

Teachers' Views on High-stakes Testing:

Implications for the Classroom

Policy Brief

by

Lisa M. Abrams

Boston College

Education Policy Research Unit (EPRU)

Education Policy Studies Laboratory

College of Education

Division of Educational Leadership and Policy Studies

Box 872411

Arizona State University

Tempe, AZ 85287-2411

February 2004

EPSL | EDUCATION POLICY STUDIES LABORATORY
Education Policy Research Unit

EPSL-0401-104-EPRU

<http://edpolicylab.org>

Education Policy Studies Laboratory

Division of Educational Leadership and Policy Studies

College of Education, Arizona State University

P.O. Box 872411, Tempe, AZ 85287-2411

Telephone: (480) 965-1886

Fax: (480) 965-0303

E-mail: epsl@asu.edu

<http://edpolicylab.org>

**Teachers' Views on High-stakes Testing:
Implications for the Classroom**

Lisa M. Abrams

Boston College

Executive Summary

There is an appealing logic associated with current models of test-based accountability: the interplay among content standards, state tests, and accountability is a powerful tool to improve the quality of schools. However, when high-stakes consequences are attached to test results for schools, teachers, and students, unexpected consequences may outweigh the intended benefits. To explore the policy impact of Florida's state testing and accountability program on classroom practices, teachers, and students as perceived by educators, this brief presents the results of a national survey in which the responses of Florida teachers are compared with those of practitioners in other states using high-stakes exams. The findings reveal that, compared to their counterparts in other high-stakes states, teachers in Florida perceived a more pronounced impact of the state test.

Recommendations

1. Florida should undertake a long-term evaluation and monitoring program to assess the impact of the Florida Comprehensive Assessment Test (FCAT) and the A+ Accountability program. This evaluation and monitoring program should be conducted by an external organization or research institution. Its purpose is to determine if the state testing program is achieving its intended goals. The evaluation should also examine the unexpected consequences of the FCAT and A+ Accountability program on the educational process and on key stakeholders.
2. Florida testing policy should adhere to the recognized professional standards regarding test development and to the appropriate use of test results as

described in the *Standards for Educational and Psychological Testing*, published jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.

3. Florida policy makers should not make highly consequential decisions about students (such as deciding whether a student is promoted to the next grade or is awarded a high school diploma) by means of test scores alone. Given the evidence pointing to weaknesses in the testing system, it is important to use other sources of information in conjunction with state tests.

Teachers' Views on High-stakes Testing: Implications for Classroom Practice and Student Learning¹

Section 1: The Issue

Accountability-based testing policies have evoked heated debate, especially as states realize full implementation of their education reform policies. The 2002-2003 school year marked the first time that students' scores on the Florida Comprehensive Assessment Test (FCAT) factored into decisions about grade promotion and graduation. Headlines in Florida newspapers were ripe with accolades, reports of failure, threats of legal action, calls for reform and organized protests of the state's education accountability system—"Students Feel Sting of Reform," "Critics of Graduation Exam Threaten Boycott in Florida," "1,300 Third-Graders in Palm Beach County Held Back by FCAT Scores," and "Governor Jeb Bush Announces Biggest Improvement Ever on FCAT" are just a few examples. The debate over how to improve schools, education quality, and student academic performance is one that many policymakers, legislators, administrators, teachers, students, parents and advocacy organizations are grappling with—not just in Florida, but across the nation.

A growing body of evidence suggests that high-stakes testing can be a driving force behind fundamental change within schools.² There is a difference of opinion as to whether this change is improving the quality of education, however. For example, whereas some contend that the guarantee of rewards—or the threat of sanctions—is essential to promote quality teaching and to encourage higher standards of achievement, others maintain that high-stakes tests limit the scope of classroom instruction and student learning in undesirable ways.³ Regardless of one's position on this issue, it is impossible

to deny that statewide testing policies influence classroom instruction and student learning. This brief explores how public school teachers in Florida perceive the effects of the state test and accountability system on their instruction and their students.

Section 2: Background

Current Testing Landscape

Current testing policies grew out of standards-based reform begun in the early 1990s. This initiative called for a rigorous and demanding curriculum that, in addition to requiring students to demonstrate their command of basic content knowledge, also asked them to exhibit higher-level cognitive processes (e.g., application, problem solving, inquiry). The shift from basic skills to high standards has given rise to the current state-level accountability systems designed to hold schools, administrators, teachers (and sometimes students) responsible for meeting these raised expectations. These systems have four main components: *content standards* that communicate the desired content knowledge and skills; *tests* that measure progress toward achieving the content standards; *performance targets* that identify criteria used to determine whether schools, students, or both have reached the desired level of achievement; and *incentives*, such as the consequences (rewards and sanctions), or stakes, that reinforce the attainment of performance targets.⁴

Over the last decade, test-based accountability systems have become widespread; every state except Iowa has instituted curricular standards or frameworks.⁵ Moreover, every state uses a test to measure the degree to which students have mastered the knowledge and skills expressed in these standards.⁶ While on the surface it might appear that state testing policies are becoming increasingly similar, there are substantial

differences in test content, item format, and how test results are used, especially for accountability purposes.⁷

The focus on state assessment requirements was further emphasized in the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), known more commonly as the No Child Left Behind Act (NCLB). This far-reaching legislation seeks to raise the level of achievement for all students and to reduce the gap in the level of performance of students from different backgrounds. At the heart of NCLB are the assessment and accountability requirements, which will substantially increase the extent to which students are tested.⁸ The law includes school accountability provisions; however, states still retain the authority to determine how, or if, students will be held responsible for test performance.

The implementation of NCLB has expanded a majority of current state assessment programs, requiring testing at more grade levels. The federal law requires that states annually administer reading and math tests to all students in Grades 3-8, and in one year between Grades 9-12 starting in 2005-2006.⁹ This requirement affects at least 25 million students annually.¹⁰ Arguably, NCLB is one of the most aggressive federal efforts to improve elementary and secondary education and marks a major departure from the traditionally noninterventionist role of the national government in forming state education policy.

Education Reform in Florida

The state testing and accountability program in Florida has met with controversy since the early 1970s. Originally, the state test targeted basic skills. Although this type of minimal competency testing was common, Florida was the first in the nation to tie

students' test results to graduation—a policy that was broadly criticized, with its legitimacy ultimately decided in the federal courts.¹¹ Throughout the 1980s, education policy at both the federal and state levels was moving away from emphasizing basic skills—and toward requiring students to meet higher or “world class” content standards—in an attempt to ensure that the United States continued to have a competitive edge in the growing world market. State testing policies in Florida reflected this shift.

