

8-9-2004

# Education Policy Analysis Archives 12/39

Arizona State University

University of South Florida

Follow this and additional works at: [http://scholarcommons.usf.edu/coedu\\_pub](http://scholarcommons.usf.edu/coedu_pub)

 Part of the [Education Commons](#)

---

## Scholar Commons Citation

Arizona State University and University of South Florida, "Education Policy Analysis Archives 12/39" (2004). *College of Education Publications*. Paper 501.

[http://scholarcommons.usf.edu/coedu\\_pub/501](http://scholarcommons.usf.edu/coedu_pub/501)

This Article is brought to you for free and open access by the College of Education at Scholar Commons. It has been accepted for inclusion in College of Education Publications by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Copyright is retained by the first or sole author, who grants right of first publication to the **Education Policy Analysis Archives**. EPAA is a project of the Education Policy Studies Laboratory. Articles are indexed in the Directory of Open Access Journals ([www.doaj.org](http://www.doaj.org)).

Volume 12 Number 39

August 9, 2004

ISSN 1068-2341

---

## **Voices from the Frontlines: Teachers' Perceptions of High-Stakes Testing**

**Brett D. Jones  
Robert J. Egley  
University of South Florida, St. Petersburg**

Citation: Jones, B. D. & Egley, R. J. (2004, August 9). Voices from the Frontlines: Teachers' Perceptions of High-Stakes Testing, *Education Policy Analysis Archives*, 12(39). Retrieved [date] from <http://epaa.asu.edu/epaa/v12n39/>.

### **Abstract**

The purpose of this study was to investigate whether teachers perceived Florida's high-stakes testing program to be taking public schools in the right direction. More importantly, we sought to understand *why* teachers perceived the tests to be taking schools in the right or wrong direction. Based on the survey results of 708 teachers, we categorized their concerns and praises of high-stakes testing into ten themes. Most of the teachers believed that the testing program was not taking schools in the right direction. They commented that the test was used improperly and that the one-time test scores were not an accurate assessment of students' learning and development. In addition, they cited negative effects on the curriculum, teaching and learning, and student and teacher motivation. The positive effects cited were much fewer in number and included the fact that the testing held students, educators, and parents accountable for their actions. Interestingly, teachers were not opposed to accountability, but rather, opposed the manner in which it was currently implemented. Only by understanding these positive and negative effects of the testing program can policymakers hope to improve upon it. To this end, we discuss several implications of these findings, including: limiting the use of test scores, changing the school grading criteria, using alternative assessments, modifying the curriculum, and taking steps to reduce teaching to the test.

The use of high-stakes tests in schools has been questioned since they were first implemented in most states several years ago. Some have questioned the use of student test scores to measure educational quality (Popham, 1999), while others have questioned the more direct effects on students and teachers (Kohn, 2000). Yet, politicians and many in the public seem more determined than ever to hold educators accountable through the use of high-stakes tests. By “high-stakes” we are describing tests that have serious consequences for students, teachers, schools, and/or school systems, such as student retention, school ratings, and monetary incentives.

Studies conducted soon after the implementation of high-stakes testing programs indicated that many teachers were not supportive of the use of high-stakes tests. Teachers noted several negative effects on education including a narrowing of the curriculum, increased teaching to the test, lower teacher morale, increased student and teacher stress, and other negative effects on students and teachers (Jones et al., 1999; Smith, 1991). We wondered whether teachers’ perceptions of testing had changed over the past few years. For instance, have teachers begun to adapt to this new era of testing in education and come to understand how testing has or can improve education? Have the initial negative reactions against testing subsided as teachers have had a chance to work in this new testing climate and better understand how it affects them and their students?

The purpose of this study was to answer these questions by asking Florida teachers about their perceptions of testing near the end of the fourth year of high-stakes testing in Florida. The specific purpose of this study was to investigate whether teachers perceived Florida’s high-stakes testing program to be taking Florida’s public schools in the right direction. More importantly, we sought to understand *why* teachers perceived the tests to be taking schools in the right or wrong direction. Based on their perceptions, we developed a framework to organize teachers’ concerns and praises of high-stakes testing. While other studies have described teachers’ perceptions of testing, none have used qualitative data from hundreds of teachers from many schools and districts to systematically identify and categorize these perceptions. Only by understanding the positive and negative effects of testing can policymakers hope to improve upon current testing programs.

## Background

### **The Florida Comprehensive Assessment Test (FCAT)**

Florida is an interesting state to assess teachers’ perceptions of high-stakes testing because it is a large state with a wide range of urban and rural schools. In addition, Florida’s testing program, called the Florida Comprehensive Assessment Test (FCAT), was developed under the leadership of Governor Jeb Bush and appears to be consistent with the type of testing being promoted at the national level by President Bush’s No Child Left Behind Act of 2002. This act requires students nationwide in the third through eighth grade to be tested in the basics of mathematics, reading or language arts, and (beginning in 2005) science.

The FCAT was first administered in Florida’s public schools and used for accountability purposes in the spring of 1999. The present study was conducted near the end of the fourth year of testing in the spring of 2002. Starting in the spring of 1999, schools were assigned a letter grade ranging from “A” (making excellent progress) to “F” (failing to make adequate progress) based on several criteria: a) the percentage of students scoring above certain levels in reading, writing, and math (the percentages and levels varied for each subject); b) the percentage of students making learning gains in reading and math compared to the previous year; c) the percentage of the lowest 25% of students who made adequate progress; and d) the

percentage of students completing the test (e.g., 95% of eligible students were required to complete the test for the school to receive an “A”) (Florida Department of Education, 2002a).

School grades were directly linked to accountability rewards and sanctions (Florida Department of Education, 2001). Schools graded an “A” or that had improved at least one grade level were eligible for monetary incentives. Students attending schools graded an “F” for two years in a four-year period were eligible for scholarships to attend another public or private school. Student retention decisions were made by the local school boards, although students were required to pass the reading and math FCAT in tenth grade starting in 2002-2003 to graduate from high school.

The test consisted of a criterion-referenced test that measured the state standards in reading, writing, and mathematics and a norm-referenced test that measured student performance against national norms (Florida Department of Education, 2001). The reading and math tests were given in grades 3 through 10 and the writing test was given in grades 4, 8, and 10. The FCAT consisted of multiple-choice items at all grade levels tested and “performance items” (requiring a written answer) in reading in grades 4, 8, and 10 and in math in grades 5, 8, and 10. Test results were provided at the student, school, district, and state level.

### **Effects of Testing on Teachers and Students**

Initial research into the effects of testing on teachers in states such as Arizona (Smith, 1991) and North Carolina (Jones et al., 1999) indicated that teachers had many concerns about using high-stakes tests as a mechanism for teacher accountability. In North Carolina, 76% of teachers surveyed reported that the testing program would *not* improve the quality of education in their schools (Jones et al., 1999). Similarly, when teachers in Virginia were asked whether the testing program was taking Virginia in the right direction, 39% said no, 38% said they were uncertain, and 22% said yes (Kaplan & Owings, 2001). While most of the effects reported by teachers have been negative, some positive outcomes of testing have also been reported. In this section, we discuss some of the major positive and negative effects that high-stakes testing have had on teachers and students.

One of teachers’ major concerns regarding high-stakes testing was that it “narrowed the curriculum” by forcing teachers to teach only the subjects that were tested to the exclusion of the non-tested subjects such as science, social studies, and health. As Smith (1991) describes: [Some teachers] began discarding what was not to be tested and what was not part of the formal agenda and high priorities of the principal and district administrators. One can imagine a kind of evolutionary process at work, with those teachers who correctly narrow curriculum and maximize scores being those that prosper or escape punishment. (p. 10)

A related concern was that the testing caused teachers to teach to the test by organizing their instruction around illustrative items that were the same as, or look like, actual test items. This type of item teaching can cause test score pollution by giving students an unfair advantage over students who have not been privy to item teaching (Haladyna, Nolen, & Haas, 1991; Popham, 2000a).

On the other hand, the testing has forced some teachers who might not have been teaching the state curriculum to re-assess what they are teaching. As an example, Ohio teachers reported that “testing has helped the school system align curriculum between grade levels, has helped educators identify curricular weaknesses, and has made educators more conscious of educational outcomes” (DeBard & Kubow, 2002, p. 396). Providing an impetus for teachers to review how the state curriculum aligns with what they are teaching has to be considered a positive outcome of testing.

Test preparation and administration have also been blamed for reducing the amount of time available for instruction (Jones, Jones, & Hargrove, 2003). For instance, one study of Texas educators found that test preparation occurred during the entire year and that teachers

spent from 8 to 10 hours a week on test preparation (Hoffman, Assaf, & Paris, 2001). Teachers have complained that students spend a lot of time practicing test taking strategies rather than engaging in learning. As one teacher commented, “Just think what you could do if you took all that time spent on testing and preparing for testing and used it to teach. There’s way too much testing” (Barksdale-Ladd & Thomas, 2000, p. 392).

The effects of testing on teachers’ teaching practices has been mixed. There is a growing consensus that high-stakes testing has a positive effect on some teachers’ teaching practices, a negative effect on some teachers’ practices, and little to no effect on others teaching practices (Cimbricz, 2002; Jones, Jones, & Hargrove, 2003). Others have found that the pressure through testing has more of an effect on the content taught than the teaching practices (Firestone & Mayrowetz, 2000).

