36,000 teachers consisting the total. The sample of the study was provided by VEA who used a computer program to generate a simple random sample of 800 teachers who are listed as members of the Association. Furthermore, VEA provided mailing labels with the names and addresses of the members that had been drawn for the purposes of the study.

Survey Development

Self-Designed Survey

To explore the questions for study through a survey research design, an appropriate pre-existing survey instrument was sought, but was not identified. Therefore a self-designed instrument was developed, the Teacher Survey: On the SOL Reform. The instrument, in keeping with most survey research, did not focus upon a singular issue. Rather, the measure was designed to collect information on multiple issues relevant to the SOL reform and its influence upon the classroom. The survey research of Kubow and DeBard (2000), the extensive interview inquiries of Barksdale-Ladd and Thomas (2000), and the comprehensive work of Heubert and Hauser (1999) served as decisive references during the construction of the questionnaire. A relevant pilot-project (Bolt, 1999) was utilized as were numerous policy statements posted on Virginia's Department of Education Web page (<http://www.pen.k12.va.us/>). Furthermore, informal conversations with classroom teachers and protracted consultations with measurement experts were heavily drawn upon.

Scale Construction

In developing the study's instrument, Likert-type scale construction was utilized (1 = *strongly agree*, 6 = *strongly disagree*). In keeping with Likert scaling (Pedhazur & Schmelkin, 1991), the first procedure of the study's scale construction involved developing an item pool generated from the sources discussed in the previous section. Accordingly, approximately equal numbers of favorable and unfavorable statements were amassed that pertained to the SOL program, which served to reduce response sets (Pedhazur & Schmelkin, 1991, p. 122). Response sets may result from participants indicating responses that are

similar to their other responses due to convenience or the participants' inattention to the specific content of each item, for example. In such cases, the responses do not reflect the participants' belief related to a particular item, therefore do not add useful information to the study. By amassing approximately equal numbers of favorable and unfavorable statements pertaining to the SOL program, this helped guard against the development of response sets and hold the participants' attention as they responded to the survey.

The item pool included roughly 160 items addressing the following salient themes of the SOL reform: the adequacy of curriculum, the diversity of instructional strategies, teachers' sense of professional autonomy, the level of teacher tension in the classroom, student learning, and test-score use. The items were appropriately matched under five headings described shortly. The item pool was refined almost daily over a period of several months, with substantial refinements made to the items at least weekly. This process, requiring the amalgamated expertise of test specialists, teachers, and the relevant literature, yielded 80 acceptable items. Eight demographic items were added to the pool. Through 12 iterations of refinements, the 88 items of the questionnaire were configured in a clear format resulting in the semi-final draft.

The items comprising the instrument were appropriately matched under the following five headings that organized the survey: (1) the Standards of Learning: curriculum objectives that indicate "what students should know," (2) the SOL Testing Program: SOL tests and the use of test scores, (3) Performance Standards for Students: minimum score of 600 to pass SOL tests; "what students should be able to do," (4) Performance Standards for Schools: at least 70% of students in a school must pass SOL tests, and (5) Accountability: explicit consequences for students and schools when they meet or fail to meet the performance

standards; and teachers' perceptions of the classroom effects of SOL-related accountability. Moreover, the definitions provided above were based upon those described on Virginia's Department of Education's Web page (<http://www.pen.k12.va.us/>) and were included on the Teacher Survey in order to clarify the intended meaning of each segment of the SOL program.

Open-Ended Questions

To help enlarge the teachers' responses to the Likert-type items of the survey instrument, three open-ended questions were added to the 88 items of the survey. The open-ended questions asked teachers to describe their views on the use of SOL test-scores, their views on the notion of school accountability, and their beliefs about their roles and responsibilities as teachers working within the SOL system. The addition of the open-ended questions provided a second data set that helped expand and strengthen the credibility of the quantitative results.

Item Content Review

A small group of elementary school teachers and high school teachers in Radford, Virginia was asked to critique the questionnaire items (Dillman, 1978, pp. 155-159). Given the critiques, the questionnaire was further refined and the final draft was completed. Following this, the negatively worded items of the questionnaire were reverse coded so that the responses could be interpreted in the same manner as the favorably worded items.

Procedures

Initial Contact

The fieldwork for this study ensued in early January of 2002 when first contact was made with the VEA teacher-members by means of a mailed note card, which is displayed in Appendix A. The note card alerted teachers to the Teacher Survey: On the SOL Reform that was soon to be mailed to their home address. Further, it explained that the teachers had been selected at random, assured them of complete confidentiality should they participate in the study, and offered a copy of the study's results once the study was complete (Dillman, 1978). Of the 800 mailed note cards, 24 were returned primarily due to teachers relocating and no available new address. One teacher, however, was recently deceased as indicated by the teacher's relative whereas another was not a classroom teacher but a counselor. Thus, contact was made with 776 of the 800 teachers sampled from the VEA membership list.

Survey Administration

Each copy of the survey instrument was numbered only so that the researcher could eventually send follow-up mailings to those who had yet to return the survey. In this way, cost was minimized and follow-up communiqués were not mailed to those who had already returned the questionnaire. The instrument, which was organized in booklet form, was mailed to the teachers' residence, including a cover letter and a self-addressed return envelope (Dillman, 1978). A 19% ($n = 147$) response rate was obtained following the survey administration. The instrument is presented in Appendix B.

Follow-up Communiqués

On January 29 a second note card was mailed to the teachers who had yet to respond to the Teacher Survey, and the note card reminded them to please complete and return the questionnaire. Approximately one week following the mailing of the second note card, a second copy of the Teacher Survey was mailed to teachers who had not responded. An additional 27% ($n = 208$) response rate was obtained.

Response rate. Having made four contacts with the teacher-members during the survey administration process, a 46% ($n = 360$) response rate was obtained, comprising 1% of the VEA teacher-member population. In general this was a modest response rate. However, when considering the length of the Teacher Survey the response rate is better understood. To complete the 88 items and three open-ended questions of the questionnaire, a prolonged effort and commitment of the teachers was required. This requirement likely reduced the response rate of the study. However, the length of the questionnaire helped strengthen the reliability of the survey and provided a means by which to explore teachers' views on a wide variety of issues related to the SOL program.

Data Entry

Quantitative data entry. Following the administration of the survey, the quantitative data were entered into a spreadsheet using the Statistical Package for the Social Sciences (SPSS), version 10.0. The quantitative variables were labeled by name and their values were labeled as well. All negatively worded Likert-type items were reverse coded so that they might be interpreted similar to the positively worded items during data analyses.

Qualitative data entry. The teachers' responses to the open-ended questions of the instrument were transcribed from the hard copies of the surveys into a Microsoft Word (2000) document. All written responses were recorded verbatim and demographic information was documented for each of the three responses for each teacher-participant.

Notably, there were only three open-ended questions included on the Teacher Survey and these followed 88 Likert-type items, yet 80% ($n = 300$) of the participants responded to the open-ended questions, often writing at length and with great emphasis through the use of double-underlined words and capitalized words or phrases. The transcription of these teacher narratives resulted in a 230-page document for thematic analysis.

Validity and Reliability

Scale Validity

Factor analysis. Principal components analysis (PCA) was used to examine the factors of the Teacher Survey, or to reduce the data into manageable meaningful parts. Resulting from the PCA, neither the component matrix nor the scree plot offered useful information in terms of identifying clear factors in the data. Roughly half of the survey items loaded on 17 factors, which did not add meaning to these data.

Domain construction. To separate the items into meaningful content domains, an examination of the item-frequencies, item content, and inter-item correlations was conducted. Provided the strength of the inter-item correlations and the item content, the following eight domains, or scales, were constructed as follows: Adequacy of the Standards of Learning (ADSOL), Support for Teacher Professional Autonomy Within the SOL Reform (TPROF), Level of Teacher Tension Within the SOL Reform (TTEN), Diagnostic Value of SOL Test

Scores (DIAGV), Adequacy of SOL Pass Rates to Signify School Quality (PASSQ), Ability of the SOL Program to Meet Diverse Student Needs (DIVERSE), Adequacy of SOL Testing and SOL Test Scores to Hold Schools Accountable (ADTEST), and the Adequacy of SOL Cut-Scores to Signify Student Learning (CUTSCR). The construction of these domains helped organize the data under meaningful headings that served the discussion of the survey results. In the construction of these domains, two items (items 31 and 32) were omitted due to their inability to meaningfully connect with any of the domains.

Reliability

The reliability of each scale was estimated using Cronbach's alpha, a measure of internal consistency. The overall alpha reliability for the Teacher Survey was .96, a notably high reliability estimate. Further, for all but one of the eight scales, alpha indicated strong internal consistency. Chapter four describes the scales in detail, including the descriptive statistics for the items and scales as well as the content for each item. However, for the purpose of the present chapter, a brief description of the scales has been provided.

ADSOL scale. First, Cronbach's alpha was estimated for ADSOL and resulted in a high coefficient, .92, thus indicating the scale's high internal consistency, the highest reliability of the eight scales of the instrument. ADSOL was comprised of eight items that queried the teachers' views on the reasonableness of the SOLs as guidelines for teaching and learning as well as the ability of the Standards to improve the quality of teaching and learning. Relevant to the quality of teaching and learning, the scale also asked teachers their views on the SOLs' ability to provide an education that is equitable for all students.

TPROF scale. The second survey scale consisted of 11 items, TPROF, which questioned the participants' views on how the SOL reform affected their sense of professional autonomy. This included how the reform affected teachers' professional decision-making in the classroom, how it influenced community and parent confidence in teachers' work, and how the SOL program affected participants' quality of teaching and ability to meet students' educational needs. Cronbach's alpha indicated the scale's high reliability, .89.

TTEN scale. Next, an alpha reliability coefficient of .89 was estimated for the TTEN, indicating high internal consistency. The 14 items of the scale asked teachers about the SOL program in relation to tension, or frustration, they might experience. The scale queried teachers about the effects of a classroom emphasis upon SOL test scores, the public comparisons of schools' SOL test scores, and the level of stress that teachers and students might experience as a result of teaching and learning SOL subject matter, respectively.

DIAGV scale. The fourth scale of the Teacher Survey, DIAGV, showed strong internal consistency, with alpha = .86. The scale was comprised of nine items oriented toward how teachers use the SOL test scores to identify weak instructional practices, to diagnose students' learning difficulties, and to determine student-related decisions such as course placement, grade retention, and student high school graduation.

PASSQ scale. The survey's PASSQ dimension consisted of eight items that queried the participants' views on the adequacy of a 70% pass rate on SOL tests to indicate school quality, thus the adequacy of the pass rate to qualify a school for State accreditation. The estimate for the scale's reliability was strong, with Cronbach's alpha = .85.

DIVERSE scale. DIVERSE, the sixth survey scale, indicated good internal consistency ($\alpha = .84$). The 11 items of this domain addressed teachers' perspectives on how the SOL program affects classroom curriculum, the use of diverse instructional strategies, students' opportunities to think critically about the subject matter, and temporal allowances in the classroom.

ADTEST scale. Cronbach's alpha was estimated for ADTEST and as a result indicated a reliability coefficient of .83, suggesting good scale reliability. The scale's nine items queried teachers' views on various ways the testing tool of the SOL system attempts to spur higher performance from teachers and students in their efforts of teaching and learning, respectively. Accordingly, the scale asked participants about their views on test-score comparisons, high school graduation requirements, and school accreditation requirements.

CUTSCR scale. CUTSCR was the eighth scale for which Cronbach's alpha was estimated. This five-item scale showed adequate reliability ($\alpha = .77$). The content of the scale asked the teachers about their views on the appropriateness of the level at which the SOL test cut-scores are set. Further, this scale queried teachers about the meaning of the cut-scores, that is, if the cut scores seemed arbitrary or if they clearly distinguished students with acceptable subject knowledge from those without acceptable knowledge. Although the scale indicated adequate internal consistency, the CUTSCR reliability coefficient was lowest among the eight scales of the Teacher Survey, and the number of teachers ($n = 303^{**}$) responding to the items of CUTSCR was least among all the scales. This scale, which queried the participants' views on a more technical element of the SOL reform, elicited fewer teacher responses than the other scale dimensions. Therefore, it might be reasonable to

conclude that the content of CUTSCR extended beyond the knowledge or experience of the teacher-participants and perhaps fell into the realm of test specialists or psychometricians.

Summary. The estimates of coefficient alpha for the eight scales of the Teacher Survey indicated, overall, strong reliability of the instrument, with most of the reliability coefficients falling between .83 and .89. These results indicated that the teachers' responses to the items of the questionnaire were consistent, thus reliable representations of their views on the SOL reform.

Data Analyses

Quantitative Data Analysis

Frequencies. For each of the 80 continuous variables, histograms were examined to determine the distributional shape of the data. Further, coefficients of kurtosis and skewness were computed for each of these variables to assess whether the data violated the assumption of normality.

Item-level descriptive statistics. Having observed the frequencies for each of the Likert-type items, the percentages of response for each of the items were examined, with particular attention to cumulative percentages. Also, the range, mean, mode, and standard deviation were computed. Also, the content of five items with the lowest means was examined, as was the content of the five items with the highest means thereby providing item-level insight into the teachers' perspectives on the SOL reform.

Crosstabulations. The third procedure of the analysis involved crosstabulations, which indicated the number of participants who belonged to various combinations of groups, or categories. The categories examined are as follows: (1) status on Virginia subject

endorsement in the primary subject taught by the income-level of teachers' school locality, (2) status on Virginia endorsement in the primary subject taught by school type, and (3) status on Virginia endorsement in the primary subject taught by teachers' status on teaching a SOL-tested subject. Furthermore, in terms of whether the participants taught a subject for which there was a SOL test by their school type and their State endorsement status, a layered crosstabulation of frequencies for these variables was conducted.

Accordingly, the crosstabulated frequencies were used to provide a thorough description of the study's participants.

Data aggregation. Next, the data for two continuous variables were aggregated to render three categories for each of the variables (The two variables included item 82 = *Please check (✓) the grade level you teach* and item 87 = *Please estimate the percentage of students eligible for free or reduced price lunch in your school*). For *grade level taught*, the categories elementary, middle, and high school were developed and then subsumed under the variable *school type*. For *percentage of students eligible for free or reduced price lunch*, which was a proxy variable used to indicate the income-level of a given school locality, the categories low-, mid-, and high-income were generated. These three categories were subsumed under the variable *income level of school locality*. Both of the aggregated variables, *school type* and *income level of school locality*, were used to assess differences with teachers' views based on school type and income level of school locality.

Scale-level descriptive statistics. Eight composite variables were computed to represent the eight dimensions of the Teacher Survey. Consequently, descriptive statistics were estimated for these composites. These composites, their means, modes, and standard deviations are described in chapter four.

Analysis of variance. The sixth procedure of the quantitative data analysis involved statistical tests of significance. Using analysis of variance for each of the eight scales of the instrument, tests for differences between the means for *school type* and for *income-level of school locality* were conducted using a one-way analysis of variance. In the event of statistically significant results of an omnibus test, the Tukey *a* procedure was performed. Additionally, a Bonferroni adjustment of alpha was applied to the Tukey tests to reduce the chances of committing a Type I error.

Independent-samples t-test. For each of the eight scales of the Teacher Survey, an independent-samples *t*-test was conducted to test for differences between the means for *teachers' status on teaching a SOL-tested subject*. In conducting this test for each of the domains of the instrument, the results indicated whether the views of participants who teach SOL-tested subjects differ statistically from participants who teach subjects that are not tested by SOL examinations.

Qualitative Data Analysis

Initial analysis. The analysis first involved reading the data record in its entirety, beginning with the first participant's responses to questions one through three, the second participant's responses to questions one through three, and so on. General notes were written as broad overviews of the data. Next, the data was divided into three sections, comprised by all responses to *question one*, *question two*, and *question three*, respectively. The data record was read again, on this occasion section by section, and notes were written based upon each of the section readings.

Thematization. Subsequently, a third reading of the data was undertaken for the purpose of physically grouping similar ideas described within each section. These groupings were read and further analyzed and then were more finely grouped according to their content. The resulting groups, or themes, were coded (using numbers corresponding to content labels) and examined for sub-themes, which also were coded according to their content (Erickson, 1986; Fortune, 1999). The results from the qualitative data analysis served to underscore and elaborate upon the quantitative results, thus strengthen the validity, or dependability, of the study's results. Chapter four reports the findings of the study.

CHAPTER FOUR

Results

This chapter presents the results of the Teacher Survey: On the SOL School Reform. First the survey participants are described and an item-level analysis of the survey data is provided. Next a scale-level analysis of each of the eight domains of the Teacher Survey is discussed and inferential results are reported. Also the thematic trends in the qualitative data are described alongside each of the eight domains of the survey instrument. Instead of presenting the qualitative results separately, the decision was made to present the qualitative results based on the open-ended questions after each section of the quantitative results. The reason was to highlight the correspondence between the qualitative and quantitative findings. Finally a summary of the quantitative and qualitative parts of the survey is provided for each domain.

Participants

Teachers' Status on Teaching a SOL-Tested Subject

Nearly three-fourths of the teacher-participants of the study indicated their primary teaching assignment included a subject for which there is a SOL test (73%, $n = 260$) while 95 participants indicated they did not teach a SOL-tested subject. Five teachers did not respond to the *SOL-tested subject* item, thus the above values were based upon a sample size of 355.

School Type in Which Participants Teach

Of the practitioners responding to the *Teacher Survey*, 44% ($n = 154$) taught children in elementary school and comprised the majority of the sample. Thirty-five percent ($n = 123$) of the participants taught in high school while teachers of middle school children consisted of 22% ($n = 77$) of the sample. Six participants did not indicate the type of school in which they taught, therefore the values for *school type* were based upon $N = 354$.

Income-Level of Participants' School Locality

To describe the income-level of the school locality in which the teachers worked, low-, middle-, and high-income areas were designated by aggregating the teachers' estimates of the percentage of children in their school who receive free or reduced price lunch and resulted in the following categories: 56% or more, 21% to 55%, and 1% to 20% on free or reduced price lunch, respectively. Most of the classroom teachers of the study taught in schools located in middle-income areas (45%, $n = 135$) while 28% ($n = 84$) worked in high-income locales, and 27% ($n = 83$) taught children in low-income areas. Fifty-eight participants did not indicate the income level of the school locality in which they taught, therefore the percentages of *income level of school locality* were based upon a sample of 302 teachers.

Crosstabulations for Categories of Participants

To describe the participants in terms of their status on teaching a SOL-tested subject by the school type in which they teach, frequencies for these variables were crosstabulated and Table 1 displays these results. In comparison to their elementary school and high school

counterparts, and proportionate to the total number of teachers comprising each group, the middle school teachers represented the greatest percentage of those who teach SOL-tested subjects.

Table 1

Frequency of School Type by Teachers' Status on Teaching

a SOL-Tested Subject

School type	Status on teaching a SOL-tested subject		Total
	Yes	No	
Elementary school	116 (76.)	37 (24.2)	153
Middle school	65 (84.4)	12 (15.6)	77
High school	75 (62.5)	45 (12.9)	120
Total	256	94	350

Note. The values in parentheses are percentages based upon the row totals.

Summary

Three-fourths of the participants of the study taught subjects for which there is a SOL test ($n = 260$). Elementary school teachers and high school teachers together comprised 79% of the sample, with $n = 154$ and $n = 123$, respectively. Forty-five percent of the participants taught in schools in middle-income localities, with the remaining percent of participants approximately equally split between low-income and high-income school locales.

Although the middle school group accounted for the smallest number of participants within the *school type* variable, in proportion to their number the group comprised the largest

percentage of participants teaching a SOL-tested subject ($n = 77$, 84%). Seventy-six percent of the elementary school group ($n = 153$) and 62% of the high school group ($n = 120$) taught a SOL-tested subject.

Item-Level Analysis

Assessment of Data Normality

Skewness and kurtosis. The quantitative data from the *Teacher Survey* was examined to determine whether the data set approximated the shape of a normal distribution, and thus whether standard statistics could be used for further analyses (Huck, 2000, p. 35). In doing so, frequency distributions and the coefficients for skewness and kurtosis were inspected for each of the 80 continuous variables. Of these, nine distributions appeared to be both positively skewed and leptokurtic, with coefficients of skewness ranging from 1.53 to 2.55 and coefficients of kurtosis ranging from 2.02, representing a modestly peaked distribution, to 7.09, extremely peaked. In addition to these items, eight other variables displayed modest positive skewness, with coefficients of skewness ranging from 1.06 to 1.58. Finally platykurtic distributions were identified for two variables, with kurtosis coefficients of -1.28 and -1.37 . These out-of-range values (values for kurtosis that exceeded -2 to 2 and values for skewness that exceeded -1 to 1) occurred in expected directions described in prior research.

The out-of-range values underscored that the survey participants believed students' family background bore upon student SOL test performance (91%, $n = 329$) and that the SOL test instruments themselves were not adequate to measure students' knowledge and skills in a given subject (88%, $n = 326$). While the participants tended to believe that a 40-60 item SOL

test cannot adequately measure students' subject knowledge and skills, overwhelmingly the teachers indicated that as a result of State policy they felt pressured to place too much emphasis on their students' SOL test scores and felt considerable pressure to help their students improve SOL scores. Others have found similar results where teachers tend to believe that scores needed to pass a test are threats, or stressors, to teachers and sometimes students (Gordon & Reese, 1997; Hoffman et al., 2001; McNeil, 2000; Nolen et al., 1989; Shepard & Dougherty, 1991; Smith, 1991; Urdan & Paris, 1994).

Furthermore, the out-of-range values for skewness and kurtosis indicated that the participants' most pervasive disagreement centered upon three particular uses of SOL test scores. First, the teachers disagreed with the use of SOL scores as the primary criteria to determine what students have learned, and second, to determine whether students should graduate from high school. Third, the use of SOL test scores as the basis for school accountability purposes elicited nearly complete disagreement, with 98% ($n = 358$) of the teacher-participants contending such test-score use. Heubert and Hauser (1999) have documented trends similar to these three points. Finally, the out-of-range values suggested that teachers tended to view the classroom curriculum as being narrowed by the Standards while teachable moments are stifled as teachers rush to cover the Standards. This too has been described in relevant literature, therefore was not an unexpected response (Gordon & Reese, 1997; Haney, 2000; Hoffman et al., 2001; Jones et al., 1999; Kubow & DeBard, 2000; McNeil, 2000; Nolen et al., 1989; Shepard & Dougherty, 1991; Smith, 1991; Urdan & Paris, 1994; Wideen et al., 1997).

Conclusion. It was concluded that the out-of-range values for skewness and kurtosis suggested teacher beliefs about high-stakes accountability that have been previously reported

in the relevant literature. Therefore, the data from the Teacher Survey was not grossly non-normal but reflected expected trends, thus standard statistics could be used for further analyses (Huck, 2000, p. 35).

In addition to the participants' strong opinions that were indicated by the out-of-range values for skewness and kurtosis and the high- and low-mean items, the remaining items of the questionnaire also elicited strong responses with modest variability. The standard deviations ranged from .79 to 1.81 (item 52 = *For school accountability purposes, my school's quality should be based on SOL test scores alone* and item 60 = *The SOL testing system has caused me to consider leaving the teaching profession*, respectively.). Table 9 presents the descriptive statistics for the 80 Likert-type items (1 = *strongly disagree*, 6 = *strongly agree*) in the sequence in which they appeared on the Teacher Survey, although in abbreviated form. Each item can be reviewed in its entirety in Appendix B. Further, Table 9 displays a column for *percent of agreement*, which represents the cumulative percent of the teachers' agreement with each of the items, which is comprised by *tend to agree* + *agree* + *strongly agree*. It should be noted that the responses to the negatively worded items were reverse coded so that the responses could be interpreted in the same manner as the favorably worded items. Accordingly, the recoded values are provided within parentheses in the table.

Missing Values

Seven variables (items 24, 25, 26, 33, 35, 37, and 43) had relatively high numbers of missing values (46 to 57 missing responses), which can be noted in Table 7 under the column for *n*. These items were examined and divided into three areas of content. First, item 24 queried to what extent the participants agreed that the quality of their teaching is reflected in

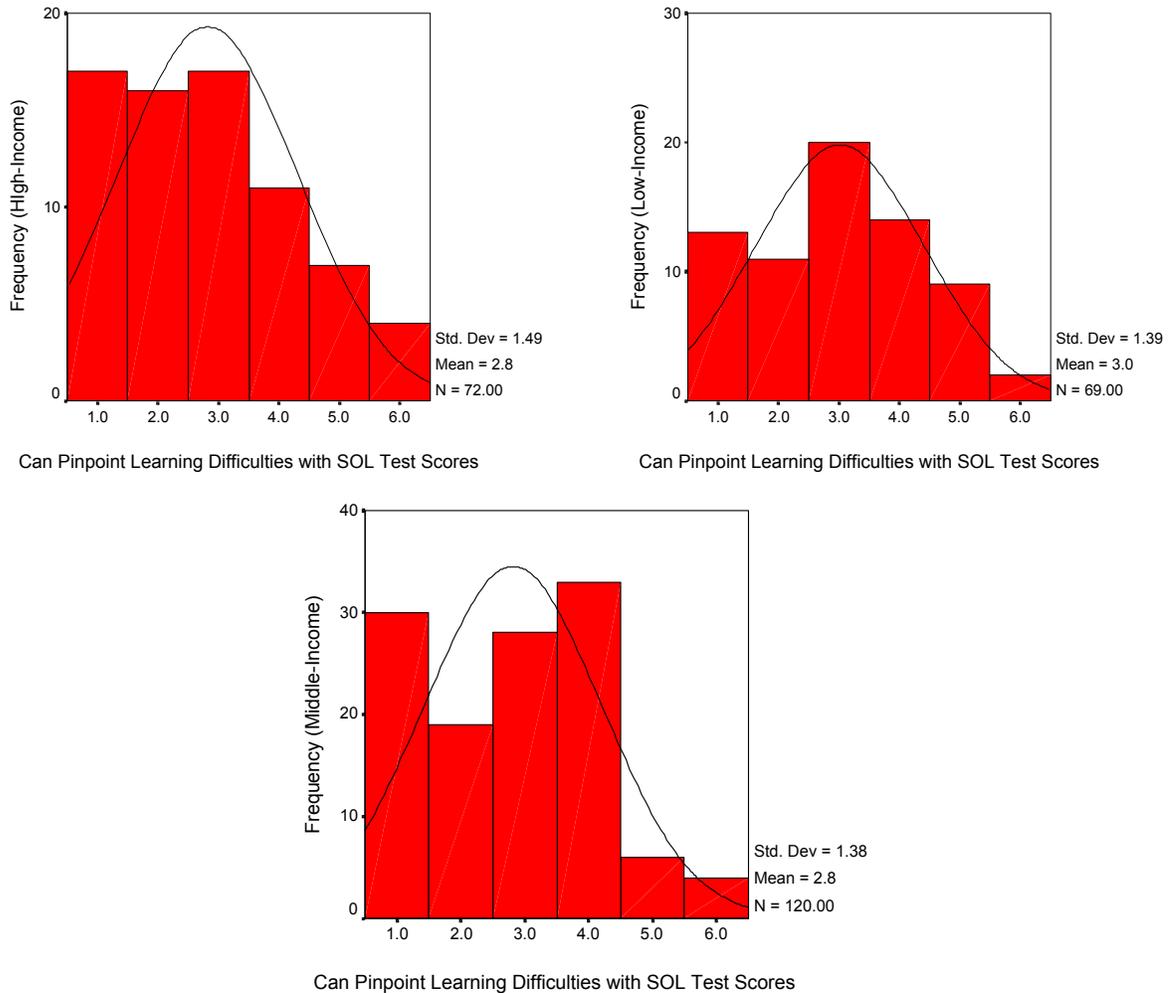
their SOL test scores. The second area of interest included items 37 and 43 that queried teachers' views on the reasonableness of SOL cut-scores (minimum passing scores), a more technical element of the SOL system that teachers may have viewed lie outside their sphere of expertise. The third area, however, addressed the diagnostic value of SOL test scores, a topic with which most teachers are familiar. Why larger numbers of teachers did not respond to items related to this topic was perplexing.

To probe possible explanations for the lower response items, the response-frequencies for each of these items was compared to the numbers of teachers comprising various groups of the following categories: status on teaching a SOL-tested subject, school type, and income level of school locality. Having examined the three categorical variables first listed, there appeared to be no pattern by which the items with lower response could be explained. For each of the items, the frequencies of lower response appeared to be similar across all groups of the categorical variables, including teachers who taught SOL-tested subjects as well as those who did not.