In 1995, the Florida Commission on Education Reform and Accountability called for the creation of rigorous state curriculum frameworks and recommended changes in the state assessment program. The FCAT, designed to measure the extent to which students had achieved the Sunshine State Standards in reading, writing and math, was first administered in 1998 (in Grades 4, 5, 8, and 10) and became a high school graduation requirement for the class of 2003. Currently, 21 other states use test results to make similar high-stakes decisions for students; by 2008, 24 states will base graduation and grade promotion on state test results.¹²

In 1999, as part of Governor Bush's A+ Accountability Plan, the state assessment program was expanded to include more students at more grade levels. Currently, students in Grades 3-10 are tested annually in reading and math. Writing tests are administered in Grades 4, 8, and 10, and science exams in Grades 5, 8, and 10.¹³ Since 1999, student performance on the FCAT has been used, in part, to assign schools grades (A–F) which are used to monitor performance, identify unsuccessful schools, and leverage rewards and sanctions. The A+ Accountability plan also includes a major reform of the accountability system by providing “opportunity scholarships” or vouchers to students in consistently low-performing or F schools that allows them to transfer to

alternative higher-performing public or private schools.¹⁴ Schools can move out of the F category by improving FCAT scores. Recent NCLB legislation contains provisions that require states to implement similar “choice” options, such as opportunity scholarships, for students in schools that consistently do not meet adequate yearly progress goals. In many ways, how teachers respond to the requirements of the FCAT and the A+ Accountability Plan may foreshadow reactions in other states.

The Effects of High-Stakes Testing

In order to examine how Florida’s testing and accountability policies are affecting classroom instruction and student learning, it is first necessary to summarize the findings of research studies in this area, thus providing a frame of reference. Numerous studies have investigated the effects of state-mandated testing programs, particularly those with high stakes attached to test results.¹⁵ The majority have addressed a variety of issues related to the effects on teaching and learning, specifically on the content of instruction: the strategies used to deliver instruction; the impact of the format of the state test on classroom practices; the test preparation; and the psychological impacts of the test on both teachers and students (e.g., pressure, morale, and motivation), as well as on student learning in general.

Impact on Classroom Practices

Much of the research addresses the effects of state testing programs on what is taught. A common finding is that teachers report giving greater attention to content areas on which students will be tested. For example, of the 722 Virginia teachers surveyed, when the state test was first implemented, more than 80 percent indicated that the Standards of Learning (SOL) test had affected their instruction, especially the content

focus of daily lessons.¹⁶ It only makes sense that increased attention to tested content would result in decreased time spent on other areas of the curriculum. In Kentucky, 87 percent of teachers agreed that their state test, the Kentucky Instructional Results Information System (KIRIS), had “caused some teachers to de-emphasize or neglect untested subject areas.”¹⁷ Results from a national survey of 4,200 teachers confirm these state-level findings—76 percent of the responding teachers indicated that they have increased the amount of time they spend on tested content areas, while more than half (52 percent) indicated they had decreased the amount of class time devoted to content areas not covered by the state test.¹⁸

The impact of the state test on instructional strategies is mixed. Studies in states that require students to formulate and to provide written responses to test questions show an increased emphasis on writing and higher-level thinking skills.¹⁹ For example, in Kentucky, 80 percent of the fourth and eighth grade mathematics teachers increased instructional emphasis on problem solving and writing as a result of the portfolio-based state test.²⁰ In contrast, teachers decreased the use of more time-consuming instructional strategies and lengthy enrichment activities.²¹ A more a recent study found that the format of the state test may adversely affect the use of technology for instructional purposes. For instance, one-third of teachers in high-stakes states were less likely to use computers to teach writing because students were required to construct handwritten responses on the state test.²²

Emphasis on Test Preparation

The pressure to respond to increased demands of the state test often requires teachers to place more emphasis on test preparation. In Maryland, 88 percent of teachers

surveyed felt they were under “undue pressure” to improve student performance on the state test.²³ When asked the same question, 98 percent of Kentucky teachers responded similarly.²⁴ Ninety percent of teachers surveyed nationally reported feeling pressure from their district superintendent to raise test scores, while 79 percent indicated feeling pressured by their building principal to improve student performance.

In light of this pressure, teachers may put greater emphasis on preparing students for the state test. Of the 470 elementary teachers surveyed in North Carolina, 80 percent reported that “they spent more than 20 percent of their total instructional time practicing for the end-of-grade tests.”²⁵ Similarly, a survey of reading teachers in Texas revealed that, on average, teachers spent 8 to 10 hours per week preparing students for the Texas Assessment of Academic Skills (TAAS).

Overemphasis on specific test preparation activities has given rise to concerns about the validity of test scores as accurate measures of students’ levels of achievement. Specific preparation activities such as coaching, teaching test-taking skills, and instruction geared toward the test can yield invalid test results.²⁶ One would expect that if students’ scores improve on the state test (from year to year), scores on other tests that measure the same content, skills, or both, should show similar improvement. When trends in student performance levels on similar standardized tests are not consistent, the accuracy of a particular test as an indicator of student achievement is questionable. For example, 40 percent of teachers surveyed nationally reported that they had found ways to raise state test scores without really improving learning. Similarly, 50 percent of responding Texas teachers did not think that the rise in TAAS scores “reflected increased learning and high quality teaching.” Based on written comments, the study authors

concluded that “teachers regarded improvement on the TAAS as a direct result of teaching to the test.”²⁷

Student performance on a highly consequential test may not generalize to other measures of achievement. Several studies have compared students’ performance on the state test with performance on other standardized tests that assess similar information. Researchers found that gains on the KIRIS math test were substantially larger than improvements for Kentucky students on the math portion of the National Assessment of Educational Progress (NAEP). This suggests that improved performance on the KIRIS math test does not necessarily reflect broader gains in student knowledge.²⁸ Further, when student performance on several different standardized tests in 18 states with high-stakes testing programs was systematically compared, the findings were inconclusive: there was not a strong link between the implementation of the state testing program and improvement in student achievement.²⁹

Impact on Motivation and Morale

High-stakes tests may motivate certain teachers and some students to achieve optimal performance levels. However, researchers have cautioned that placing a premium on student test performance can lead to instruction that is reduced primarily to test preparation, thus limiting the range of educational experiences for students and constraining the pedagogical skills of teachers.³⁰ Studies have also shown that high-stakes assessments increase stress and decrease morale among teachers. More than 77 percent of North Carolina teachers surveyed indicated decreases in their morale; 76 percent reported that teaching was more stressful since the implementation of the North

Carolina state testing program.³¹ In Texas, 85 percent of teachers surveyed agreed with the statement “some of the best teachers are leaving the field because of the TAAS.”³²