Teachers have reported feeling shame, embarrassment, guilt, and anger from the publication of student test scores (Smith, 1991). Part of teachers’ frustration has been that they do not believe that the tests adequately capture the complexity of students’ learning and are being used in ways that are invalid (Hoffman, Assaf, & Paris, 2001). Yet, others have pointed to the fact that the results can be used by teachers in planning their curriculum and instruction (Borko & Stecher, 2001).

Teachers have repeatedly reported that they feel pressure to improve test scores (Koretz, Mitchell, Barron, & Keith, 1996). Some claim that the pressure might cause teachers to leave the profession. In fact, a survey of Texas educators found that 85% of teachers agreed that some of the best teachers are leaving the profession “because of the restraints the tests place on decision making and the pressures placed on them and their students” (Hoffman, Assaf, & Paris, 2001, p. 488). However, some pressure might be what is needed to coerce some teachers into re-evaluating their curriculum and instruction. A principal in Danielson’s (1999) study reported that the testing “provided the ‘leverage’ needed to move some teachers who were not ‘risk takers’ into seeing the necessity for change. Not only can the [testing] become the ‘catalyst for change,’ [the principal] believed it could also ‘support the change process’” (p. 75).

Teachers have also reported many negative effects of the testing on students. Some have cited concerns about the emotional effects of the testing on children such as increased stress and anxiety (Elliott, 2000). The pressure can be especially difficult for lower-performing students who might already have low self-concepts and self-esteem. As Gordon and Reese (1997) found: “Many of the teachers lamented that they had worked hard to build up at-risk students’ self-concepts and help them to achieve some measure of academic success, only to have the students’ progress wiped out by the [test] failure” (p. 357).

One of the goals of this study was to determine whether teachers’ perceptions had changed after several years of testing. Moreover, we wanted to systematically categorize teachers’ concerns to better understand which aspects of the testing program were of greatest concern to teachers. Teachers’ perceptions of testing are important because teachers are on the frontlines and in the best position to help policymakers understand how the testing policies are affecting teaching and learning.

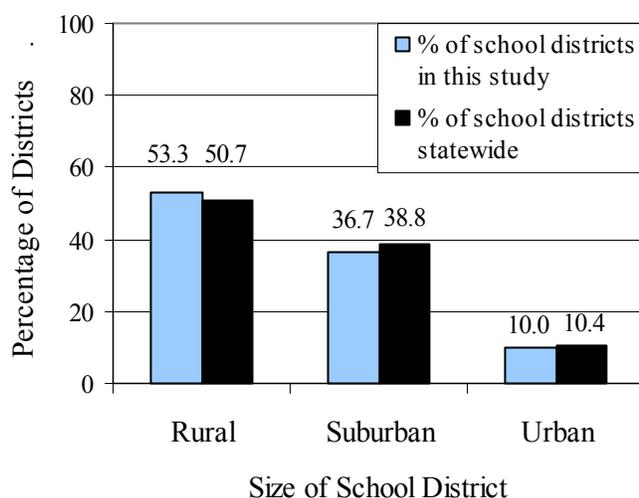
## Method

### Participants

We surveyed third, fourth, and fifth grade teachers in Florida because the state testing program begins in the third grade (third, fourth, and fifth grade students take the FCAT reading and mathematics tests; in addition, fourth graders also take the FCAT writing test). All 67 Florida school districts were invited to participate in this study because we wanted to include the voices of all teachers who wanted to be heard. Of the 67 districts, 34 districts (50.7% of all districts) agreed to participate; that is, we received approval from the

superintendent's office or research department in those districts. We contacted the principals at all of the elementary schools in the districts agreeing to participate a total of three times: twice by email and once by letter. In the email correspondence we asked principals to tell their teachers about the survey and to provide them with the Web site URL for the survey. In the letter correspondence, we included copies of a one-page flyer with an explanation of the study and the Web site URL for the survey and asked the principals to distribute the flyers to their third, fourth, and fifth grade teachers.

We received completed surveys from 708 third, fourth, and fifth grade teachers from 30 school districts (45% of all districts) in Florida. We identified 16 (53.3%) of the districts as rural (less than 15,000 Pre-K to Grade 12 students), 11 (36.7%) as suburban (15,000 to 100,000 students), and 3 (10.0%) as urban (more than 100,000). The percentage of participating districts in each of these categories appears to be similar to the percentage of districts statewide in each category (50.7% rural, 38.8% suburban, and 10.4% urban; see Figure 1).

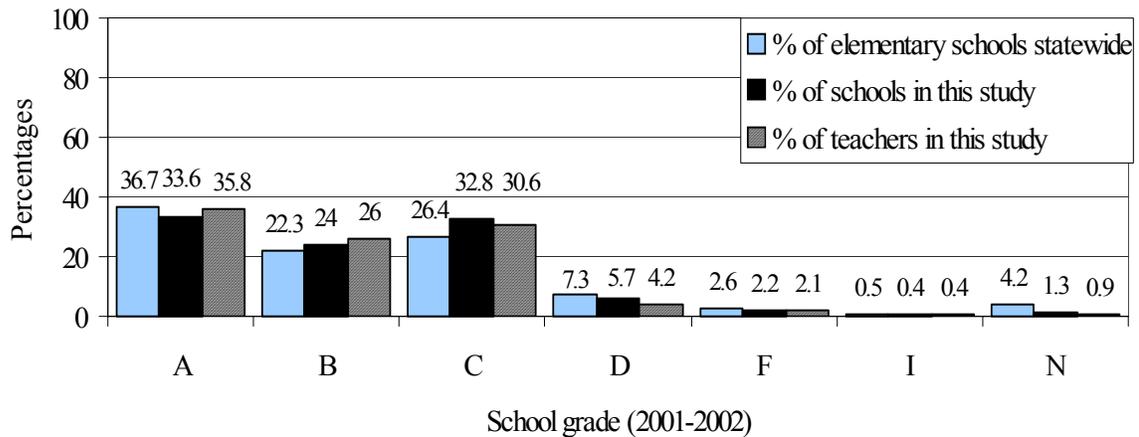


**Figure 1. Percentage of Districts in this Study and Statewide by Size of School District**

Of the 631 elementary schools in the participating districts, we received surveys from teachers in 235 different schools (37.2% of schools). For the average participating school, 52.9% ( $SD = 22.1$ ) of their students were eligible for free or reduced-price lunch, which is similar to the 52.3% of students eligible statewide (Florida Department of Education, 2002b). One-eighth (12.3%) of the schools had 25.0% or less students eligible for free or reduced-price lunch, 31.7% had 25.1-50.0% of students eligible, 41.3% had 50.1-75.0% of students eligible, and 14.7% had 75.1-100% of students eligible.

The teacher response rate for the 235 participating schools was 23.8% (708 participating teachers). For the 2001-2002 school year, 35.8% of the teachers participating in this study taught at schools graded an "A" and 33.6% of the schools participating in this study were graded an "A." These percentages are similar to the 36.7% of elementary schools statewide receiving an "A" grade for the 2001-2002 school year. See Figure 2 for the percentage of teachers and schools at the other school grade levels. This figure shows that the percentage of teachers and schools participating in this study appears to be very similar to the statewide percentage of elementary schools at each school grade level. This comparison is important because it shows that the sample of teachers in this study does not consist of a disproportionate number of teachers from lower-performing schools who might be more likely

to complain about the inequities of testing. Rather, the highest percentage of teachers (35.8%) in this study taught at schools rated an “A.”



**Figure 2. Percentage of Schools Statewide Classified by School Grade and Schools/Teachers in this Study Classified by School Grade (Note: “A” is the highest grade (“making excellent progress”) and “F” is the lowest (“failing to make adequate progress”). “I” indicates “incomplete” and “N” indicates “not previously graded.”)**

Most of the teachers were female (88.5%) and White or Caucasian (91.0%), while 5.3% were Black or African-American, 2.6% were Hispanic, and 1.1% were of another race/ethnicity. Teachers ranged in age from 22 to 68 years old ( $M = 41.2$  years old,  $SD = 10.4$ ) and had taught school an average of 13.4 years (ranging from one to 45 years,  $SD = 9.6$ ), which is similar to the Florida state average of 13.0 years (Florida Department of Education, 2002b). Thirty percent of the teachers had taught 5 years or less, 15.9% had taught from 6 to 10 years, 17.0% had taught from 11 to 15 years, 12.5% had taught 16 to 20 years, 10.4% had taught 21 to 25 years, and 14.2% had taught 25 years or more. A quarter (25.2%) of the teachers taught third grade, 37.4% taught fourth grade, 28.9% taught fifth grade, and 8.5% taught in a multiage classroom with at least some students in the third, fourth, or fifth grade.

### Survey Instrument

Teachers completed an online questionnaire that required approximately 15-20 minutes to complete. To limit the possibility of having ineligible individuals complete the questionnaire, teachers entered a unique school code assigned to them by us. The questionnaire queried teachers about their demographic information, their current teaching practices, and their beliefs about the FCAT.