Although the response frequencies for the lower-response items presented no patterns based on teacher categories, an examination of the distributional shapes for these items indicated some unique features. The items comprising the domain related to the diagnostic value of SOL test scores had larger numbers of missing values than the other domains of the Teacher Survey. Two items will be discussed. For example item 25 asked the participants to what extent they agree that they can pinpoint students' learning difficulties by reviewing the students' SOL test-scores. The distribution of responses for middle school teachers ($M = 2.9$, $SD = 1.27$) and high school teachers ($M = 2.8$, $SD = 1.35$) both appeared somewhat platykurtic with slight positive skewness. However, elementary school teachers indicated a

bimodal response primarily clustered around *strongly disagree* and *tend to disagree* ($M = 2.7$, $SD = 1.48$). When the item-distributions for this same variable were examined based on the income-level of school locality, the teachers in low- and high-income areas showed a unimodal response of disagreement with the pinpointing usefulness of SOL test scores (with means of 3.0 and 2.8 and standard deviations of 1.39 and 1.49, respectively), but teachers in middle-income areas indicated a clear bimodal response that tended toward disagreement ($M = 2.8$, $SD = 1.38$). Figure 1 displays the item-distributions for the *pinpointing* variable by the income-level of school locality.

Figure 1. Response frequency for usefulness of SOL test scores for pinpointing students' learning difficulties by income level of school locality (*high-, middle-, and low-income*).



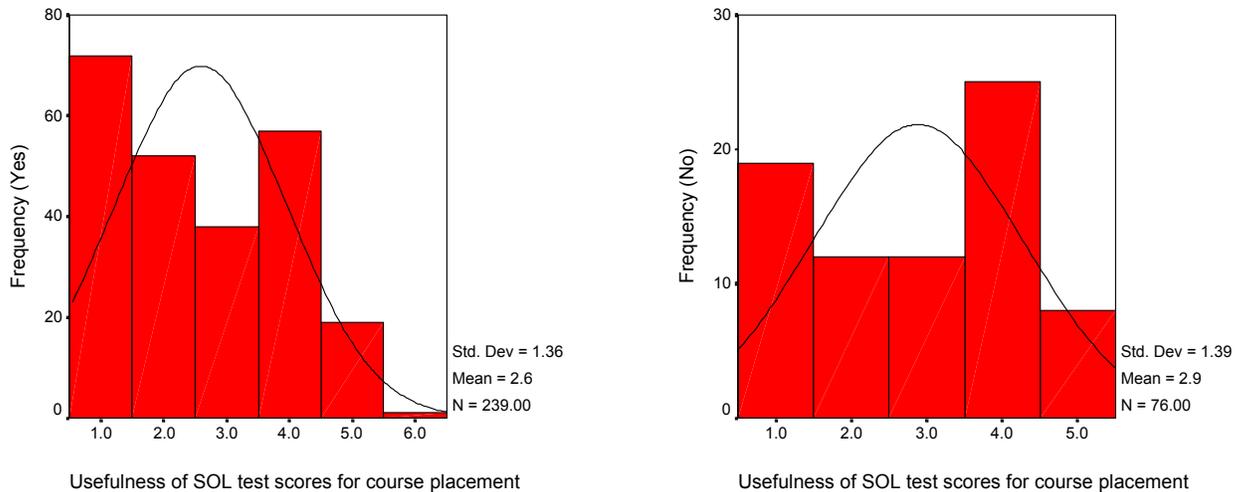
Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

Additionally, for *usefulness of SOL test scores for pinpointing students' learning difficulties*, the response pattern appeared different for the group who teaches a SOL-tested subject and for those who do not. For those who teach a SOL-tested subject, the response pattern appeared somewhat platykurtic and slightly positively skewed while those who did

not teach a SOL-tested subject showed a strong peaked response that clustered around *tend to disagree* and *tend to agree*. Thus, the teachers who did not teach SOL-tested subjects indicated more agreement, although small, with the pinpointing usefulness of SOL test scores than did teachers who taught a SOL-tested subject.

A second example, item 35, asked teachers to what extent they agree with the usefulness of SOL test scores to help determine course placement. The responses to this item presented disparate response patterns based upon whether the participants teach a SOL-tested subject. Figure 2 shows that teachers of SOL-tested subjects indicated disagreement with the usefulness of SOL test scores in determining students' course placement, with some dissent indicating a tendency to agree. However, the teachers who do not teach a SOL-tested subject appeared divided on the issue. Their responses clustered around *strongly disagree* and *tend to agree*.

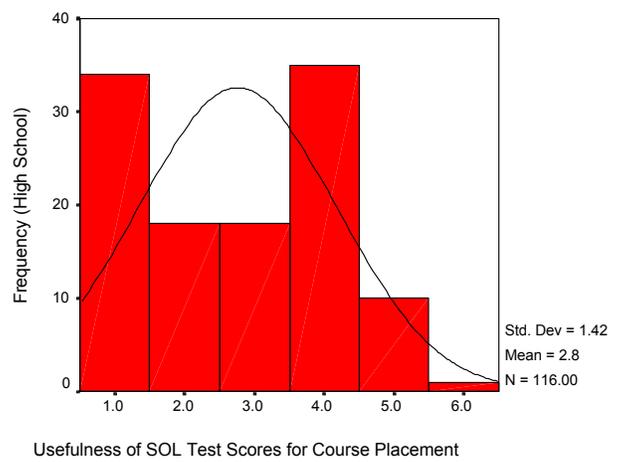
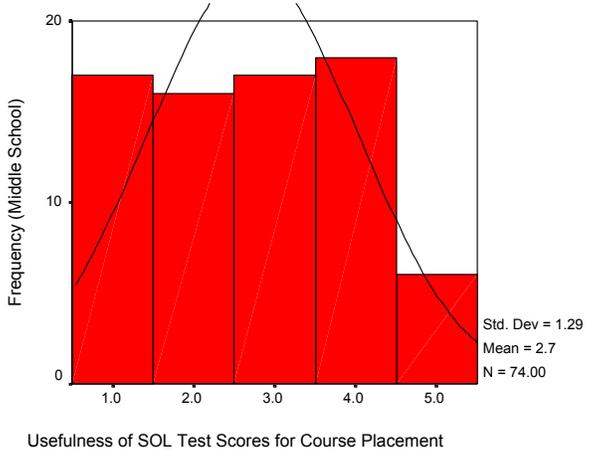
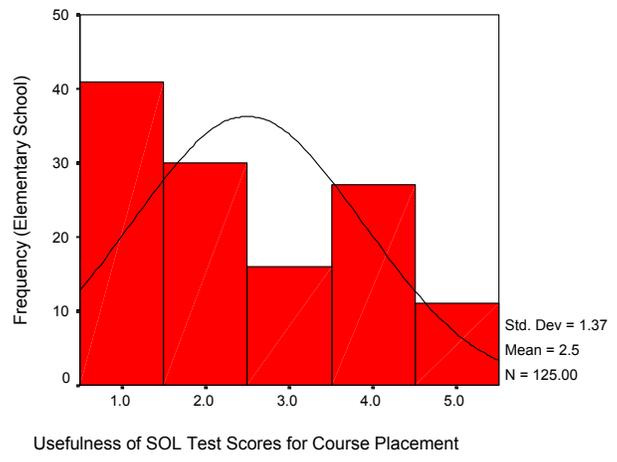
Figure 2. Response frequency for usefulness of SOL test scores for determining students' course placement by teachers' status on teaching a SOL-tested subject (*yes, no*).



Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

Furthermore the participants' responses to the *course placement* item presented different response patterns based upon *school type*. For the teachers working in middle school, the distribution appeared largely leptokurtic across four response-options ranging from *strongly disagree* to *tend to agree* ($M = 2.7$, $SD = 1.29$) whereas elementary school teachers indicated a positively skewed, somewhat bimodal response ($M = 2.5$, $SD = 1.37$), with responses clustering around *disagree* and *tend to agree*. For the high school teachers, the responses distinctly evidenced bimodality, with clustering around *strongly disagree* and *tend to agree* ($M = 2.8$, $SD = 1.42$). Thus, inspection of the distributions for each teacher-group yielded several distinct patterns, patterns that were not distinguished given the item means. Figure 3 displays the frequency distributions for *course placement* based on *school type*.

Figure 3. Response frequency for usefulness of SOL test scores for determining students' course placement by school type (*elementary, middle, or high school*).



Note. Responses were based upon Likert-type scaling (1 = *strongly disagree*, 6 = *strongly agree*).

The lower-response items not only seemed to elicit varied patterns of response given the categories of teachers, but elicited rather distinct bimodal responses across some of the groups of teachers within those categories. It seems reasonable to suggest, then, that the lower-response items may have presented a more difficult challenge to the participants as they sought to decide upon a response-option.

Scale-Level Analysis

Mean Differences Based on Income-Level of School Locality, School Type, and Status on Teaching a SOL-Tested Subject

Income-level of school locality. To test the hypothesis that the mean score on the dependent variable for each of the eight domains is the same in each of three teacher populations represented by *income-level of school locality* (low-, middle-, and high-income), a one-way analysis of variance was performed. Consequently, the omnibus F -test was not statistically significant for the eight scales ($\alpha = .05$), which suggested that there are no differences between the means of the three populations of teachers in low-, middle-, and high-income school locales. Moreover, Levene's test of homogeneity of variance ($\alpha = .05$) was employed to test the hypothesis of equal variances for the three populations of teachers for each of the eight domains and the hypothesis was supported, which suggested the variance for the three populations of teachers was homogeneous.

School type. Also conducted for the eight scales of the Teacher Survey was the omnibus test of the main effect of *school type* (elementary, middle, and high school), which was not statistically significant at the alpha level of .05 for each of the domains except for ADTEST (Adequacy of SOL Testing and Test Scores to Hold Schools Accountable), $F(2,$

351) = 3.89, $p = .021$. Aside from ADTEST, the results indicated that the teachers of the study, regardless of whether they teach in elementary, middle, or high school, held similar views on Virginia's SOL program. Further, for each of the scales with the exception of CUTSCR (Adequacy of SOL Cut-Scores to Signify Student Learning) ($p = .002$, $\alpha = .05$), these data did not violate the assumption of equal variances as suggested by Levene's test of homogeneity of variance ($\alpha = .05$). To address the unequal variances of CUTSCR, the harmonic mean was used, thus the Type I error level was not guaranteed. This uncertainty did not seem problematic given that the omnibus test of the main effect of *school type* did not yield a statistically significant result.

Status on teaching an SOL-tested subject. Finally, to test for mean differences concerning the participants' *status on teaching a SOL-tested subject* (yes, no), an independent-samples *t*-test was performed. For each of the survey scales with the exceptions of TPROF (Support for Teacher Professional Autonomy Within the SOL Reform) and TTEN (Level of Teacher Tension Within the SOL Reform), the results of this test at the alpha level of .05 indicated no statistically significant difference between the means for the population of VEA teachers who teach an SOL-tested subject and for the population who do not.

Further, to test the hypothesis that these data do not violate the assumption of equal variances, Levene's test of homogeneity of variance was conducted. For each of the scales, except TPROF and TTEN, Levene's test did not reach statistical significance at the alpha level of .05, thus supported the null hypothesis that the variances were equal.

Taken together, the inferential results suggested that for most of the survey scales the teachers' views on the SOL program are consistent across all groups within the three teacher-

categories explored. Those scales that indicated difference of opinion within the categories are described under the headings for the domains of the survey.

Adequacy of the Standards of Learning (ADSOL)

Scale Description and Statistics

As discussed in chapter three, the ADSOL scale (Cronbach's alpha = .92) queried the teachers' views on the reasonableness of the SOLs as guidelines for teaching and learning and the adequacy of the Standards to provide a high-quality, equitable education for all students. Table 2 presents the scale items, item frequencies, and descriptive statistics for ADSOL.

Table 2

*Item Frequencies and Descriptive Statistics for Adequacy of the
Standards of Learning*

Item	Frequency of teacher response						<i>n</i>
	SD	D	TD	TA	A	SA	
1. (1) SOLs represent reasonable guidelines for teaching	33 (9.3)	49 (13.8)	75 (21.1)	114 (32.1)	61 (17.2)	23 (6.5)	355
2. (2) SOLs represent reasonable guidelines for learning	39 (10.9)	56 (15.7)	74 (20.7)	114 (31.9)	52 (14.6)	22 (6.2)	357
3. (6) SOLs specify most important skills for learning	36 (10.2)	67 (18.9)	94 (26.6)	103 (29.1)	39 (11.0)	15 (4.2)	354
4. (7) Alignment to the SOLs improves quality of education	53 (15.2)	55 (15.8)	100 (28.7)	77 (22.1)	48 (13.8)	16 (4.6)	349
5. (10) Alignment to the SOLs improves quality of teaching	76 (21.8)	65 (18.6)	100 (28.7)	65 (18.6)	36 (10.3)	7 (2.0)	349
6. (12) SOLs specify what students must know to be competitive in a global economy	50 (14.7)	67 (19.6)	109 (32.0)	72 (21.1)	29 (8.5)	14 (4.1)	341
7. (13) SOLs help students gain a world-class education	59 (17.2)	70 (20.3)	126 (36.6)	61 (17.7)	23 (6.7)	5 (1.5)	344
8. (14) SOLs help ensure students receive an equitable education	28 (8.2)	54 (15.7)	96 (28.0)	100 (29.2)	52 (15.2)	13 (3.8)	343

 Descriptive statistics for scale

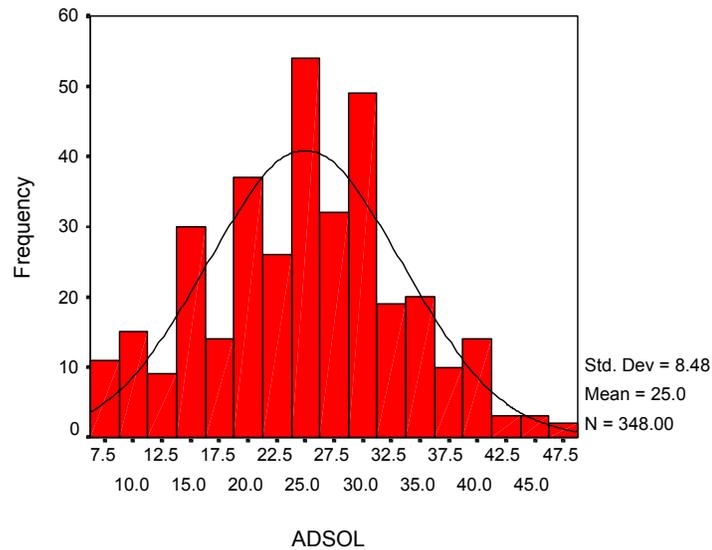
	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
Adequacy of the Standards of Learning	3.18	1.32	.92	8

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items.

Teacher Response to ADSOL

For ADSOL, the participants' responses to the scale items were normally distributed, as can be observed in Figure 4 ($M = 3.18$, $SD = 1.32$), with two of the 348 teachers indicating extreme responses.

Figure 4. Teachers' Response Frequency for ADSOL.



Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

The item frequencies and percents of agreement indicated that the teachers tended to disagree with most of the items (*3 = tend to disagree* on the 6-point Likert-type scale of the Teacher Survey), thus suggesting that the teachers believed the Standards lacked adequacy to accomplish their intended goals. The ADSOL scale was dichotomized so that the teachers' cumulative percent of agreement and the cumulative percent of disagreement could be examined. The results showed that when the teachers considered the Standards of Learning in terms of what they were originally intended to provide students—a world class education—a modest 26% of the participants believed the SOLs help accomplish this aim ($n = 344$). Furthermore, 34% viewed the SOLs as helpful in specifying what students must know to be competitive in a global economy ($n = 341$).

However when considering the items of ADSOL that addressed the equity and the reasonableness of the Standards, in terms of representing the content that teachers should

teach and that students should learn, the participants were somewhat divided. Fifty-six percent ($n = 355$) agreed that the SOLs represented reasonable *guidelines* for what teachers should teach; and twelve percent fewer (44%, $n = 354$) believed that the SOLs were successful in *specifying* the most important skills and knowledge that should be taught in a given subject area. Still slightly less of the participants (41%, $n = 349$) believed that *aligning* the classroom curriculum to the SOLs improves the quality of one's teaching. Furthermore, a little less than half (48%, $n = 343$) of the participants agreed that the Standards help ensure students receive an equitable education.

A trend was noted in the above percentages, suggesting the teachers' agreement with the reasonableness of the Standards decreased with each item that presented the Standards as having increased governing power over what the participants teach and the students learn. Furthermore, of all the scales of the Teacher Survey, ADSOL elicited the most equally divided responses between the teachers' agreement and disagreement. This seemed to suggest that the VEA teacher-members were somewhat uncertain as to the reasonableness of the content of the Standards, the reasonableness of the Standards given varying degrees of classroom influence, and the provision of equitable educational opportunities for students.

Qualitative Results

As indicated by the quantitative results of ADSOL, the teacher-participants tended to disagree that the Standards represented reasonable content for what teachers should teach and for what students should learn. The teacher narratives helped underscore the nature of this trend. Across the three categories of teachers, the participants contended that the SOLs represent enrichment rather than minimum competency objectives. As such, the elementary

teachers described this view in terms of the Standards lacking developmental or age-appropriateness while middle and high school teachers discussed the same view in terms of the Standards representing subject matter that is suitable for college bound students.

Difficulty of the Standards. “We are beating our heads against a wall trying to teach young children before they are mature enough to learn some concepts” wrote a third grade teacher in a middle-income locality. Although this perspective persisted throughout the teacher narratives, the teachers also expressed that the Standards were helpful in providing a unified curriculum across the State. “There is nothing wrong with centralized standards” observed a fifth grade teacher in a high-income locale; “they best serve as a safeguard against local extremes and as a source of welcome consistency.”

While the teachers tended to favor the idea that the Standards might help provide a more equitable curriculum, they tended to disapprove of the Standards’ high difficulty. On this point, a second grade teacher of history, math, and science explained

As a 2nd grade teacher in the inner city, I do not have a problem with the concept of the SOL Standards. I believe that all students, no matter what level, should have objectives or standards to learn. However, I feel that these Standards should be appropriate for the chronological and developmental age of the students... I am required to teach concepts that I didn’t learn until middle school, such as investments, bonds, checking/savings (economics) and estimation, dormancy and adaptation (science).

Narrowed scope of education. In addition to this, many middle school and high school teachers contended that the Standards are too difficult and supplant a diversity of academic offerings, including vocational education programs that might better suit students with interests or needs outside the college bound track. A high school teacher noted

As a math teacher, I believe we have developed a college prep math curriculum for all students because of the SOLs. We need to realize that many students do not go to college and have other needs and interests. We need to reinstate courses such as consumer math, technical math and business math for students with those interests and allow them to count as SOL courses. If we do that, we might begin to design a curriculum that meets the needs of all VA students and the SOLs might have some value.

A high school English teacher asserted similar views as those expressed above:

SOLs have narrowed the scope of education at a time when systems should instead be able to diversify and create courses/programs to transition all students successfully into the “world of work” or college (small percentage). Schools need to be released from the stress of everyone passing the SOL tests and allowed to offer a curriculum suited to a particular group of students (vocational, business, marketing/sales, college bound, etc.). In this way, each student graduates with marketable skills and that is the real meaning of a school’s success. If the only thing a student can say is, “I

passed six end-of-course tests,” what have we given her/him? What have we equipped the students with who don’t pass—failure, the welfare line?? What do you have in mind for them?

Moreover, many teachers expressed concern over students who, given the difficulty of the Standards and the lack of alternative course offerings, will likely be denied a high school diploma. In light of this concern, one high school teacher recommended that the State Department of Education examine the number, rather than percent, of sophomores in each county who have not passed at least three SOL tests. “This prognosticator showing the number of students at risk of not earning a diploma will, some teachers feel, be a valuable eye-opener that may save the Department of Education and the state of Virginia an embarrassment and multiple civil suits in 2004.” Representing a salient pattern in the teacher narratives, the same high school teacher contended that any student who earns the Carnegie units required to graduate deserves a diploma and not a certificate of attendance.

It seemed unreasonable to the participants to track all students along a similar college bound path, where the subject matter seemed too difficult and diverse educational opportunities seemed to be devalued. A sixth and seventh grade history teacher in a school where 80% of the students receive free or reduced price lunch explained, “Why throw out or devalue vocational training, as ALL students LEARN but all students do not excel in all subjects.” The same teacher contended that the variety of program or course availability does not meet the diversity of student needs and thus questioned, “What happens to the plumber, auto mechanic, etc., that our society desperately needs TOO?”

Students with special education needs. For students with learning difficulties or those needing special education, the difficulty level of the SOLs seemed to present to the teachers an even more complex problem. Special education teachers also tended to concur that the Standards offer very little to these students. “Students with learning problems are being left out of the equation,” wrote a high school science teacher, and an eighth grade teacher of learning disabled students offered the following characteristic response:

We are going to lose a generation of young people who know they’ll never pass enough SOLs to graduate. I have heard young teenaged kids say that they plan to drop out at 16 if they can’t get a “real” diploma. These are kids who struggle with academics, are not college bound, but want to be good workers and good citizens. If all the energy, time, and money had gone into developing excellent vocational programs, effective alternative schools, and community outreach to the children of its families—well, schools could then *provide* for every child, *prepare* every child—however troubled, however gifted—and give teachers the opportunity to do their jobs. Teach every child. NOT Test every child.

Summary

The teachers’ responses to the open-ended questions underpinned a prominent trend found in the scale-level results. Both data sets indicated the classroom practitioners tended to disagree that the Standards represent reasonable content for teaching and learning. As found from the qualitative data analysis, this seemed in large part due to the difficulty level of the

SOLs. Teachers asserted that the Standards tended to be unreasonable in terms of their developmental and age-appropriateness or in terms of their precedence over opportunities for less difficult coursework and vocational education availability for non-college bound learners.

Also explicated in the teacher text was the general concern for students with learning difficulties and those in need of special education. The participants tended to believe that coursework that focused upon the Standards was overly difficult for these students and that some seemed likely to drop out of high school once they reached 16 years of age. However given the teachers' view that the SOLs seemed too difficult for many students, roughly half of the teachers indicated on the ADSOL that the Standards were reasonable in that they helped provide a more equitable education for students. Although the teachers did not write about equity per se, some used language that supported the SOLs' power to provide a "unified curriculum" across the State. In so doing, approximately half of their narratives seemed to sustain the view that the Standards help ensure an equitable education for all students.

Ability of SOL Program to Meet Diverse Student Needs (DIVERSE)

Scale Description and Statistics

The DIVERSE (Cronbach's $\alpha = .84$) domain of the Teacher Survey addressed the teachers' perspectives on how the SOL program affects the classroom curriculum in terms of the breadth of the Standards, the SOL system's influence upon teachers' use of diverse instructional strategies, and the system's effect upon students' ability to master difficult

subject matter. Table 3 displays the item content, item frequencies, and descriptive statistics for DIVERSE.

Table 3

*Item Frequencies and Descriptive Statistics for Ability of the SOL Program
to Meet Diverse Student Needs*

Item	Frequency of teacher response						<i>n</i>
	SD	D	TD	TA	A	SA	
1. (3) Classroom curriculum is narrowed by the Standards	14 (3.9)	12 (3.4)	20 (5.6)	61 (17.1)	87 (24.4)	162 (45.5)	356
2. (8) Attempting to teach all the SOLs stifles opportunities for teachable moments	14 (4.0)	22 (6.2)	29 (8.2)	52 (14.7)	80 (22.7)	156 (44.2)	353
3. (11) Alignment to SOLs limits diverse instructional strategies	16 (4.6)	30 (8.6)	59 (16.9)	58 (16.6)	85 (24.4)	100 (28.7)	349
4. (16) SOLs are too extensive to be taught adequately in a given school year	8 (2.3)	31 (9.0)	53 (15.4)	80 (23.2)	74 (21.4)	99 (28.7)	345
5. (23) Covering SOLs has restricted use of diverse instructional strategies	13 (3.8)	30 (8.8)	53 (15.6)	65 (19.1)	81 (23.8)	98 (28.8)	340
6. (70) With the SOL system, students are required to think critically about subject	80 (23.6)	60 (17.7)	77 (22.7)	75 (22.1)	38 (11.2)	9 (2.7)	339
7. (71) Not enough time in school day to teach all SOLs	5 (1.5)	30 (9.1)	32 (9.7)	66 (19.9)	76 (23.0)	122 (36.9)	331

Item	Frequency of teacher response						<i>n</i>
	SD	D	TD	TA	A	SA	
8. (72) Students have enough time to master difficult concepts in the Standards	115 (33.2)	88 (25.4)	89 (25.7)	27 (7.8)	21 (6.1)	6 (1.7)	346
9. (73) Teaching is rushed because I try to cover SOLs	8 (2.4)	11 (3.3)	29 (8.6)	66 (19.6)	87 (25.9)	135 (40.2)	336
10. (74) With the SOL system, teaching is focused on basics over higher-order thinking	12 (3.6)	12 (3.6)	42 (12.5)	79 (23.4)	93 (27.6)	99 (29.4)	337
11. (80) Attempting to cover SOLs prevents me from providing opportunities for students to think critically about the subject matter	11 (3.2)	17 (5.0)	53 (15.6)	99 (29.2)	66 (19.5)	93 (27.4)	339

Descriptive statistics for scale

	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
Ability of the SOL Program to Meet				
Diverse Student Needs	2.45	1.37	.84	11

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree

[D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the

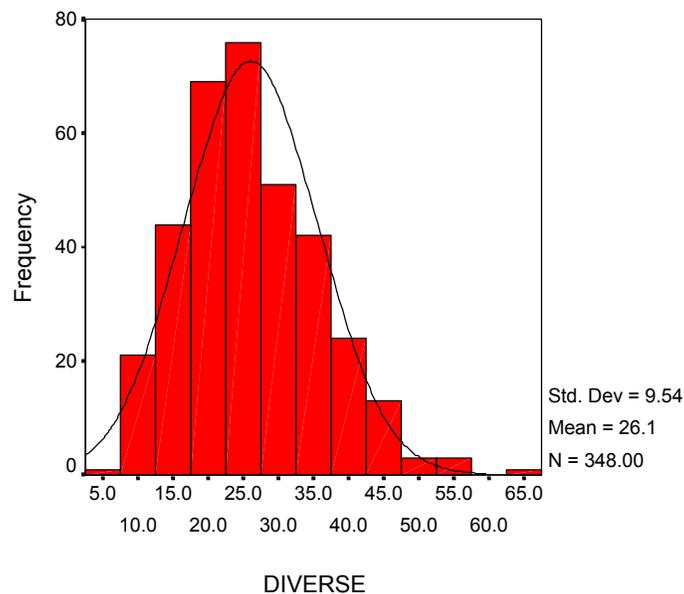
Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items.

Items 3, 8, 11, 16, 23, 71, 73, 74, 80 were recoded prior to computing the descriptive statistics for the scale.

Teacher Response to DIVERSE

The participants' responses to the scale appeared leptokurtic and slightly positively skewed, with seven participants showing extreme responses. Figure 5 displays these results.

Figure 5. Teachers' Response Frequency for DIVERSE.



Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

The descriptive statistics for the 11 items comprising DIVERSE indicated that the teachers tended to agree with most of the items of the scale (*4 = tend to agree* on the 6-point Likert-type scale of the instrument), which suggested the SOL system lacks the ability to meet the diverse intellectual needs of students. The DIVERSE scale was dichotomized so that the teachers' cumulative percent of agreement and the cumulative percent of

disagreement could be examined. The results showed that the participants agreed that the SOLs were too extensive to be taught adequately in a given school year. Almost three-fourths of the participants (73%, $n = 345$) believed that the Standards were too broad. Given this view, 80% ($n = 331$) of the teachers of the study concurred there is not enough time in the school day to teach all the Standards and 86% ($n = 335$) believed that their teaching is rushed from trying to cover the SOLs. Accordingly, 84% ($n = 346$) agreed that students do not have enough time to master difficult concepts in the Standards.

In attempting to teach all the skills and knowledge specified by the Standards, the participants indicated that the opportunities for spontaneous “teachable moments” are stifled (82%, $n = 353$) and that the teachers are prevented from providing opportunities for students to think critically about the subject matter (76%, $n = 339$). Participants reported they must focus their teaching on basics rather than higher-order thinking (80%, $n = 337$).

In short, almost ninety percent of the classroom teachers (87%, $n = 356$) believed the classroom curriculum was narrowed, or limited, in their attempt to cover the Standards. In keeping with this trend, the impact on teachers’ use of diverse instructional strategies was made clear. Teachers reported they must restrict the use of diverse instructional strategies in order to cover the Standards (72%, $n = 340$). Finally, a rather small percentage of the survey participants (36%, $n = 339$) agreed that the SOL system requires students to think critically about the subject matter studied in the classroom.

Qualitative Results

The hurried classroom. The most salient theme evidenced in the teachers’ responses to the open-ended questions was that the breadth of the Standards is entirely too great, hence

creating a plethora of problems at the classroom level. Trying to teach an over-extensive body of content seemed to compress and expedite the classroom activities in such a manner that teachers across all categories wrote at length on the matter. “In primary grades,” wrote a third grade teacher of history, math, reading, and science, “the emphasis should be math [and] language arts ONLY. We leap from one topic to the next trying to complete the required elements [of the SOLs].” “There is little time to study any subject in depth,” added a first grade history and science teacher.