Increased levels of anxiety, stress, and fatigue are frequently reported effects on students that may prohibit optimal performance on the state exam. In a national teacher survey, 75 percent reported that students were under intense pressure to perform well on their state test; 76 percent indicated that they perceived students to be extremely anxious about taking the exam.³³ One third of Kentucky teachers reported that student morale has declined in response to the KIRIS.³⁴ Other research conducted in Chicago involving 102 low-achieving 6th and 8th grade students illustrates the positive impact the test can have on student motivation. The results suggest that high-stakes testing encouraged a majority of students to work harder and in many cases led to score improvements.³⁵

Accountability and Meaning of Test Results

Not only do the results of state tests provide information about the progress of individual students, they are often aggregated to evaluate the performances of both schools and districts. In 2002, 17 states offered schools rewards for high or improved test scores; at least 19 attached sanctions for schools consistently exhibiting poor student performances on the state test; and 8 permitted students to transfer out of low-performing schools.³⁶ In addition to losing accreditation if students perform poorly on the state test, schools may also lose funding and face the possibility of closure or reconstitution.

Several studies have examined teachers’ views on accountability and found both positive and negative results. In North Carolina, 76 percent of teachers reported that they “believed that the accountability program would not improve the quality of education in their state,”³⁷ as compared with a majority of teachers surveyed in Kentucky who held

positive perceptions of the impact of the state test on instruction.³⁸ Research conducted in Maine and Maryland suggests that teachers' perceptions within the same state were not always consistent. In other words, the intended effects of the rewards and sanctions tied to test performance may be influenced by other factors specific to schools and districts, such as the availability of resources and professional development opportunities. As a result, state-testing policies may produce inconsistent and varied effects across schools and districts.³⁹

Research on the effects of state-mandated testing programs, as perceived by teachers, has revealed mixed effects: testing policies have both positive and negative impacts on instruction and learning, as well as on the teachers and students themselves. Therefore, it is necessary to determine if the benefits outweigh the potentially negative results, which, although unintended, may distort the educational process and are associated with great human costs. In order to address these complex issues, it is necessary to determine whether state testing policies are having the intended effect—not just by relying on information provided by test scores. A critical look at changes in schooling, as well as in teacher and student behaviors, is crucial. It is necessary to include those closest to the educational process in this endeavor.

Section 3: Data

In an effort to determine how state testing programs are affecting instruction and learning, the National Board on Educational Testing and Public Policy⁴⁰ sought the opinions of classroom teachers. The board's 2001 survey of teachers in 47 states included questions about the following: time spent on instruction in tested and non-tested areas; alignment of state standards and tests with teachers' instruction and classroom

assessments; pressure to improve test scores; school climate and morale; teachers' views on accountability; and their general opinions about state testing programs.⁴¹ From a sample of 12,000 public school elementary and secondary teachers, about 4,200 responded to the mail survey (35 percent). Of the responding teachers, 167 were from Florida.

This section compares the responses of Florida teachers with those of teachers in other states using high-stakes testing⁴² -- a useful comparison considering the federal push toward accountability and the increasing number of states using test results for grade promotion or graduation decisions. High-stakes policies affect the majority of the nation's public school teachers and students. The states using high-stakes testing are generally the most populous; by 2008, state exit exams will affect 7 out of 10 public school students.⁴³ These comparisons provide insight into the similarity between the perceptions of teachers in Florida and other states of the effects of their states' testing and accountability policies. As shown in Table 1, Florida teachers were slightly more diverse with regard to race/ethnicity and teaching experience than were teachers in other high-stakes states, but otherwise these are relatively similar groups.

Table 1: Profile of Responding Teachers from Florida and Other High-stakes States⁴⁴

Respondent Characteristics		Percent in Florida	Percent in Other High-stakes States
Gender	Female	81	83
	Male	19	17
Age	Over 40	59	67
Race/Ethnicity	African American	11	9
	Hispanic	13	6
	White	77	84
Years of teaching experience	1-12	50	38
	13-20	9	24
	Over 20	41	38
School type	Elementary	53	59
	Middle	25	20
	High	22	21

Source: Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College.

The survey results are organized into several broad areas: general views on the state standards and tests; impact on classroom practices; pressure on teachers and test preparation; impact on motivation and morale; views on the use of tests for purposes of accountability; and the meaning of test scores. Tables 2 and 3 present a summary of the results.

Views on State Standards and the State Test

Compared to teachers in other high-stakes states, Florida teachers held more positive views on state standards and the compatibility of the state test with their classroom instruction. Roughly three out of every four Florida teachers reported that the FCAT is based on a curriculum framework that all teachers should follow; a significantly smaller percentage of teachers in similar settings held this view (59 percent). Moreover,

79 percent of teachers in Florida, compared with 64 percent in other high-stakes states, found the state test compatible with their daily classroom practices. A greater percentage of Florida teachers also found the state test to have had a positive impact on the political agenda by bringing much needed attention to education issues than did teachers in states with similar types of testing programs (58 percent v. 42 percent). Florida's teachers' responses were similar to those in other high-stakes states: a majority in both groups reported that the state test measures high standards of achievement and that if they taught to the curriculum frameworks or standards, students would be successful.

Impact on Classroom Practices

Survey results as to the impact of the state test on classroom practices were consistent with research findings in this area. Teachers in Florida reported in larger percentages than did teachers in other high-stakes states that they had greatly increased time spent on tested material (62 percent v. 42 percent). Not surprisingly, a majority of teachers from both groups reported that they had decreased instructional time devoted to non-tested content (67 percent v. 58 percent). Florida teachers were almost twice as likely as teachers in states with similar testing programs to report fewer classroom enrichment activities, although the percentages were fairly small—23 percent compared with 12 percent, respectively.

With regard to the use of technology, Florida teachers reported in significantly larger percentages (than did teachers in other high-stakes states, 42 percent v. 32 percent) that they did not use computers when teaching writing because the state writing test requires students to provide handwritten responses. A consideration: tests that require

handwritten responses underestimate the abilities of students who are accustomed to writing with computers.⁴⁵

The overwhelming majority of teachers reported that the state testing program has led them to teach in ways contrary to their own ideas of sound educational practices. Florida teachers, though, were significantly more likely to hold this view than were teachers in other states with high-stakes policies (90 percent v. 75 percent). Survey results suggest that teachers in high-stakes settings gear the content of instruction to that of the state test, and that they are also modeling their own classroom assessments based on the test's format. Florida teachers reported these changes in greater percentages than did teachers in other high-stakes states. Furthermore, a great percentage of educators in Florida perceive that the instructional changes—required to meet the demands of the state testing programs—do not translate into good practice.