This article discusses the results of three of the survey items. The first item asked teachers “Is the FCAT program taking Florida’s public schools in the right direction?” and they responded either “Yes” or “No.” The second item was an open-ended item that asked teachers to “Please explain your answer to the previous question of ‘Is the FCAT program taking Florida’s public schools in the right direction?’” Teachers were provided with an online text box into which they could type a response of any length. The third question asked teachers “Do you believe that it is fair to assign grades to schools based on the FCAT scores?” and they responded either “Yes” or “No.”

## Procedure

We conducted descriptive statistics for the two items that required a “Yes” or “No” response. For the open-ended item, the overall analysis strategy involved a microanalysis of the teachers’ responses based on a grounded theory approach to qualitative data (Strauss & Corbin, 1998). We conducted this analysis to generate initial categories, and in doing so, we allowed the data to “speak” and we “listened closely” to what the teachers were trying to tell us (Strauss & Corbin, 1998, p. 65).

Three researchers developed the initial coding scheme for the open-ended item after reading 60 randomly-selected responses, identifying themes, and creating coding categories within the themes. After developing 112 coding categories that we grouped into 11 themes, we independently coded two-thirds of the responses so that all of the responses were coded by two researchers. Disagreements in coding between the two researchers were settled by the third researcher who had not originally coded the response.

After coding the responses, we re-analyzed the coding categories and re-read the responses within each category to ensure that none of them were redundant or overlapped in function. As a result of this re-analysis, we either eliminated or re-categorized 48 of the 112 original coding categories, which left us with a total of 64 final coding categories. Eight of the original coding categories were eliminated completely because only one teacher provided a response in that category. Forty of the original coding categories were re-categorized or combined with other coding categories to which they were very similar. The inter-rater reliability rate after the re-categorization was 92.2%.

## Results and Discussion

After several years of high-stakes testing in Florida, teachers’ perceptions of the effects of testing remain more negative than positive. This is evidenced by the fact that most teachers (79.9%) reported that the FCAT program was *not* taking Florida’s public schools in the right direction. Moreover, the preponderance of teacher responses to the open-ended item described the *negative* effects that the testing has had on education in Florida, not positive effects. Interestingly, 47.3% of the teachers who reported that the FCAT was taking schools in the right direction also provided at least one negative comment about a concern they had with the FCAT. Further, almost all (93.7%) of the teachers believed that it was *not* fair to assign grades to schools based on the FCAT scores. These results suggest that there is much room for improvement with the current implementation of the high-stakes testing program in Florida.

Of the 708 teachers who completed the survey, 610 teachers provided responses to the open-ended item asking them to explain their answer to whether the FCAT is taking Florida’s public schools in the right direction. On the broadest level, we placed the 64 coding categories into three groups: one that described the reasons why the FCAT was *not* taking schools in the right direction (54 categories, 84.4% of all categories); another that described the reasons why the FCAT was taking schools in the right direction (9 categories, 14.1% of all categories); and a third that was neither negative nor positive (1 code, 1.6% of all categories).

Some teachers’ responses were coded with only one coding category, while other responses were coded with as many as 20 coding categories. No teacher’s response was coded more than once with the same coding category. Each teacher’s response was coded with an average of 3.3 coding categories. We used the 64 coding categories a total of 2026 times: 1807 (89.2%) of which described reasons why the FCAT was *not* taking schools in the right direction; 156 (7.7%) of which described reasons why the FCAT was taking schools in the right direction; and 63 (3.1%) of which were neutral.

To better understand the broader issues and to help summarize our findings, we grouped the 64 coding categories into one of ten themes. The first five themes described the

negative effects of testing in that they included reasons why the FCAT was not taking schools in the right direction (see Table 1). The next four themes described the positive effects of testing in that they included reasons why the FCAT was taking schools in the right direction (see Table 2). The final theme was neither negative nor positive; therefore, it warranted a separate coding category. In Tables 1 and 2, we present the number of teacher responses for each coding category, as well as the total number of teacher responses within each theme. Because some teacher responses were placed in more than one category within a theme, the total number of teacher responses in each theme is less than the sum of the teacher responses in all categories.

**Table 1**  
**Number of teacher responses per category that describe why the FCAT is not taking schools in the right direction**

Type of Response	n	%
<b>Theme 1: Negative comments concerning the use and accuracy of the test</b>	<b>321</b>	<b>52.6</b>
Improper use of test (in general, no specifics given)	12	2.0
Unfair to compare students (in general)	14	2.3
Unfair to compare students because of their differences (due to their: socioeconomic status; existing cognitive abilities; emotional stability; cultural values and norms; and community size and location)	90	14.8
Unfair to compare students because some students will never perform well on standardized tests (they are not good test takers)	23	3.8
Test results do not reflect teachers' ability	23	3.8
Students and parents are not held accountable (parents and home life are part of the problem)	18	3.0
Grading system is not fair (criteria for grades unreasonable; unfair to judge a school based on test scores)	84	13.8
Rules (criteria for success) related to testing and accountability change every year	21	3.4
Unfair to give money to high-scoring schools and not low-scoring schools (money should not be tied to test scores)	20	3.3
Test does not accurately measure learning and development (in general)	96	15.7
Student learning cannot be measured by a one-time test	80	13.1
Test should not be used for retention or graduation decisions	18	3.0
Test is not developmentally appropriate or the test is too difficult	47	7.7
Test does not reflect knowledge and skills of students with disabilities or English speakers of other languages	21	3.4
Some students do not perform well on the day of test (because of sickness, nervousness, home issues, etc.)	26	4.3
Test results do not match levels on national tests or the test ignores tests given elsewhere in the nation	10	1.6
<b>Theme 2: Negative effects on curriculum</b>	<b>115</b>	<b>18.9</b>
"Narrows the curriculum" (forces teachers to ignore or reduce some subjects or some topics within a subject because the material is not tested) or it does not promote well-rounded students because it does not cover everything that is important for a good education or to survive in today's society	80	13.1
Curriculum is too broad and shallow	30	4.9
Timing of the test is too early in the school year	29	4.8

<b>Theme 3: Negative effects on teaching and learning</b>	<b>215</b>	<b>35.2</b>
Test takes time and/or focus away from learning	38	6.2
Forces “teaching to the test” and test preparation	142	23.3
Negatively affects teaching practices (in general)	12	2.0
Stifles teachers’ creativity (forces a formulated approach; not free to do what they deem appropriate)	29	4.8
Forces a focus on lower-level objectives such as knowledge and comprehension	8	1.3
Does not allow teachers to meet the learning needs of students	17	2.8
Forces teaching that is not developmentally appropriate	24	3.9
Stifles student creativity (pushes students into a mold)	22	3.6
Does not provide results usable for student remediation or teaching improvement (test is not diagnostic)	21	3.4
<b>Theme 4: Negative effects on student and teacher motivation</b>	<b>283</b>	<b>46.4</b>
Student motivation		
Too much pressure on students (stress, anxiety, worry, fear)	154	25.2
Students do not enjoy school (learning unpleasant; decreased love of learning; school is less fun or interesting)	73	12.0
Students feel labeled	14	2.3
Students more likely to dropout of school	5	0.8
Teaches students that success in public education is equal to performance on a test	16	2.6
Teacher motivation		
Too much pressure in general (stress, anxiety, worry, fear)	13	2.1
Too much pressure on teachers	137	22.5
Too much pressure on administrators	16	2.6
Too much pressure on parents	10	1.6
Teachers do not enjoy school (teaching is less fun or interesting)	25	4.1
Teachers are more likely to leave the profession	19	3.1
Teachers are more likely to transfer from low-performing school to high-performing school	2	0.3
Lowers teacher morale	19	3.1
Teachers are not respected or valued (are degraded or humiliated)	14	2.3
Test does not cause teachers to work harder	9	1.5
People are less likely to go into the teaching profession	3	0.5
<b>Theme 5: Other negative effects on education</b>	<b>166</b>	<b>27.2</b>
Too much emphasis on test scores or too high of stakes	95	15.6
Tests or accountability system were created by non-educators	28	4.6
Testing is a political game or political tool	24	3.9
Test program takes away money from more critical needs	19	3.1
Test program is costly to implement	7	1.1
Test promotes competition between students, teachers, or schools	23	3.8
Stigma is attached to lower performing schools	6	0.9
Parents and the public blame teachers and schools	10	1.6
Grading system leads the public to incorrect conclusions about schools	4	0.7
Test program creates a negative image of public education	7	1.1

**Table 2**  
**Number of teacher responses per category that describe why the FCAT is taking schools in the right direction**

Type of Response	n	%
<b>Theme 6: Positive comments concerning the use and accuracy of the test</b>	<b>57</b>	<b>9.3</b>
Test holds students, educators, or parents accountable	32	5.2
Test results provide useful information about students	17	2.8
Test is adequately fair or it is a good test in general	9	1.5
<b>Theme 7: Positive effects on curriculum</b>	<b>40</b>	<b>6.6</b>
Gives teachers a target/guideline/standard to teach to	28	4.6
Standardizes the curriculum across the state	15	2.4
<b>Theme 8: Positive effects on teaching and learning</b>	<b>37</b>	<b>6.1</b>
Students learn more (teaches students valuable knowledge and skills)	18	3.0
Encourages learning of higher-order thinking skills	13	2.1
Positively affects teaching or causes teachers to rethink teaching practices	11	1.8
<b>Theme 9: Positive effects on student and teacher motivation</b>	<b>13</b>	<b>2.1</b>
Causes higher expectations or motivates students or teachers	13	2.1

More than half (52.6%) of the teachers reported a negative comment concerning the use and accuracy of the test (Theme 1). Theme 4 was the second largest theme with 46.4% of teachers reporting a concern related to the negative effects of testing on student or teacher motivation. About a third of teachers (35.2%) made a comment about the negative effects of testing on teaching and learning (Theme 3), a quarter of teachers (27.2%) made a comment regarding other negative effects on education (Theme 5), and 18.9% of teachers made a comment regarding the negative effects of testing on the curriculum (Theme 2). Fewer teachers made positive comments regarding the testing: 9.3% made positive comments concerning the use and accuracy of the test (Theme 6), 6.6% made positive comments concerning the effects on the curriculum (Theme 7), 6.1% made positive comments relating to teaching and learning (Theme 8), and 2.1% made positive comments with respect to student and teacher motivation (Theme 9).