The fast pace of teaching that the teachers believed to be required of them, presented negative outcomes not only for most learners whom teachers believed were denied opportunities to study a topic in depth, but presented greater negative consequences for students with learning difficulties. “Slow learners are left behind—pace is too fast!” wrote an eighth grade science teacher. Teachers across *school type* described slower learners, those working below grade level, and students requiring special education as those who are most adversely affected by the speed of classroom activities, being without adequate time to understand the subject matter.

Curriculum considerations. The breadth of the SOLs and the pace of instruction appeared to be a primary concern for most of the teacher-participants, and on this point a second grade teacher offered the following observations:

A decision also needs to be made as to whether emphasis should be on the basic subjects of reading, writing, [and] math with a small emphasis on science and Soc. Studies in grades K-2, or should we continue with the SOLs and teach too much information to young children who need a focus

on reading, writing, [and] math? My school system requires that we teach an hour of math, science, [and] Soc. St. daily. I believe those standards should be reduced in content to allow more time for reading.

As might not be expected, some high school teachers posited a similar view to that as elementary teachers in terms of how the breadth of the SOLs affects the elementary school classroom. A high school math teacher in a middle-income school locality wrote

Some of the elements of the [SOL] objectives need to be reevaluated. Do elementary schools really need to cover as broad an area of topics as they do because it takes away from the time needed in learning basic skills—reading—writing—and math. If we focus more on building strong basic skills [and] leaving some of the broader topics for later years we might build stronger students that are more successful. We currently are requiring too much in broad areas from our early years in schools.

Some high school teachers coupled this view with their strong opinions about the difficulty some high school students encounter when trying to read about complex subject matter. The teachers explained that some of their students read on a third, fourth, or fifth grade level. Consequently, the teachers favored more time spent on reading and writing during the early grades, which teachers believed might eschew students' reading comprehension problems in later grades. Mathematics too, some participants at the high school level contended, need attention for longer periods of time in primary grades, thus requiring the SOLs for the early

grades to be reduced in sheer volume (and lessened in difficulty). As a result these teachers asserted the curriculum should present a narrower subject focus—including largely reading and mathematics.

Also, middle school teachers tended to believe the Standards were much too extensive. “Local history and geography of the communities within Virginia are sacrificed for Mesopotamia, ancient Egypt, or Rome (as examples). Less content should be taught in grades K to 8 with more emphasis on math concepts, reading skills, and critical thinking (seventh and eighth grade science teacher in a middle-income school locale). Although the teacher narratives suggested that the participants believed the Standards to be over-extensive, they also believed the SOLs lacked information as to what more or less should be emphasized given the Standards’ breadth. A veteran teacher of 23 years exemplified this point:

If we are going to continue to test students, then the SOLs need to be more specific. For example: One SOL in earth science states: “The student should understand geologic processes including plate tectonics. Key concepts include—faulting, folding, volcanism, metamorphism, weathering, erosion, deposition [and] sedimentation [and] their resulting features and -----.” Earthquakes [as a concept] is not listed but I’m sure is understood by E.Sci teachers. In most textbooks just this part is 8-9 chapters. A vocabulary list would help us, as teachers, to know how much detail to go into. The old [standards] from 20 yrs. ago did this [and] that is still how I teach today.

Newer teachers may not be going into enough detail [and] just skim the surface.

Covering the Standards and teaching to the test. Without sufficient information as to what should be more or less emphasized in the Standards, the teachers articulated their efforts to *cover* the SOL material, thereby being compelled to narrow the content of the classroom to that of the Standards, and doing so at an intense pace that generally is not amenable to students critically thinking about the subject matter or to the study of subject interests that lie outside the scope of the Standards (e.g., current events).

From the teacher narratives, a distinct pattern was found for the participants to limit, or narrow, the classroom curriculum to that which might be tested on a SOL test. A third grade teacher in a low-income school locality expressed, “I don’t have time to work on analytical skills and higher order thinking skills because [my students] have to be able to regurgitate facts—to pass the SOLs and keep my school accredited. There is only so much I can do,” Teachers across *income-level of school locality* wrote of similar experiences; however elementary school teachers tended to be most prolific in describing such classroom experiences. Further, in comparison to middle- and high-income areas, the negative impact of a narrowed classroom curriculum, at least for some teachers, appeared to be exacerbated in low-income area schools. “In the at-risk schools a bad situation [is] made much worse by the SOLs. We skip all of the important holistic teaching in the lower grades to teach isolated SOL skills and leave the students with large gaps in their education” (a third grade teacher in a school where 92% of the students receive free or reduced price lunch).

“More and more, teachers are turning toward ‘teaching to the test’ so that their students score well,” offered an eighth grade reading and music teacher in a high-income area. On this point, some teachers used strong and emphatic language, as exemplified by the following middle school teacher’s assertion: “Students are bombarded then prepped for SOL tests. Much of this content is then forgotten soon after the testing, especially students who are regarded as average or below average.” The same teacher stated that students are prepped for SOL tests at the expense of their opportunities to think critically about the subject matter. “Critical thinking...[is] bypassed so that content is drilled into the students. Rote learning is stressed with the SOL reform.” Such a narrow focus, the teachers contended, restricts the use of diverse instructional strategies. A first grade teacher observed “Isolated facts and [drill-oriented] learning has taken the place of sequential learning and critical thinking in the classroom.” In sum, an overwhelming majority of the teachers expressed that their students were learning many facts that lacked connections to a broader context.

Summary

The thematic trends in the teacher narratives were similar to the quantitative results of DIVERSE and underscored the pattern that the participants tended to rush their teaching (items 71 ad 73) in an effort to cover over-extensive Standards (item 16). Furthermore, the focus upon the SOLs narrowed the classroom curriculum (item 3), often compressing it to that which would be tested on a SOL test. Teachers found that teachable moments tended to be stifled (item 8) and the use of diverse instructional strategies was reduced (items 11 and 23). Consequently, as the teacher narratives suggested, the abovementioned factors seemed to

coalesce and negatively affect the opportunities for students to think critically about the subject matter (items 70, 74, and 80).

The adverse influence upon learners' opportunities to study subject matter in depth and to critically examine it points up one of the most consequential undesirable effects of the SOL reform described by the teacher-participants and calls into question the quality of Virginia public schools. Although the example is lengthy, a high school English teacher in a high-income locale captured the crux of the participants' concern, and with it this section is closed:

Last week I asked a class of 20 juniors *where* the Pilgrims landed. Silence! I prompted, thinking they misunderstood the question. More silence. Finally one said, "Florida." Another offered, "Mexico." I was stunned. We are on Block scheduling and at least half the class took VA/US History last semester. Our scores were in the 80-96 percentile. Their answer when I pointed this out: "We had to know *when*. We didn't have to know *where*!" I've taught for 25 years, and I still hold to the seemingly out-of-date notion that the priority of public schools should be to teach children *how* to *think* so they can become confident, life-long, independent learners. We are concentrating so hard on the "trees" that our kids have no concept of the "forest."

The same teacher continued

By the way, my students have been scoring in the 88-94 percentile, so my complaints do not come out of frustration over low scores. *However*, this system is *so* flawed that I do not allow anyone to suggest that these scores reflect my teaching effectiveness any more than I would accept their criticism of my teaching ability if my scores were low.

Appropriateness of SOL Cut-Scores to Signify Student Subject Mastery (CUTSCR)

Scale Description and Statistics

The CUTSCR scale (Cronbach's alpha = .77) describes the participants' views on the appropriateness of the minimum score required for students to pass a given SOL test. Table 4 presents the item content, item frequencies, and descriptive statistics for CUTSCR.

Table 4

*Item Frequencies and Descriptive Statistics for the Appropriateness of
SOL Cut-Scores to Signify Student Subject Mastery*

Item	Frequency of teacher response						<i>n</i>
	SD	D	TD	TA	A	SA	
1. (37) Cut-scores are set at appropriate levels	54 (17.1)	60 (19.0)	97 (30.8)	82 (26.0)	19 (6.0)	3 (1.0)	315
2. (39) Pass score of a SOL test clearly distinguishes students who have acceptable level of knowledge and skills from those who do not	74 (21.4)	91 (26.3)	102 (29.5)	55 (15.9)	20 (5.8)	4 (1.2)	346
3. (40) Pass scores are arbitrary, little to do with mastery of a subject	11 (3.2)	21 (6.2)	71 (20.8)	105 (30.8)	74 (21.7)	59 (17.3)	341
4. (41) Pass scores on SOL tests are set too high	15 (4.9)	24 (7.9)	87 (28.5)	103 (33.8)	49 (16.1)	27 (8.9)	305
5. (43) SOL cut-scores are set at reasonable levels	29 (9.6)	57 (18.8)	85 (28.1)	109 (36.0)	20 (6.6)	3 (1.0)	303

 Descriptive statistics for scale

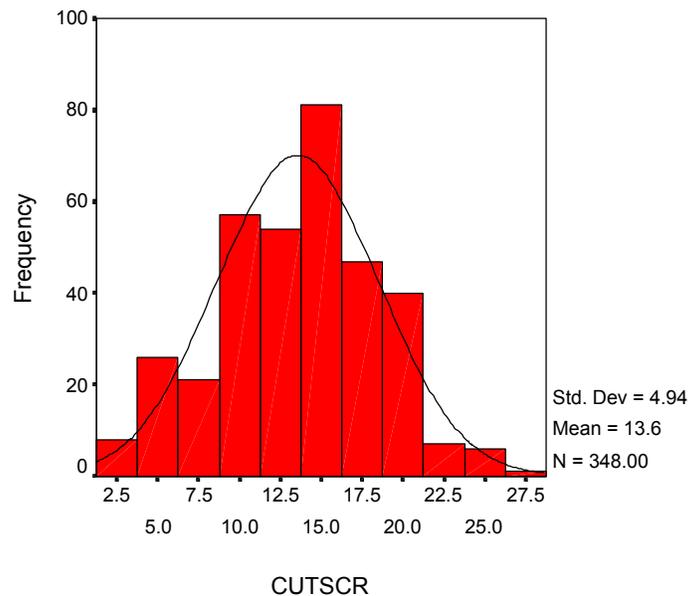
	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
Appropriateness of SOL Cut-Scores to Signify Student Subject Mastery	2.95	1.21	.77	5

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items. Items 40 and 41 were recoded prior to computing the descriptive statistics for the scale.

Teacher Response to CUTSCR

As can be observed in Figure 6, the teachers' response to the scale appeared roughly normally distributed and slightly leptokurtic while one outlier was observed for these data.

Figure 6. Teachers' Response Frequency for CUTSCR.



Note. Responses were based upon Likert-type scaling (1 = *strongly disagree*, 6 = *strongly agree*).

The CUTSCR scale was dichotomized so that the teachers' cumulative percent of agreement and the cumulative percent of disagreement could be examined. A majority (70%, $n = 341$) of the participants agreed that the cut-scores of the SOL tests, or minimum passing scores, are arbitrary and have little to do with mastery of a subject. Responses to a second similar question underpinned this perspective, as 77% ($n = 346$) disagreed that "The score that students must attain to pass a SOL test clearly distinguishes the students who have an acceptable level of knowledge and skills from students who do not." Further, the means for these items (items 39 and 40) indicated that the participants disagreed (2 = *disagree* on the questionnaire's 6-point scale) with the ability of the SOL cut-scores to clearly signify student learning. This pattern of response suggested that most participants lacked confidence in the SOL passing scores.

Questions of the survey instrument that directly addressed the participants' views on the appropriateness of the cut-score levels, however, prompted somewhat varied responses. For example teachers were asked to what extent they agreed with the statement "The cut-scores (minimum passing scores) of the SOL tests are set at appropriate levels." Thirty-three percent ($n = 315$) indicated agreement. When a similar statement was posed, "The SOL cut-scores are set at reasonable levels," the percentage of agreement reached 44% ($n = 303$). Then, when asked to what extent they believed the SOL passing scores are set too high, 59% of the participants concurred ($n = 305$).

As indicated by the results of CUTSCR, although the participants tended to view the cut-scores as arbitrary indicators of subject knowledge, they seemed somewhat uncertain as to the appropriateness of the level of the cut-scores. Additionally as discussed in the Missing Values section of the chapter, it was noted that more teachers omitted the item that queried them on the reasonableness of the level of cut-scores (item 43, $n = 303$) than any other item of the Teacher Survey. Analysis of the teacher narratives did not lend explanation to this point.

Qualitative Results

From examining the thematic trends in the teachers' responses to the open-ended questions of the survey, it was noted that the text was void of any comment regarding the SOL cut-scores. Although the teachers wrote voluminously on the possible applications of *test* scores, they evidenced little interest in the technical merit of the cut-scores. As such, the teachers did not seem to differentiate cut-scores, test scores, or passing rates.

The notion of *scores* appeared to take on a general meaning of that which is used to represent student learning and school quality and that which is used to make decisions about students and schools. Scores, in this manner, took on significant yet indefensible power according to the teachers. However this point moves beyond the description of the cut-score domain of the questionnaire, therefore, it is reserved for discussion in the subsequent section of the chapter.

Summary

Because the teacher-participants of the study did not address the topic of SOL cut-scores in their written responses to the open-ended questions of the survey, it was not possible to explicate their views any further than the item-responses would allow. However the item-responses indicated that the participants had some uncertainty as to whether the cut-scores are set at appropriate levels (items 37, 41, and 43) and that the cut-scores are arbitrary and have little to do with students' mastery of a subject.

Adequacy of the SOL Pass Rates to Signify School Quality (PASSQ)

Scale Description and Statistics

The PASSQ domain (Cronbach's $\alpha = .85$) of the Teacher Survey was comprised of eight items that asked the participants to what extent they agreed with the adequacy of 70% pass rate on SOL tests to indicate school quality. In doing so, PASSQ addressed three relevant areas of interest. First, the scale queried to what extent the participants agreed that the SOL pass rates provided an accurate representation of school quality. Second, the scale asked the teachers to what extent they agreed that requiring

70% of the students in a school to pass each SOL test helps improve the quality of teaching in the school. The third area of the scale addressed the appropriateness of requiring a 70% pass rate for the purpose of school accreditation. Table 5 displays the item content, item frequencies, and the descriptive statistics for PASSQ.

Table 5

*Item Frequencies and Descriptive Statistics for Adequacy of SOL Pass Rates
to Signify School Quality*

Item	Frequency of teacher response						<i>n</i>
	SD	D	TD	TA	A	SA	
1. (27) Pass rates accurately portray my school's quality	136 (40.1)	82 (24.2)	68 (20.1)	36 (10.6)	12 (3.5)	5 (1.5)	339
2. (28) Pass rates are appropriate for school accreditation	150 (44.4)	88 (26.0)	60 (17.8)	29 (8.6)	7 (2.1)	4 (1.2)	338
3. (44) 70% of students passing ensures my school is high quality	100 (28.7)	81 (23.2)	100 (28.7)	46 (13.2)	17 (4.9)	5 (1.4)	349
4. (45) 70% pass rate ensures I teach to rigorous standards	85 (24.9)	73 (21.3)	86 (25.1)	64 (18.7)	27 (7.9)	7 (2.0)	342
5. (46) 70% passing each test improves quality of teaching	112 (32.9)	87 (25.6)	78 (22.9)	43 (12.6)	15 (4.4)	5 (1.5)	340
6. (47) 70% pass rate for school accreditation is inappropriate	39 (11.2)	33 (20.7)	42 (32.9)	57 (49.3)	79 (72.0)	97 (72.0)	347
7. (53) Basing school accreditation on 70% pass rate is appropriate	147 (42.4)	80 (23.1)	65 (18.7)	40 (11.5)	9 (2.6)	6 (1.7)	347

Item	Frequency of teacher response						<i>n</i>
	SD	D	TD	TA	A	SA	
8. (54) Important for schools to be held accountable by SOL tests	93 (26.3)	61 (17.2)	78 (22.0)	96 (27.1)	20 (5.6)	6 (1.7)	354

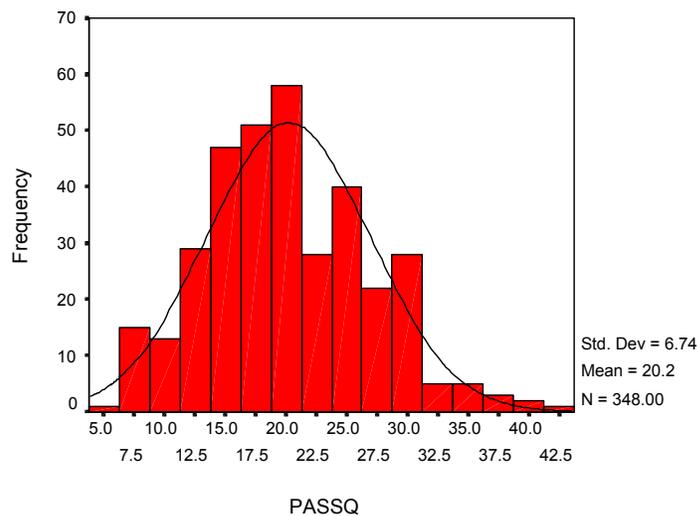
	Descriptive statistics for item			
	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
Adequacy of SOL Pass Rates to Signify School Quality	2.43	1.32	.85	8

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items. Item 47 was recoded prior to computing the descriptive statistics for the scale.

Teacher Response to PASSQ

As can be viewed in Figure 7, the classroom teachers' responses to PASSQ were approximately normally distributed and slightly leptokurtic. Three outliers were identified for the scale responses.

Figure 7. Teachers' Response Frequency for PASSQ.



Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

The VEA teacher-members disagreed with every item of PASSQ (*2 = disagree* on the survey's 6-point scale). Examination of the items' cumulative percent of agreement provided another means by which to inspect the teachers' general disapproval of the adequacy of the SOL pass rates to signify school quality. The PASSQ scale was dichotomized so that the teachers' cumulative percent of agreement and the cumulative percent of disagreement could be examined. The results indicated consistent disagreement over two of the areas queried by the scale. However, the area concerning the appropriateness of pass rates for the purpose of school accreditation produced some inconsistency in the results.

While 34% ($n = 354$) of the participants believed it is important for schools to be held accountable by the SOL tests (item 54), less than half of that percentage (16%, $n = 339$) believed that passing rates on the SOL tests provide an accurate representation of school quality (item 27). In terms of the quality of teaching, 71% ($n = 342$) of the teachers disagreed

that setting a 70% pass rate for schools ensures that teachers teach to rigorous standards (item 45). Ten percent more of the teachers, 81% (n = 340), disagreed that the quality of teaching is improved by requiring 70% of a student population to pass each SOL test (item 46).

Finally, the classroom practitioners indicated some inconsistency in their views on the appropriateness of a 70% pass rate for the purpose of school accreditation. On this point, items 28 and 53 showed only 12% and 16% agreement, respectively; however, item 47 showed about twice as much agreement concerning the appropriateness of a 70% pass rate for schools to become accredited by the State. It was noted that item 47 was recoded, thus was negatively worded on the Teacher Survey whereas items 28 and 53 were not. Even so, the means for these three items suggested that the teachers of the study clearly disagreed that the SOL pass rate was appropriate for school accrediting purposes.

Qualitative Results

Similar to the content comprising the CUTSCR domain, the specific discussion of pass rates appeared to be virtually nonexistent in the teacher narratives. However, the participants discussed their views on how scores in general are used to signify school quality; and this topic is taken up in the next section of the paper.

Summary

Without amplification in the qualitative data regarding teachers' views on SOL pass rates, the results of PASSQ could not be elaborated upon. However, the results from scale-level data of PASSQ showed the teachers' disapproval of the pass rates as being appropriate

for improving teaching in the classroom, for representing school quality, and for determining school accreditation.

Diagnostic Value of SOL Test Scores (DIAGV)

Scale Description and Statistics

The DIAGV (Cronbach's alpha = .86) domain consisted of nine items that queried the participants on how they use the SOL test scores to identify weak instructional practices, to diagnose students' learning difficulties, and how the scores are use to determine course placement, grade retention, and student high school graduation. Table 6 displays the item content of the scale, the item frequencies, and the descriptive statistics for the scale.

Table 6

Item Frequencies and Descriptive Statistics for Diagnostic Value of SOL Test Scores

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
1. (25) SOL scores help me pinpoint learning difficulties	73 (23.5)	58 (18.7)	73 (23.5)	68 (21.9)	26 (8.4)	12 (3.9)	310
2. (26) SOL scores help me identify weak instruction	70 (22.5)	58 (18.6)	83 (26.7)	73 (23.5)	16 (5.1)	11 (3.5)	311
3. (29) SOL scores should be single most important factor in determining high school graduation	258 (75.0)	44 (12.8)	26 (7.6)	10 (2.9)	2 (.6)	4 (1.2)	344
4. (33) SOL scores give information unavailable from Stanford-9	96 (30.6)	75 (23.9)	68 (21.7)	54 (17.2)	15 (4.8)	6 (1.9)	314
5. (34) SOL scores are trustworthy for determining grade retention	139 (41.1)	86 (25.4)	67 (19.8)	34 (10.1)	11 (3.3)	1 (.3)	338
6. (35) SOL scores are useful for determining course placement	92 (28.8)	66 (20.6)	51 (15.9)	82 (25.6)	28 (8.8)	1 (.3)	320
7. (48) SOL scores should be primary to judge student learning	175 (49.0)	87 (24.4)	65 (18.2)	19 (5.3)	8 (2.2)	3 (.8)	357

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
8. (49) Scores should be deciding factor for high school graduation	216 (60.8)	80 (22.5)	44 (12.4)	9 (2.5)	4 (1.1)	2 (.6)	355
9. (51) Students not passing tests should be denied a high school diploma	173 (49.3)	76 (21.7)	70 (19.9)	22 (6.3)	6 (1.7)	4 (1.1)	351

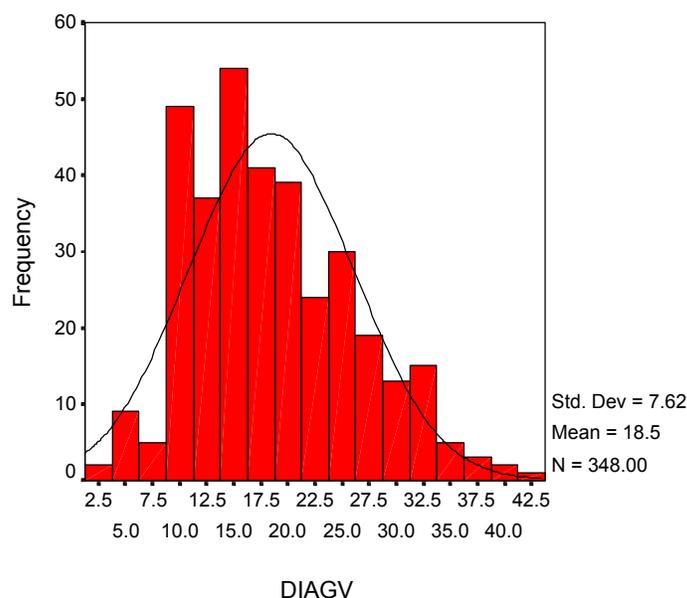
	Descriptive statistics for scale			
	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
Diagnostic Value of SOL Test Scores	2.20	1.19	.86	9

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items.

Teacher Response to DIAGV

Figure 8 displays the distributional shape of DIAGV and shows that the scale was slightly positively skewed and slightly leptokurtic. However, overall, the responses appeared to be roughly normally distributed, with one outlier identified.

Figure 8. Teachers' Response Frequency for DIAGV.



Note. Responses were based upon Likert-type scaling (1 = strongly disagree, 6 = strongly agree).

The DIAGV scale was dichotomized so that the teachers' cumulative percent of agreement and their cumulative percent of disagreement could be inspected. The results indicated that in terms of the use of SOL test scores to determine whether students should graduate from high school, the participants overwhelmingly contended this use. Items 29, 49, and 51 that addressed this topic each resulted in no more than 9% ($n = 351$) of the classroom teachers approving. Slightly more, 14%, ($n = 338$) believed that SOL test scores are trustworthy indicators of whether students should be retained in grade. On the point of whether the SOL test scores are useful for determining course placement, however, the percentage of teachers agreeing increased to 35% ($n = 320$).

When the teachers considered the use of SOL test scores for pinpointing learning difficulties and for identifying instructional weaknesses, the cumulative percent of agreement

was similar to that for *course placement*, only slightly less at 34% ($n = 310$) and 32% ($n = 311$), respectively. Also not quite one-fourth (24%, $n = 314$) of the teacher-participants believed the SOL test scores offer distinct information from other tests such as the Stanford-9.

Although the teachers indicated very modest agreement with each of the items comprising DIAGV, the participants indicated the largest support for the use of SOL test scores for the purposes of pinpointing learning difficulties and instructional weaknesses as well as for determining students' course placement. The diagnostic value of the SOL test scores drew even less support as the teachers considered using the scores to help determine grade retention. Even less support was shown for the use of the scores when students' high school graduation is at stake. Thus the more important the decision was to be made about students, the less useful the SOL test scores were for making the decision.

Qualitative Results

The scale-level results of DIAGV indicated that the teachers did not find the SOL test scores to be diagnostically valuable or appropriate for making the high-stakes decision of whether students should graduate from high school. The participants contended that the scores seemed to possess little power to assist them in “pinpoint[ing] individual students' ‘areas of opportunity’ in order to focus more effort on problem subjects” (high school history teacher). The teachers desired the SOL test-scores to render a practical power at their disposal, enabling them to better diagnose students' learning difficulties and to help focus and improve teachers' classroom instruction. “The tests [test scores],” wrote a fifth grade math teacher in a school of moderate SES, “could be used to identify areas for

remediation...They could be used to monitor student progress through an academic subject area. At present, tests aren't used to measure student performance. Administrators and the public transfer student performance to an evaluation of teacher efforts. Low scores are always interpreted as bad teaching." This comment was representative of the teachers' persistent distinctions between the limited diagnostic value of the SOL test scores and the scores' potential for improved value, and as such, this distinction pointed up the teachers' desire to assess students' academic growth.

Desire to assess students' academic growth. The teachers described the SOL test scores as being insufficient to accomplish the teachers' desire to improve, or cultivate, students' academic *growth*. The idea of student growth emerged as the centerpiece of the teachers' rationale for their firm view that SOL test-score data are impractical for diagnosing students' learning difficulties, improving classroom instruction, and refining the quality of student learning. Many teachers attempted to address this problem, as they defined it, and wrote that they would like to see SOL tests administered at every grade level at both the beginning and the end of the school year or course.

The teachers proposed that by pre- and post-testing at every grade level, a baseline of information could be provided that might inform teachers' planning of curriculum and instruction for a given classroom of students. A fifth grade teacher added, "Test results should be used to measure students' learning from beginning to end of a school year. It should be used as a progress report (accountability can be held here) and a diagnosis for curriculum planning for the following year."

Fairness in student assessment. The notion of fairness permeated the qualitative data. Participants of the study believed that measuring student growth was the only fair method of

assessing student learning. “Let’s be *fair*,” urged a fourth grade teacher of English, reading, and science, “and measure students’ growth. They aren’t all coming with the same background, support or ability. So how could one test measure how far they’ve come??” Similarly, a sixth grade teacher of English, history, and reading offered, “At least let us begin where the students are, not where they *are supposed to be*. Measure gain and improvement, not some preset point that is impossible for the majority of students to reach.”

Need for SOL test scores to help guide classroom planning. In terms of SOL testing occurring at each grade level, teacher-participants believed this practice would assist them to better plan and organize each school year’s subject matter, identifying where and to what extent students might need subject-area review or remediation. Given the present situation, the teachers viewed the large lapses in time between testing as thwarting optimal practical use of SOL test scores, thus preferred the use of the scores for practical planning purposes rather than for penalizing students. The scores, as suggested by a high school history teacher, should be used “as simply a guide to planning curriculum” while a sixth grade math and science teacher affirmed, “The results should be used to assess the curriculum.” Still, a fifth grade teacher wrote, “We should be testing math, reading [and] writing at the elementary level. Scores should be used to plan curriculum [and] support for students, not to penalize students, schools, [and] teachers.” Given these pervasive responses by the teachers, the notions of student academic *growth*, identification of students’ learning needs, and curriculum planning stood in sharp contrast to what the teachers believed to be the *penalizing* power of the SOL test scores.

Desire for multiple modes of assessment. In addition, the participants discussed the singular power of the SOL test scores and argued against the use of the scores as the only

indicator for making important decisions about students. For instance the VEA teacher-members deemed that SOL test scores “should be one indicator, along with grades in determining placement. It is impossible to be fair to students or teachers using one assessment tool,” (seventh grade English teacher in a high-income school locale). Further, most participants wrote that the SOL test scores, as one of several indicators, should be used in determining whether students should be retained in grade. “If test scores are returned earlier,” noted a fifth grade teacher, “scores can be a part of determining retention.”

On the point of high school graduation decisions, a high school science teacher offered, “I like the idea of having my students accept responsibility for their own learning. I object to telling a student they can’t graduate because they had a bad testing experience. A bad day/bad week should not prohibit graduation.” In keeping with the aforementioned statement, a fourth grade teacher posed, “I feel the test should be used to measure the student’s progress but not their knowledge of this inappropriate age-level material. Also, some students are good students but do not always do well on standardized tests and I do not feel their eligibility to graduate should be solely based on these tests.”