Pressure on Teachers

Compared with teachers in other high-stakes settings, those in Florida were more likely to feel pressure to raise test scores, especially from sources external to their school. Teachers in Florida were twice as likely to report test-related pressure coming from the district superintendent (in contrast to principals, 80 percent v. 40 percent). In general, Florida teachers reported feeling more pressure. Florida teachers also were significantly more likely than their counterparts in other high-stakes settings to strongly agree that the pressure for high scores on the state test was so intense that they had little time to teach anything not on the test (63 percent v. 40 percent). Test preparation took additional time for Florida teachers—54 percent of instructors reported this to be so, as compared with 43 percent of the educators in other high-stakes settings. The ways in which teachers in

both Florida and other high-stakes states prepared students were fairly similar, however. The use of commercially available or state-developed materials and released items were most common among both groups when getting students ready for the state test, although teachers in Florida were more likely to use prepared resources (75 percent v. 63 percent). Overall, when compared with teachers from other states that rely on high-stakes testing, Florida educators—in significantly greater percentages—felt pressure to raise students’ test scores. Florida instructors spent more time on test preparation—upwards of 30 hours per year—toward this effort.

Table 2: Percentage Agreement of Teachers in Florida and Other High-stakes States⁴⁶

Survey Items	Florida	Other High-stakes States
Views on Standards and the State Test		
The state test is compatible with my daily instruction.	79*	64
The state test is based on a curriculum that all teachers in my state should follow.	71*	59
If I teach to the state standards or frameworks, students will do well on the test.	57	54
The state test measures high standards of achievement.	54	48
The state testing programs has brought much needed attention to education issues in my state.	58*	42
Impact on Classroom Practices		
Greatly increased time spent on instruction in tested areas.	62*	42
Decreased time spent on instruction in non-tested areas.	67	58
Greatly decreased time spent on class enrichment activities.	23*	12
My tests are in the same format as the state-mandated test.	61*	50
I do not use computers when teaching writing because the state test is handwritten.	42*	32
The state testing program has lead teachers to teach in ways that contradict their own ideas of good educational practice.	90*	75
Pressure on Teachers to Prepare Students for the State Test		
Strongly agreed they felt pressure from their district superintendent to raise scores on the state test.	80*	56
Strongly agreed they felt pressure from their building principal to raise scores on the state test.	48	40
Strongly agreed that there is so much pressure for high scores on the state-mandated test that they have little time to teach anything not on the test.	63*	40
Spent more than 30 hours per year preparing students specifically for the state test.	54*	43
Prepared students for the state-mandated test throughout the year.	80	70
Used test preparation materials developed commercially or by the state.	75*	63
Used released items from the state test to prepare students.	43	44
Taught test taking skills to prepare students.	85	85

Source: Pedulla et al., 2003.

* Percentage differences are significant at alpha =.001.

Impact on Motivation and Morale

With regard to the impact of the state test on teachers' and students' respective levels of motivation and morale, the survey results for teachers in Florida and in other states with a high-stakes testing policy are consistent with prior research in this area. An overwhelming percentage of teachers in both settings—roughly 95 percent—reported that their schools' atmospheres were conducive to learning. A large majority of teachers in both settings also reported having high academic expectations for students (roughly 90 percent) and that student morale was high in their school (approximately 70 percent). Yet, less than a majority of teachers in both Florida and other high-stakes states indicated that teacher morale was high (about 45 percent). Moreover, a greater percentage of Florida teachers, compared with those in other high-stakes states, reported that teachers in their schools wanted to transfer out of the grades in which the state test is administered (49 percent v. 38 percent).

Florida teachers were also significantly more likely to report the perception that students were extremely anxious about the state test (92 percent v. 79 percent) and were under intense pressure to perform well (92 percent v. 79 percent). Even though a large majority of teachers in both Florida and other high-stakes states indicated that many of their students do try their best (approximately 90 percent), a smaller but sizable percentage, roughly 50 percent, reported that many students in their classes feel that, no matter how hard they try, they will still do poorly on the state test. Last, a significantly greater percentage of Florida teachers indicated that the testing program has led students to drop out of high school (38 percent v. 27 percent). A similar percentage of teachers in both groups, however, reported that the state test has contributed to an increase in grade

retention, roughly 30 percent. In general, Florida teachers were significantly more likely to perceive that students experienced high degrees of both test-related anxiety and pressure, compared with teachers in other states using high-stakes exams. Respondents in Florida and other high-stakes states gave similar responses with regard to teacher and student morale, school atmosphere, and expectations held of students; however, teachers in Florida were more likely to express a preference to relocate to grades in which the state test was not administered.

Views on Accountability and the Meaning of Test Scores

With the exception of their opinions on the use of test results to award school accreditation, the responses of Florida teachers were similar to those of practitioners in other high-stakes states. By and large, the majority of both groups of teachers regarded the use of test results for school and student accountability purposes as inappropriate. For example, a sizable percentage of teachers found using test results to award school accreditation inappropriate; Florida teachers were significantly more likely to hold this view (77 percent v. 65 percent). Roughly 60 percent of teachers in both groups reported that using scores on the state test to make decisions about grade promotion, or about retention, was inappropriate. Yet the decision to grant a student a high school diploma can in fact be determined by tests: a majority of teachers in Florida and in other states using high-stakes testing held this view (approximately 55 percent).

Florida teachers and their counterparts in other high-stakes states held especially strong views about the capacity of the state test to serve both as an accurate measure of student achievement and as an indicator of school quality. Roughly 80 percent disagreed with the statement that the state test is as accurate a measure of student achievement as a

teachers' judgment. A similar percentage of teachers in both groups also disagreed that the scores on the state test accurately reflect the quality of education students have received. Moreover, Florida teachers reported—in greater percentages than did teachers in similar high-stakes settings—that they had found ways to raise students' scores on the state test without really improving learning (59 percent v. 39 percent); thus suggesting that improvements on the state test may not show up on other tests that measure similar content and skills.