In the next section, we discuss the coding categories within each of the 10 themes. To do so, we compare the negative themes from Table 1 with the corresponding positive themes from Table 2. For instance, we discuss the results of Theme 1 (Negative comments concerning

the use and accuracy of the test) with the results of Theme 6 (Positive comments concerning the use and accuracy of the test). To allow teachers' voices to be heard in their own words, we included several quotations in "bulleted" form. After each quotation, we provided the grade of the school in which the teacher taught during the year of this study. These quotations are representative of the types of comments that teachers made within each of the categories.

### **Themes 1 and 6: Comments Concerning the Use and Accuracy of the Test**

The major concerns expressed by teachers in Theme 1 were that the tests did not accurately measure student learning and development and that the testing system and use of the test scores were unfair. That is, the concerns in this theme related to the reliability and validity of the test scores, both of which are the cornerstones of a quality test and its use. These concerns are legitimate and consistent with position statements of national educational organizations (AERA, 2000). It is beyond the scope of this work to discuss these types of measurement issues in detail and others have already done so (Messick, 1994; Popham, 2000b). However, in this section, we discuss several teacher concerns within these themes.

Teachers reported that the tests were being improperly used in many ways. First, 20.9% said that it was unfair to compare students and listed reasons such as: students come from different backgrounds and that some students do not perform well on standardized tests.

- "What this test is doing to our already hard to reach students is an atrocity... It is absurd to think that they should be given the same test on the same day and be expected to produce the same quality of knowledge. All people talk at different ages, they walk at various ages, and they are going to learn at different times." (Grade C school)

The teachers' major concern regarding the comparison of students was that inferences were being made about teachers and schools based on test scores, when in fact, students' backgrounds were not the same. Teachers cited several other factors beyond the teachers' or schools' control that played an important role in test scores, such as students': socioeconomic status, existing cognitive abilities, emotional stability, cultural values and norms, and community size and location. They felt these factors made it unfair to compare students using a standardized test such as the FCAT.

- "Grading teachers and schools can never, and I mean never, be done fairly. Every teacher has a different group of students. Some students will score high no matter what. Other students will show growth and some may never show growth on the areas tested on the FCAT. The scores of FCAT depend on many factors and it should not reflect the ability of the student or the teacher." (Grade B school)
- "It is ridiculous to expect low socioeconomic schools with high mobility to compete with schools from affluent areas. It is much easier to teach wealthy kids with highly involved, educated parents." (Grade C school)
- "Many things affect test scores and teachers are expected to take some students who belong in T-ball all the way to the major leagues. If that doesn't happen, we are considered poor teachers." (Grade A school)
- "Some children do not test well, yet can produce fine work when asked to perform in other ways. I believe a complete and more accurate evaluation of a child would involve an equal percentage of factors such as teacher observation, student product, parental input, and standardized assessment." (Grade A school)

These teachers' concerns appear reasonable and consistent with the findings of other studies. For instance, researchers have found that students who come from families of poverty have different needs than students that come from well-to-do families (Comer, 1988). For instance, students of poverty are regarded as having deficiencies in their language, behavior patterns, and values as compared to their middle-class counterparts. In addition, students of poverty are likely to have parents that did not have a successful formal education (Holman, 1997) and are less likely to use academic skills outside the school (Knapp & Shields, 1990). Popham (1999) has also noted the importance of students' out-of-school learning: "If children come from advantaged families and stimulus-rich environments, then they are more apt to succeed on items in standardized achievement test items than will other children whose environments don't mesh as well with what the tests measure" (p. 13).

Related to teachers' concerns about comparing students, 6.8% of teachers were concerned that it was unfair to use students' test scores as a measure of their teaching ability. They cited factors out of their control, such as students' parents and home life, that contributed to a student's achievement. Therefore, they believed that it was improper to use the test scores to judge their teaching ability.

- "How do I force a child to practice and use the skills and strategies I have taught him to use on the FCAT? I can't, yet their score directly points to me and how I have taught. What about the accountability of parents and the students?" (Grade not available)

One of the biggest concerns teachers had with the testing was that the tests were not a valid measure of school quality. Some teachers (13.8%) found that the test scores were used to unfairly judge and make improper decisions about teachers and schools.

- "The grading of schools by using this pathetic test should be a crime." (Grade B school)
- "I think grading of schools is awful because it pits one school against another and not all schools are able to teach students from good socioeconomic areas." (Grade C school)

The problem of holding teachers accountable for uncontrollable variables is exacerbated by the fact that the schools are graded and the results are made available to the public (often through the media). This type of public reporting of scores and grades implies a cause-and-effect relationship between the quality of the teachers and the school rating. In other words, lower-rated schools are assumed to have lower-quality teachers and *visa versa*. Teachers, however, do not believe that this is always the case. Instead, lower-performing schools might have students that come from lower socioeconomic communities, have highly transient student populations, and/or have a high percentage of English as a second language students. In these cases, the lower school rating might not accurately reflect the quality of teaching and learning that takes place within the school. Measurement experts have also noted that standardized tests should not be used to evaluate the quality of education (Popham, 1999). In this regard, measurement experts and teachers agree: student test scores should not be used to make inferences about the quality of education provided by teachers and schools. Considering these negative outcomes of rating schools, it is no wonder that when teachers were asked on the "yes/no" item about whether it was "fair to assign grades to schools based on the FCAT scores," 93.7% of them believed that it was *not* fair to assign grades based on the FCAT scores.

Another concern of a few (3.4%) was that the testing rules kept changing each year. Teachers perceived this as a moving target that made it difficult to compete in this high-stakes "game." Furthermore, the practice of distributing money to higher-rated schools and not lower-rated ones was seen as unfair by 3.3% of teachers.

- “I teach at a fabulous A+ school, yet I know that the grading system is terribly unfair and biased, not to mention changing, with nobody knowing where it’s going.” (Grade A school)

Some teachers (15.7%) reported that the test did not accurately measure student learning and development.

- “The format of various questions in reading and math seem to trick students rather than accurately test their knowledge.” (Grade B school)
- “I do not believe the FCAT is always scored so that it shows student growth and achievement. For example, I have had students score higher than I think they should have. They had not demonstrated to me that they deserved a four on [the writing test]. Also, I have had students score lower than what they have demonstrated in class.” (Grade C school)

Of particular concern was the use of scores from a one-time test to make inferences about students, teachers, and schools. In fact, several teachers (13.1%) said that student learning cannot be measured by a one-time test.

- “There is too much emphasis on the results of the FCAT as the only judge of a student’s ability. We need to consider other ways of determining how well a student is performing and learning. One test doesn’t achieve that objective.” (Grade C school)
- “FCAT is a small picture of a child. The whole picture is what I see that child do each and every day in class: his portfolio; my narrative; and his self-reflection of his work.” (Grade A school)

As a result of teachers’ concerns about a one-time test, 3.0% of teachers said that the test should not be used for retention or graduation decisions.

- “We work very hard all year and one test should not determine whether or not a student is retained in the same grade. The FCAT makes the work we do all year in the classroom seem insignificant.” (Grade B school)

Some teachers (7.7%) said that the tests were not developmentally appropriate or that the test was too difficult. Some teachers (3.4%) specifically commented that the tests did not accurately measure the learning and development of students with disabilities or those who were learning English as a second language (ESL).

- “The focus in teaching, in my opinion, has shifted, from teaching to meet the individual needs of each child, to forcing each child, regardless of his/her individual differences/needs, to perform for FCAT.” (Grade A school)
- “Many of my students are not reading on grade level, so asking them to take a test that is well above their independent or instructional level is unfair.” (Grade A school)
- “I feel that the FCAT test is not valid for children who are only two years out of the ESOL program. It takes more than seven full years of education for a child that speaks Spanish to fully understand, write, and comprehend in English. Therefore, the scores given to a predominantly high ESOL population should be given other consideration. The other schools in our district, with the exception of a few, have an advantage to getting a better grade because their children can read and write in English and ours are still learning.” (Grade C school)

- “The FCAT focuses on too difficult of concepts for many 3rd graders - and it makes children feel like they are failures in math and they're only in the 3rd grade! Many concepts that we are now expected to teach (like decimals) are very difficult for children because they are not developmentally appropriate. I just taught my class a whole unit on decimals and they could pass the final test - but they didn't really understand that a decimal is less than one! They shouldn't have to - they are only 8-9 years old! They are not developed enough with their abstract thinking to truly understand some math concepts that the FCAT tests. I can teach them to jump through the hoops to pass the test but true understanding is not happening - and it really demotivates me as a teacher.” (Grade B school)

This finding is similar that of Pedulla et al. (2003) who found that 9 in 10 teachers did not regard their state test as an accurate measure of what ESL students know or can do. These findings raise several questions about the reliability and validity of test scores for special population students, including: Can the existing types of high-stakes paper-and-pencil tests accurately measure the knowledge and skills of disabled and ESL students; and should these types of students be allowed to receive help during the test, and if so, how much? These types of questions relate to fundamental measurement issues that must be addressed by the designers of testing programs.