The teachers tended to view the SOL scores as one source, a partial source, of information that can inform whether a student should graduate. Accordingly some provided rather specific ideas on how the process might proceed, as did the following classroom teacher: “[SOL test scores should be used] as a part of whether children should graduate from high school. The day-to-day grades should be the other part. For instance, count the SOL as 25% of a student’s grade [and] the class grade as 75%. One test shouldn’t count more than a year’s work.” In sum, “Due to the many factors that can affect the outcome of the test, the

results should not be used exclusively to determine students' learning, pass/fail, or graduation" (third grade teacher).

Summary

Themes identified in the teacher narratives were similar to the results of DIAGV, underscoring the pattern that the participants strongly contended the use of SOL test scores as the primary criteria for what students have learned and whether they should graduate from high school (items 29, 48, 49, and 51). Furthermore, little value was placed upon the use of SOL test scores to determine course placement and whether students should be retained in grade (items 34 and 35). The singular power of the test scores, according to the teachers, lacked fairness toward students, particularly since the teachers believed the scores do not adequately reflect student academic growth. The participants posited that the SOL test scores held punitive power rather than being diagnostically useful.

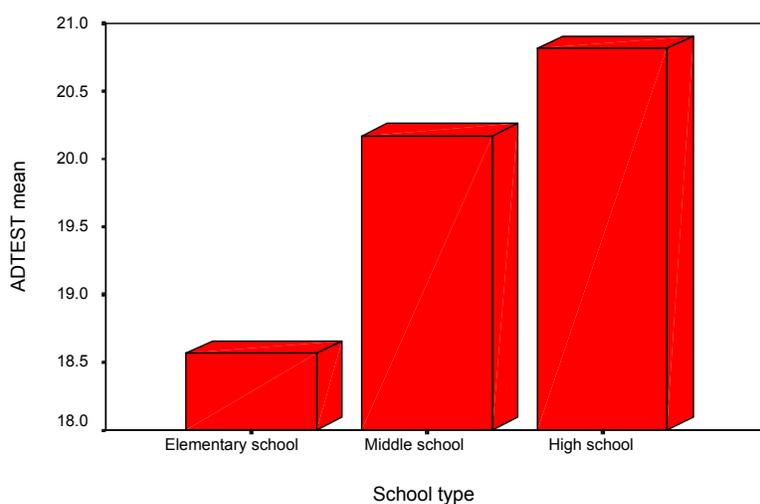
Adequacy of SOL Testing and SOL Test Scores to Hold Schools Accountable (ADTEST)

Inferential Results

The omnibus test of the main effect of *school type* (elementary, middle, and high school) was performed for ADTEST and yielded a statistically significant result, $F(2, 351) = 3.89, p = .021$. Subsequently, post hoc tests were performed to better understand why the omnibus test yielded a significant F -ratio. The conservative Tukey procedure was employed, and to guard against an inflated Type I error, a Bonferroni adjustment was applied. As a result, the comparison of the means for elementary school teachers ($M = 18.56, SD = 6.93$)

and high school teachers ($M = 20.81$, $SD = 7.02$) reached statistical significance ($p = .019$). No statistically significant difference was found between middle and high school teachers or between elementary and middle school teachers. This outcome suggested that VEA elementary school teachers, on the average, are more likely to report disagreement with the adequacy of SOL testing for school accountability purposes than VEA high school teachers. Figure 9 displays a bar graph for these results.

Figure 9. Comparison of ADTEST Means for School Type.



Note. Responses were based upon Likert-type scaling ($1 = strongly disagree$, $6 = strongly agree$).

Scale Description and Statistics

The ADTEST (Cronbach's alpha = .82) scale addressed the issue of school accountability and queried the participants as to what extent they believed in the adequacy of a given 40-60 item SOL test to measure students' knowledge. Additionally the scale was concerned with whether the teachers believed SOL test scores should be used for school accountability purposes and whether the test requirements help students take responsibility for learning as they might be motivated by the requirements to pass a given SOL test.

The scale also addressed to what extent the participants believed the requirements to pass a given test improve the quality of students' learning, thus ensure a world-class quality of education. Moreover, ADTEST queried whether the teachers believe that the quality of teaching is reflected in student SOL test scores and whether the comparisons between schools' scores on the SOL tests help hold schools accountable. Table 7 displays the item content, item frequencies, and descriptive statistics for the scale.

Table 7

*Item Frequencies and Descriptive Statistics for Adequacy of SOL Testing
and SOL Test Scores to Hold Schools Accountable*

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
1. (24) SOL test scores reflect the quality of my teaching	114 (36.3)	65 (20.7)	67 (21.3)	39 (12.4)	17 (5.4)	12 (3.8)	314
2. (30) A 40-60 item SOL test cannot adequately measure the skills and knowledge in the subject I teach	9 (2.8)	9 (2.8)	21 (6.4)	36 (11.0)	68 (20.9)	183 (56.1)	326
3. (36) Requiring a passing score on SOL test assures student competence in subject area	81 (23.1)	86 (24.5)	97 (27.6)	55 (15.7)	29 (8.3)	3 (.9)	351
4. (38) Requiring students to pass a SOL test ensures they gain a world-class education	135 (39.1)	87 (25.2)	83 (24.1)	33 (9.6)	5 (1.4)	2 (.6)	345
5. (42) A single pass score cannot motivate all my students	4 (1.3)	4 (1.3)	8 (2.5)	58 (18.2)	80 (25.1)	165 (51.7)	319
6. (50) Requiring students to pass SOL tests helps them take responsibility for learning	65 (18.3)	55 (15.4)	73 (20.5)	114 (32.0)	35 (9.8)	14 (3.9)	356

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
7. (52) School quality should be based on SOL test scores alone	251 (70.1)	61 (17.0)	40 (11.2)	3 (.8)	3 (.8)	0 (0)	358
8. (64) Comparisons between my school and other schools' SOL test scores help hold my school accountable	71 (20.6)	70 (20.3)	67 (19.4)	98 (28.4)	32 (9.3)	7 (2.0)	345
9. (79) Quality of students' learning is improved by requiring them to pass SOL tests to graduate from high school	107 (30.7)	78 (22.4)	79 (22.7)	65 (18.7)	13 (3.7)	6 (1.7)	348

Descriptive statistics for scale

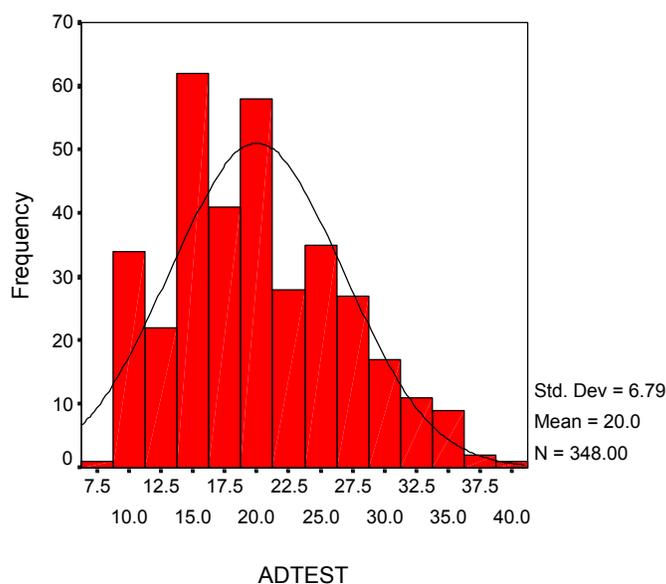
	M	SD	Cronbach's <i>alpha</i>	n
Adequacy of SOL Testing and SOL Test Scores to Hold Schools Accountable	2.31	1.22	.83	9

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items. Items 30 and 42 were recoded prior to computing the descriptive statistics for the scale.

Teacher Response to ADTEST

The teachers' responses when considering the scale's focus on school accountability, were roughly normally distributed, with slight positive skewness and one discernable outlier. Figure 10 displays the distributional shape of these data.

Figure 10. Teachers' Response Frequency for ADTEST.



Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

From the item means it was observed that the VEA teacher-members believed that the adequacy of SOL testing and resulting test-score data were inadequate to hold schools accountable. However, in light of the item means and the ANOVA results for *school type*, it appeared that the elementary school teachers were more likely to indicate strong disagreement with the adequacy of SOL testing for school accountability than were the high school teachers who indicated disagreement (*1 = strongly disagree, 2 = disagree* on the 6-

point Likert-type scale of the study). The ADTEST scale was dichotomized so that the teachers' cumulative percent of agreement and their cumulative percent of disagreement could be examined. The results showed that both elementary and high school groups, along with middle school teachers, disagreed with the adequacy of the current State assessment of school quality.

Also, the teachers indicated general disapproval of Virginia's test-based school accountability effort, again with elementary school teachers being more likely to indicate *stronger* disagreement than the high school teachers. First, 88% ($n = 326$) of the participants believed the 40-60 item SOL test instruments are inadequate to measure students' knowledge. This point alone suggested the direction of the teachers' opinion for the remaining items of the scale. If the test instrument is not perceived to be adequate for its purposes, it seems difficult for a system of accountability that is based upon the test to be perceived as adequate for its purposes. The teachers' responses to the scale items suggested this logic. Only 2% ($n = 358$), which was the smallest percent to agree with a given item on the entire questionnaire, agreed that school accountability should be based on SOL test scores alone.

Second, although the teachers believed SOL tests and the resulting scores were not adequate for the purpose of holding schools accountable, a little less than half of the participants (46%, $n = 356$) believed that the SOL test requirements helped students take responsibility for their learning. Even so, 95% ($n = 319$) of the classroom teachers indicated that a single pass score for a SOL test cannot motivate all students. Nor was the quality of students' learning improved by requiring them to pass SOL tests to graduate, reported three-fourths of the participants (76%, $n = 348$).

Third, 88% ($n = 345$) concurred that the above requirement did not ensure students receive a world-class education. Similar results were found when the participants considered whether requiring students to achieve a pass score on a SOL test assures students' competence in the subject tested. Given this point, 75% disagreed ($n = 351$). Finally, as to the point that the quality of teaching is reflected in students' SOL test scores, 22% of the classroom teachers agreed ($n = 314$). Approximately twice this percent (40%, $n = 345$), however, believed that comparisons between their school and other schools help hold their school accountable. These results, combined with those above, suggested that the teachers disapprove of SOL testing for school accountability on the grounds of (1) an inherently inadequate test measure (2) the inadequacy of SOL test scores to signify the quality of teaching and student learning, and (3) the inadequacy of the test requirements to motivate all students and to improve learning.

However amid these views, a subtle distinction was observed: Forty percent of the teachers (albeit well below half) concurred that comparisons between their school's SOL test scores and other schools' scores help hold their school accountable. Furthermore, it bears restating that 46% of the participants believed that requiring students to pass SOL tests assisted students in taking greater responsibility for their learning. Therefore, from these data it appeared that while the teachers, collectively, evidenced disapproval over the SOL testing technology, elementary school teachers were more likely to *strongly* disagree than were the high school teachers. A little less than half of the participants agreed that test requirements encourage greater student responsibility and that comparing SOL test scores between schools helps improve school accountability. These subtle distinctions were noted in the teacher narratives as well.

Qualitative Results

Multiple modes of assessment. Overall, the quantitative results were supported by the thematic trends identified in the teacher narratives. In reference to the adequacy of the SOL testing technology as a means by which to hold schools accountable, the resounding trend in the teacher text was that the teachers intensely advocated for an expanded notion of school quality and for a diversity of means by which to assess that quality. Given these changes in state-level assessment, the teachers contended that school accountability might be attempted in a manner that is defensible. In brief, the teachers believed that the SOL testing technology was an inadequate mechanism for holding schools accountable. They defended this position by describing what they believed to be the inadequate singularity of the test-orientation of the current school reform.

The participants wrote that good teaching requires diverse teaching strategies to meet diverse student needs, and as such, it seemed reasonable to the teachers that diverse modes of assessment accompany diverse instructional practices. Further, the diverse instructional strategies and diverse modes of assessment were a logical approach to teaching and assessing diverse student populations among schools, according to the teachers. The participants described the current State assessment as “a one size fits all.” “The state needs to be aware of the diversity and individual dynamics of each school and its population,” wrote a fourth grade history teacher in a middle-income school.

Complexity of school quality. Because SOL test scores were viewed as indicators of school quality that did not take into account the diversity of school populations, the singular power of SOL test scores to accurately and justly signify school quality was indefensible to the teachers. School quality, as most teachers concurred, is much more complex than what

can be represented by SOL test scores; it must consider both socio-economic influences upon the classroom as well as students' intellectual, or learning, needs. An eighth grade teacher in a low-income area wrote, "I teach students who live in shelters, who live with abusive parents, who raise themselves because their parents work 2 or 3 jobs. And my primary focus is supposed to be state exams?" On the point of social-economic influences upon student learning, a high school English teacher added

A 21st century teacher faces a challenge that reaches beyond the textbooks. Teaching the material will always be a top priority, but today's teachers must also incorporate interpersonal skills to bridge the gap between school and the lack of guidance/support from home that could once be taken for granted.

Teachers viewed that because social-economic factors appear to influence student learning, thus SOL test-scores, school quality (and the quality of teaching) is not justly assessed by SOL tests. Furthermore, the participants of the study wrote that although teacher accountability should not be dismissed, they asserted that SOL test scores are largely dependent upon "students' ability, background, experiences, [and] emphasis of education in the home" (sixth and seventh grade reading teacher in a middle-income school locality). While the teachers expressed their recognition of such student diversity, they tended to question the State on this point. "The state needs to be aware of the diversity and individual dynamics of each school and its population," admonished a fourth grade history teacher in a middle-income school locale.

Often the inferences made from schools' SOL test scores, according to the participants, suggest that the non-passing schools, often in low-income areas, are of low quality, hence have low quality teachers. The teachers of the study contended that in general this is an unwarranted assertion. A high school science teacher amplified this theme identified in the teacher text:

First of all, what everyone knows but is too scared to say is that the schools that are doing well are succeeding mostly because they have *good* students [with ample family and school resources]. The teachers at unsuccessful schools are just as good, they just have a harder job and more remediation to do. I can say this because I come from a fully accredited school and I know teachers at less fortunate ones. Accountability of the schools is *not* always the teachers' fault. (We have some pretty lousy ones [with] SUPER SOL scores.) The test is not an accurate measure for many reasons.

Often factors that teachers perceive to lie beyond their control affect students' SOL test scores, and consequently are not taken into account by the current evaluation of school quality, according to the teachers.

The participants cited student mobility as a primary factor that was beyond the teachers' influence. Students in schools located in low-income areas, the teachers observed, tend to be more mobile than their counterparts in more affluent locales. The participants discussed that as a result of student mobility teachers may begin a school year with a particular class of students, teach the Standards, and before SOL testing begins in the spring

many of the students with whom they began the school year have moved and new students have replaced them. A high school teacher exemplified this theme in the survey data, “Emphasis on SOL scores is absurd. Testing all students is an utopian ideal. Students who transfer in just one week before the test are tested. Students who have been identified with major education weaknesses are tested. Kids with absentee ratios of over 50% are tested.” Consequently, many teachers questioned how they could be held accountable for conveying SOL subject matter to students when students are highly mobile or when they do not attend class regularly.

In addition to citing student mobility and absenteeism as factors that lie beyond the teachers’ control, some participants catalogued a variety of factors that they perceived were not taken into account when school quality was signified by SOL test-scores. Although lengthy, the following example, written by a reading specialist who teaches in an elementary school where 50% of the students receive free or reduced price lunch, provides a comprehensive catalogue of factors, many of which concerned the participants of the study:

At our school, we are working above and beyond the call of duty. We go early, we stay late, we attend continuing [education] classes, we have staff development constantly, we tutor, we go to school on [Saturday], we involve the community, etc. etc. But our test scores, I fear, will never be “up to par.”

“Why?” this teacher questioned rhetorically,

Because we have students, most [of] whom meet the following profile—
Mom in jail, Dad never existed, grandma tired of raising grandchildren, no
one in family educated [schooled], many illiterate, no family car, dirt floors,
no running water, obesity, junk food diets, rotten teeth, no prenatal care,
little doctor care, no fruits, no vegetables, fast food, family on the run from
social services, domestic violence, alcohol, drugs, cigarettes, poverty,
terrible attitudes, reading isn't important, school attendance isn't important,
disrespect to authority & adults, etc., no responsibility, poor attitudes,
believe teacher is always wrong, believe child is perfect, sexual abuse, child
never talked to, child doesn't know own name, child won't talk, ADHD,
exposed to pornography, no problem-solving skills, no manners, loud,
aggressive behavior, lack of values, no religion, no prior [subject-related]
knowledge, limited experiences. SOL reform will never fix the “real”
problems.

Teachers across school types tended to elaborate upon the diversity of needs that students primarily in schools of low-income areas bring with them to the classroom, including a need for adult homework support, access to home computers, and a parental presence to impose a suitable bedtime. Such student needs, the teachers believed, are reflected in schools' SOL test scores, thus inadequately and unfairly signify school quality.

Student growth. As mentioned earlier in this section of the chapter, the teacher-participants expressed that most schools with student populations from low-income areas generally do not pass the SOL tests, thus do not meet the 70% passing rate required by the

State to attain school accreditation. In reference to this point, however, the teachers believed that within these schools small gains and student improvements throughout the course of a school year should not be dismissed. Accordingly when school quality, hence school accreditation, is at stake, the teachers overwhelmingly concurred that the gains, or growth, of students should be the primary signifier of school quality.

According to the participants, only by measuring students' academic growth from year to year, that is, comparing students' SOL test scores to the same students' scores the following school year, could the SOL reform demonstrate fairness across diverse school populations. "Why do we compare last year's SOL 7th grade history results to those scores of this year's 7th grade (not the same students/children)" observed a sixth and seventh grade history teacher. Similarly, the participants questioned the fairness of comparing SOL test scores from one school to the scores of another school that might possess significant differences in student populations, particularly given students' diverse family backgrounds and various economic levels of the school locales. The purpose of SOL test scores, asserted an elementary teacher of English and reading, should be "to chart each school's progress...not to compare one school against another (apples and oranges) since school populations differ." "They should not be used as a 'threat' to schools or to teachers," added an eighth grade history teacher, "There are numerous factors that should be analyzed when determining success, not just the SOL tests [test scores]."

Student responsibility. Although the teachers recognized that students' diverse family backgrounds and various learning needs should be taken into account when school quality is assessed, the teachers also believed that students should take a degree of responsibility for learning the SOL subject matter. "I believe that schools should be held accountable for what

they teach, not for their students' scores on SOL tests. Socio-economic background plays too large a role. Making sure that school and teachers provide the essential resources and information is crucial. After that the responsibility for learning falls in the hands of the students," wrote an eighth grade science teacher in a high-income school locale.

In addition to describing their views on school accountability in terms of students' diverse family and income-levels, the teachers indicated their view on accountability in terms of students' motivation and discipline. The participants did not believe it was fair for teachers to work hard trying to meet diverse student needs while the students do not seem to care about learning the material. Then in the end, once SOL test scores are reviewed, the school/teachers are held responsible for any shortcomings in the test results. A high school English teacher wrote, "How can I—and why should I—be held accountable for students who do not care, are not interested in learning, and only come [to school] for social reasons? ...[Can] I take aluminum and make it into stainless steel[?] No one can do that. What teachers do is teach; what students should be held accountable for is learning. They aren't held accountable yet."

Given the theme of student responsibility in the teacher text, it was observed that while participants regardless of school type advocated increased student responsibility for learning SOL subject matter, the high school teachers of the study responded 2.5 times more frequently than elementary school teachers and middle school teachers alike. Although the high school teachers wrote more frequently, elementary school teachers and middle school teachers wrote no less emphatically. An elementary school teacher's view exemplified this point: "[The role of] teachers [is] to teach—not be a social worker, babysitter, nurse, secretary...as well as teach" (second grade history teacher in a middle-income area). In

conclusion, although the teachers expressed understanding of the influence of students' family background upon student learning, the participants clearly did not share this understanding or empathy for students who appear to be unmotivated and/or oppositional in the classroom. The teachers asserted there is "no way" they could be held accountable for the student who is "totally unmotivated and a discipline problem" (high school teacher).

Supplemental Statistical Analysis

Because the teacher narratives did not supply evidence to support the inference suggested by the results of the post hoc tests for the school type variable, it seemed possible that only one or two items of ADTEST influenced the results of the follow-up tests. To explore this possibility, for each variable (item) comprising ADTEST, a one-way analysis of variance was performed. Moreover, to guard against an inflated Type I error, alpha was divided in half for each successive test that was performed.

As a result of the nine one-way ANOVAs, only one variable (item 50 = Requiring students to pass SOL tests helps them take responsibility for their learning.) reached statistical significance, $F(2, 347) = 5.056, p = .007$, at the alpha level of .05. To test the hypothesis that the variances were equal, Levene's test of homogeneity of variance was employed and did not reach statistical significance at the alpha level of .05 ($p = .908$), thus supported the hypothesis that the variances were equal. Subsequently, post hoc tests were conducted using the Tukey procedure to understand why the omnibus test yielded a significant F -ratio. Consequently, the comparison of the means for elementary school teachers ($M = 25.74, SD = 12.24$) and high school teachers ($M = 30.33, SD = 12.6$) reached statistical significance ($p = .006$).

From the statistically significant outcome it was inferred that VEA elementary school teachers, on the average, are more likely to indicate disagreement with the position that requiring students to pass SOL tests helps them take responsibility for their learning than VEA high school teachers. Perhaps more importantly, however, was that the results of the nine omnibus *F*-tests showed that a single item influenced the scale mean for ADTEST, thus presented the illusion that teachers of the three school types tended to differ in terms of their overall beliefs about the adequacy of SOL testing and SOL test-scores for the purpose of holding schools accountable. Rather, *item 50* alone accounted for the difference observed between elementary school teachers and high school teachers. This threw light upon why a departure of opinion was not identified in the teacher narratives for the elementary and high school teachers in relation to their general beliefs about the adequacy of the current assessment of school quality. Moreover, other than high school teachers writing more frequently on the topic of student responsibility than elementary school teachers and middle school teachers, the teacher narratives did not lend support to the claim that the high school teachers and elementary school teachers differ in their opinion concerning student responsibility for learning (item 50). Both groups expressed a desire for increased student responsibility in learning SOL subject matter.

Summary

Taken together, the results of the scale-level analysis and the qualitative data were similar to each other and suggested that the teachers of the study did not believe SOL testing and SOL test-scores adequately represent school quality. First, the SOL test itself was viewed to lack the ability to measure student knowledge. Second, the test scores alone were believed

to be singular in nature, thus insufficient to adequately reflect the quality of teaching/schools given schools' diverse student populations. Third, requiring students to pass SOL tests neither motivates all students nor improves the quality of their learning, asserted the teachers. Finally, while about half of the participants believed the SOL test requirements help students take responsibility for learning, less than half believed that comparisons between the SOL test scores of their school and other schools help hold their school accountable.

In conclusion, the inadequacy of SOL test-scores to take into account the diversity of student populations when representing school quality was intensely frustrating to the teacher-participants, who tended to perceive the test-scores' representation as an affront to their professional autonomy. This topic is taken up in the subsequent section of the chapter.

Teacher Professional Autonomy (TPROF)

Inferential Results

Status on teaching a SOL-tested subject. An independent-samples t-test was performed to test for mean differences on teachers' *status on teaching a SOL-tested subject* (yes, no). The test yielded statistically significant results at the alpha level of .05, $t(114.03) = 3.54, p = .001$, indicating that the mean difference is larger than expected by chance in the population of VEA teachers who teach an SOL-tested subject ($M = 36.07, SD = 5.58$) and the population of VEA teachers who do not ($M = 32.18, SD = 10.13$). As such, this outcome suggested that VEA teachers of SOL-tested subjects, on average, are more likely to report disagreement with the level of support for teacher professional autonomy within the SOL reform than VEA teachers who do not teach a SOL-tested subject. To test the hypothesis of equal variances, Levene's test of homogeneity of variance was employed and reached

statistical significance ($p = .000$, $\alpha = .05$). Therefore the harmonic mean was used to test for mean differences.

Scale Description and Statistics

The TPROF (Cronbach's alpha = .89) consisted of 11 items, which questioned the participants' views on to what extent the SOL reform affects their sense of professional autonomy and expertise. This included to what extent the reform affects teachers' professional decision-making in the classroom, influences community and parent confidence in teachers' work, and the SOL program's effect on the participants' quality of teaching and their ability to meet students' educational needs. Table 8 presents the item content for TPROF, the item frequencies, and the descriptive statistics for the scale.

Table 8

Descriptive Statistics for Support of Teacher Professional Autonomy Within the SOL Reform

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
1. (4) SOLs complement my professional decision-making	34 (9.9)	69 (20.2)	80 (23.4)	95 (27.8)	53 (15.5)	11 (3.2)	342
2. (5) SOLs help teachers focus on subject matter	24 (7.1)	27 (7.9)	56 (16.5)	118 (34.7)	81 (23.8)	34 (10.0)	340
3. (17) SOLs help community/parents respect teachers' work	91 (26.7)	78 (22.9)	81 (23.8)	52 (15.2)	30 (8.8)	9 (2.6)	341
4. (18) SOLs help community confidence in teachers' decisions	82 (24.6)	77 (23.1)	81 (24.3)	61 (18.3)	24 (7.2)	8 (2.4)	333
5. (22) SOLs restrict my professional decision-making	10 (2.9)	21 (6.1)	56 (16.4)	68 (19.9)	93 (27.2)	94 (27.5)	342
6. (58) SOL system is consistent with my view of my teaching role	124 (35.6)	75 (21.6)	76 (21.8)	49 (14.1)	14 (4.0)	10 (2.9)	348
7. (61) Importance placed on SOL test scores constrains my professional expertise	14 (4.0)	21 (6.1)	47 (13.6)	62 (17.9)	84 (24.3)	118 (34.1)	346

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
8. (66) SOL system helps my students respect me as a teacher	170 (50.0)	75 (22.1)	64 (18.8)	25 (7.4)	6 (1.8)	0 (0)	340
9. (76) Able to meet educational needs of students with SOL system	80 (23.5)	65 (19.1)	89 (26.2)	63 (18.5)	33 (9.7)	10 (2.9)	340
10. (77) My professional autonomy is restricted by SOL reform	15 (4.4)	20 (5.9)	46 (13.6)	81 (23.9)	85 (25.1)	92 (27.1)	339
11. (78) School accreditation requirements diminish the quality of teaching	21 (6.3)	43 (12.9)	72 (21.6)	97 (29.1)	53 (15.9)	47 (14.1)	333

Descriptive statistics for scale

	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
--	----------	-----------	---------------------	----------

Descriptive Statistics for Support of Teacher

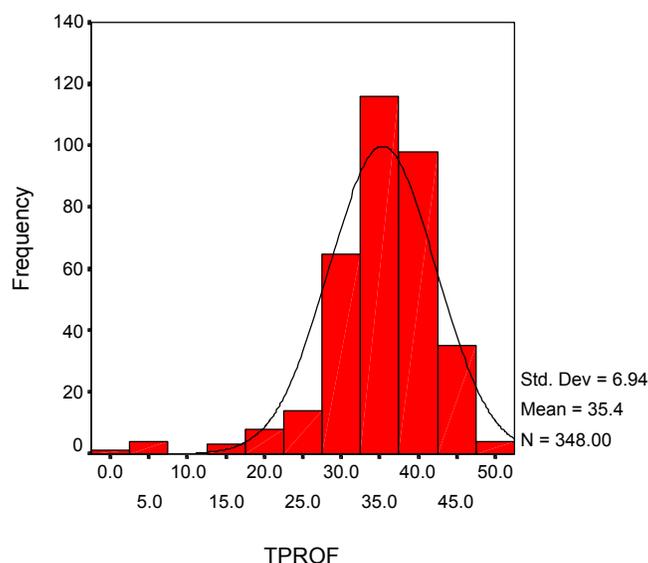
Professional Autonomy Within the SOL Reform	2.76	1.34	.89	11
---	------	------	-----	----

Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, *n* = the number of scale items. Items 22, 61, 77, and 78 were recoded prior to computing the descriptive statistics for the scale.

Teacher Response to TPROF

The responses to the items of the scale were negatively skewed and leptokurtic, with no extreme cases identified. Figure 11 displays the distributional shape of these data.

Figure 11. Teachers' Response Frequency for TPROF.



Note. Responses were based upon Likert-type scaling (*1 = strongly disagree, 6 = strongly agree*).

The item means of the scale indicated that the participants in general did not believe that the SOL program supported their professional autonomy in the classroom. Accordingly, all but one of the item means indicated either the teachers' disagreement or their tendency to disagree (*2 = disagree, 3 = tend to disagree* on the 6-point scale of the survey) that the SOL reform supported their professional autonomy.