Table 3: Percentage Agreement of Teachers in Florida and Other High-stakes States

Survey Items	Florida	Other High-stakes States
Impact on Motivation and Morale		
My school has an atmosphere conducive to learning.	95	92
Teachers have high expectations for the academic performance of students in my school.	92	91
Teachers in my school want to transfer out of the tested grades.	49*	38
Teacher morale is high in my school.	46	44
Strongly agreed that students are extremely anxious about taking the state test.	92*	79
Students are under intense pressure to perform well on the state test.	92*	79
The majority of students try their best on the state test.	91	83
Student morale is high in my school.	71	64
Many students in my class feel that, no matter how hard they try they will still do poorly on the state test.	44	52
The state test has caused many students in my district to drop out of high school.	38*	27
The state test results have led to grade retention in my district.	30	27
Views on Accountability and the Meaning of Test Scores		
Inappropriate to use test results to award school accreditation	77*	65
Inappropriate to use test results to promote or retain students in grade	62	59
Appropriate to use test results to award high school diplomas	52	57
Teachers in my school have found ways to raise state-mandated test scores without really improving learning.	59*	39
The state test is as accurate a measure of student achievement as a teacher's judgment.	14	19
Scores on the state test accurately reflect the quality of education students have received.	13	20

Source: Pedulla et al., 2003.

*Percentage differences are significant at alpha = .001.

Quality of Available Data

All measures that rely on self-reported data rest on the assumption that the participants respond truthfully and accurately; unfortunately, it is almost impossible to test this assumption. Efforts at the developmental stage, as well as statistical analyses, suggest that the survey instrument used in the national study and the results described herein are reliable and valid. In an effort to produce a high quality measure of teachers' opinions, the survey was based in part on other surveys used to gather teachers' opinions about the effects of state testing programs.⁴⁷ In addition, analyses revealed a high level of consistency among teachers' responses to the survey items. This indicates that the measurement tool produced reliable results.⁴⁸

Another consideration when assessing the quality of survey data is whether the respondents are actually representative of the population to which results are generalized. The teachers who completed the National Board survey were comparable to the national teaching force in terms of their ages, races/ethnicities, the types of schools in which they taught, and their years of teaching experience.⁴⁹ In addition, the responding group of Florida teachers was similar to the larger teaching population in the state with regard to sex, race/ethnicity, and type of school.⁵⁰ The available data present a national picture of how teachers are responding to different types of testing programs. Although the sampling design for the national survey was not intended for disaggregating and reporting results at the state level, the reasonably sufficient number of responding teachers from Florida provides for useful comparisons. These results, however, should be interpreted with some caution and viewed within the limitations of self-reported data.

Section 4: Findings

Florida teachers' survey responses reflect mixed views on the state testing and accountability program. Generally, the most positive views were directed toward the state's curriculum frameworks. For example, a large majority reported that the Sunshine State Standards should be followed, and that the state test was compatible with their daily instructional habits. About half regarded the state test as a measure of high standards of achievement. An alternative way of interpreting teachers' responses to the questions about their views on standards is that they have accepted the content of the state test as reflecting that of the state standards, and agree that if they teach to those standards, their students will do well on the test. This is one view that may account for the divide between teachers' more positive opinions toward the content standards and more negative views on the state test. However, research has consistently shown that teachers generally hold positive views toward content standards and find that they provide instructional focus, ensure greater homogeneity across classrooms with regard to instructional content, and facilitate collaboration with their colleagues.⁵¹

With regard to classroom practices, Florida teachers' responses were consistent with prior research findings. The majority reported that they had reallocated instructional schedules, allowing for more time to be spent on tested content while decreasing the time ordinarily devoted to material that would not appear on the FCAT. In addition, a fourth of the respondents indicated that they had also decreased time spent on enrichment activities in order to prepare students for the state test.

Teachers' responses suggest that the content of the state test influences their instruction, as does the format of the FCAT: a sizable minority indicated that they do not

use computers for writing instruction because students are required to provide handwritten responses to questions on the test. In addition, a majority of teachers also changed their assessment practices by modeling their own classroom tests after the format of the state test. Although a substantial percentage of teachers reported changes in their classroom practices, it was clear that they were not comfortable with making them: 90 percent indicated that the state test required them to teach in ways contrary to their own ideas of good practice.

Results also suggest that for many teachers, schools are highly stressed environments where a premium is placed on improving student test performance. Florida teachers held especially strong opinions about the pressure to raise test scores; 63 percent indicated the pressure was so great that they had little time to teach anything that would not appear on the test. Furthermore, a majority reported that they had found ways to raise test scores without improving learning. Taken together, these findings suggest that pressure to raise test scores has forced educators to fixate on short-term, immediate goals, perhaps at the expense of developing skills that encourage long-term, independent learning.

Results also suggest that the state testing and accountability program in Florida may have potentially negative implications for the teaching profession. Less than half of the teachers surveyed reported that morale was high at their school; further, a majority indicated that teachers wanted to transfer out of the grades in which the test was given.⁵² Since the National Board survey was conducted, the number of grades in which the FCAT is given has increased; now, all Grades 3-8 and 10 are tested in at least one subject of the FCAT. As a result, there are few classroom practitioners unaffected by the state

test; instead of making changes to their teaching assignments within their respective schools or districts, they may choose to leave the profession altogether.

The pressure to succeed on the state test did not rest solely with teachers; in Florida, students were also affected by the FCAT. More than 90 percent of respondents believed that students were under intense pressure and were highly anxious about the test. While moderate levels of test-related anxiety can actually improve motivation and test performance, an unmanageable amount can have an adverse effect. Students who view the test as an insurmountable barrier are also more likely to give up.⁵³ A sizable percentage of teachers reported that many of their students have adopted the view that passing the state test may be a hopeless goal: 45 percent of teachers indicated that no matter how hard they tried, many students felt that they would still do poorly on the FCAT. The fact that almost 40 percent of teachers reported that the state testing program had contributed to students' decisions to leave high school is disquieting.

In order to meet high school diploma requirements, legislative provisions have allowed students to use scores on alternative assessments such as the Scholastic Achievement Test (SAT), the American College Test (ACT), and the College Placement Tests in lieu of 10th grade FCAT scores.⁵⁴ Although this was an immediate solution to the potentially large numbers of Grade 12 students in Florida who failed to meet the FCAT requirements and who thus would not receive high school diplomas, it does not solve the problem in the long term. Only 6 out of 67 school districts had a majority of Grade 12 students pass the reading section of the 2003 FCAT; 10 districts boasted similar results on the math portion.⁵⁵ Tenth grade results from the previous year (2002) suggest that the picture is not as bleak, but it is still very troublesome: 73 percent passed the math

portion and 59 percent passed the reading test. When the results are disaggregated by race, however, that rate of success is far lower for Black and Hispanic students than for their White or Asian counterparts.⁵⁶ These results suggest that for many students, especially for those of minority groups, failure on the FCAT is a real possibility, thus increasing the likelihood that they may leave the public school system altogether.