Some teachers (4.3%) noted that some students do not perform well of the day of the test because of sickness, nervousness, home issues, etc.

- “The FCAT measures student performance in a timed manner where anxiety then plays a big role in actual student performance.” (Grade A school)
- “Some students are very intelligent, but become very nervous and just cannot perform on standardized tests.” (Grade B school)

A few teachers (1.6%) were concerned that the test results did not match levels on national tests or that the testing ignored tests given elsewhere in the nation.

Despite the concerns of many teachers related to the reliability and validity of the testing, the Florida Department of Education (2001) maintains that the reliability indices for all of the grades are above 0.90 and that “therefore, the tests are reliable” (p. 9). Similarly, they state that the FCAT has content and concurrent validity. A few teachers did agree, as 1.5% of teachers believed that the test was adequately fair. Other positive comments related to the use and accuracy of the test included teachers who reported that the test held students, educators, or parents accountable (5.2% of teachers) and that the test results provided useful information about students (2.8% of teachers).

- “I believe that the FCAT has made teachers accountable for teaching the Sunshine State Standards. We had the Sunshine State Standards, but until there was the accountability, not all teachers were using them.” (Grade C school)
- “Everyone feels more accountable and you have an actual number to show parents when their child is struggling versus just our professional opinion.” (Grade A school)
- “I believe the FCAT is helpful in gauging levels of performance of students in my classroom, but like any assessment, the test should be considered a part of a complete picture about any given student, not the whole picture.” (Grade A school)

### **Themes 2 and 7: Effects on the Curriculum**

Teachers expressed concern with how the testing had affected the curriculum. Specifically, 13.1% indicated that testing “narrows the curriculum” by causing them to spend more time on subjects and topics tested. Because of this, they were concerned that the test did

not take into account the whole child or provide the students with the knowledge and skills required to survive in today's society. In other words, the test doesn't cover everything that is important for a well-rounded education. This finding has also been reported in other states (Firestone, Mayrowetz, & Fairman, 1998; Jones et al., 1999) and shows that this issue remains an important concern for teachers.

- “The FCAT is teaching teachers to stay within the narrow confines of the FCAT. Too many times I've been told, when going beyond the confines (especially in math): ‘Why are you teaching that? It isn't on the FCAT.’” (Grade C school)
- “Our total curriculum is focused on reading, writing, and math. There is no extra time for students to study the arts, have physical education, science, or social studies. Our curriculum is very unbalanced.” (Grade D school)
- “While it is a way of testing some components of standards based performance, it leaves many gaps in the educational process. If we just ‘teach to the test’ which many teachers in our district are pressured to do, then the students are left with HUGE educational gaps that have not been covered in their education. Students deserve a well-rounded education, not just bits and pieces that are presented on a state test.” (Grade C school)
- “Before FCAT I was a better teacher. I was exposing my children to a wide range of science and social studies experiences. I taught using themes that really immersed the children into learning about a topic using their reading, writing, math, and technology skills. Now I'm basically afraid to NOT teach to the test. I know that the way I was teaching was building a better foundation for my kids as well as a love of learning. Now each year I can't wait until March is over so I can spend the last two and a half months of school teaching the way I want to teach, the way I know students will be excited about.” (Grade C school)

The narrowing of the curriculum concerned 4.9% of teachers because they felt that it had negative effects on students' understanding. Their main concern was that the curriculum was too broad and shallow (i.e., that the curriculum lacked a more in-depth exploration of the topics), which caused teachers to cover the material too quickly prior to the test.

- “I believe that the FCAT is pushing students and teachers to rush through curriculum much too quickly. Rather than focusing on getting students to understand a concept fully in math, we must rush through all the subjects so we are prepared to take the test in March. This creates a surface knowledge or many times very little knowledge in a lot of areas. I would rather spend a month on one concept and see my students studying in an in-depth manner.” (Grade C school)
- “It is impossible to teach all the Sunshine State Standards. We teach so many different standards that it is not possible for the children to learn them well. Should we teach a curriculum that's a mile wide and an inch deep, or concentrate on developmentally appropriate concepts and teach them well? Do you know what stem and leaf math is? We waste time teaching a lot of things that children are not ready to understand. They can memorize a formula but have no conceptual understanding if it. For example, long division is inappropriate for fourth graders.” (Grade A school)
- “I feel that our students are becoming ‘jacks of all trades’ and masters of none. Our curriculum must be taught in a condensed time span, which is stressful to all concerned. We are teaching them to perform tricks, like monkeys in a circus.” (Grade C school)
- “Our FCAT ‘dumps’ stringent requirements on all students, without allowing any exception for the child who just needs more time to develop basic concepts. We have

to rush along, not mastering anything, but exposing to everything. What a sad thing to do to both students and teachers.” (Grade B school)

- “Sometimes we cannot linger longer on topics that need in-depth discussion and instruction.” (Grade A school)
- “Subjects like science and social studies get left in the dust because they are not tested. I am not saying the answer is to test them, too. I find myself stressing that the students learn how to answer multiple choice lessons after reading a piece because that is how the FCAT is. Enjoying a really great book, or spending a lot of time on a certain theme is out because I have to teach ALL the standards for the whole year BEFORE March! There is not time to do everything, and a lot of kids, especially those from backgrounds that are not as advantaged do not do as well in school, and are not ready for this test. I think that it does not belong in third-grade. There is way too much emphasis on FCAT, FCAT, FCAT, and not enough time to develop creativity, social skills (yes, these days, the teacher has to teach social skills and manners) and science.” (Grade not available)

This problem of a broad and shallow curriculum appears to be exacerbated in Florida by the fact that the test is administered in February and March, well before the end of the school year in May. A few teachers (4.8%) were concerned that the timing of the test was too early. The early test administration forced teachers to teach a year’s worth of curriculum in less than one year which created an unrealistic teaching expectation. Teachers were frustrated with the expectation of having to teach a year’s worth of curriculum and the reality of having less than a year in which to do it.

- “Learning occurs over an entire school year, not just from August to March. Students are expected to master a year’s worth of growth by the testing date. To do this, educators are ‘hopping’ around the curriculum to ensure that their students have been exposed to (not mastered) every topic (i.e., math).” (Grade A school)

Teachers reported that the positive effects on the curriculum were that it gave some teachers a guideline or standard to teach to (4.6% of teachers) and/or standardized the curriculum across the state (2.4% of teachers).

- “Our state curriculum is clear and we as teachers know exactly what we are responsible for teaching.” (Grade C school)
- “I know that there are probably teachers who don’t even make lesson plans according to the standards, so it puts the pressure on them to actually be teaching the mandated curriculum.” (Grade C school)
- “Having set standards puts all the teachers on the same page. It is a goal that everyone should reach. It informs a teacher where they should be in the curriculum.” (Grade A school)

This finding is consistent with reports from teachers in Ohio who reported that the testing had helped them to more closely evaluate their curriculum (DeBard & Kubow, 2002). These points are interesting because the curriculum in Florida has been standardized since 1996, three years prior to the implementation of high-stakes testing in Florida. Therefore, the FCAT is not responsible for standardizing the curriculum, but rather, it might have served as an impetus for coercing teachers to include more of the state curriculum in their teaching.

### **Themes 3 and 8: Effects on Teaching and Learning**

Teachers reported many instances of how the testing had negatively affected their teaching practices. Some (6.2%) reported that testing took time and focus away from learning; and instead, placed the focus on other areas such as the tests and rewards.

- “So much of what I spend my time on, at school and home, is geared toward accountability. I spend more time trying to justify and prove what I’m doing than actually doing it.” (Grade A school)
- “Florida needs to be relieved of such a burden and focus on higher education at all times.” (Grade C school)

However, the most frequent complaint of the effects on teaching (23.3% of teachers) was that they had to spend a lot of time preparing for the tests and “teaching to the test.” Teachers said that they were teaching knowledge and skills that they wouldn’t otherwise have taught or that they were teaching content that would be the same as on the test.