The TPROF was dichotomized so that the teachers' cumulative percent of agreement and their cumulative percent of disagreement could be inspected. The results showed 76% (*n*

= 339) of the participants agreed that their professional autonomy is restricted (item 77) by the SOL system. Similarly, 75% ($n = 342$) indicated that, more specifically, the Standards restrict their professional decision-making in the classroom (item 22) while 76% ($n = 346$) believed that the importance placed upon SOL test scores constrains their professional expertise (item 61). From these results it was observed that three-fourths of the teachers indicated that the SOL reform as a whole and its two major components, Standards and test scores, do not support the teachers' professional autonomy.

Just under half (46%, $n = 342$) of the VEA teacher-members believed the Standards complement their professional decision-making in the classroom (item 4), and slightly less than one-third (31%, $n = 340$) agreed that they were able to meet the educational needs of students with the SOL system (item 76). These percentages appeared rather low in comparison to the percent of participants indicating that the Standards help them remain focused on the subject content (68%, $n = 340$, item 5).

Finally, the responses to TPROF showed that the teachers agreed the SOL system does not help students or parents and the community to respect the teachers and to promote confidence in their classroom decisions (items 66, 17, and 18). Nine percent ($n = 340$) agreed that the SOL system helped students respect the teachers while roughly 28% ($n = 333$) believed that the Standards help parents and the community to respect the teachers and their work as well as promote confidence in the participants' decisions.

Qualitative Results

As indicated by the results of TPROF, the teacher-participants across two categories (*income level of school locality* and *school type*) agreed that the SOL system did not seem to

support the professional autonomy of the teachers. In view of this, the teacher narratives helped further support this trend. However, given the statistical results for TPROF and the category for *status on teaching an SOL-tested subject*, the teachers appeared to differ in their extent of agreement that their professional autonomy was supported by the SOL reform. Teachers of a SOL-tested subject were more likely to report less agreement than teachers who do not teach a SOL-tested subject. However provided the qualitative data, this distinction was not further supported. Though a notable difference was observed based upon the sheer volume of teacher narratives for these two groups.

Of the participants who wrote on the topic of teacher professional autonomy, three times as many teachers of a SOL-tested subject wrote as did teachers who did not teach a SOL-tested subject. Also the participants who are not responsible for teaching a SOL-tested subject wrote narratives that were similar in content to their counterparts who do teach a SOL-tested subject, but those who are not responsible for teaching SOL material wrote considerably less frequently. Teachers who do not teach a SOL-tested subject wrote as though expressing the views of “the collective body of teachers” rather than describing their individual classroom experience.

Across the three categories of teachers, the participants contended that the State-level implementation of the SOL reform is related to four overarching negative consequences that are associated with the concept of teacher professional autonomy. According to the teachers of the study, the SOL program lowers trust in teachers, constrains teachers’ classroom decisions over curriculum and instructional strategies, reduces teachers’ opportunities for classroom creativity, as well as reduces and dulls class discourse and activities, respectively.

Discounted teacher trust. Arguing that they not only hold expert subject knowledge, but that they know their students' learning needs best, the teachers viewed the SOL system as one that undercuts their professional autonomy by means of discounting a trust in the teachers. A fifth grade teacher asserted, "Teacher [or school] accountability has taken away trust in the teaching profession. Does the State believe we are not doing our best, that we're not trained well enough to teach our children?" A first grade history and science teacher (who does not teach an SOL-tested subject) added to this point:

If teachers are entrusted to teach and nurture children, they should have the responsibility to evaluate each child on an individual basis (subject grade, participation, ability to communicate through writing, etc.). State mandated tests do not evaluate children—they evaluate teachers and school systems. "Big Brother" is not helping our children—it is only policing our profession.

Further the teachers not only suggested that they were under some degree of surveillance, but that they opposed the methods used to carry out this measure. On this point a library media specialist wrote, "Teachers should be accountable for doing a good job—the best they can—through observation [and] evaluation by peers [and] administrators. Doctors [and] dentists are not held accountable by how many cavities their young patients have, or how soon they get well! Why should teachers be judged by student performance!"

The participants' reasoned that the State's low trust of teachers was exemplified in the State's lack of opportunities for teachers to engage in active and meaningful participation

in the process of policy development that might have a direct bearing upon curricula and pedagogical practices. In the teachers' lexicon, they desired more "input," more input, more input. The teachers asserted that even if they were not provided the opportunities to have had a greater bearing upon the development of the Standards, at the very least the teachers believed they should have the opportunity to have more input into what Standards should be more or less emphasized or whether to include all the Standards in a course. "The extent to which the current reform has been imposed top down...has continually eroded any hope for 'professional' decision-making," added a teacher of a SOL-tested subject.

The teachers suggested they should be held accountable through the SOL program if they are permitted to develop the criteria by which school quality is assessed and accountability is expressed (high school teacher of science [SOL-tested subject]). A high school teacher of science (SOL-tested subject) listed several points that underscore this trend in the qualitative data: Accountability is ambiguous; some teachers teach only students in advanced courses whereas others teach only students in remedial; and some teach students primarily living in low socio-economic areas. "How can you have one scale for all?" questioned the same science teacher.

Constraint of teacher decisions over curriculum and instructional strategies. The teachers suggested that the SOL system is a set of tools exercised *on* teachers rather than a program into which the teachers might have ongoing input. Although the participants expressed a desire for more input at the policy level, their particular concern was more immediate to the classroom condition: that is, how policy is translated into classroom pedagogical practice and to what extent the policy (the SOL program) seems to compel the teachers to abide by the Standards, to rush through the specified curricula, and to teach to the

test. As such, the teacher narratives suggested that the participants did not believe they have ample autonomy to make decisions related to curriculum and instruction, decisions that are based upon the teachers' expertise and therefore were described by the teachers to be well-suited to the diverse learning needs of students.

Should the State give a broader berth to teachers, the participants believed they could better provide a more hospitable learning environment for students, where the teachers have "rational, studied *control* over classroom activity" (fifth grade teacher) and where the teachers could balance the necessity of teaching rigorous and important subject matter with students' relevant interests and intellectual needs. Paradoxical to the initial intent of the SOL program, the power of the SOL system over teachers' curricular and instructional decision-making, according to the participants, does not ensure that teachers are meeting students' educational needs and teaching important subject matter. Rather, the teachers believed that the system was contributing to the opposite outcome, and in so doing, compressing teachers' creativity in the classroom.

Reduction of teacher creativity. "The SOLs as they are now have virtually eliminated the opportunity for teachers to create exciting, educational units for children. No longer do we have time to expand on the excitement of the children's learning. We are teaching that we don't have time to further explore new ideas, as we must move on the next SOL, regardless of the value or cost" (first grade social studies teacher). Given the breadth and speed of teaching required to cover the Standards as described by the survey participants, creative units or diverse teaching methods they would like to employ are ultimately constricted or delayed until the end of the school year. A high school math teacher's view exemplified this point:

I do feel that the [SOL] tests should not be administered until the very end of the year, if they are going to be an accurate measure of what [students] have learned that year; however, I will say at least [that having the SOL tests administered earlier] gives me a few weeks of opportunities [at the end of the course] to use diverse methods of teaching and technology (other than the TI-85 calculator).”

Teachers tended to associate the constraint upon their classroom creativity as a direct shot at their “individuality” as teachers, often describing significant changes in their pedagogical practices, although generally not in keeping with their better judgment. An eighth grade teacher of “learning disabled” students wrote on this point,

My whole teaching style has change[d]. I feel I teach the skill in a rote manner because the kids don’t get it quickly. The school division is focus[ed] on SOLs—The quote is—Teach the test, not the extras— Everything is SOLs. Teach the skills—Make sure you teach everything. If a large percentage of the school is low-income and low academically, it is hard to get them to pass.

Diminished class discourse. The teachers’ concentration upon upcoming SOL tests and that the SOL program constricts the nature of their pedagogy was a persistent theme in the qualitative data, which appeared to introduce another negative outcome of the SOL reform: diminished class discourse. Without many opportunities for lively or thought-

provoking discussion among the teachers and students, the classroom activities were dulled, according to the teachers. In describing this classroom condition, a high school history teacher wrote simply, “Yawn.”

The combination of what teachers described as over-extensive Standards, often overly difficult Standards, and the speed or pace at which the classroom activities must transpire largely accounted for the diminished discourse in the classroom. Teachers felt compelled to move quickly through the Standards, trying to cover them, and in so doing provided slim opportunities for students to discuss the subject matter as a whole class or in cooperative groups. The teachers believed that the students experience rather dull classroom activities that do not cultivate students’ intellectual curiosity or ability to think critically. How this situation tends to affect the participants’ view of their role as a teacher and their level of tension in the classroom will be discussed in the next major section of the chapter.

Summary

The themes identified in the teacher narratives helped support the results of TPROF and underpinned the pattern that the participants agree that the nature of the Standards, the importance placed on SOL test scores, and the SOL system in its entirety constrains the teachers’ professional expertise and conflicts with their desire to meet the diverse educational needs of students (items 22, 61, and 77). The teachers did not comment on the ability of the SOL system to help students, parents, or the community to respect teachers’ professional decision-making (items 17, 18, and 66); therefore, the results of the scale-level data on this point could not be extended. This was also the case for item 5 of the TPROF, which elicited 68% ($n = 340$) of agreement that the Standards help the teachers remain focused on the

subject matter. However it was clear from the teacher narratives as well as the TPROF that the participants believed that the methods through which Virginia schools are held accountable, diminish the quality of teaching (item 78), and in so doing, diminish teachers' sense of professional autonomy.

Level of Teacher Tension Within the SOL Reform (TTEN)

Inferential Results

Status on teaching a SOL-tested subject. To test for mean differences concerning the participants' *status on teaching a SOL-tested subject* (yes, no), an independent-samples *t*-test was performed. The results of this test yielded $t(129.47) = 4.13, p = .000$ at the alpha level of .05, indicating a statistically significant difference between the means for the population of teachers who teach an SOL-tested subject ($M = 56.73, SD = 10.96$) and for the population who do not ($M = 49.56, SD = 15.59$). Levene's test of homogeneity of variance was conducted and reached statistical significance at the alpha level of .05 ($p = .000$), and therefore the harmonic mean was employed.

Scale Description and Statistics

The TTEN's 14 items (Cronbach's alpha = .89) asked the participants of the study about the level of tension they might experience in the classroom as a consequence of the SOL reform. This included querying the teachers concerning their beliefs about the effects of a classroom emphasis upon SOL test scores, the public comparison of schools' SOL test scores, and the level of stress that teachers and students might experience as a result of teaching and learning SOL subject matter.

Table 9

*Item Frequencies and Descriptive Statistics for Level of Teacher Tension**Within the SOL Reform*

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
1. (19) SOLs, as a guide for student learning, I feel no stress	137 (39.7)	76 (22.0)	71 (20.6)	32 (9.3)	17 (4.9)	12 (3.5)	345
2. (20) Teaching SOLs increase my work-related stress	11 (3.2)	19 (5.6)	34 (9.9)	53 (15.5)	70 (20.5)	154 (45.0)	342
3. (21) Attempting to SOLs is an overwhelming job	15 (4.5)	38 (11.3)	66 (19.6)	87 (25.8)	68 (20.2)	63 (18.7)	337
4. (55) Feel considerable pressure to help raise scores	6 (1.8)	6 (1.8)	14 (4.2)	53 (15.8)	85 (25.3)	172 (51.2)	336
5. (56) Feel pressured to overemphasize test scores	5 (1.5)	7 (2.1)	22 (6.7)	38 (11.5)	62 (18.8)	196 (59.4)	330
6. (57) Do not feel troubled in effort to raise scores	80 (24.3)	76 (23.1)	78 (23.7)	57 (17.3)	22 (6.7)	16 (4.9)	329
7. (59) Teachers of SOL subjects/grades have excessive stress	11 (3.1)	3 (.9)	14 (4.0)	46 (13.1)	52 (14.8)	226 (64.2)	352

Item	Frequency of teacher response						n
	SD	D	TD	TA	A	SA	
8. (60) SOL test system caused me to consider leaving the teaching profession	53 (15.7)	48 (14.2)	50 (14.8)	50 (14.8)	48 (14.2)	88 (26.1)	337
9. (62) Feel undue pressure to help students raise SOL scores	10 (2.9)	16 (4.7)	35 (10.3)	60 (17.6)	86 (25.2)	134 (39.3)	341
10. (63) Feel considerable stress over public comparisons of schools' scores	11 (3.2)	19 (5.4)	39 (11.2)	78 (22.3)	64 (18.3)	138 (39.5)	349
11. (65) Students have excessive stress in trying to pass SOL tests	14 (4.3)	17 (5.3)	40 (12.4)	78 (24.2)	84 (26.1)	89 (27.6)	322
12. (67) SOL system minimizes students' disruptive behavior	223 (64.3)	73 (21.0)	35 (10.1)	11 (3.2)	4 (1.2)	1 (.3)	347
13. (68) Feel frustrated teaching within the SOL reform	13 (3.8)	27 (8.0)	47 (13.9)	92 (27.2)	70 (20.7)	89 (26.3)	338
14. (75) My school system overemphasizes SOL test scores	8 (2.3)	11 (3.2)	39 (11.2)	79 (22.7)	84 (24.1)	127 (36.5)	348

Level of Teacher Tension Within the SOL Reform	Descriptive statistics for scale			
	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>n</i>
	2.32	1.33	.89	14

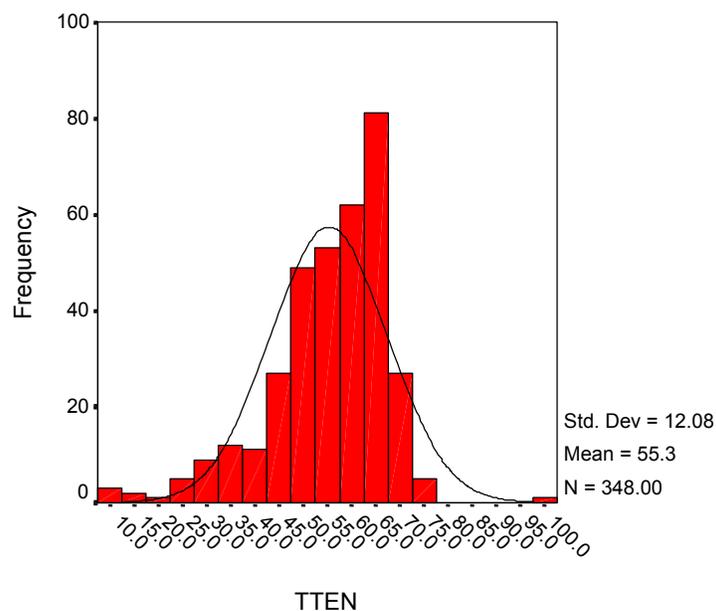
Note. Responses were made on a 6-point scale (1 = strongly disagree [SD], 2 = disagree [D], 3 = tend to

disagree [TD], 4 = tend to agree [TA], 5 = agree [A], 6 = strongly agree [SA]). The numbers in parentheses under Item indicate the number of each item as it appeared on the Teacher Survey. The values in parentheses under Frequency of Teacher Response represent the teachers' percent of agreement. Under the Descriptive Statistics for Scale, n = the number of scale items. Items 20, 21, 55, 56, 59, 60, 62, 63, 65, 68, and 75 were recoded prior to computing the descriptive statistics for the scale.

Teacher Response to TTEN

The item-responses of TTEN were negatively skewed and leptokurtic, with one extreme case identified. Figure 12 displays the distributional shape of these data.

Figure 12. Teachers' Response Frequency for TTEN.



Note. Responses were based upon Likert-type scaling (1 = strongly disagree, 6 = strongly agree).

As suggested by the scale mean, the participants in general believe that the SOL reform is responsible for the teachers experiencing considerable frustration and stress in the classroom. As such, all but one of the item means (item 60) indicated either the teachers' strong disagreement or their disagreement ($1 = \textit{strongly disagree}$, $2 = \textit{disagree}$ on the 6-point scale of the instrument) that the SOL reform did not affect their level of frustration and tension.

The TTEN was dichotomized so that the teachers' cumulative percent of agreement and their cumulative percent of disagreement could be examined. The results showed that both the teachers of SOL-tested subjects and those who do not teach SOL-tested subjects agreed that the SOL program applies considerable stress to their work in the classroom. The teachers of SOL-tested subjects were more likely to indicate *stronger* disagreement than those who do not teach SOL-tested subjects. First, 90% ($n = 330$) of the participants believed that as a result of Virginia policy they feel pressured to place too much emphasis on students' SOL test scores (item 56) and 92% ($n = 336$) agreed that they feel considerable pressure to help students improve their SOL test scores (item 55). The teachers' responses to items 57, 62, and 75 indicated similar beliefs about the tension they experience within the SOL program. Thus the teachers tended to feel considerable pressure as a consequence of State policy and focusing on helping improve students' SOL test scores. Although the teachers as a whole agreed on these points, the teachers of a SOL-tested subject were more likely to indicate a higher level of agreement than teachers who do not teach SOL-tested subjects.

Second, although the teachers overwhelmingly agreed that the focus upon improving SOL test scores places considerable pressure upon them, slightly fewer teachers (82%, $n = 345$) agreed that they feel stress using the Standards as a guide for what students should learn

(item 19). Underscoring this point the teachers indicated that teaching the Standards increases their work-related stress (item 20) and that attempting to cover the Standards is an overwhelming job (item 21), with 81% ($n = 341$) and 65% ($n = 337$) of the participants agreeing, respectively. This trend was consistent for all categories of teachers, with the distinction that teachers of SOL-tested subjects were more likely to indicate stronger agreement than those who do not teach SOL-tested subjects. The general consensus of the teachers was that they experience undue pressure as a result of the SOL reform (scale mean = 2.32, 2 = *disagree* on the 6-point scale of the survey instrument).

Third, 92% ($n = 352$) of the participants of the study believed that teachers of subjects or grade levels for which there is a SOL test are under excessive stress (item 59) while 74% ($n = 338$) felt frustrated teaching within the SOL reform (item 68). Over half (55%, $n = 336$) of the participants indicated that the SOL testing system has caused them to consider leaving the teaching profession (item 60).

Finally, as to the point that the SOL system helps minimize students' disruptive behavior (item 67), only 5% ($n = 347$) of the teachers agreed. However, more than three-fourths of the teachers (78%, $n = 322$) believed students seem to experience excessive stress in trying to pass the SOL tests (item 65). These data suggested that teachers as well as students experience extensive stress in the classroom as a consequence of the SOL program and its ostensible focus upon students' SOL test scores.

Also, it should be borne in mind when considering these data, that while the teachers as a whole agreed that the SOL system places undue stress upon the teachers, the teachers of SOL-tested subjects were more likely to *strongly* agree than were the teachers who do not

teach a SOL-tested subject. However this distinction was not sustained by the themes identified in the qualitative data.

Qualitative Results

Differences among teacher categories. As a result of the analyses of variance for *income-level of school locality* and *school type*, the teachers of the study did not seem to differ in their opinions concerning the tension they reported to experience as a consequence of the SOL reform, for these categories of teachers the thematic analysis of the qualitative data rendered similar results. However, for *status on teaching a SOL-tested subject*, the independent-samples *t*-test suggested statistically significant differences. The content of the teacher narratives did not sustain this trend. The only salient difference that was observed was that the participants who teach a SOL-tested subject wrote 2.76 times more frequently about the tension they believe results from the SOL reform than did their counterparts who do not teach a SOL-tested subject. Even so, the content of the teacher narratives suggested that both of these groups seemed to be negatively affected by the level of tension the teachers reported to result from the SOL reform.

Those who do not teach a SOL-tested subject seemed to share the stress of teachers as a collective body affected by the reform, writing as forcefully as those who teach a SOL-tested subject. Across all categories of teachers, the participants expressed considerable tension that they believe is brought about by the State's current accountability system that assesses school quality by means of SOL test-score results. Due to the similar content of the teacher narratives across the three categories of teachers, the description of the thematic

trends associated with the teachers' reported tension is described in terms of the entire sample of teachers.

Sources of Teacher Tension

Relevant to the teachers' level of tension within the SOL reform, the themes identified in the teacher narratives seemed to consist of three broad areas that the participants believed to be negative consequences of the SOL program. Whereas much of what the teachers described paralleled the content of the TTEN, the qualitative results also expanded this content. According to the teachers of the study the primary sources of their tension included issues of time and materials; parent participation and accountability; teacher professional autonomy and the joy of teaching. Subsequently, these thematic trends are described.

Time and materials. Given the teachers' reported stress over trying to help students' raise SOL test scores, the absence of enough time to "cover" the Standards seemed to be a primary stressor for the teachers. "Teachers are breaking their necks trying to cover all the SOLs" expressed a middle school teacher of a SOL-tested subject.

A high school teacher who does not teach a SOL-tested subject concurred that there is little time for teachers to cover the curriculum: "The time crunch in 11th grade U.S. History is unbelievable. If teachers were only responsible for up to 1945, or if the test were in June, teachers would have more time to be creative. It's hard to teach a year's worth of content, plus stop a week or so ahead to review, by an early May test date."

In keeping with the lack of time to cover the Standards, the participants expressed that teachers are compelled to teach students in a manner of teaching to the test so that

students might memorize as many facts as possible for SOL tests. A fourth grade teacher on this point wrote, “The SOLs force teachers to teach to the test. We are forced to teach students to be able to recite [facts]. We are so pressed for time to teach the students all of the Standards that there is little time for anything else.”

The teachers contended that in attempting to teach all of the Standards it does not ensure that students are learning the content of the Standards. A high school teacher of students with learning and/or emotional difficulties explained “Some of our students are telling teachers, ‘You are going too fast; I’m not getting this!’” The same teacher added, “Our teachers are told what to teach, when to teach it, and how to teach it—Students and teachers alike are becoming frustrated—.” Students’ opportunities to better understand and review the subject matter are being supplanted by the fast pace at which “teachers *must* [double underline] cover as much SOL material as possible”(high school business education teacher in a low-income school locality).

Also of negative consequence within the SOL program as described by the participants was that teachers of SOL-tested subjects expressed they do not have adequate SOL-related teaching materials with which to teach the Standards. The VEA teacher-participants who do not teach a SOL-tested subject did not comment on this issue. Those who do teach a SOL-tested subject expressed that if the schools are headed in the current direction aimed at assessing school quality based upon SOL test scores, then it seemed reasonable to provide teachers with materials that support the Standards. An elementary school teacher exemplified this trend in the teacher narratives:

I feel that the State needs to provide the funding for the proper curriculum in the areas of history and science. Much of my material is attained by me, off my home computer, my internet [sic] access and on my own time. Many of my colleagues are doing the same and I believe that this issue must be resolved!

Others articulated stronger views on the need for SOL-related materials, as did a third grade teacher who wrote “I think that it is absolutely absurd that the State can *demand* that we teach the Standards without giving the appropriate materials to teach them. I have spent *countless* hours searching for information and lesson ideas to effectively teach these Standards.” The teachers of the study used words such as *absurd* and *ludicrous* to characterize their frustration over not being provided with instructional resources that are similar across the State. This view was consistent across various school types and income-levels of school locales. For the teachers, an identical SOL test for all children across the State necessitated the provision of similar SOL teaching materials. On this point an eighth grade science teacher asserted

I think it is a good idea to have a statewide curriculum. This ensures that teachers of the same grade are on the same page. However, I find it ludicrous that the state of V[irginia] states that we must teach specific standards and that every student will take the exact same SOL test at the end of the year. Yet we don't have the same resources. If we all use the same

test, then shouldn't we all be using the same set of classroom materials (books)? Create resources that are based on the SOL!

The absence of enough time and SOL-related teaching materials seemed to place considerable pressure upon the participants as they attempted to cover an extensive curriculum. Some teachers desired practical materials such as vocabulary lists that might help them prepare students for SOL tests with unfamiliar language while others desired books to accompany the SOL content. The participants questioned how they were supposed to provide students with a high quality education as measured by the SOL tests when there seemed to be few SOL-relevant and easily accessible teaching materials available. Thus, the teachers did not believe they were treated fairly by the State, that they were held unjustly responsible for teaching an over-extensive curriculum with inadequate teaching materials to do so. However, the absence of teaching materials and adequate time to teach the breadth of the Standards were not the only stressors upon the teachers. Also of great teacher concern, was that students' parents did not appear to take appropriate responsibility for their children's learning the Standards. On this point, the teachers wrote more voluminously than any other point described in the narratives, with the one exception that will be discussed later in this section of the chapter.

Parent participation and accountability. A clear pattern in the teacher narratives was the teachers' strong desire for greater parent responsibility and accountability for their children in learning the Standards. Overwhelmingly, the teachers viewed the current assessment of school quality as being unjustly weighted, with the teachers shouldering most

of the responsibility for trying to help students' improve their SOL test scores. A high school English teacher in a high-income school locality explained

This system holds teachers and school accountable—But I can't make a child learn. I can motivate, plead and do cartwheels—but the teacher does not produce or cause learning to take place. We structure [and] model but do not “make” a student learn. Where are the parents and the students held accountable? When will attendance be blamed on the student and parent instead of the school?

The teachers described that an unfair burden was placed upon them to improve SOL test scores, which in effect was highly frustrating to them. Many agreed with the idea of school accountability; however, they argued against the current system because it “does not equally distribute the ‘accountability’” (fifth grade teacher). Often, the teachers wrote despairingly that they doubted their school would attain accreditation without the assistance of parents to provide “supervision, structure and support for their children so that they come to school prepared to learn” (sixth grade science teacher). No matter how good a teacher may be, explained the teachers, when students do not attend class, nor complete homework, and remain awake until the wee hours of the morning and come to school tired, teachers should not be blamed for poor SOL test scores. “The teacher,” wrote an elementary school teacher, “becomes the one who shoulders the whole responsibility in terms of the student—presenting material, making sure of understanding, PLUS test-taking skills, emotional well-being, and

self-confidence.” “Students [and] parents should also be held accountable. Who will dare to tell them they are a part of the equation?” (fourth grade teacher of math and science).

The teachers expressed that students of any school—regardless of their local income-level—need to be held accountable through the assistance of their parents. Consistently, the participants of the study explained the need for parents, students, and teachers to work together toward improving student learning and accountability. This included, as described by the participants, parents checking to make sure their child completes his/her homework, reinforcing concepts that have been taught at school, and seeing that their child attends class. After all, noted the participants, it is not fair to hold teachers accountable for students’ performance of SOL tests when the student does not attend class regularly or complete assignments.

In sum, the teachers described intense frustration and pressure stemming from their efforts to teach students who sometimes do not seem to be as invested in passing SOL tests as the teachers are invested in helping the students to pass. Consequently, the teachers did not argue against the notion of accountability itself. Rather, they tended to believe that the schools should not be held more accountable for students’ SOL test scores than the students and parents are held accountable. A high school English teacher explained

I don’t believe that schools should be *more* accountable than *parents* and *students*. We are all in this together! It’s like a 3-legged stool. It won’t stand if one of the legs is missing, and right now, schools are accountable for holding up the whole structure. If each component (school-parent-student) meets its *responsibility*, accountability will take care of itself.

Because the teachers observed such a variety of intellectual and social factors that appeared to be beyond their control, which influenced students' SOL test scores, the participants argued for change in the current SOL system. "Teachers alone cannot remedy the ills of our culture. The SOLs are a shot in the arm. We need fundamental change in the entire system," asserted a high school English teacher in a low-income school locality.

The notion that "equally distributed accountability" is necessary for a fair assessment of school quality permeated the teacher narratives. A high school special education teacher (in a low-income school locality) provided a lengthy but typical description of the teachers' unmet desire for equally distributed accountability:

My biggest problem with the emphasis on SOL scores is that it focuses on teacher—not student or parent—accountability. I know the saying "It takes a village to raise a child" is quoted by many people, but it seems that many parents are waiting for everyone else to be the village. Until schools become a place where students are educated instead of baby-sat, [and where]...teachers no longer spend a quarter of their salary buying school supplies, daily treats to bribe students into behaving and teachers' main purpose is to educate—not be a social worker, surrogate parent, baby-sitter, primary care-giver and mentor, etc., SOL scores will continue to be a detriment to schools that do not have the money and outside resources to make up for the academic and emotional deficits students bring to school with them.

In addition to the participants discussing at length the need for a shared responsibility among students, parents, and teachers for a fair assessment of school quality, the teachers argued for a more fair assessment of student learning by means of measuring students' academic growth. According to the teachers, students might be given a fair opportunity to demonstrate that indeed they are learning and making progress, although they may not yet be passing the SOL tests. In this manner, students may have a fair opportunity to escape high-stakes consequences such as failing to pass a grade level or failing to graduate from high school. Furthermore, the teachers conceded that student growth might be assessed by measuring the same group of student gains from the beginning to the end of a given school year or by measuring the same group of student gains from one year to the next. The focus upon growth, or gains, as viewed by the teachers, might serve a twofold need: to provide a more fair assessment of student learning, thus provide a more fair application of consequences that result from students passing or failing to pass the SOL tests; also, to provide a fair rendering of school/teacher quality where factors beyond the teachers' influence could be better taken into account.