Already the reality of decreased graduation rates and increased rates of grade retention is evident. Data (1984-2000 state enrollment statistics) reveal that Florida has a 63 percent graduation rate, one of the lowest in the nation. This figure is found by comparing enrollments for Grade 8, four and a half years earlier, with the number of students who eventually graduated. Florida also has the highest Grade 9 to 10 attrition rates. One out of every four 9th grade students in the 1999-2000 academic year did not progress to the 10th grade the following school year (2000-2001).⁵⁷ These statistics, coupled with the fact that 30-40 percent of 10th grade high school students did not pass either the math or the reading portion of the 2002 FCAT, are alarming and are indicative of the widespread human costs associated with high-stakes testing programs.

With regard to using test scores for accountability purposes, the majority of teachers did not support using test results to hold schools accountable, but their opinions about student accountability were mixed: a majority supported using test scores as a criterion for graduation. Teachers were less supportive of using results to determine grade promotion or retention, however. These findings may be explained, in part, by a belief on the part of teachers that accountability for student test performance should be shared among schools, teachers and students. Florida teachers held much more decisive views on the meaning of test scores. Overwhelmingly, educators did not consider

students' scores on the FCAT accurate indicators of either student achievement or school quality, and regarded their own professional judgments as more precise measures of student learning.

Perhaps teachers recognize that like any measurement tool that produces a number—such as blood pressure gauges, complex laboratory tests, radar detectors, and breathalyzers—test scores are fallible. Most state laws, however, do not consider margin of error when interpreting a student's scores. Misguided executive decisions, poorly conceived policies, understaffing, unrealistic deadlines, and unreasonable progress goals can cause numerous errors in test scores. The extent to which human errors in scoring, programming, and reporting are widespread has been thoroughly documented.⁵⁸ Students have been wrongfully denied high school diplomas, mistakenly thought to have failed state exams, and erroneously required to attend summer schools as a result of test score errors. For example, in Florida, when the tests of 71 third-graders who had scored below the passing cut-off score were reviewed and scored by hand, the process unearthed a scoring machine error—leading a parent to comment, “Instantly, Raven was transformed from a 3rd-grade dufus to a state-certified 4th grader.”⁵⁹ Given these limitations and potential for serious harm, the *American Educational Research Association* has put forth a position statement that cautions policy makers about the use of high-stakes testing and advocates that test scores should not be used in isolation to make important decisions about “students’ life chances or educational opportunities.”⁶⁰

When the survey responses of Florida teachers were compared with those of teachers in other high-stakes states, clear differences emerged. Teachers in Florida were more likely to hold positive views on the state standards and on the compatibility of the

testing program with their instruction than were teachers in other states using high-stakes tests. In addition, comparisons indicated that the state test had a greater impact on the classroom practices in Florida, as evidenced by the larger percentages of the state's teachers who reported spending more time on tested content, and less time on enrichment activities; modeling the format of the state test in their own assessments; providing writing instruction in longhand rather than with computers (because computers were not permitted for writing on the state test); and teaching in ways that contradicted their own beliefs about good practice.

Florida teachers were more likely than their counterparts in other high-stakes states to report feeling pressure to raise test scores. They also were more likely to teach to tested content and to spend instructional time on test preparation. Teachers in Florida were, in addition, more likely to perceive negative effects on students, compared with teachers in other high-stakes states, as they reported in greater percentages that their students experienced high levels of test-related anxiety and pressure. Further, a larger percentage of Florida teachers reported that the state test has led students to drop out of high school, compared with teachers in other states that rely on high-stakes exams. Although a larger majority of teachers in Florida held the state standards in higher regard compared to their counterparts in other high-stakes states, they were also more likely to oppose the use of test results for school accountability purposes. However, Florida teachers' views regarding student accountability were similar, with a majority in both groups supporting the use of test results to award high school diplomas.

Overall, the survey results indicate that, when compared to teachers in other states that use high-stakes testing, the state testing program in Florida is having a more

pronounced effect, in both positive and negative ways, on classroom practice, on teachers, and on students. In addition, according to Florida teachers, the extent to which FCAT scores are accurate measures of student achievement is questionable: the pressure to raise test scores may have distorted the educational process by requiring educators to over-emphasize test-specific preparation, and to implement what they view as unsound instructional practices.

Section 5: Recommendations

The importance of shared accountability in efforts to improve the quality of education and to enhance student learning is undeniable; the value of information imparted by measures of achievement is irrefutable. When teachers and students are the main parties on which rewards or sanctions are leveraged, however, research indicates that the educational process becomes distorted. In the attempt to improve indicators on which highly consequential decisions are based, efforts are directed toward the rapid raising of test scores, often at the expense of more effective educational practices. The current survey findings suggest that this is happening in Florida to a greater extent than in other states that have implemented high-stakes testing programs. These findings, considered together with the large body of research conducted on the effects of high-stakes testing on both teaching and learning, point to the following recommendations:

1. Florida should undertake a long-term evaluation and monitoring program to assess the impact of the Florida Comprehensive Assessment Test (FCAT) and the A+ Accountability program. This evaluation and monitoring program should be conducted by an external organization or research institution. Its purpose is to determine if the state testing program is achieving its intended

goals. The evaluation should also examine the unexpected consequences of the FCAT and A+ Accountability program on the educational process and on key stakeholders.

2. Florida testing policy should adhere to the recognized professional standards regarding test development and to the appropriate use of test results as described in the *Standards for Educational and Psychological Testing*, published jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.
3. Florida policy makers should not make highly consequential decisions about students (such as deciding whether a student is promoted to the next grade or is awarded a high school diploma) by means of test scores alone. Given the evidence pointing to weaknesses in the testing system, it is important to use other sources of information in conjunction with state tests.

Notes and References

¹ This brief is based on a larger study of the impact on state testing programs conducted by the National Board on Educational Testing and Public Policy at Boston College. The original study was generously supported by a grant from the Atlantic Philanthropies. The findings and conclusions described herein are those of the author and do not necessarily reflect the views of the Atlantic Philanthropies.

² See, for example:

Abrams, L. & Madaus, G. (2003). The lessons of high-stakes testing. *Educational Leadership*, 61(3), 31-35.

Koretz, D., Linn, R., Dunbar, S., & Shepard, L. (1991, April). *Effects of high-stakes testing on achievement: Preliminary findings and generalization across tests*. Paper presented at the annual meeting of the American Education Research Association. Chicago, IL.

Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.) *Critical issues in curriculum* (pp. 83-121). Chicago: University of Chicago Press.

McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Smith, M. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.

³ See, for example:

Stecher, B. & Barron, S. (1999). *Quadrennial milepost accountability testing in Kentucky* (CSE Technical Report 505). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Stecher, B., Barron, S., Chun, T., & Ross, K. (2000). *The effect of Washington state education reform on schools and classrooms* (CSE Technical Report 525). Los Angeles: National Center for Research on Evaluation, Standards and Student Testing (CRESST).