- “I can say one thing, if my kids learn one thing in third grade, it is this: how to pass a standardized test even if you are not familiar with the material. Now is that what our goal is? Perhaps we should revisit it.” (Grade not available)
- “I have seen that schools are teaching to the test (how can you not?) and that is not a true reflection of student abilities. This is only a reflection of the abilities of each school to teach effective test-taking strategies, not academics.” (Grade B school)
- “Schools aren’t improving their academics as students score better on the FCAT. They are just taking more time to teach to the test and unfortunately, away from real learning. We aren’t getting smarter students, we are getting smarter test takers. That is NOT what we are here for! They can say what they want about if we teach the SSS then students will score well. The schools who score well are focusing on teaching to the test at a very high cost to their students.” (Grade C school)

The fact that some teachers are teaching to the test should not be surprising, however, given that in a “Technical Assistance Paper,” the Florida Department of Education stated that “It is desirable for students to be given a certain amount of practice so they will be familiar with the format of the test questions and the materials that will be used with the statewide and district assessments” (State of Florida, 2000, p. 6). They further state that: “To prepare students for the future assessments, teachers can...have students practice taking short and extended response, gridded response, multiple-choice, and essay items so they will become familiar with the test formats; structure activities that require students to work against fixed time limits; and help students practice with mark-sense answer sheets.” (p. 6, 7)

Although the “Technical Assistance Paper” notes the dangers of “teaching to the test,” some teachers apparently find it impractical or unrealistic to provide a “certain amount of practice” related to the test format without teaching to the test. There appears to be a fine line between providing practice for the test and teaching to the test. When high-stakes are attached to the test scores, we believe that most teachers will err on the side of caution and teach to the test instead of risking the possibility of low test scores.

Other findings were that some teachers (2.0%) found the testing to negatively affect their teaching practices. Specifically, 4.8% reported that the test stifled their teaching ability and creativity in that it limited their freedom and forced them to use a formulated approach.

- “I feel that the FCAT is taking the learning styles and teaching styles away from students and teachers. The flexibility to teach the best way to meet the needs of the students is eliminated.” (Grade A school)
- “I believe that the FCAT hinders teachers from being creative with their teaching. The programs that we have implemented at our school to help improve FCAT scores has caused us to teach like robots. Everything is scripted. Maybe our students will be robots too!” (Grade C school)

Teachers reported three different ways in which the tests hindered their ability to meet the learning needs of the students. First, it forced them to teach in ways that were not developmentally appropriate (3.9% of teachers). Teachers claimed that students were often not ready for the knowledge and skills that were being taught, but that the fast pace of the curriculum was necessary due to the testing content and timing. Second, 3.6% of teachers said that they were less able to foster student creativity. Third, 3.4% of teachers noted that the test results were not usable for student remediation or teaching improvement. That is, the results could not be used to help meet students' learning needs or to improve their own teaching.

- “Some of the FCAT skills I HAD to teach were not presented to me until high school or college. My students deserve the opportunity to learn in a developmental progress suited for them. If they cannot master basic skills, when why must they be forced to learn something they are not ready for? Because it is going to be on the test and I must at least expose them!!!” (Grade B school)
- “Mainly, I feel that the child is not really the important issue, because if you have ANY experience at all with children, you would know that all children learn at different rates AND MATURE AT DIFFERENT TIMES, and to expect so much from our children is putting way too much pressure on them at such a young age.” (Grade A school)

A few teachers (1.3%) reported that it forced them to use a lower quality of teaching. For instance, it forced them to focus on lower-level objectives such as knowledge and comprehension rather than higher-order thinking.

- “Florida’s public schools are going to become nothing more than places of drill and skill rather than places of quality learning.” (Grade A school)
- “Problem solving skills and upper-level questioning seems to be evaporating.” (Grade A school)

In contrast to this finding and the often-cited criticism that testing forces lower-level learning (Kohn, 2000), 2.1% of teachers reported that the test encouraged the learning of higher-order thinking skills which suggests that the FCAT might be different from other types of tests that focus more on knowledge and comprehension than on analysis, synthesis, and evaluation. Because the FCAT tests are not available for the public to view at this time, it is impossible for us to verify whether the tests focus on lower- or higher-level objectives. However, teachers’ perceptions of the test are important and we believe that they are based on practice tests and sample items that they have seen. This finding is encouraging and suggests that it might be possible to develop high-stakes tests that promote higher-order thinking, which is generally viewed as an important outcome of education.

Other positive testing outcomes reported by teachers included that the testing led to an increase in student learning (3.0% of teachers) and that it had a positive affect on their teaching practices (1.8%).

- “I believe the design of the FCAT helps students to explain their thought processes and hopefully alleviates guessing on standardized tests.” (Grade A school)
- “The positive benefits are that a lot of the teaching done for the tests is used in the real world.” (Grade B school)
- “Gone are a lot of the ‘fun/not academically related’ activities; in are more thought provoking activities which stimulate children to think and solve problems...FCAT has made me look at my teaching skills and work to improve them. Prior to the emphasis on accountability I simply did what I had been doing for many years because it was easy. I don’t know that it was always best for the students though.” (Grade C school)

### Themes 4 and 9: Effects on Student and Teacher Motivation

The themes regarding student and teacher motivation, as much as any other, were more heavily weighted towards the negative than the positive. Of all the reasons teachers provided as to why the FCAT was not taking schools in the right direction, two of the three highest percentage of responses were found in this theme. Whereas politicians and the public often focus on the *achievement* of students in public schools, teachers appear to be as concerned about the impact of testing on student and teacher *motivation*.

#### Student motivation

A quarter (25.2%) of the teachers reported that the testing had caused students to feel too much pressure and stress. We defined stress in the same manner that Kyriacou (1989) described it: as the experience of tension, frustration, anxiety, anger, and depression resulting from work. Because researchers have found that high student anxiety can have detrimental effects on student performance (Everson, Smoldaka, & Tobias, 1994), these concerns must be taken seriously and not simply pushed aside as evidence that students and teachers need to “work harder” or “toughen up.”

- “Too much pressure is put on this one aspect of education! A fourth grade student should not have to feel all the stress put on them by the FCAT. High school or college pressure on a fourth grader is not good! Many burn out by the time the FCAT is finished! Being held accountable is fine, but don’t put school fund raising on the backs of the students.” (Grade A school)
- “In our school I heard of some students crying in the morning or vomiting on the test because of so much pressure. It is ridiculous!” (Grade not available)
- “I wonder if we’re going to burn out this generation from education. Will this have an effect on the amount of future college students we will have? Or, are we going to make our students so stressed about education that we get the emotional problems the Japanese have at such young ages (high suicide rates)? I am not saying that we should not have high expectations for our children. I have children myself, but I feel that we are trying to create ‘miniature adults’ instead of remembering that we are dealing with children.” (Grade A school)

Teachers (12.0%) also noted that the testing negatively affected students’ enjoyment of school or interest in school. Further, 2.3% believed that students felt labeled as a result of the test scores and grades and 0.8% claimed that students might be more likely to drop out of school in the future.

- “School is becoming a drudgery for teachers and students alike. Yes, standards are important and schools should work to ensure every child’s success, however, not at the expense of the love of learning.” (Grade N school: “not previously graded”)
- “I think we are forcing children to grow up too quickly. Of course we should encourage higher-order thinking, but more importantly we should be teaching children to love learning. That is how we’re going to motivate them to stay in school.” (Grade B school)
- “FCAT has made children feel like failures when the truth is, they just haven’t ‘bloomed’ according to our legislator’s time line.” (Grade A school)
- “Students are supposed to learn and to show growth each year, but to continually add more stress to these students is wrong. I believe the state will eventually see an increase in dropout rates due to students hating school earlier each year in the elementary grades.” (Grade C school)

In contrast to the negative effect on students' enjoyment of or interest in school, none of the teachers said that the testing made school more enjoyable. We viewed this finding as a measure of their intrinsic motivation because individuals who are intrinsically motivated participate in activities for their own sake; that is, they enjoy or are interested in the activity itself (Pintrich & Schunk, 2002). Because researchers have found that intrinsic motivation facilitates learning and achievement (Gottfried, 1985; Ryan, Connell, & Plant, 1990), reducing students' intrinsic motivation likely has a negative effect on students' achievement as well.

The final negative effect on student motivation cited was that the testing program was teaching students that success in public education was synonymous with performing well on a test (2.6% of teachers).

- "We are teaching our kids that it is not as important what you do throughout the school year as long as you perform well on the test." (Grade A school)
- "The children seem to understand that the only thing that is important is their performance on the test." (Grade A school)

### **Teacher motivation**

Many teachers (22.5%) said that they were feeling stress from the pressure of the tests, as were administrators (2.6% of teachers) and parents (1.6% of teachers).

- "The pressure to perform is cruel and unusual punishment for both the students and the teachers." (Grade B school)
- "As a new teacher, I have noticed many 'veteran' teachers who have negative attitudes toward this profession and to what it has become. I think this is due to the pressures of FCAT and this is very discouraging to me as a new teacher." (Grade C school)

In fact, 4.1% of teachers reported enjoying teaching less as a result of the tests.

- "The pressure of the scores leading to school grades takes a lot of the joy out of teaching, and I LOVE teaching." (Grade A school)

The final group of responses in this category related to how the test had negatively affected teacher motivation and the teaching profession. Some teachers (3.4%) felt that their motivation to remain teachers had decreased and that teachers were more likely to leave the profession or transfer to a higher-performing school as a result of the testing. Others said that teacher morale at their school was lower (3.1% of teachers), that they felt less respected or valued (2.3% of teachers), that the tests did not cause them to work harder (1.5% of teachers), and/or that people were less likely to go into the teaching profession (0.5% of teachers).