Teacher professional autonomy and the joy of teaching. While the teachers of the study agreed that time and SOL-related materials were lacking for their disposal as was lacking parent accountability for student learning, the teachers articulated a final area that appeared to give rise to the tension they expressed. The teachers discussed that they lack the opportunities to creatively design classroom instruction that best meets the intellectual needs of their students and satisfies the teachers' desire to utilize their professional expertise. The pressure to focus on raising students' SOL test scores negatively transformed their pedagogy, thus intensified their tension, and depleted their joy of teaching.

The narrow focus upon SOL content and the narrowing of diverse instructional strategies, as discussed by the teachers, result from the SOL program's focus upon test-score data by which the schools might be assessed for quality. The qualitative data suggested that the teachers do not wish for curriculum or instructional strategies to be narrowed; however, the power of SOL test scores to signify student learning and school (or teacher) quality tended to compel the teachers, against their better judgments concerning pedagogy, to engage in narrower and less creative teaching practices. A high school science teacher (in a high-income school locality) with numerous accolades amplified this trend in the data:

Please. Get rid of the SOLs...suffice it to say I have been recognized [with] awards for my teaching ability (which is not the point)...The SOLs have been a big cause in having teachers become stressed, to cease or decrease their creativity in the classroom ([because] hey—WE HAVE TO PASS THESE TESTS). If SOLs existed when I was in school, I probably wouldn't be here. You'd think I was an idiot. Instead, I hold a master's in biology...am a GTE Gift Scholar...my county's Gifted education teacher of the year, and I am my school's Agnes Meyer Teacher of the Year nominee (Washington Post). Thank God no one ever made my "stick-me-in-a-room-for-2-hours-and-take-a-test-my-life-depends-on" scores as an indication of who I would become.

Teachers across the three categories examined expressed similar views similar to that above, emphasizing the pressure associated with overly powerful, overly significant SOL test scores

and adding that their emphasis takes away teachers' joy of teaching. A fifth grade math teacher (in a middle-income school locale) offered a characteristic response:

Teachers are losing the joy of teaching. This is easy to understand. We are constantly battered because of scores that are low or lower than those of another school or system. If the scores are good, there's constant pressure to raise them even higher. The only thing important in Virginia education today is the almighty test score. It doesn't matter how stressed the teachers and students are. It doesn't matter how unhappy the teachers are. When I see a child crying during a test, I know that this system is failing him.

The survey participants explained that the focus upon SOL test-scores applied considerable pressure upon not only teachers but students as well. Teachers noted that students seem "burned-out" because the "SOLs become a 24-7 obsession for teachers. It's a performance on everyone's behalf with little to no consideration for the student's emotional well-being. When you're in the classroom everyday, with these students you see it...SOLs create an overwhelming amount of excess stress on everyone and I have seen it manifest itself in many unhealthy ways in our young people" (high school art teacher in a middle-income school locality). Other teachers added that students and teachers are physically sick and "mentally unmotivated" because they "work too long and have little fun" (third grade teacher in a middle-income school locale).

According to the participants, their loss of joy of teaching has led some to consider leaving the field of teaching. Recall that in response to the TTEN scale, over half (55%, $n =$

336) of the participants indicated the SOL testing system caused them to consider leaving the field of teaching. On this point, an elementary school reading specialist wrote, “It stinks! A lot of excellent [and] caring teachers are leaving the profession due to SOLs—that’s all we hear—I only teach reading 3 months during a school year—the rest of my time I am testing, testing, testing (per orders of the school admin.). Our children need to be TAUGHT.” The same teacher continued, “There is no excuse for kids not learning how to read...I’m leaving the profession—Too much BS!! TEACH THE CHILDREN!”

Given the theme of teacher tension, some participants voiced unambiguous intentions similar to that above, and in so doing, indicated a severe and negative influence of the SOL program upon VEA teacher-members across the State. A participant (fifth grade teacher) who taught 25 years in Montana expressed

I am terribly disappointed in teaching in Virginia. It has taken my love of teaching away. I know many young teachers that are quitting! They don’t want a teaching career and I don’t blame them. I have 5 years to be vested in VA Retirement and I’m quitting! I might work in a private school where they have no SOLs and honor their teachers. Something needs to be done about the SOLs! [The participant drew a frowning face at the conclusion of his/her response.].

Although the teachers tended to view that the State was “losing too many good teachers because of the demands put on them due to the SOLs” (high school teacher), clearly, the power and significance of the SOL test scores was the central source of teacher tension,

frustration, and unrest. Consistently, the participants noted that “true” student learning could not be measured alone by the SOL tests, nor could the resulting test scores as they are currently used signify the quality of student learning and school/teacher quality.

Additionally, as for the Standards, a math and science technology teacher (of an elementary school) offered a plain summary of the teachers’ perspectives, “High standards and high expectations are held by all great teachers. [The] SOL school reform does not make poor teachers have high standards; it simply makes good teachers have high stress and doubts about their career choice.”

Lengthy narratives suggesting the teachers’ level of tension underscored some teachers’ intent to retire early or leave the field of teaching. A high school history teacher (in a middle-income school locale with a student population representing 22 countries) expanded this trend in the qualitative data:

I am taking early retirement this year (50/26) partly due to SOL requirements. It consumes all school and department meetings, has added stress for all my colleagues and has caused bitterness between the elective and core teachers, who feel the burden. School is not a lot of fun. Students are extremely stressed. The tests invalidate the hardworking honor student’s success who fails a test, and this is quite common. I shudder to think what the state has spent and Harcourt Brace has made on this fiasco. The stress of teaching and accountability both increase while teachers fall farther behind the national average in salary. Field trips and school assemblies have been curtailed so students have more time to cram.

In conclusion, the participants expressed an undue burden and excessive stress in trying to meet the demands of the SOL program while observing what they believed to be the system's inadequacies and unjust treatment of teachers and students alike. While the teachers as a whole elucidated their frustration within the current system, some experienced such a high level of tension that they intended to leave the field of teaching, although they indicated a past joy in practicing their craft.

Summary

The themes identified in the teacher narratives supported the results rendered from the analysis of TTEN and corroborated the pattern that as a result of Virginia policy, the participants agree they feel pressured to place too great an emphasis upon students' SOL test scores. Thus the teachers expressed considerable pressure teaching within the system and helping students improve their SOL test scores (items 20, 55, 56, 57, 59, 62, 68, and 75). The qualitative data added to this trend, explaining that the "power of the almighty test score" as well as the extent and use of the Standards (items 19 and 21), constrained the teachers' creative use of curriculum and diverse instructional strategies. As a result, the participants described their diminished joy of teaching.

Additionally, in terms of the 5% ($n = 347$) of the participants indicating that the SOL system helps minimize students' disruptive behavior (item 67), the teacher narratives did not lend specific support. However, from the narratives it might be suggested that this point is loosely connected to the teachers' frustration over student absences (as well as a lack of student motivation to learn the Standards and a lack of homework completion). As such, the teacher narratives might suggest that "disruptive" student behavior could include that which

opposes what the SOL system appears to demand. Whether this is the case, the quantitative data as well as the qualitative data lent validity to the view that the teachers believed students experienced considerable stress in trying to pass the SOL tests (item 65).

Finally, the issue that was not captured in the quantitative data but that became clearly discernable from the analysis of the qualitative data was the considerable tension that teachers experienced over their perceived lack of parent participation and accountability for their children learning the Standards. Teachers articulated the view that the responsibility associated with school quality and accountability should be shared among the student, parent, and teacher. Anything less magnified the teachers' tension. The unjust degree of responsibility placed upon the teacher-participants to help improve students' test scores while their professional autonomy and creativity are constrained seemed to provide an environment in which some teachers appeared to be pushed beyond their limits. Consequently some considered taking early retirement or leaving the field of teaching altogether (item 60, 55%, $n = 336$).

CHAPTER FIVE

Teachers' Perspectives on the SOL School Reform

This chapter begins with a brief summary of the study's findings about the VEA teachers' perspectives on the SOL school reform. The chapter then presents the conclusions of the study in a dialogic format whereby the teachers' views are juxtaposed to multiple perspectives within the public discourse on high-stakes accountability-based school reform. Theoretical and practical considerations that relate to the study's results are discussed as well. Finally, the limitations of the present research are discussed and suggestions are made for future research.

Summary of Findings

Adequacy of the Standards and the Diversity of Instructional Strategies

The teachers of the study reported that the Standards did not represent reasonable content for teaching and learning, which appeared in large part due to the difficulty level of the SOLs. Furthermore the Standards did not help students gain a world-class education and prepare them for a global economy. However about half of the teachers believed that the SOLs helped provide a more equitable education for students. Teachers held this view while noting that students experience excessive stress in trying to pass the SOL tests.

The Standards, according to the study's participants, were unreasonable in terms of their developmental and age-appropriateness or in terms of their precedence over vocational education and optional coursework for non-college bound learners. Many teachers viewed the Standards as college bound or enrichment material only, and as such the SOLs did not

meet the learning needs of most students. Also the participants were concerned for students with learning difficulties and young people in need of special education. The coursework was too difficult for these students and some seemed likely to drop out of high school at the age of 16. In addition to viewing the Standards as too difficult for most students, the participants believed that the SOLs were too broad in scope.

The breadth of the Standards compelled the teachers to rush their teaching in an effort to cover the SOLs. This effort narrowed the classroom curriculum, compressing it to that which would likely be tested on a SOL test. Consequently, teachable moments were stifled and teachers felt compelled to reduce their use of diverse instructional strategies. Teachers reported that the hurried classroom impeded students' opportunities to think critically about the subject matter. Student learning was not of high quality. In sum, this appeared to be due to the emphasis upon too difficult, over-extensive Standards that ordered a hurried pace and high-pressure classroom environment where teachers and students leapt from one topic to the next. As a result of the hurried, high-pressure classroom, children were unable to learn material better suited to their intellectual needs or interests. Some teachers suggested that the more technologically advanced era of the 21st century should mean that students are provided with more—rather than fewer—diverse learning opportunities.

Testing and Accountability

Although the teachers asserted that the Standards of Learning inadequately specify what students should learn, they also indicated that the SOL testing technology was more problematic than the SOLs, for a number of reasons. First, SOL tests were viewed to lack the ability to measure student knowledge. The cut-scores that students must attain to pass the

tests were thought to be arbitrary and have little to do with students' mastery of a subject. As such, the teachers strongly contended the use of SOL test scores as the primary criteria for what students have learned and whether they should graduate from high school. Little value was placed upon the use of the SOL test scores to determine course placement and whether students should be retained in grade. According to the participants, the singular power of the SOL test scores lacked fairness toward students, particularly since the teachers deemed that the scores did not adequately reflect student academic growth. The teachers argued for multiple measures of student assessment as well as measurement of all students' academic growth from year to year. The teachers viewed the SOL test scores as possessing punitive power rather than diagnostic usefulness.

Similar to their view that the SOL test cut-scores were arbitrary and inadequate to represent student learning, the participants believed that the pass rates were inappropriate for representing school quality, for improving teaching in the classroom, and for determining school accreditation. The SOL test scores were thought to be a singular and simplistic representation of student learning and school quality. The scores were insufficient to adequately reflect the quality of schools, provided the diverse populations of students from school to school. However, it is important to note that there was some difference in opinion between the elementary school teachers who *strongly disagreed* with the adequacy of SOL testing for school accountability purposes and the high school teachers who *disagreed*. Moreover, less than half of the teachers believed that comparisons between the SOL test scores of their school and other schools help hold their school accountable. The inadequacy of the SOL test scores to take into account the diversity of student populations when

representing school quality was intensely frustrating to the participants, as will be discussed in the subsequent section of the chapter.

Teacher Professional Autonomy and the Level of Teacher Tension

The participants of this study asserted that the SOL system in its entirety constrains their professional expertise and conflicts with their desire to meet the diverse educational needs of students. There was evidence to suggest, however, that teachers of a SOL-tested subject experience this constraint more than teachers who do not teach a SOL-tested subject, although both groups indicated substantial constraint on their professional autonomy. This suggested that the SOL program not only had the power to diminish the professional autonomy of teachers who teach a SOL-tested subject, but that the program penetrated the work of teachers who did not teach a SOL-tested subject, and they too sensed diminished autonomy. In addition to this undesirable outcome of the SOL program, a smaller number of teachers argued that the SOL system generated conflict between teachers who do and those who do not teach a SOL-tested subject. Some teachers of SOL-tested subjects, especially at the high school level, asserted that their colleagues who did not teach a SOL-tested subject are sometimes critical and unsympathetic toward the others' high-pressure work. Thus, the time that the teachers spend meeting with one another is sometimes contentious rather than productive.

There were four major issues related to the SOL program that characterize the teachers' diminished professional autonomy. First, the program lowers trust in teachers. Second, it constrains practitioners' decisions over curriculum and instructional strategies. Third and fourth, the program reduces teachers' opportunities for classroom creativity, as

well as reduces and dulls class discourse and activities. However, sixty-eight percent of the participants noted that the Standards help teachers remain focused on the subject matter while they did not believe that the SOL system as a whole helped students, parents, or the community to respect teachers' professional decision-making.

In addition, the participants contended that the methods by which the schools are held accountable in Virginia diminish the quality of teaching and thus diminish teachers' sense of professional autonomy. The participants' sense of diminished autonomy contributed to the excessive tension they experience in the classroom. The teachers reported that they feel pressured to place too great an emphasis upon students' SOL test scores and helping students try to improve their scores. This point was underscored by what a teacher described as the "power of the almighty test score." Furthermore, enormous tension was produced by the absence of enough time to "cover" the Standards and the lack of adequate SOL-related teaching materials with which to teach the SOLs.

The teachers reported that an unfair burden was placed upon them to improve SOL test scores. While many agreed with the idea of school accountability, they argued against the present SOL program because it "does not equally distribute the accountability." The teachers strongly desired greater parent responsibility and accountability for their children in learning the Standards. They noted that parents could help discourage student absences/cutting class and encourage the completion of homework and an adequate night's sleep for their child. If the school, parent, and the student were jointly responsible for student learning, the teachers believed that an accountability system that links test scores with rewards and sanctions would be better justified.

Finally, because the teachers lacked opportunities to creatively design classroom instruction that best meets the intellectual needs of their students, the participants' joy of teaching was depleted. The teachers reported that an unjust degree of responsibility is placed upon them to help improve students' test scores while their professional autonomy and creativity are constrained. Given the undue tension that emerges from this classroom environment, some teachers indicated they were pushed beyond their limits. As a result, more than half considered taking early retirement or leaving the field of teaching altogether.

Conclusions: Juxtapositions of Perspectives

Chapter one underscored the problem that teachers' perspectives have been largely excluded from the public discourse on contemporary accountability (Goodlad, 2000; Rapp, 2000; 2001). Further, it was discussed that there is need for fuller disclosure of the possible undesirable effects of the SOL program (Turner, 2001, B7). To add breadth and clarity to the public discourse pertaining to these problems, the present chapter juxtaposed the results of the study to multiple perspectives within the public debate on high-stakes accountability-based school reform. Thus in the following section, teachers' views on the SOL program were compared to the perspectives of policymakers and business leaders, the public at large, parents, and scholars in the field of education.

Perspectives on the Adequacy of the Standards and the Diversity of

Instructional Strategies

Content standards and equity. The participants of the study indicated that the Standards of Learning do not represent a reasonable set of standards for teaching and

learning, rather are over-extensive in scope and are too difficult for most students; they represent enrichment material. Nor did the participants believe that the Standards help prepare students for the global economy of the 21st century. Further, less than half of the teachers (48%) suggested that the SOLs help ensure that students receive an equitable education. This view appeared to be grounded in the teachers' rationale that teaching uniformly high SOLs does not accommodate or meet the diverse intellectual needs of all students.

Robert L. Linn writes in *Assessments and Accountability*, "It is quite possible to have high standards without the standards being common for all students. High standards of performance for a given grade level do not necessarily mean common standards for all students" (2000, p. 10). Added to this view, the teachers suggested that the notion of equity does not simply denote uniformity. Equity denotes evenhandedness, justice, and impartiality. Students arrive in the classroom with a diversity of learning needs, and the role and responsibility of teachers is to know the learning needs of their students so that the teachers might better transform the curriculum to meet the diversity of students' academic needs and interests. These teacher perspectives stand in sharp contrast to the views of many policymakers and business leaders.

Policymakers and business leaders cite the Third International Mathematics and Science Study (TIMSS) (Schmidt & McKnight, 1998) as justification for the new thrust toward state-specified rigorous standards. TIMSS indicates, "that, relative to students in other nations, American elementary students are among the leaders, our middle school students are in the middle of the pack, and our high school students bring up the rear" (<http://www.achieve.org>). Also the report suggests that these "results confirm that the problem is with the

system, not with our students, and that we must continue to raise expectations and performance across the board” (<http://www.achieve.org>). To address the weakened state of American public schools, according to many business leaders and policymakers, the schools must focus upon core academics in mathematics, science, and language skills.

Virginia policymakers, specifically the Virginia Department of Education and the Virginia School Board, along with the support of many business leaders assert that the uniformly high Standards of Learning are a desirable substitute for a previous statewide “system of inconsistent, often low, standards” (Christie, 1999, p. 37). The Standards, provided this view, represent reasonable and rigorous curricula for all students and will improve equity, increase academic achievement, and prepare students for the global economy of the 21st century. For these goals to be accomplished, teachers should align the classroom curriculum to the Standards regardless of whether this requires “teaching to the test.”

Juxtaposed to the perspectives of teachers and the views of policymakers and business leaders, parents in general seem to support the notion that high academic standards and good skills taught in public schools should improve equity. A national survey conducted by Educational Testing Service (ETS, 2001, <http://www.ets.org>) to describe the views of public, parent, educator, and policymaker attitudes on [high-stakes] education reform adds to the dialogue on standards and equity. The study involved focus-group discussions and a survey including approximately 200 primary- and secondary-school teachers among its 1, 054 participants of adults and parents of school-aged children (<http://www.ets.org>). Provided the ETS report, *A Measured Response*, a New York parent stated, “I think education should be the equalizer in our community. I think that good skills, regardless of where you come from, will allow you to be competitive.”

Diversity of instructional strategies and subjects not tested on high-stakes tests.

Although parents support high standards within the contemporary reform, they are concerned that subjects that are not tested do not receive adequate attention in class and that teachers emphasize test-taking skills rather than broader educational activities (<http://www.ets.org>; Rose & Gallup, 2000). Results from this study's Teacher Survey indicated that 82% of the participants frequently use practice tests to prepare students for SOL tests. According to these data, it seems that parents have a legitimate concern about the overemphasis on test preparation.

The Virginia Department of Education (VDOE) contends that there is “nothing wrong with ‘teaching to the test’ if the underlying academic standards and material taught are sound (<http://www.pen.k12.va.us/VDOE/PolicyPub/Parents/>). The Department buttresses its contention with evidence that suggests the SOLs are among the best in the United States. Outside groups such as the American Federation of Teachers (AFT), the Fordham Foundation, and Education Week's *Quality Counts* annual national study indicate that the SOLs are at or near the top of all states' standards (<http://www.pen.k12.va.us/VDOE/PolicyPub/Parents/>).

The public at large adds to the discourse concerning the adequacy of high academic standards by questioning whether a single focus upon academics is best for students. The public believes that high academic standards are important, but there is evidence to suggest that the public is almost twice as concerned with the character of young people (Song, 1999, p. 62), which is not tested on high-stakes examinations. The 1999 LIFE Education Poll indicates that the public desire for schools to contribute to the development of young people's character has remained relatively unchanged since 1950 (p. 62).

However, since 1950 a shift in public opinion has occurred regarding the importance of students having an academic background versus vocational training. Song (1999, p. 62) reports that in 1950 41% of the public believed vocational training was the most important thing “young people should get out of high school,” while in 1999, 21% of the public believed vocational training was most important. In addition to this trend, the LIFE poll (p. 62) indicates that currently there appears to be a relatively even split between the public’s desire for students to acquire an academic background (25%) and for them to receive vocational training (21%). The teachers of the study share the public’s perspective that vocational education and some subjects that lie outside the content of the SOLs are important and necessary to help meet a diversity of student learning needs.

Worth of the curriculum and the quality of education. With the increased emphasis upon academics and basic skills as evidenced in high-stakes accountability programs such as the SOL reform, policymakers and business leaders might suggest that “schoolchildren today are taught more worthwhile and useful things than children were 20 years ago”; however, the public seems less confident in the value of what is being taught today (Song, 1999, p. 62). In 1950, 67% of the public indicated that schoolchildren were being taught more worthwhile things than students 20 years earlier, whereas in 1999 just 26% of the American public believed this to be so (p. 62). That many policymakers and business leaders attach great importance to uniformly high academic standards, with little emphasis upon vocational education and subjects not tested on high-stakes examinations, suggests a considerable departure from the views of the public at large, the teachers of this study, parents, and scholars in the field of education.

Similar to other scholars in the field of education who have discussed the adequacy of content standards within the context of the contemporary reform (Linn, 2000; Resnick & Resnick, 1992; Stake, 1999), Madaus contends that the classroom curriculum is narrowed “and encourages undue, even exclusive, attention by teachers and students to the material covered in the exam” (1991, p. 228). Madaus continues

Instructional and study time spent on various aspects of the curriculum correspond to the likelihood of their appearance on the test. And worthwhile educational objectives and experiences may be excluded from the classroom simply because we do not know how to test them properly. Teachers and students employ a number of strategies and tactics whose sole purpose is to improve test performance. In effect, the examinations may come to determine the shape of the curriculum rather than the curriculum determining the shape of the examinations (p. 228).

Parents and the public at large articulate similar views to the aforementioned. However, many policymakers and business leaders believe that state-specified content standards offer no cause for concern in the classroom. Rather, contemporary content standards supply the classroom with a rigorous, uniform curriculum for all students, whereby teaching to the test is an acceptable consequence of the reform effort. Contending this perspective, the teachers of the present study asserted that the very content objectives set by the State to improve the quality of education for all students, in effect, undermine the quality of students' education because a narrow focus upon the Standards does not adequately

address the diversity of student learning needs in the classroom. Also the hurried pace at which the teachers must cover the over-extensive SOLs reduces teachers' diversity of instructional strategies and impedes students' sustained and critical study of the more difficult Standards.

Testing and Accountability

Test validity. Overwhelmingly, the teachers of the study believed the SOL tests lack the ability to measure student knowledge and assess school quality. The participants made a case against the tests by asserting that the SOL tests were too brief to assess the extensive SOLs, and that a single measure was insufficient to assess student competency in a subject area and to judge the quality of schools. To counterbalance the arbitrary nature of the tests' cut-scores and to improve the validity of student and school assessment, the teachers believed that multiple measures of assessment should be used. The Virginia Parent Teacher Association agrees that multiple criteria should be used to determine student achievement because "There is absolutely no way a 40- or 50-item multiple-choice test is going to fully assess all of the standards in a curriculum" (The Roanoke Times, 2000, p. B8). Cross (2000), a measurement expert and author of *The Standards of Learning (SOL) Reform: Real Consequences Require Real Tests*, affirms the parents' position: "And test length does matter! Not only is the test able to sample more adequately the skills and knowledge to be tested, but a fundamental fact about test scores is that reliability increases with test length."

High-stakes and the quality education. The teachers of the study asserted that multiple modes of assessment were imperative particularly when the results were used to make high-stakes decisions related to student high school graduation and school

accreditation. Clearly distinct from the teachers' perspective on the use of SOL test scores, many policymakers and business leaders assert that young people's scores on high-stakes examinations adequately represent student academic attainment, therefore the scores should be the decisive factor in whether students graduate from high school. Similarly test scores are legitimate indicators of whether schools receive state accreditation (Achieve on the Web). Mark Christie, a member of the Virginia Board of Education notes "[T]o decouple the SOL tests from the consequences linked to them would diminish, if not destroy, the powerful incentive now driving the schools to increased focus on student academic achievement for the broadest possible range of students" (1999, p. 36).

VanDerwerker, organizer of the 5,000-member parent organization Parents Across Virginia United to Reform SOLs (PAVURSOL), contends that higher SOL test scores and pass rates do not signal a higher quality education (Turner, 2001). Countering this claim is the group Parents and Students Supporting the Standards of Learning (PASS-SOL), which defends the SOL policy (Turner, 2001). In considering these two perspectives represented by the two parent organizations it seems that the former is gaining momentum, as suggested by a survey of Virginia parents conducted by Bonfadini (2000).

Bonfadini (2000) showed that the parents surveyed believed that SOL tests do not lead to a significant improvement in the quality of education for their children. The parents indicated that the stakes associated with the SOL tests are punitive and that the test scores should not be used as the sole criteria for high school graduation (2000). Amy Proctor, a parent of Quicksburg, Virginia believes the SOL tests are more harmful than helpful: "I think the fundamental reasoning behind it is a good thing. Someone should be accountable for what our children are learning....But we focus so much on the tests that part of the fun has

been taken out of class” (The Bristol Herald Courier, 2000, 12A). Moreover, Bonfadini (2000) reports that the parents surveyed tended to believe that the SOL tests are more politically motivated than they represent a high quality, or sound, educational strategy (2000). “With recent polls,” including a poll conducted by the Washington Post, “showing that a majority of Virginians think the SOLs are not working, VanDerwerker and [PAVURSOL] are pressing state legislators to halt the use of the tests alone to assess schools and students” (Roanoke Times SOL face test in gen assem; Bristol Herald Courier, 2000).

Fairness of SOL testing. To address parents’ suggestion that the SOL tests are flawed, the VDOE asserts that the tests have received high marks from outside testing experts and that SOL test-score data is similar to that of the Standford-9 and the NAEP. Even so, parents suggest that it is unfair to base a student’s future on one SOL test when the student could have had a bad day at the test. However, since the SOL tests are not timed, argues the VDOE, students do not experience a time pressure, which presumably alleviates the brunt of the “bad test day” phenomenon. Also students are allowed to take a failed test over if they need the test in order to graduate from high school. On this point one should bear in mind that students must wait until the next SOL test administration period to receive this opportunity.

For some time the VDOE has assuaged parents’ concern about students not able to graduate due to failing a SOL test by reminding parents that SOL test scores will not be used to make this high-stakes decision until 2004 ([http://www .pen.k12.va .us/VDOE/Policy Pub/Parents/](http://www.pen.k12.va.us/VDOE/PolicyPub/Parents/)). At present, the year 2004 does not lie far ahead and parents appear concerned about the fairness of testing transfer students who are “subject to the same requirements for verified units of credit for graduation as students attending Virginia schools since kindergarten” (<http://www.pen.k12.va.us/VDOE/PolicyPub/Parents/>). Further, they are

concerned about the negative effects that the SOL system might generate for poor, minority or immigrant students. Parents wonder if schools with large numbers of poor, minority or immigrant children will be able to meet the requirements of the high-stakes accountability system. In response to this query, the VDOE indicates that such a question only suggests a belief that children cannot achieve to high standards because they are of a certain race, ethnicity, or income level. Juxtaposed to this point, the participants of the study stressed that the Virginia accountability program does not improve equity for school children. Rather, the SOL program diminishes the quality of education for students attending schools in low-income areas or schools with ethnically diverse student populations and treats these schools unfairly because the program does not account for all students' academic growth from year to year.

Like the teachers of this study, Cross (2000) contends that "Under the present system, where SOL test scores measure achievement across several school years, it is not clear who is accountable. With testing at every grade level, accountability can be based on an improvement model rather than the current model that requires all students and schools to attain the same uniformly high performance standards that are unrealistic for many." Cross adds,

Waiting until the third grade to impose high-stakes tests is too late to help those having difficulties with the three Rs. And when children with deficiencies are identified in the first grade or before, our leaders in Richmond should be held accountable to do all they can to help these students, their teachers, and their schools, rather than propose policies that

blame the teachers, retain the students, and “dis” accredit the schools. The deeply rooted social and economic problems associated with poor test performance cannot be resolved by shaming the victims or the schools.

Many policymakers and business leaders, however, support the coupling of state-specified standards to tests that are associated with serious consequences for schools and students. In Virginia, business leaders and policymakers view the SOL tests as valid instruments, therefore yielding accurate portrayals of student academic attainment and public school quality. Christie argues that the state of Virginia has “gone to extraordinary lengths to ensure that our SOL tests are correlated with the material actually being taught in Virginia classrooms.” Further, he asserts that the State does not “automatically accept the test items prepared by the test contractor”; rather, the State subjects “all proposed test items to a rigorous screening process in which Virginia classroom teachers play the largest role” (1999, p. 37).

Content-review committees that are “dominated by public-school educators and bias-review committees that include groups such as the NAACP” screen proposed SOL test items for association with the Standards, reading level and age-appropriateness, as well as screen the importance of the subject content and remove any ethnic or gender bias (Christie, 1999, p. 37). Also, the Virginia School Board, being informed by a SOL test advisory committee appointed by school board president Kirk Schroder, seeks continual improvement of the SOL testing program (p. 37). Christie suggests that these efforts to make the SOL tests as fair and effective as possible warrant trust in the tests to assess the academic achievement of students

and the quality of Virginia schools. As such, according to this perspective, the consequences associated with the SOL test scores are justifiable.