⁴ Hamilton, L., Stecher, B., & Klein, S. (Eds.) (2002). *Making sense of test-based accountability in education*. Washington, DC: RAND.

⁵ Iowa has not developed state content standards; instead each district was responsible for developing its own content standards and benchmarks (Quality Counts, 2001). Local district content standards rather than state standards were also included as part of Iowa's accountability plan submitted to the U.S. Department of Education in 2003 (www.state.ia.us/educate).

⁶ Edwards, V. (Ed.) (2003, January 9). *Quality Counts 2003: If I can't learn from you...* (Education Week Special Report), 12(17). Bethesda, MD: Editorial Projects in Education.

⁷ See, for example:

Edwards, V. (Ed.) (2003, January 9). *Quality Counts 2003: If I can't learn from you...* (Education Week Special Report), 12(17). Bethesda, MD: Editorial Projects in Education. Kentucky, Vermont, and Washington use test results to hold schools accountable. In New York, Massachusetts, Texas, Virginia, and Florida test results are used to make highly consequential decisions at both the school and student levels.

⁸ Previous reauthorization of the ESEA (e.g. the 1994 Improving America's Schools Act) left the states largely responsible for dealing with low- and under-performing schools. In contrast, NCLB requires states to meet Adequate Yearly Progress (AYP) goals to ensure school accountability for student achievement on

state tests; if AYP goals are not consistently achieved, schools can face increasingly demanding corrective actions (e.g. replacement of school staff, implementation of new curriculum, extension of the school day or academic year, restructuring). See:

Joftus, S. & Maddox-Dolan, B. (2003). *Left out and left behind: NCLB and the American high school*. Washington, DC: Alliance for Excellent Education. Retrieved May 1, 2003 from <http://www.all4ed.org>

⁹ Under NCLB provisions states will be required to administer a science exam annually starting in the 2007-2008 academic year. Students will be tested in science less frequently than in reading and math. The law requires that the science test be administered in each of the following spans: 3-5, 6-9, and 10-12 (www.nclb.gov).

¹⁰ U.S. Department of Education, National Center on Education Statistics (2002). *Digest of education statistics 2001*. Washington, DC: Author.

¹¹ A history of the state assessment and accountability program in Florida is available on the state department of education website (<http://www.firn.edu/dow/sas/hsap9000.htm>). The core issue in the landmark case Debra P. v. Turlington was the use of a high school graduation test to award diplomas. The courts ruled in favor of the State, allowing the State to deny a diploma to any student who had not passed the High School Competency Test provided the State could provide evidence that all students had been given instruction on the content and skills tested. Students in class of 1983 were the first required to pass the competency test to earn and high school diploma.

¹² Edwards, V. (Ed.) (2003, Jan. 9). *Quality Counts 2003: If I can't learn from you...*(*Education Week Special Report*), 12(17). Bethesda, MD: Editorial Projects in Education.

¹³ Florida State Department of Education, *FCAT fact sheet*. Available at <http://www.firn.edu/doe/sas.fcat.htm>

¹⁴ Office of Program Policy Analysis and Government Accountability (2001). *Justification review: Kindergarten through twelfth grade public education program* (Report No. 01-22). Available at <http://www.oppaga.state.fl.us/reports/pdf/0122.pdf>

¹⁵ The majority of these studies gathered information from those closest to the educational process, teachers and administrators, by using surveys, interviews, classroom observations and various combinations thereof. Most studies have focused on a single state and have been conducted in Arizona, Colorado, Florida, Kentucky, Maine and Maryland, North Carolina, Texas, Vermont, Virginia and Washington. See, for example:

Smith, M., et al. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program (ASAP): Equity and teacher capacity building* (CSE Technical Report 425). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught and what gets lost* (CSE Technical Report 588). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95-117.

Koretz, D., Barron, S., Mitchell, K., & Keith. S. (1996). Perceived effects of the Kentucky instructional results information system (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: Rand.

- Lane, S., Parke, C., & Stone, C. (1998, April). *Consequences of the Maryland school performance assessment program*. Paper presented at the annual meeting of the National Council of Measurement in Education, San Diego, CA.
- Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), 199-203.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved December 15, 2003, from <http://epaa.asu.edu/epaa/v8n41>
- Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482-494.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5-16.
- McMillan, J., Myran, S., & Workman, D. (1999, April). *The impact of mandated statewide testing on teachers' classroom assessment and instructional practices*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada, p. 10.
- Stecher, B., Barron, S., Chun, T., & Ross, K. (2000). *The effect of Washington state education reform on schools and classrooms* (CSE Technical Report 525). Los Angeles: National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College. The results of this more recent national study of teachers' perceptions of the impact of state mandated testing programs frequently confirm state-level research findings.
- ¹⁶ McMillan, J., Myran, S., & Workman, D. (1999, April). *The impact of mandated statewide testing on teachers' classroom assessment and instructional practices*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada, p. 10.
- ¹⁷ Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996a). The perceived effects of the Kentucky Instructional Results Information System (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: RAND.
- ¹⁸ Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College
- ¹⁹ Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught and what gets lost* (CSE Technical Report 588). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- ²⁰ Koretz, D., Barron, S., Mitchell, K., & Keith, S. (1996). Perceived effects of the Kentucky instructional results information system (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: Rand.
- ²¹ Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College.
- Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught and what gets lost* (CSE Technical Report 588). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

- ²² Russell, M. & Abrams, L. (in press). Instructional uses of computer for writing: The impact of state testing programs. *Teachers College Record*.
- ²³ Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996b). *Perceived effects of the Maryland school performance assessment program* (Technical Report 409). Los Angeles: National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- ²⁴ Koretz, D., Barron, S., Mitchell, K., & Keith, S. (1996). Perceived effects of the Kentucky instructional results information system (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: Rand.
- ²⁵ Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), p. 201.
- ²⁶ See, for example:
- Haladyna, T., Nolen, S., & Haas, N. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Koretz, D., Linn, R., Dunbar, S., & Shepard, L. (1991, April). *Effects of high-stakes testing on achievement: Preliminary findings and generalization across tests*. Paper presented at the annual meeting of the American Education Research Association. Chicago, IL.
- Linn, R. (2000). Assessment and accountability. *Educational Researcher*, 29(2), 4-16.
- Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.) *Critical issues in curriculum*, pp. 83-121. Chicago: University of Chicago Press.
- ²⁷ Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482-494.
- ²⁸ Koretz, D. & Barron, S. (1998). The validity gains on the Kentucky Instructional Results Information System (KIRIS) (MR-1014-EDU). Santa Monica, CA: RAND.
- When students' scores on the TAAS were compared with NAEP test results, researchers found a similar trend. See:
- Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2000). *What do test scores in Texas tell us?* (IP-202). Santa Monica, CA: RAND.
- ²⁹ Amrein, A. & Berliner, D. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved April 13, 2002, from <http://epaa.asu.edu/epaa/v10n18/>
- ³⁰ McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Smith, M. (1991). Put to the test: The effects of external testing on teachers. *Education Researcher*, 20(5), 8-11.
- ³¹ Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), 199-203.
- ³² Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482-494.
- ³³ Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College.