- "The morale in our school is the lowest I have ever seen in my 25 years of teaching." (Grade A school)
- "Dedicated teachers are dropping out like flies because we can't handle the stress anymore." (Grade C school)
- "When teachers feel their salaries will one day be based on student performance, many of us say that will be the day when we will walk out on the profession. A teacher can't force a child to perform to the best of their ability on the test." (Grade B school)
- "Many teachers are requesting transfers from C schools to A schools. I would be willing to bet if you looked at turnover rates in C or lower schools, you would find that it is higher than in A schools." (Grade B school)
- "You can't abuse a person in any other profession the way teachers are abused by students, parents, even principals, and our legislators!" (Grade C school)

The positive effects on student and teacher motivation, cited by 2.1% of the teachers, were that the tests caused higher expectations and that it motivated students and/or teachers.

- “I do feel the higher expectations have served to improve the focus and effort in Florida’s schools.” (Grade A school)
- “The low-performing students were not expected to achieve, therefore, they were not exposed to challenging information. Now, more teachers are saying all children can learn.” (Grade D school)
- “It seems to be an effort in the right direction, that of providing students with motivation to do better and to learn better.” (Grade A school)
- “I think that to some degree it influences some teachers to do a better job of teaching. Some teachers need to have extra motivation to come to school and do the best job they can do of educating students. You have some people who need outside motivation because they don’t have the motivation on their own.” (Grade C school)

### **Theme 5: Other negative effects on education**

The final negative theme included several different categories that did not fit into any of the other themes. We believed that it was important to include these categories in the framework because these issues were important to several of the teachers. First, 15.6% of teachers thought there was too much emphasis on the tests scores in general. This belief is likely exacerbated by the fact that the test scores are used to retain students, to rate schools, and to distribute money to higher-performing schools. Second, 4.6% of teachers challenged the accuracy of the tests because they perceived that they were created by non-educators.

- “If you want to go in the right direction in education, try asking the teachers who went to school to learn this and live with it daily, instead of having just the politicians decide that have no background knowledge.” (Grade C school)
- “The people that come up with these ideas should have to spend at least one month in each school in their district. My bet is it would be a REAL eye opener.” (Grade B school)
- “It makes me wonder if these people in Tallahassee have ever met any children from a poor, rural community.” (Grade C school)
- “Legislators have NEVER seriously asked teachers for their input with the intention of using it. They sit in an office and pass legislation to test students, retain them, and hold teachers accountable without once looking ahead at the long term consequences.” (Grade C school)

Teachers perceived their voices to be largely unheard by policymakers and complained that they had not been a part of the process of creating the accountability program. To ignore teachers’ voices is to ignore their ideologies. Moreover, this lack of a voice appears to have created a resistance and silent controversy to the testing program. As Matthews and Crow (2003) explain, “Although not all problems you face can be solved by giving people a listening ear, refusing to hear or ignoring individuals and groups that want to be heard is likely to aggravate the situation and intensify the negative aspects of the conflict” (p.206). This sentiment is consistent with the teacher voices we heard in this study.

Further, 3.9% of teachers believed that testing was a political game or was used as a political tool to serve the interests of politicians.

- “FCAT is just a political tool that the state uses to make them feel like they are doing something good for education.” (Grade A school)

- “I believe that the legislature is doing a great deal of harm to our students...I feel that the money we need is not being given to the schools for two reasons. The first is to somehow dismantle the public school system (through vouchers), and secondly, to create an elite system run by private interests. I have worked in both business (law, engineering, and banking) and say without reservation, education is the most efficient in the use of both man power and dollars. The FCAT is nothing more than the politicians ploy to say either ‘See, we’ve fixed the system’ or ‘See, they’re not doing the job and we need to step in.’ All for the next election!” (Grade B school)
- “Florida’s public schools have long been the target of ambitious, power-hungry politicians. This is just another political move to discredit the public schools and repay political contributors with vouchers for expensive private schools that their children already attend. Between the FCAT tests, vague Sunshine State Standards, school grades, and mathematically impossible required gains in test scores, it seems that the politicians’ goal is to eliminate public education from the state of Florida.” (Grade A school)
- “My personal belief is that the FCAT is a political football and that given the current climate in Tallahassee, its real mission is not to provide accountability to families, communities, etc. or to help schools discern better instructional techniques for students. Rather, the mission is to diminish public education, advance a special interest agenda for charter schools and private education, and advance political careers.” (Grade C school)

Politicians were perceived as making their own decisions, possibly for their own gain or as a political tool to achieve other purposes. Because most teachers get into the profession to help children grow and learn (Ornstein & Levine, 2000), it is easy to understand why they would be opposed to a testing system that they view as doing little to promote student growth and learning. Instead, some teachers see the political motives for the testing as incongruent with their personal view of education that centers around doing what is best for the children. Taking the focus from the children and placing it on politics is understandably troubling for some of these teachers.

Another concern related to the amount of money being spent on testing. Some thought that it took away money from more critical needs (3.1% of teachers) and/or that it was costly to implement (1.1% of teachers).

- “I believe a lot of money is going towards these tests, grading them and implementing them and that money should be sent towards reading programs that are simple and that work. Primary grades need teacher assistants. We need the money spent in a more productive fashion.” (Grade B school)

The fact that the test promoted competition between students, teachers, and/or schools was also seen as a negative effect of the testing (3.8% of teachers).

- “Why ‘grade’ schools by these tests alone, pitting schools and even grade levels and individual teachers against each other. There used to be an atmosphere of sharing; now, if we help someone, they might get a ‘higher’ FCAT score which makes us feel less capable.” (Grade A school)
- “I feel the FCAT test has a negative impact on schools by creating competition between them as a result of the grading system...Instead of taking public schools in the right direction, the state has pitted schools against each other. Instead of working together as a whole, it’s survival of the fittest.” (Grade B school)

A few teachers (0.9%) said that there was a stigma attached to lower performing schools. Others said that the test results led parents and the public to blame teachers and schools (1.6% of teachers) or that the grading system led the public to incorrect conclusions about schools (0.7% of teachers). Finally, 1.1% of teachers believed the testing program created a negative image of public education.

- “The FCAT seems to be a way to make teachers scapegoats for problems plaguing society. It serve the purpose of creating a great deal of negativity.” (Grade C school)
- “The FCAT makes schools look bad instead of celebrating many of their successes.” (Grade A school)
- “Some days, I just can’t stand to read the editorials on ‘What’s wrong with our schools?’ when my children, staff, and parents are working so diligently!” (Grade A school)

### **Theme 10: Accountability is good or necessary but...**

The final theme was reported by 63 teachers (10.3%) who claimed that accountability was good, necessary, or that they were in favor of accountability. These teachers indicated that the FCAT was *not* taking schools in the right direction, yet they believed in accountability, just not in the manner that it was currently being implemented. A lot of these responses started out in favor of accountability and then said “but...” and described why the FCAT was not effective in holding people accountable.

- “I feel accountability is important on all levels, but this system is tearing hard working, dedicated teachers down into total burnout.” (Grade C school)
- “While I support accountability and assessment, I feel the focus on using the FCAT for the purpose of assigning school grades undermines the potential positive effects, such as focusing on higher-level critical thinking.” (Grade B school)
- “I do agree that accountability is extremely important, but who’s to say that the FCAT is the right tool to measure students’ abilities or progress????...No, I do not feel that the FCAT is the right tool.” (Grade A school)

We find this result noteworthy, especially because none of the teachers reported that they were against accountability. This finding leads us to believe that teachers understand the importance of accountability in the teaching profession.

### **Summary**

The negative comments provided by teachers about the effects of testing appear to far outweigh the positive comments. This finding is consistent with prior research (e.g., Jones et al., 1999) and suggests that several years of testing have not drastically changed teachers’ concerns regarding testing. Issues that remain problematic for teachers, include: the unfairness of comparing students, teachers and schools based on test scores; the negative effects of increased teaching to the test; the large amount of pressure felt by students and teachers; and the lack of reliability of a one-time test. We have attempted to further explain, clarify, add to, and categorize these types of concerns in the current work.

In addition, we present some important findings that have not received as much attention in prior studies. Perhaps most importantly, teachers indicated that they are not against being held accountable, only that they are not in favor of the current means by which they are being held accountable. The results of other studies might lead one to believe that teachers can be characterized as complainers who do not like testing because it holds them accountable for doing a job that they are not doing. On the contrary, the results presented here show that teachers are in favor of accountability or believe that accountability is necessary. This is an

important finding because it shifts the discussion from whether or not teachers should be held accountable to a discussion of *how* teachers should be held accountable.

Although the findings presented here did not specifically address how teachers envisioned a revised testing program that would hold them accountable, there are several implications that can be derived from their comments. In the following section, we present some of the most important implications for improving upon Florida's current testing program.

### Implications

This study provides policymakers with evidence that after four years of high-stakes testing in Florida, teachers continue to express concerns and frustrations with Florida's testing program. The purpose of this study was to document teachers' concerns with the hope that policymakers would use this information to improve upon the current testing program. We agree with others (e.g., Grant, 2000) that for teachers to support a testing program, they need to have their voices heard by policymakers and be a part of developing the testing program. DeBard and Kubow (2002) have also noted: "What is needed is a policy shift that emphasizes inclusion of constituents. The end result will not be a reduction of accountability but rather an assumption of greater responsibility" (p. 403). This belief is also in concert with the comments of teachers in this study, one of which reported: "Legislators have NEVER seriously asked teachers for their input with the intention of using it." These types of comments indicate that teachers continue to resent the manner in which testing has been thrust upon them without their input or acceptance.