Diagnostic use of SOL test scores. Moreover the state of Virginia suggests that SOL test scores are diagnostically useful not only in helping teachers pinpoint students' learning difficulties and teachers' weak instruction, but also are useful for making decisions concerning students' course placement and whether students should be retained in grade (<http://www.pen.k12.va.us>). In contrast to this perspective, the teachers viewed the SOL test scores as possessing punitive power rather than diagnostic usefulness. Teachers found little use of the test scores to help them make decisions about course placement and whether students should be retained in grade. Less than 35% of the participants believed the scores were useful for pinpointing students' learning difficulties and for identifying weak instructional practices. Robert Stake (1999, p. 670), author of *The Goods on American Education*, agrees that high-stakes assessments and their test results are not used diagnostically and contends that the overemphasis on testing "erodes confidence in legitimate teaching competence."

Purposes of high-stakes testing. The SOL tests and the consequences related to the test results are deemed by the business community and policy makers as a necessary system of high-stakes accountability that directs and oversees public schooling so that the desired results of these constituencies are accomplished. According to Christie, the goal of the SOL reform is to ensure "that all our public schoolchildren get the opportunity for a better future by giving them the skills they need to be successful in a global economy and to be informed, responsible citizens of our commonwealth" (Christie, 1999, p. 37). SOL tests and associated consequences provide the "powerful incentive" to do this (p. 36).

The public at large holds a similar perspective to many policymakers and business leaders in that it believes there should be a system of accountability. These groups seem to share the view that academic standards coupled with accountability mechanisms are warranted on the basis that “taxpayers have every right to expect a measurable payback” from their investment in public education (McGinn, 1999, p. 48). The results of a national survey by ETS in 2001, *A Measured Response* (<http://www.ets.org>), underscore this point.

A Measured Response involved “perhaps the most comprehensive study done to date of public, parent, educator, and policymaker attitudes on [high-stakes] education reform” (<http://www.ets.org>). The study involved focus-group discussions and a survey including approximately 200 primary- and secondary-school teachers among its 1, 054 participants of adults and parents of school-aged children. According to the ETS report, the survey results were shared with the President’s leading advisors, Congressional staff members, business leaders, educational organizations, and the American news media. A synopsis of the study’s results follows:

This survey confirmed that education is the single most important issue related to the American quality of life. The public, politicians, and policymakers alike know this. The public also believes that we have problems in and with our schools, but its faith in public education remains strong. The people want our schools fixed, not abandoned. They want to invest in schools, teachers, and students, not punish them. They understand that money alone will not fix everything; however, they believe that real money is certainly needed to solve many problems. The investment, our

survey tells us, must come with some strings attached in the form of recognized standards and accountability. On this the public is clear—it wants to know what it is buying for the tax dollars it spends. This is a typical American response, pragmatic and centrist (<http://www.ets.org>).

According to the ETS report, “the public comes down squarely in favor of testing as one element in an overall education-reform proposal . . . Americans favor testing in the broader context of reform by 68% to 22%” (<http://www.ets.org>). Specifically, the public at large “strongly endorses annual state tests for students in grades three through eight and a nationally standardized test to confirm the comparability of the state tests” (<http://www.ets.org>). Although this public perspective supports the use of testing as one element of school reform, the ETS report also states that the public’s view is tempered by its belief that caution should be applied to test use.

Cautions applied to testing and accountability. The rationale for the public’s cautionary stance is indicated by the percentage of all adults surveyed who were “very concerned” with the following consequences that appear to be associated with testing: (1) an overemphasis on standardized test scores and their use to replace broader assessment methods (38% were very concerned); (2) teachers teaching to the test and emphasizing test-taking skills rather than broader educational activities (37%); and (3) test bias against economically disadvantaged students, minorities, or students who are not yet proficient in English (38%). Roughly similar results were found regarding the overemphasis of testing as reported in The 32nd Annual Phi Delta Kappa/Gallup Poll of the Public’s Attitudes Toward the Public Schools (Rose & Gallup, 2000).

Scholars, too, discuss concerns similar to the public and the classroom teachers of the present study (Holman, 1995; McGill-Franzen & Allington, 1993; Rothman, 1996; Stake, 1999; Suarez & Gottovi, 1992; Whitford & Jones, 2000). Holman (1995) conducted a study involving a sample of 363 fifth grade students of whom 191 did not master one or more areas of the Texas high-stakes examination, the Texas Assessment of Academic Skills (TAAS). Holman found that students' ethnicity and socio-economic status (SES) were positively predictive of TAAS status. Children from higher SES homes and White students were more likely to pass the TAAS (2000). Although these findings suggest unjust consequences of high-stakes accountability programs, the full consequences of testing and accountability are not likely to be discerned soon, posits Stake (p. 670). "Lacking an adequate knowledge base, education policy needs to be based on deliberations, on long and studied interpretations of assessment, on experience and on ideology. When professional wisdom gets little respect, as now, this is an unlikely prospect" (Stake, 1999, p. 670).

Teacher Professional Autonomy and the Level of Teacher Tension

Paradoxical consequences of the reform. While there is ample discussion of rigorous content standards, teachers' instructional strategies, and testing and accountability by groups of parents, the public at large, and policymakers and business leaders, the issues of teacher professional autonomy and teacher tension in the classroom do not seem to be discussed except by scholars in the field of education (Borko & Elliot, 1998; Cochran-Smith, 2000; Darling Hammond, 2000; Darling Hammond & Wise, 1983; Goodlad, 2000; van den Berg & Ros, 1999; Ryan & Cooper, 1998; Soder, 1986). Scholars open the dialogue on these issues and offer perspectives that tend to overlap with the views of the teachers of the current study.

For example, in discussing how educational policies affect the classroom, Darling-Hammond and Wise (1983, p. 68) report that the more prescriptive the policies, the more teachers feel constrained in meeting what they believe are students' learning needs. Teachers who feel that they are not allowed to sufficiently draw upon their professional expertise indicate dissatisfaction with teaching and are more likely to report that they will leave the profession of teaching (p. 68). "Paradoxically," note Darling-Hammond and Wise, "some policies designed to prevent incompetent teaching discourage the efforts of highly motivated and competent teachers" (p. 68). Shepard (1991, p. 27) adds that

Forcing modes of instruction via external high-stakes assessments detracts from the professional role of teachers. It trades making the worst 10% of teachers better by fiat against empowering the other 90%. Both of these concerns can be alleviated, or course, if the assessments are sufficiently broad so that tasks are not pre-specified and taught to, and there are multiple paths to successful performance. But in litigious environments these features are often negotiated out of testing programs because there is safety in specificity.

Similar to Shepard's (1991) view, some teachers of the present study asserted that the SOL system does not improve bad teachers. Rather, good teachers begin to question their career choice.

Teachers' desire for ongoing participation in policy development. According to the participants, the SOL program undercuts teachers' professional autonomy by means of

discounting trust in teachers. SOL policy overrides the professional expertise of classroom practitioners and provides little room for teachers' ongoing input into policies that have a direct bearing upon the classroom curriculum and teachers' instructional strategies. Teachers were frustrated over how the SOL policy was translated into classroom practice. Specifically the participants believed that the SOL program compelled them to rush through the Standards and to teach to the test, at the expense of high-quality teaching and learning. The teachers' autonomy was limited in that the teachers did not believe they had ample decision-making power to select curriculum and employ diverse instructional strategies to meet a diversity of student learning needs. Similarly Darling-Hammond and Wise (1983, p. 68) state that teachers believe that curriculum and testing policies are counterproductive to classroom teaching when they limit what teachers can teach.

Constraints on curriculum and instruction. From narrowed curriculum choices and a reduction in the diversity of their instructional strategies, the participants of this study reported few opportunities to create intellectually stimulating units for school children. Students' interests and excitement in a given subject area could not be expanded upon due to a lack of time and high-pressure to cover the SOLs. That the teachers were unable to provide creative units for their students went against the teachers' better judgment. Teachers noted that due to the reform their entire approach to teaching had changed in undesirable ways.

Within a system such as the SOL, teachers find it difficult to engage their students in writing projects or in classroom discussions about interesting ideas (Darling-Hammond & Wise, 1983, p. 68). They feel they have little time for educational activities that "are not geared toward discrete cognitive skills that will be tested on multiple-choice tests used for promotion purposes, tracking purposes, or accountability purposes (p. 68). Further, teachers

“cannot pursue topics of the children’s interest because they are supposed to...achieve certain objectives by the end of the classroom period. They feel constrained in their ability to meet what they see as the needs and interests of their children” (pp. 68-69).

In conjunction with few opportunities for creative units of study, the participants asserted that lively and thought-provoking class discussion was often supplanted by dull, didactic communication. There was little time for teachers to cultivate students’ intellectual curiosity or ability to think critically and deeply. Taken together, the undesirable effects upon the teachers’ professional autonomy generated considerable tension for the teachers. These tensions were exacerbated by a lack of time to cover all the Standards of a given subject and a lack of adequate SOL-related teaching materials. As noted in chapter four, one teacher stated “I think it is absolutely absurd that the State can *demand* that we teach the Standards without giving the appropriate materials to teach them. I have spent *countless* hours searching for information and lesson ideas to effectively teach these Standards.” In sum, the teachers believed they were held unjustly accountable for teaching an over-extensive curriculum with inadequate teaching materials to do so.

Similar to the report of Darling-Hammond and Wise (1983), the teachers of the present study do not suggest that having high, rigorous content standards is counterproductive. However, how one seeks to improve teaching is as important as the standards (p. 69). “Creating teacher-proof approaches for improving instruction and establishing elaborate accountability systems may control the least competent and committed teachers, but may also have the effect of driving the more competent and committed teachers out of teaching” (p. 69). Indeed more than half the teachers of the current study indicated they have considered taking an early retirement or leaving teaching altogether. They

contended that teachers do not want to work in a classroom environment of high-pressure, which is produced by the SOL program. Teachers are losing their joy of teaching and enduring considerable tension while working within a system that rubs against their better judgment of what comprises teachers' roles and responsibilities as well as high-quality student learning.

Theoretical and Practical Considerations

While the importance of the goals of equity and educational excellence cannot be overstated, the propelling force behind the contemporary accountability initiative of policymakers and business leaders appears to be economic in nature. As such, it is useful to consider the contemporary reform in terms of its organizers' larger purpose. To business leaders who are represented by groups such as the Business Roundtable, U.S. Chamber of Commerce, National Alliance of Business (Christie, 1999) and Achieve (Achieve on the Web), state-specified content standards help students gain a world-class education, be prepared to compete successfully in the global marketplace, and supply American businesses with competent workers.

Virginia School Board member, Mark Christie (1999), summarizes business leaders' need for a better-educated workforce through the states' commitment to high academic standards:

First, while many of our schoolchildren were doing quite well in our public schools, far too many others were falling through the cracks and were left unprepared for success as adults in our economy and society. Second, the

21st century will be characterized by a knowledge-based economy more competitive than ever before in history. Academic achievement levels that may have been sufficient in the past will not be enough in the future. Consequently, all our students, even our previously successful ones, will need higher levels of skills and knowledge to get the good, well-paying jobs that will flow to those countries and states that have the best-educated workforces.

Christie's statement underscores the larger goal of the contemporary reform, which is economic advancement, or social efficiency. This goal sets out to raise the level of human capital in American society (Labaree, 2000, p. 33). Accordingly the schools should help prepare students "for the full array of jobs that make up the American economy by giving them the skills they need in order to carry out these jobs productively. It is the kind of argument we often hear from Presidents, governors, and corporate leaders, who are worried about the economic consequences of inadequate education" (2000, p. 33).

From the perspective of schooling for social efficiency the focus on core subjects is beneficial to the public good, as this focus is believed to strengthen the American marketplace. Kennedy (2001, p. 455) draws upon the work of Chubb and Moe (1990), *Politics, Markets, and America's Schools*, to describe this prevailing view in terms of privatization theory. Kennedy writes

"Effective schooling" is to be measured by academic criteria only, that the "core academic mission" of our schools is to impart competency in the

math, science, and language skills “so crucial to a future of sophisticated technology and international competition.” That is, to the extent that schools are to provide a public good, privatization theory defines that good solely as achievement of a level of academic competence sufficient to sustain economic growth and make America competitive in the global marketplace.

Thus the impetus for rigorous state-specified standards seems to be grounded in the nation’s economic needs. Given this view, the standards specify the most important knowledge and skills for student learning, hence for participation in the global marketplace. In sum, the schools should provide a public good that involves the “transmission of literacy and technical knowledge sufficient to support economic growth and individual self-sufficiency” (Kennedy, 2001, p. 455).

Democracy and social efficiency. Provided the larger intent of the designers of the contemporary reform, it seems clear that there is a preference for schooling for social efficiency over schooling for cultivating democracy. There is evidence to suggest that the public tends to diverge from the designers’ preferred purpose of public schooling. Rose and Gallup (2000, p. 47) report that the public ranks *first* the purpose of public schooling that prepares schoolchildren to become responsible citizens. The public ranks *second* the purpose that schooling helps students become economically self-sufficient (p. 47). This view seems to contrast with the contemporary design of policymakers and business leaders who aim to make schools primarily a place for improving America’s position in the global marketplace.

As discussed earlier Labaree (2000, p. 33) describes the goal of social efficiency as that of raising the level of human capital in American society. Human or social capital refers

to “those features of social organization, such as networks, norms, and trust, that facilitate coordination and cooperation for mutual benefit (Kennedy, 2001, p. 454). Given this perspective, public education aims “less to educate citizens than to train productive workers” (Labaree, 2000, p. 32). According to Labaree (p. 32) “economic growth requires workers with skills that are matched to particular occupational roles. As a result, schools need to provide specialized kinds of learning for alternative career paths, sort students according to predicted future careers, and then provide them with the specialized learning they need” (p. 32).

Labaree adds that schooling for democracy includes democratic equality, where children become competent citizens (2000, p. 32). “[C]itizens need to be able to make valid judgments about the essential issues in democratic political life (as voters, jurors, and so on). At the same time, democracies also require citizens whose social differences are modest enough that they can reach agreement about the policies shaping political and social life” (p. 32). From this perspective, Labaree notes that schools provide “a shared level of competence and a common set of social experiences and cultural understandings essential for an effective democracy” (p. 32). In sum, public schooling is for developing “each individual self in the context of justice, fairness, responsibility, and mutual caring to which the Declaration of Independence and the Constitution speak so eloquently” (Goodlad, 2000, p. 87). Schooling for democratic ends requires the “creation of a political community, a process of creating unum from our pluribus” (Kennedy, 2001, p. 455).

Social mobility. The purposes of both democratic equality and efficiency are similar in that they are grounded by the notion that public schooling is a public good, “which should serve the broader public interest in producing competent citizens and productive workers”

(Labaree, 2000, p. 29). Contrasted to these larger goals of public education, Labaree characterizes a third purpose of public education, which is social mobility. Provided this perspective, education is a private good that “benefits only the owner, serving as an investment in my future, not yours; in my children, not other people’s children” (p. 32). The aim here is to “preserve the advantages and increase the distinctions that arise from the way individual consumers currently work the education system...

[This approach] tends to put special emphasis not on improving skills but on distinguishing winners from losers. The focus is on labeling rather than learning—giving gold stars to those who pass through the promotional gates, who get into the gifted program or the advanced placement class, and who win a special endorsement on their high school diploma. And giving lumps of coal to those who fail to [make] the grade in any of these ways (Labaree, 2000, p. 33).

Economic gain and reductionist outcomes. Labaree seems to present social mobility as separate from the aims of democracy and social efficiency because social mobility is primarily concerned with the individual whereas democracy and efficiency are concerned with the collective good of public schooling. However, in the context of high-stakes accountability programs it appears that the aims of social efficiency and social mobility bear resemblance in that both seem to give “lumps of coal” to those students who do not pass high-stakes tests. Those who note the lumps of coal contend a model of schooling that primarily promotes economic gain. Lumps of coal include diminished diversity of curricula

(Goodlad, 2000), also a diminished diversity of instructional strategies and reduced complexity in learning to that which will be measured on high-stakes tests (McNeil, 2000). These reductionist outcomes seem most harmful to poor students, children of color or immigrant children (McNeil, 2000, Valenzuela, 1999). Parents and the public at large indicate similar concerns over these groups of schoolchildren. Teachers, too, receive lumps of coal. These were described by the teachers of the present study in terms of the undesirable consequences of the high-stakes SOL program.

Moreover in noting the lumps of coal that seem to result from contemporary accountability, it can be contended that large-scale accountability programs such as the SOL do not adequately address complex school problems. Nor do such programs appear to cultivate democracy. “Accountability without authority is punishment,” Houston writes (2001, p. 432). “[W]e must evolve a distributed system of leadership in which the skills and the ability to make things happen and the accountability for whether they did happen are spread across a wider spectrum” (p. 432). Such inclusive participation is indicative of democracy, and the teachers of this study overwhelmingly supported this view.

Limitations of the Study and Suggestions for Future Research

Although this study utilized a simple random sample of teachers and provided evidence to support the reliability and validity of the survey instrument employed, it was limited by its sample of a specific population of teachers who were members of the VEA. Furthermore it was primarily a descriptive study intended to add breadth and clarity to the discourse of controversy on high-stakes accountability by describing teachers’ perspectives.

It was not intended to test hypotheses concerning teachers' views, although this might be suitable for future research.

There is a need for more empirical research concerning teachers' views on the contemporary reform. Presently there is a small body of research that specifically addresses teachers' perspectives, and approximately half of this literature does not report evidence to support the reliability and validity of the measures used in conducting the research. In addition to this, many of the studies that heavily employ qualitative methods such as observations and in-depth interviews do not sufficiently report their methods of data analysis. As such there seems to be insufficient information as to precisely how the researchers arrived at the results. To address these problems, future research should more thoroughly report the methods of survey construction, provide evidence to support the reliability and validity of the measures employed, and more clearly specify the analytic processes used. A reader of research that is germane to teachers' perspectives on high-stakes accountability should not have to presume the integrity of the research instruments and the modes of data analysis employed in the studies.

With the aforementioned suggestions in mind, future research might utilize a sample of teachers from a broader population than what was used in the present study. This study was based on VEA teacher-membership, therefore the results do not generalize beyond the VEA teacher population. Studies are needed where the findings are more generalizable to a larger population of Virginia teachers. Also surveying one sample of teachers at two different periods of time over the course of a school year might be helpful to better understand the consistencies or changes in teachers' views on contemporary reform as the time for high-stakes test administration draws near.

Finally, research is suggested that moves beyond descriptive purposes. While this study contributes to the small body of empirical literature on teachers' views of contemporary accountability by providing a description of teachers' perspectives, this research serves only as a beginning. The study of relationships between domains and the development of confirmatory models of teachers' views on the contemporary reform are needed. For example, several domains of the Teacher Survey: On the SOL School Reform might be identified that predict the level of teachers' tension while working within the reform. Such information might be useful to school and district administrators as well as to policymakers and business leaders who might be interested in what aspects of the SOL program seem to create the most tension and contribute to teacher attrition. Further it would be important to note a teacher profile of those teachers who leave the field of teaching. This might provide information as to the quality of teachers who might leave the field due to the SOL program. These suggestions for future research might help build a stronger case to persuade some policymakers and business leaders that teachers assert that high-stakes accountability-based school reform diminishes the quality of teaching and learning through a high-pressure classroom environment and does not support the professional autonomy of teachers.

Bibliography

- Achieve (<http://www.achieve.org>).
- Airasian, P. W. (1988). Symbolic validation: The case of state-mandated, high-stakes testing. *Educational Evaluation and Policy Analysis, 10*, (4), 301-313.
- American Federation of Teachers. (1997). Making standards matter: 1997. Washington, DC: American Federation of Teachers.
- Anderson, J. O., Muir, W., Bateson, D. J., Blackmore, D., & Rogers, W. T. (1990). *The impact of provincial examinations on education in British Columbia: General report* (Report to the British Columbia Ministry of Education. Victoria, BC: British Columbia Ministry of Education.
- Area SOL scores on the rise (2000, October, 12). *The Southwest Times*, pp. 1-2.
- Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*, (5), 384-97.
- Banes, M., Clarke, D., & Stephens, M. (2000). Assessment: The engine of systemic curricular reform? *Journal of Curriculum Studies, 32* (5), 623-650.
- Barth, P. (1998). Virginia's version of excellence. *The American School Board Journal*, 185, 41-43.
- Bates Bolt, M. A. (1997). *A Deweyan perspective on teachers' roles and responsibilities: An experiential reflection*. Unpublished master's thesis, Texas Christian University, Fort Worth.
- Bentley, M. L. (2000, June 30). Virginia's SOL program is bad education policy. *The Roanoke Times*, p. A15.

- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud and the attack on America's public schools*. Reading, MA: Addison Wesley.
- Board of Education (1983). *Virginia's vision for world class education*. Richmond: Virginia Department of Education.
- Bonstingl, J. J. (2001). Are the stakes too high? *Principal Leadership, January*, 8-14.
- Borko, H., & Elliott, R. (1998). *Tensions between competing pedagogical and accountability commitments for exemplary teachers of mathematics in Kentucky*. (ERIC Document Reproduction Service No. ED428127.
- Bracey, G. W. (1995). *Final exam: A study of the perpetual scrutiny of American education*. Bloomington, IN: TECHNOS Press.
- Brady, M. (2000). Who decides what counts on high-stakes tests? *The Education Digest*, 66 (3), 4-10.
- Chubb, J. E., & Moe, T. M. (1990). *Politics, Markets, and America's Schools*. Washington, D.C.: Brookings Institution, p. 2.
- Cochran-Smith, M. (2000a). Editorial: Teacher education at the turn of the century. *Journal of Teacher Education*, 51 (3), 163-165.
- Cochran-Smith, M. (2000b). The questions that drive reform. *Journal of Teacher Education*, 51 (5), 331-33.
- Coleman, A. (2000). None of the above. *Education Week, XIX*, (43), 42, 45.
- Coleman, A. L. (1998). Excellence and equity in education: High standards for high-stakes tests. *Virginia Journal of Social Policy and the Law*, 6 (1), 81-114.
- Coleman W. T., & Selby, C. C. (1983). *Educating Americans for the 21st century: A plan of action for all American elementary and secondary students so that their*

- achievement is the best in the world by 1995*. Washington, DC: National Science Board Commission.
- Commonwealth of Virginia (1995). *Standards of learning for Virginia public schools* (8 VAC 20-131-10 et. Seq.). Richmond: Board of Education.
- Commonwealth of Virginia (1997). *Standards for accrediting public schools in Virginia* (8 VAC 20-131-10 et. Seq.). Richmond: Board of Education.
- Cook, A., Cunningham, C., & Tashlik, P. (2000). Unmasking the low standards of high-stakes tests. *Education Week, XIX* (26), 40.
- Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park: Sage Publications.
- Cooper, M. & M. Leiter (1980). Teachers on testing. In Charles B. Stalford (Ed.), *Testing and evaluation in schools: Practitioners' Views* (pp. 33-41).
- Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks: Sage Publications.
- Cross, C. T. (1995). What business needs from schools. Washington, DC: Council for Basic Education.
- Cross, L. H. (2000). The Standards of Learning (SOL) reform: Real consequences require real tests. *Virginia Issues & Answers, 7* (1), 24-25.
- Cypress, E. J. (1980). Making reading achievement tests work for the inner-city student. In Charles B. Stalford (Ed.), *Testing and evaluation in schools: Practitioners' Views* (pp. 27-32).
- Darling-Hammond, L., & Wise, A. E. (1983). Teaching standards or standardized learning? *Educational Leadership, 41* (2), 66-69.

- Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal*, 85, 315-36.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64 (1), 5-30.
- Darling-Hammond, L. (2000). How teacher education matter. *Journal of Teacher Education*, 51, (3), 166-173.
- Delpit, L. (1995). *Other people's children: Culture conflict in the classroom*. New York: New Press.
- Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: John Wiley & Sons, Inc.
- Erickson, F. E. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.
- Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the burden of "acting white." *Urban Review*, 18, 176-206.
- Fortune, J. (1998). *Qualitative research methods*. Graduate course conducted at Virginia Tech, fall semester.
- Gehring, J. (2000). Students boycott tests in Mass[achusetts] to protest emphasis on exams. *Education Week*, XIX (32), 8.
- Gitlin, A. & Margonis, F. (1995). The political aspect of reform: Teacher resistance as good sense. *American Journal of Education*, 103, 377-405.
- Glass, G. V, & Hopkins, K. D. (1996). *Statistical methods in education and psychology*. (3rd ed.). Boston: Allyn and Bacon.

- Goodlad, J. I. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill Book Company.
- Gordon, S. P., & Reese, M. (1997). High stakes testing: Worth the price? *Journal of School Leadership*, 7, 345-368.
- Grandal, M. (1995). Not all standards are created equal. *Educational Leadership*, 52, pp. 16-21.
- Grogan, M. (2001). Positive and not-so-positive effects of preparing students to pass Standards of Learning tests in Virginia's high school classrooms. *NASSP Bulletin*, 85 (622), 5-23.
- Hadderman, M. (2000). *Standards: The policy environment*. (ERIC Document Reproduction Service No. ED444239).
- Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan*, 80 (9), 662-666.
- Haney, W. (2000). *The myth of the Texas miracle in education*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Hanson, A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.
- Hargreaves, A., & Goodson, I. (1996). Teachers' professional lives: Aspirations and actualities. In I. Goodson & Hargreaves (Eds.), *Teachers' professional lives* (pp. 1-27). London: Falmer Press.
- Harrington-Lueker, D. (2000). The uneasy coexistence of high stakes and developmental practice. *The School Administrator*, January, 6-11.

- Haycock, K. (1998). Good teaching matters . . . a lot. *The Education Trust*, 3(2), 3-14.
- Henkel, R. E. (1976). *Tests of significance*. Sage publications, Inc: Beverly Hills and London.
- Herman, J. (1992). What's happening with educational assessment? Presentation delivered at the Annual Meeting of the Southeastern Regional Vision for Education, Tampa, FL, February.
- Hess, F. M. & Brigham, F. (2000). How kids pay for high-stakes testing. *The Educational Digest*, 66 (4), 11-17.
- Hit SOLs where it hurts, bill proposes. *The Roanoke Times*, p. B8.
- J. P. Heubert, and Hauser R. M. (Eds.). (1999). *High stakes testing for tracking, promotion, and graduation*. Committee on Appropriate Test Use, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington D. C.: National Academy Press.
- Hilliard, A. G. III (2000). Excellence in education versus high-stakes standardized testing. *Journal Teacher Education*, 51 (4), 294-303.
- Hoffman, J., Pennington, J., Assaf, L., & Paris, S. (1999). *High stakes testing in reading and its effects on teachers, teaching, and students: Today in Texas, Tomorrow?* (Unpublished manuscript, University of Texas at Austin.
- Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54, (5), 482-492.
- Holloway, J., & Pearlman, M. (2001). Reform through standards: The ultimate challenge. *Principal Leadership*, January, 40-44.

- Holman, L. J. (1995). *Impact of ethnicity, class, and gender on achievement of border area students on a high-stakes examination*. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22).
- Horton, R. L. (1978). *The general linear model: Data analysis in the social and behavioral sciences*. New York: McGraw-Hill International Book Company.
- Houston, P. D. (2000). A stake through the heart of high-stakes tests. *The School Administrator*, 57 (11), 58.
- Howell, I. (2002, June 14). Salem meets goal on SOLs. *The Roanoke Times*, pp. B1, B6.
- Hoyle, E. (1980). Professionalization and deprofessionalization in education. In E. Hoyle & J. Megarry (Eds.), *World yearbook of education* (pp. 42-54). London: Kogan Page/New York: Nichols.
- Jaeger, R., & Tittle, C. K. (1980). *Minimum competency achievement testing: measures, and consequences*. Greensboro: McCutchan Publishing Corporation.
- Jehlen, A. (2001). Can the standards movement be saved? *NEA Today*, 19 (4), 8-9.
- Jennings, J. F. (1998). *Why national standards? Politics and the quest Motives, for better schools*. Thousand Oaks: Sage Publications.
- Johnson, R. (2000). *Whole school inquiry that promotes equitable student outcomes*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 24-28.
- Jones, M. G., Jopnes, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M.