- ³⁴ Koretz, D., Barron, S., Mitchell, K., & Keith, S. (1996) Perceived effects of the Kentucky instructional results information system (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: Rand.
- ³⁵ Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197-227.
- ³⁶ Edwards, V. (Ed.) (2003, Jan. 9). *Quality Counts 2003: If I can't learn from you...* (Education Week Special Report), 12(17). Bethesda, MD: Editorial Projects in Education, p. 87.
- ³⁷ Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), p. 202.
- ³⁸ Koretz, D., Barron, S., Mitchell, K., & Keith, S. (1996). Perceived effects of the Kentucky instructional results information system (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: Rand.
- ³⁹ Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95-117.
- ⁴⁰ Located in the Lynch School of Education at Boston College, The National Board of Educational Testing and Public Policy (NBETPP) is an independent organization that both monitors testing in the United States and seeks to provide information about the use and impact of educational testing.
- ⁴¹ In an effort to achieve a nationally representative sample of teachers, the selection process involved four criteria. After identifying the stakes-level classification for each state, the sample was further organized according to school type (e.g. elementary, middle, high school), subject area taught for high school teachers (e.g. English, math, science, social studies, and special education) and finally by geographic location to ensure that teachers from urban and non-urban were proportionally represented. More detailed information about the sampling process as well as the survey instrument can be found in the full report which is available on-line at <http://www.bc.edu/nbetpp> or see:
- Abrams, L., Pedulla, J., & Madaus, G. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into practice* 42(1), 18-29.
- ⁴² At the time of the survey administration states were classified according to the stakes attached to the state test at both the school and student level. This process was based on information found in state legislatures, direct contact with state departments of education, their personnel, and web sites as of January, 2001. Eighteen states were identified as having high stakes for schools and high stakes for students: Alabama, California, Delaware, Florida, Georgia, Indiana, Louisiana, Maryland, Massachusetts, Mississippi, Nevada, New Jersey, New Mexico, New York, North Carolina, South Carolina, Tennessee, Texas, and Virginia. Teachers' responses from these states, with Florida excluded, comprise the high-stakes classification.
- ⁴³ Chudowsky, N., Kober, N., Gayler, K., & Hamilton, M. (2002). *State high school exit exams: A baseline report*. Washington, DC: Center on Education Policy.
- ⁴⁴ Percentages may not total 100 due to rounding.
- ⁴⁵ Russell, M. (2002). *How computer-based technology can disrupt the technology of testing and assessment* Chestnut Hill, MA: Technology and Assessment Study Collaborative, Boston College. Available at <http://www.bc.edu/intasc>.
- Russell, M. & Abrams, L. (in press). Instructional uses of computer for writing: The impact of state testing programs. *Teachers College Record*.
- ⁴⁶ Significance testing for the results presented in Tables 2 and 3 was conducted at the item-level using chi-square tests. Differences that were significant at alpha = .001 are highlighted.

⁴⁷ The National Board survey was based in part on teacher surveys conducted in Arizona (Smith, Nobel, Heinecke et al., 1997), Maryland (Koretz et al., 1996b), Michigan (Urdu & Paris, 1994), and Texas (Haney, 2000) as well as on the National Science Foundation (NSF) study of the Influence of Testing on Teaching Math and Science Grades 4-12 (Madaus, West, Harmon, Lomax, & Viator, 1992) and a study of the Effects of Standardized Testing (Kelleghan, Madaus, & Airasian, 1980).

⁴⁸ Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College. Factor analyses were conducted and produced several different scales. The reliability index ranged from a low of .57 to a high of .91, with the majority of scales having a reliability of above .80, indicating a strong degree of internal consistency.

⁴⁹ See Appendix D and:

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M. & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College.

⁵⁰ According to demographic data available on the Florida State Department of Education website (<http://www.firn.edu/doe/eias/eiaspubs/pdf/psstaff.pdf>) the Florida teachers who responded to the National Board Teacher Survey were similar to the state's larger teaching population according to several characteristics: school type (52% elementary, 48% secondary), sex (22% male, 78% female), and race/ethnicity (67% White, 21% Black, 11% Hispanic). See Table 1 for characteristics of the 167 Florida teachers who participated in the national survey.

⁵¹ See, for example:

Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College.

⁵² Anecdotal evidence from another study in Florida found shifts in personnel from tested to non-tested grades as well in response to the state's A+ Plan. See:

Goldhaber, D. & Hannaway, J. (2001). *Accountability with a kicker: Observations on the Florida A+ Plan*. Paper presented at the annual meeting of the Association of Public Policy and Management, Washington, DC.

⁵³ Kelleghan, T., Madaus, G., & Raczek, A. (1996). The use of external examinations to improve student motivation. *American Educational Research Association*.

⁵⁴ 2003 Florida Statutes 1008.22 [9]. Available at <http://www.flsenate.gov>

⁵⁵ Florida State Department of Education (2003). *FCAT 2003 grade 12 passing percent by district*. Available at <http://www.fccat.fldoe.org>

⁵⁶ Gayler, K., Chudowsky, N., Kober, N., & Hamilton, M. (2003). *State high school exit exams: Put to the test*. Washington, DC: Center on Education Policy.

⁵⁷ Haney, W., Madaus, G., Abrams, L., Wheelock, A., Miao, J., & Gruia, I. (2004). *The education pipeline in the United States, 1970-2000*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College. Available at <http://www.bc.edu/nbetpp>

⁵⁸ Rhoades, K. & Madaus, G. (2003). *Errors in standardized tests: A systemic problem*. Chestnut Hill, MA: The National Board on Educational Testing and Public Policy, Boston College

⁵⁹ Winerip, M. (2003, July 23). Rigidity in Florida and the results. *The New York Times*, p.B9.

⁶⁰ American Educational Research Association. *Policy on High-Stakes Testing in PreK-12 Education*. Available at <http://www.area.net/about/policy/stakes.htm>. This policy statement is based the 1999 *Standards for Educational and Psychological Testing*.