- (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, November, 199-203.
- Kalton, G. (1983). *Introduction to survey sampling*. London: Sage Publications.
- Keefe, J. W., & De Leonibus, N. (Eds.) (1979). *Competency tests and graduation requirements*. Second Edition. Reston, VA: The National Association of Secondary School Principals.
- Kennedy, S. S. (2001). Privatizing education: The politics of vouchers. *Phi Delta Kappan*, 82, (6), 450-456.
- Kim, J., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. London: Sage Publications.
- Kirst, M. W. (1994). The politics of nationalizing curricula content. *American Journal of Education*, 102, 383-393.
- Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education*, 51 (4), 315-27.
- Kohn, A. (2000). High-stakes testing as educational ethnic cleansing. *The Education Digest*, 66 (4), 13-18.
- Kubow P. K., & DeBard, R. (2000). Teacher perceptions of proficiency testing: A winning Ohio suburban school district expresses itself. *American Secondary Education*, 29, 2.
- Kugelmass, J. W. (2000). Not made for defeat. *Educational Leadership*, 57 (7), 25-8.
- Labaree, D. F. (2000). Resisting educational standards. *Phi Delta Kappan*, 82, (1), 28-33.
- Lewis, A. C. (1995). An overview of the standards movement. *Kappan*, 76, 744-750.
- Lewis, A. C. (1999). Standards and testing. *Phi Delta Kappan*, November, 179-80.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing*

- research*. Cambridge: Harvard University Press.
- Linn, R. L. (1989). *Educational measurement*. New York: American Council on Education and Macmillan Publishing Company.
- Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching*. Columbus: Prentice-Hall, Inc.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29 (2), 4-16.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum. The 87th yearbook of the National Society for the Study of Education, part I* (pp. 83-121). Chicago: University of Chicago Press.
- Madaus, G. F., West, M. M., Harmon, M. C., Lomax, R. G., & Viator, K. A. (1992). *The influence of testing on teaching math and science in grades 4-12: Executive summary*. Boston: Center for the Study of Testin, Evaluation, and Edcacational Policy.
- Marzano, R., & Kendall, J. (1997). National and state standards: The problems and the promise. *NASSP Bulletin*, 81 No. 590, 26-41.
- Massell, D., & Fuhrman, S. (Eds.). (1994). Ten years of state standards: The problems and the promise. *NASSP Bulletin*, 81 (590), 26-41.
- McColskey, W., & McMunn, N. (2000). Strategies for dealing with high-stakes state tests. *Phi Delta Kappan*, 82 (2), 115-120.
- McGill-Franzen, A., & Allignton, R. L. (1993). Flunk 'em or get them classified: The contamination of primary grade accountability data. *Educational Researcher*, 22 (1), 19-22.

- McKenna, B. (1977). What's wrong with standardized testing? In *Standardized Testing Issues*. Washington, D. C.: National Education Association.
- McNeil, L. M. (1986). *Contradictions of control: School structure and school knowledge*. New York and London: Routledge.
- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York and London: Routledge.
- Meisels, S. J. (2000). On the Side of the Child: Personal reflections on testing, teaching, and early childhood education. *Young Children*, November, 16-19.
- Miller, D. W. (2001). Scholars say high-stakes tests deserve a failing grade. *The Chronicle of Education*, March, A14-A16.
- Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1989). *A survey of Arizona teachers and school administrators on the uses and effects of standardized achievement testing* (Tech. Rep. No. 89-2). Phoenix: Arizona State University, West Campus.
- Olson, J. K. (1980). Teacher constructs and curriculum change. *Journal of Curriculum Studies*, 12, 1-11.
- Olson, L. (2001a). Test debate: What counts as multiple? *Education Week*, XX (16), 1, 18.
- Olson, L. (2001b). States adjust high-stakes testing plans. *Education Week*, XX (19), 1, 18.
- Pearson, P. D., Vyas, S., Sensale, L. M., & Kim, Y. (2001). Making our way through the assessment and accountability maze: Where do we go now? *The Clearing House*, March/April, 175-182.
- Pilling, B. O. (1999). *A critical analysis of the modern standards movement: A historical*

- portrayal through archival review, written documents, and oral testimony from 1983 to 1995*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.
- Pipho, C. (1995). The search for a standards-driven utopia. *Phi Delta Kappan*, 77 (3), 198-199.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-82.
- Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56, 8-15.
- Puhlmann, K. (1991). Interpreting the mathematics curriculum. Unpublished doctoral dissertation, University of Alberta, Edmonton.
- Ramirez, A. (1999). Assessment-driven reform. *Phi Delta Kappan*, 81 (3), 204-208.
- Rea, L. M., & R. A. Parker. (1997). Designing and conducting survey research. Second Edition. San Francisco: Jossey-Bass Publishers.
- Reeves, D. B. (2001). Standards make a difference: The influence of standards on classroom assessment. *NASSP Bulletin*, 85 (621), 5-12.
- Romberg, T. A., Zarinnia, E. A., & Williams, S. R. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions*. Madison: National Center for Research in Mathematical Science Education, University of Wisconsin-Madison.
- Romo, H. D., & Falbo, T. (1996). *Latino high school graduation: Defying the odds*. Austin: University of Texas Press.
- Rothman, R. (1996). Taking aim at testing. *The American School Board Journal*,

- February, 27-30.*
- Rottenberg, C., & Smith, M. L. (1990). *Unintended effects of external testing in elementary schools*. Paper presented at the annual meeting of the American Educational Research Association, Boston.
- Ryan, K. & Cooper, J. M. (1998). *Those who can, teach (8th ed.)*. Boston: Houghton Mifflin.
- Samiroden, W. D. (1990). External examinations in science: Their effects on teaching and learning. Paper presented at the annual meeting of the Canadian Society for Studies in Education, Victoria, BC.
- Scarth, J. (1984). Teachers' attitudes to examining: A case study. In Patricia Broadfoot (Ed.) *Selection, Certification, and Control*, (pp. 83-102). New York: The Falmer Press.
- Schmidt, W. H., & McKnight, C. C. (1998). *Facing the consequences: Using TIMSS for a closer look at US mathematics and science education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Schmoker, M. (2000). Standards versus sentimentality: Reckoning—successfully—with the most promising movement in modern education. *NASSP Bulletin*, 84 (620), 49-58.
- Senator challenges the tests (2001). *NEA Today*, 19 (4), 10.
- Shepard, L. A. & Dougherty, K. C. (1991). *Effects of High-Stakes Testing on Instruction*. Paper presented at the Annual Meetings of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
- Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1989). *The role*

- of testing in elementary schools*. Los Angeles, CA: Center for Research on Educational Standards and Student Tests, Graduate School of Education, UCLA.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20 (5), 8-11.
- Smith, M. L. & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education*, 51 (5), November/December, 334-44.
- Soder, R. (1986). *Professionalizing the profession: Notes on the future of teaching*. Center for Educational Renewa, Occasional Paper Number 4. Seattle, WA: University of Washington.
- SOL revisions praised, panned (2001, January 9). *The Southwest Times*, p. 1-2.
- Stake, R. E. (1999). The goods on American education. *Phi Delta Kappan*, 80, 668-70.
- Stalford, C. B. (1980). Proceedings of the practitioners' conference. In Charles B. Stalford (Ed.), *Testing and evaluation in schools: Practitioners' Views* (p. 123).
- Strauss, A.L. (1987). *Qualitative analysis for social sciences*. Cambridge, MA: Cambridge University Press.
- Suarez, T. M., & Gottovi, N. C. (1992). The impact of high-stakes assessments on our schools. *NASSP Bulletin*, September, 82-88.
- Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal*, 31 (2), 231-262.
- The Third International Math and Science Study (2000). Retrieved from <http://ustimss.msu.edu/>.

- Turner, J. (2001, January 14). SOLs face difficult test in general assembly. *The Roanoke Times*, pp. B1, B7.
- Thurlow, M. L., & Johnson, D. R. (2000). High-stakes testing of students with disabilities. *Journal of Teacher Education*, 51 (4), 305-14.
- Tyack, D. & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge and London: Harvard University Press.
- Urdu, T. C., & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy*, 8 (2), 137-56.
- Valencia, R., Valenzuela, A. Sloan, K., & Foley, D. E. (2001). Let's treat the cause, not the symptoms: Equity and accountability in Texas revisited. *Phi Delta Kappan*, 83(4), 318-321, 326.
- Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring*. Albany: State University of New York Press.
- van den Berg, R., & Ros, A. (1999). The permanent importance of the subjective reality of teachers during educational innovation: A concerns-based approach. *American Educational Research Journal*, 36 (4), 879-906.
- Virginians give SOL tests low grades (2000, September, 12). *Bristol Herald Courier*, p. 1A, 12A.
- Wermers (2001). <http://www.timesdispatch.com/editorial/index.shtml>.
- Whitford, B. L., & Jones, K. (2000). Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts. Suny Series, Restructuring and School Change. ERIC document ED440139.
- Wideen, M. F., O'Shea, T., Pye, I., & Ivany, G. (1997). High-stakes testing and the

- teaching of science. *Canadian Journal of Education*, 22 (4), 428-444.
- Wilson, B. L., & Corbett, H. D. (1991). *Two state minimum competency testing programs and their effects on curriculum and instruction*. (ERIC Document Reproduction Service No. ED377251).
- Williams, W. M. (2001). The plus side of big tests. *Principal Leadership*, January, 21-25.
- Wise, A. E. (1979). *Legislated learning*. Berkeley and Los Angeles: University of California Press.

Appendix A

Note Card of Initial Contact With VEA Teacher-Members

Teachers' Views on the SOL School Reform

Supported by Virginia Education Association, VEA

January 10, 2002

Teachers,

The Standards of Learning (SOL) program continues to elicit discussion from the public, businesspersons, and policymakers alike. However, *teachers' views on the SOL reform need to become more visible to the public and available to policy makers who make important decisions that affect our schools*. To help make teachers' opinions more apparent, a *Teacher Survey on the SOL School Reform* has been developed. Your name has been randomly selected from the membership list of teachers in VEA to receive the survey in approximately one week.

Your participation in completing and returning the Teacher Survey will be most appreciated, as each survey is needed to ensure adequate representation of VEA teachers' views on the SOL reform. Also, *a summary of the survey results will be made available to all participants*. Please be looking for the Teacher Survey in your mailbox. Thank you.

Appendix B

Teacher Survey: On the SOL School Reform

While responding to the survey, please think about the students you are responsible for teaching, giving priority to any group you teach who must take an SOL test. For each statement, please circle the number that best describes to what extent you disagree or agree with the statement. The options for response are as follows:

Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,

Strongly Agree=6

Circle 7 if you have no basis for responding (Not Applicable=NA).

I. Standards of Learning: curriculum

objectives that indicate “what students should know”	Strongly Disagree			Strongly Agree NA			
	1	2	3	4	5	6	7
1. The Standards represent reasonable guidelines for what teachers should teach.	1	2	3	4	5	6	7
2. The Standards represent reasonable guidelines for what students should know.	1	2	3	4	5	6	7
3. The classroom curriculum is narrowed, or limited, by the Standards.	1	2	3	4	5	6	7
4. The Standards complement my professional decision-making in the classroom.	1	2	3	4	5	6	7

*Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,
Strongly Agree=6*

Circle 7 if you have no basis for responding.

5. The Standards help me remain focused on the subject content.	1	2	3	4	5	6	7
6. The Standards are successful in specifying the most important skills and knowledge that should be taught in a subject area.	1	2	3	4	5	6	7
7. Aligning the classroom curriculum to the state Standards improves the quality of education my students receive.	1	2	3	4	5	6	7
8. Attempting to teach all the skills and knowledge specified by the Standards stifles opportunities for spontaneous “teachable moments.”	1	2	3	4	5	6	7
9. As the classroom curriculum becomes more aligned to the state Standards, I expect my students’ SOL test scores to increase.	1	2	3	4	5	6	7
10. Aligning the classroom curriculum to state Standards improves the quality of my teaching.	1	2	3	4	5	6	7
11. Aligning the classroom curriculum to the state Standards limits my use of diverse instructional strategies.	1	2	3	4	5	6	7

*Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,
Strongly Agree=6*

Circle 7 if you have no basis for responding.

12. The Standards specify what my students must know 1 2 3 4 5 6 7

to be competitive in a global economy.

13. The Standards help my students gain a world-class 1 2 3 4 5 6 7

education.

14. The Standards help ensure that my students 1 2 3 4 5 6 7

receive an equitable education.

15. The Standards are too difficult for many of the 1 2 3 4 5 6 7

students I teach.

16. The Standards are too extensive to be taught 1 2 3 4 5 6 7

adequately in a given school year.

17. The Standards help the community, including 1 2 3 4 5 6 7

parents, respect my work as a professional.

18. The Standards promote parent and community 1 2 3 4 5 6 7

confidence in my classroom decisions.

19. I do not feel any stress using the Standards 1 2 3 4 5 6 7

as a guide for what my students should learn.

20. In teaching the content of the Standards, it has 1 2 3 4 5 6 7

increased the level of my work-related stress.

21. Attempting to cover the Standards is an 1 2 3 4 5 6 7

overwhelming job for me.

*Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,
Strongly Agree=6*

Circle 7 if you have no basis for responding.

22. Standards restrict my professional decision- 1 2 3 4 5 6 7
making in the classroom.

23. I have had to restrict the use of diverse instructional 1 2 3 4 5 6 7
strategies in order to cover the Standards.

II. SOL Testing Program: SOL tests and the use of test scores

24. The quality of my teaching is reflected 1 2 3 4 5 6 7
in my students' SOL test scores.

25. I can pinpoint learning difficulties by reviewing my 1 2 3 4 5 6 7
students' SOL test scores.

26. The SOL test scores help me identify instructional 1 2 3 4 5 6 7
weaknesses in my teaching.

27. Passing rates on the SOL tests provide an accurate 1 2 3 4 5 6 7
representation of my school's quality.

28. Passing rates on the SOL tests are appropriate for 1 2 3 4 5 6 7
determining school accreditation.

29. SOL test scores should be the single most important 1 2 3 4 5 6 7
factor in determining whether or not students graduate
from high school.

Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5, Strongly Agree=6

Circle 7 if you have no basis for responding.

30. I believe that a 40 to 60 item SOL test cannot 1 2 3 4 5 6 7

adequately measure my students' knowledge and skills in the subject I teach.

31. My students' performance on SOL tests is unrelated 1 2 3 4 5 6 7

to their family background.

32. My students' performance on SOL tests is 1 2 3 4 5 6 7

unrelated to their race or ethnicity.

33. Scores from an SOL test provide me with useful 1 2 3 4 5 6 7

information that cannot be obtained from other standardized tests such as the Stanford-9.

34. SOL test scores are trustworthy indicators of 1 2 3 4 5 6 7

whether students should be retained in grade.

35. SOL test scores are useful for helping 1 2 3 4 5 6 7

determine course placement for students.

III. Performance Standards for Students: minimum score of

600 to pass SOL tests; "what students should be able to do"

36. Requiring students to achieve a passing score on a 1 2 3 4 5 6 7

SOL test assures that students are competent in the subject matter.

*Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,
Strongly Agree=6*

Circle 7 if you have no basis for responding.

37. The cut-scores (minimum passing scores) of the SOL tests are set at appropriate levels. 1 2 3 4 5 6 7

38. Requiring students to pass a SOL test ensures that students gain a world-class education. 1 2 3 4 5 6 7

39. The score that students must attain to pass a SOL test clearly distinguishes the students who have an acceptable level of knowledge and skills from students who do not. 1 2 3 4 5 6 7

40. Passing scores on SOL tests are arbitrary and have little to do with mastery of a subject. 1 2 3 4 5 6 7

41. Passing scores on SOL tests are set too high. 1 2 3 4 5 6 7

42. A single passing score cannot motivate all of my students. 1 2 3 4 5 6 7

43. The SOL cut-scores are set at reasonable levels. 1 2 3 4 5 6 7

IV. Performance Standards for Schools: at least 70% of students in a school must pass SOL tests

44. Requiring 70% of the students to pass each SOL test ensures that my school is of high quality. 1 2 3 4 5 6 7

Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5, Strongly Agree=6

Circle 7 if you have no basis for responding.

45. Setting a 70% pass rate for schools ensures that I 1 2 3 4 5 6 7
am teaching to rigorous standards.

46. Requiring 70% of the students to pass each SOL 1 2 3 4 5 6 7
test helps improve the quality of my teaching.

47. Requiring a 70% pass rate for school accreditation 1 2 3 4 5 6 7
is an inappropriate use of test scores.

V. Accountability: explicit consequences for students and schools when they meet or fail to meet the performance standards; and consequences that teachers may perceive in the classroom

48. For student accountability purposes, SOL test 1 2 3 4 5 6 7
scores should be the primary criteria used to
determine what students have learned.

49. SOL test scores should be the deciding factor 1 2 3 4 5 6 7
used to determine whether students should
graduate from high school.

50. Requiring students to pass SOL tests helps them 1 2 3 4 5 6 7
take responsibility for their learning.

*Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,
Strongly Agree=6*

Circle 7 if you have no basis for responding.

51. A high school diploma should be denied to those 1 2 3 4 5 6 7
who cannot pass the SOL tests.

52. For school accountability purposes, my school's 1 2 3 4 5 6 7
quality should be based on SOL test scores alone.

53. Basing school accreditation on a 70% pass rate is 1 2 3 4 5 6 7
an appropriate use of test scores.

54. I believe that it is important for schools to be 1 2 3 4 5 6 7
held accountable by the SOL tests.

55. I feel considerable pressure to help my students 1 2 3 4 5 6 7
improve their SOL test scores.

56. As a result of state policy, I feel that I am pressured 1 2 3 4 5 6 7
to place too much emphasis on my students' SOL
test scores.

57. I am not troubled by my efforts to raise my 1 2 3 4 5 6 7
students' SOL test scores.

58. The SOL system is consistent with my view of what 1 2 3 4 5 6 7
my role as a teacher should be.

59. Teachers of subjects or grade levels for which 1 2 3 4 5 6 7
there is a SOL test are under excessive stress.

*Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,
Strongly Agree=6*

Circle 7 if you have no basis for responding.

60. The SOL testing system has caused me to consider leaving the teaching profession.	1	2	3	4	5	6	7
61. Importance placed on SOL test scores constrains my professional expertise.	1	2	3	4	5	6	7
62. I feel undue pressure to help students raise SOL test scores.	1	2	3	4	5	6	7
63. I feel considerable stress over public comparisons between my school and other schools' SOL test scores.	1	2	3	4	5	6	7
64. Comparisons between my school and other schools' SOL test scores help hold my school accountable.	1	2	3	4	5	6	7
65. My students seem to experience excessive stress in trying to pass the SOL tests.	1	2	3	4	5	6	7
66. The SOL system helps my students respect me as a teacher.	1	2	3	4	5	6	7
67. The SOL system helps minimize students' disruptive behavior.	1	2	3	4	5	6	7
68. I feel frustrated teaching within the SOL school reform.	1	2	3	4	5	6	7

Strongly Disagree=1, Disagree=2, Tend to Disagree=3, Tend to Agree=4, Agree=5,

Strongly Agree=6

Circle 7 if you have no basis for responding.

69. I frequently use practice tests that prepare students 1 2 3 4 5 6 7

for the SOL tests.

70. With the SOL system, students are required to 1 2 3 4 5 6 7

think critically about the subject matter.

71. I do not have enough time in the school day to 1 2 3 4 5 6 7

teach all the Standards.

72. Students have enough time to master difficult 1 2 3 4 5 6 7

concepts in the Standards.

73. My teaching is rushed because I try to “cover” 1 2 3 4 5 6 7

the Standards.

74. With the SOL system, I focus my teaching more on 1 2 3 4 5 6 7

basic knowledge than on higher-order thinking skills.

75. I feel my school system puts too much emphasis 1 2 3 4 5 6 7

on SOL test scores.

76. I am able to meet the educational needs of my 1 2 3 4 5 6 7

students with the SOL system.

77. My professional autonomy as a teacher is restricted 1 2 3 4 5 6 7

by the SOL reform.

78. The requirements for school accreditation diminish 1 2 3 4 5 6 7

the quality of my teaching.

82. Please check (✓) the subject you thought about while responding to this survey.

English	Specialization, please specify below
History/Social Studies	_____
Mathematics	Art
Reading/Language Arts	Music
Science	Physical Education

83. Are you Virginia State Certified as a teacher?

Yes No

84. Do you hold Virginia endorsement in the primary subject you teach?

Yes No

85. Are you a National Board Certified Teacher?

Yes No

86. Does your primary teaching assignment include subjects for which students are mandated to take a SOL test?

Yes No

87. Please estimate the percentage of students eligible for free or reduced lunch in your school.

_____ %

88. For each group of students below, please indicate the approximate percentage attending your school.

_____ % African American _____ % Caucasian _____ % Latino

_____ % Asian American _____ % Please write group as needed:

VII. Three open-ended questions conclude this survey. Please provide brief remarks regarding these questions.

1. According to a recent poll conducted by a national television news program, 63% of those polled believe that students should be tested (by tests such as the SOL tests) in order to measure students' learning. If such tests should measure students' learning, how should the test results be used?

Curriculum Vitae

Melanie A. Bolt

EDUCATION

- 2003 ***Ph.D., Educational Research and Evaluation***
Virginia Polytechnic Institute and State University,
Blacksburg, VA
Cognate: Statistics
- 1997 ***M.A. Ed., Educational Research and Collaboration***
Texas Christian University, Fort Worth, TX
Master's thesis: A Deweyan Perspective of Teachers' Roles
and Responsibilities: An Experiential Reflection
- 1994 ***B.S., Elementary Education***
University of Southern Mississippi, Hattiesburg, MS
- Other**
Educational Research and Evaluation Special Seminars,
Virginia Tech, Blacksburg, VA
Specialized training in methodological issues of quantitative
and qualitative research
- 1999 ***Women Writers Conference and Workshop,*** Virginia Tech
Blacksburg, VA
Instruction on strategies for publishing writing
- 1998 ***Focus Groups as Qualitative Research: Workshop at Virginia***
Tech, Blacksburg, VA
Training in methods and theoretical underpinnings of focus
group inquiry
- 1997 ***Educational Research and Evaluation Laboratory,***
Virginia Tech, Blacksburg, VA
Training in research and evaluation consultation and the use of
specialized computer software
- 1996 ***New Jersey Writing Institute,*** Bedford, TX
Certified writing instructor

1994 ***Overseas Student-Teaching Scheme*** through the University of Southern Mississippi, Hattiesburg, MS
Selected by the Dean of Education and Psychology as one of only four pre-service teachers at the University to be trained as a classroom teacher in Redhill, England

PROFESSIONAL EXPERIENCE

- 2001 ***Instructor of Doctoral Level Course***
Methods of Qualitative Research, Virginia Tech
Supervising professor: Dr. Jim C. Fortune
- 2001 ***Co-Instructor of Doctoral Level Course***
Statistics for the Behavioral Sciences I, Virginia Tech
Supervising professor: Dr. Kusum Singh
- 1997-2001 ***Graduate Research and Evaluation Assistant,***
Program of Educational Research and Evaluation, Virginia Tech
- Conceptualized, performed, and reported a study of rural Southwest Virginia teachers while apprenticed by Dr. Jimmie Fortune
 - Conducted and reported meta-analyses regarding mathematics achievement
- 1997-2000 ***Educational Research and Evaluation Consultant,***
Blacksburg, VA, Educational Research and Evaluation Laboratory at Virginia Tech
- Consulted master's and doctoral students regarding research methods and content areas of research and program evaluation
 - Instructed students on how to manage and analyze data using computer software such as SPSS, NUDIST, and LISREL
- 2000 ***Session Chair***
Eastern Educational Research Association, Clearwater, FL
- 1999 ***Guest Lecturer, Doctoral Level Class***
Program of Educational Research and Evaluation, Virginia Tech
Topics of reliability and validity in applied research and program evaluation

- 1999 ***Program Evaluator***
- Requested by Virginia Tech to conduct focus group interviews of male and female University students who were currently enrolled as engineering majors
 - Analyzed qualitative data and prepared a report of the results for University administrators
- 1999 ***Sessions Chair***
Excellence in Education Conference, Virginia Tech
- 1998 ***Program Evaluator***
- Requested by Virginia Tech to conduct focus group interviews of University students regarding Virginia Tech's new graduation requirements
 - Analyzed qualitative data and prepared a report of the results for University administrators
- 1994-1997 ***Classroom Teacher of Sixth Grade Students***
Hurst-Eules-Bedford Independent School District, Hurst, TX
Twenty-two race-ethnicities represented in a school of 600 students
- Assumed all responsibilities of a sixth grade teacher
 - Evaluated by administrators as clearly outstanding/master teacher
- 1995 ***District-Wide School Restructuring Consultant***
Hurst-Eules-Bedford Independent School District, Hurst, TX
- Reviewed and reported current research on school restructuring
 - Worked alongside community members to devise a school restructuring plan
- 1993 ***Child Literacy Teacher***
University of Southern Mississippi, Hattiesburg, MS
- Customized instruction for children with literacy difficulties

AWARDS AND HONORS

Mississippi Professional Educators (MPE) Scholarship Award, 1994
Presented by the president of MPE in Hattiesburg, MS

Most Outstanding Elementary Education Major, 1994

Presented by the Dean of Education and Psychology at the University of Southern Mississippi

Student-Teacher Abroad, 1994

Selected by a review team and the Dean of Education and Psychology at the University of Southern Mississippi

Kappa Delta Pi, 1993

Membership granted based upon academic excellence by Kappa Delta Pi, University of Southern Mississippi

Gamma Beta Phi Society, 1993

Membership granted based upon academic excellence by Gamma Beta Phi Society, University of Southern Mississippi

COMMUNITY SERVICE

1999

Invited Guest Speaker

The Round Dozen, a women's community group, Pulaski, VA

- Topic on traditions of qualitative and quantitative research

1994-1997

Student Mentor

Hurst-Eules-Bedford Independent School District, Hurst, TX

1994

Adult Literacy Teacher

Literacy Volunteers of America, Hattiesburg, MS

- Designed and presented literacy lessons to adults

RESEARCH INTERESTS

School improvement policies and programs

High-stakes school and student accountability in K-12 settings

the adequacy of content standards

the diversity of teachers' instructional strategies

the quality of teaching in the classroom

the quality of student learning, particularly for student

populations of diverse race-ethnicities and/or

low-income areas

the psychometric properties of high-stakes tests

Teachers' perspectives on student assessment

Teachers' views on assessment of school quality

Teacher professional autonomy, teacher tension and attrition

PRESENTATIONS

- Bolt, M. A. (2003). *Classroom consequences of contemporary accountability: Teachers' views of statewide school reform in Virginia*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Bolt, M. A., Fortune, J. C., & Cachaper, C. (2000). *Attitudes toward student assessment: Perspectives of rural teachers*. Paper presented at the annual meeting of the Eastern Educational Research Association, Clearwater, FL.
- Bolt, M. A. (2000). *Teachers' voices: On the impact of standardized testing of students*. Paper presented at the annual meeting of the Eastern Educational Research Association, Clearwater, FL.
- Bolt, M. A. (2000). *Southwest Virginia teachers' views on student assessment*. Paper Presented at the Third Annual Inquiring Voices, Applied and Methodological Studies in the College of Human Resources and Education, Virginia Tech, Blacksburg, VA.
- Bolt, M. A. (2000). *Minimum-competency testing: Teachers' perspectives*. Paper Presented at the Third Annual Inquiring Voices, Applied and Methodological Studies in the College of Human Resources and Education, Virginia Tech, Blacksburg, VA.
- Bolt, M. A. (2000). *A critique of the Virginia Department of Education report, "Standards of Learning Tests: Validity and Reliability Information, Spring 1998 Administration."* Paper presented at the Third Annual Inquiring Voices, Applied and Methodological Studies in the College of Human Resources and Education, Virginia Tech, Blacksburg, VA.
- Bolt, M. A., & Fortune, J. C. (1999). *Are we getting more than we are paying for?* Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, SC.
- Bolt, M. A., & Singh, K. (1999). *Achievement variability in mathematics and science: An emphasis on the affective domain*. Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, SC.
- Bolt, M. A., Cachaper, C., & Fortune, J. C. (1999). *Teacher-student classroom interactions and cultural diversity: A decade review of the literature*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.
- Bolt, M. A., Cachaper, C., & Fortune, J. C. (1999). *Using USR-Nudist to facilitate the process of writing a literature review*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.

- Cachaper, C., Singh, K., & Bolt, M. A. (2000). *An application of ecological systems Theory and SEM: African-American college achievement*. Paper presented at the annual meeting of the Eastern Educational Research Association, Clearwater, FL.
- Bolt, M. A. (1999). *Response options on questionnaires: Is there bias based on nationality?* Paper presented at the Second Annual Inquiring Voices, Applied and Methodological Studies in the College of Human Resources and Education, Virginia Tech, Blacksburg, VA.
- Bolt, M. A. (1999). *Teachers' perceptions of student assessment: A cross-cultural study*. Paper presented at the Second Annual Inquiring Voices, Applied and Methodological Studies in the College of Human Resources and Education, Virginia Tech, Blacksburg, VA.
- Bolt, M. A. (1998). *To purpose and pursue: An ethical dilemma in teaching*. Paper presented at the First Annual Inquiring Voices, Applied and Methodological Studies in the College of Human Resources and Education, Virginia Tech, Blacksburg, VA.

PROFESSIONAL AFFILIATIONS

American Educational Research Association
Eastern Educational Research Association
Mid-South Educational Research Association
American Educational Studies Association
Phi Delta Kappa