In this section, we provide some implications for changing the testing program based on teachers' comments. The recommendations provided in this section are based on teachers' perceptions of the testing program in Florida. We recognize, however, that teacher perceptions might be different from those of administrators, parents, or students. Understanding teachers' concerns is important, however, because they have the most direct effect on students' learning and motivation.

One message in Theme 1 was clear: the *use* of test scores needs to be limited. Some teachers perceived that the test scores could be used effectively to help inform their teaching practices and improve student learning. However, the test scores were not perceived as being valid when used to make comparisons between students, teachers, or schools. Almost all of the teachers noted that it was not fair to assign grades to schools based on the test scores. These comments suggest that policymakers should eliminate the school grading or change the criteria for grading to make it more fair. Under the current testing program, half of the points for the school grade are based on students meeting certain performance standards. As a result, schools that serve students who come to school more cognitively developed in reading, writing, and mathematics receive higher scores, and thus, an unfair advantage. Teachers are justified in their complaints that it is unfair to compare teachers and schools based on students' scores because the scores reflect other influences on students besides those of the school and teacher.

One way to make the school grades more fair would be to adjust the scores for the socioeconomic status of the students (which is generally correlated with achievement) or to test students' cognitive abilities at the beginning of the year and compare these scores with their end-of-year scores. Doing so would more directly measure the effects of student learning during that year. In response to whether the school grades are adjusted for the socioeconomic status of students, the state has responded that:

Schools are responsible for teaching all students, regardless of socioeconomic status. All students are capable of learning and making adequate progress. There are no double standards in the FCAT program. All students and schools will be held to challenging performance standards. (Florida Department of Education, 2001, p. 4)

While we believe that most teachers would agree to holding all students to high standards, these types of statements do little to address teachers' concerns of fairness. In other words, the schools continue to be graded on an uneven playing field.

To somewhat rectify the inequity of an uneven playing field, part of the school grade is computed using the gains students make during the year. This component of the school grade appears to be more consistent with teachers' comments in that it more directly measures the effects of a particular teacher and school on the student during that year. Using student gains as a more prominent component of the testing program might help alleviate some of the teachers' concerns. Furthermore, it would reduce the likelihood that teachers would use students' backgrounds as an excuse to accept lower expectations.

Another major concern of teachers in Theme 1 related to the use of a one-time test to accurately measure students' learning and development. The logical implication would be to conduct the test more than once a year. Of course, this option would take away more instructional class time and be more costly. Another option would be to use an alternative type of assessment such as portfolios. Portfolio assessments are a collection of student works and generally include a student's classroom work, revisions, assessments, and reflections on his or her learning. Some teachers have found that portfolios can positively impact their teaching methods and are essential to holding teachers accountable (Borko & Elliott, 1999). Bridge, Compton-Hall, and Cantrell (1997) examined the use of portfolios for writing and found: "Portfolio assessments more nearly match what is known about the development of writing; they enable both teachers and students to evaluate the progress students have made over time and their ability to bring a given piece from first draft to final form" (p. 168). Unfortunately, the cost of grading portfolios is generally much more than grading standardized tests, and the reliability (consistency of scores) had been shown to be poor (Koretz, McCaffrey, Klein, Bell, & Stecher, 1993). Further research into the use of these types of alternative assessments would be useful in developing less expensive assessments that more accurately reflect students' learning and development with greater reliability and validity.

To address the concerns raised by teachers in Theme 2, the curriculum needs to be modified to include fewer topics within each subject (become less "broad and shallow"). For too long the U.S. curricula has been unfocused and "a mile wide and an inch deep" (Schmidt, McKnight, & Raizen, 1996). In addition, the test should be given later in the school year or only include topics that can reasonably be taught before the test is administered. During the year of this study, the tests were administered in February and March, a couple of months before the end of the school year. Teachers felt so much pressure from the early testing date that they rushed to teach all of the curriculum topics before the testing began. Because rushing through the curriculum is not consistent with current learning theories (National Research Council, 2000), the testing appears to be hindering student learning. Revising the curriculum to address this concern would help teachers to manage their instructional time more effectively, resulting in increased student learning.

Based on the findings reported in Theme 3, steps need to be taken to prevent teachers from teaching to the test. The challenge for policymakers is to create a system that encourages teachers to engage in curriculum teaching without promoting item teaching. The difference is that item teaching includes "teachers who organize their instruction, for instance, teacher-explained illustrative items or items-based practice activities – either around the actual items found in a test or around a set of look-alike items" (Popham, 2000a, p. 2). In contrast, curriculum teaching is directed toward the specific domain of content knowledge or skills, but not limited to the specific items within the domain tested.

We believe that the testing program itself does not cause the teachers to teach to the test. Rather, a variety of factors, including pressure from others (parents, other teachers, administrators) to achieve and the fear of sanctions (a low school grade, less state money),

contribute to the way in which teachers internalize these pressures. Corbett and Wilson (1991) noted that even the same sanctions can have different meanings to different people. Others have shown that when teachers feel pressured and responsible for ensuring that their students perform up to standards, they become more controlling (Flink, Boggiano, & Barrett, 1990), which can lead to a reduction in students' intrinsic motivation (Ryan & Grolnick, 1986). More research needs to be conducted to better understand how the pressure is interpreted within different political climates and school contexts. Reducing teachers' perceived pressure might reduce the likelihood that they would engage in item teaching.

Some of the concerns cited by teachers in Themes 4 and 5 would likely be lessened if some of the recommendations provided in this section were implemented. For instance, eliminating or changing the grading of schools would likely lessen the emphasis on test scores and reduce the amount of pressure felt by teachers and students. What remains to be seen is how the elimination of the grading and/or rewards and sanctions would affect the higher expectations that supposedly accompany the high stakes.

### **Conclusion**

Teachers provided many powerful insights regarding high-stakes testing and its effects on teachers and students. Although teachers do not believe that Florida's testing program is taking schools in the right direction, they are not afraid of being held accountable. In fact, teachers appear to be in favor of accountability or at least recognize the need for it. The framework that we developed based on teachers' comments can be used as a means to evaluate the strengths and weaknesses of the testing program as perceived by teachers. Furthermore, these comments can be used to improve upon the existing testing program. Until policymakers take teachers' concerns seriously and make an effort to address them, teachers will not likely support reform through high-stakes testing. Without the support of teachers, high-stakes testing will likely become just another failed education reform. However, with the input of those on the frontlines and some vital and well-conceived changes, testing programs are likely to have a more positive effect on the teaching and learning processes.

## References

- AERA (2000, July). *AERA Position Statement Concerning High-Stakes Testing in PreK-12 Education*. Retrieved April 11, 2002 from <http://aera.net/about/policy/stakes.htm>
- 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*, 384-397.
- Bridge, C. A., Compton-Hall, M., & Cantrell, S. C. (1997). Classroom writing practices revisited: The effects of statewide reform on writing instruction. *The Elementary School Journal, 98*(2), 151-170.
- Borko, H., & Elliott, R. (1999). Hands-on pedagogy versus hands-off accountability: Tensions between competing commitments for exemplary math teachers in Kentucky. *Phi Delta Kappan, 80*, 394-400.
- Borko, H. & Stecher, B. (2001, April). *Looking at reform through different methodological lenses: Surveys and case studies of the Washington state education reform*. Paper presented at the meeting of the American Educational Research Association, Seattle, WA.
- Cimbricz, S. (2002, January 9). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives, 10*(2). Retrieved April 11, 2002 from <http://epaa.asu.edu/epaa/v10n2.html>.
- Comer, J. P. (1988). Educating poor minority children. *Scientific American, 259*(5), 42-47.
- Danielson, M. L. (1999). *How principals perceive and respond to a high-stakes accountability measure*. Dissertation Abstracts International, 61, 03A. (UMI No. 9967692)
- DeBard, R. & Kubow, P. K. (2002). From compliance to commitment: The need for constituent discourse in implementing testing policy. *Educational Policy, 16*(3), 387-405.
- Elliot, J. (2000). *NCAE: ABCs Accountability Statement*. Retrieved March 3, 2003 from, <http://www.ncae.org/news/000717pr.shtml>
- Everson, H., Smodlaka, I., & Tobias, S. (1994). Exploring the relationship of test anxiety and metacognition on reading test performance: A cognitive analysis. *Anxiety, Stress and Coping: An International Journal, 7*, 85-96.
- Firestone, W. A., & Mayrowetz, D. (2000). Rethinking "high stakes:" Lessons from the United States and England and Wales. *Teachers College Record, 102*(4), 724-749.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.
- Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. *Journal of Personality and Social Psychology, 59*(5), 916-924.
- Florida Department of Education (2001). *FCAT Briefing Book*. Retrieved July 13, 2002, from [http://www.firn.edu/doe/sas/fcat/pdf/fcat\\_brief.pdf](http://www.firn.edu/doe/sas/fcat/pdf/fcat_brief.pdf)
- Florida Department of Education (2002a). *Grading Florida Public Schools 2001-2002*. Retrieved August 9, 2003, from [http://www.firn.edu/doe/schoolgrades/pdf/guide02\\_bw.pdf](http://www.firn.edu/doe/schoolgrades/pdf/guide02_bw.pdf)
- Florida Department of Education (2002b). *Florida School Indicators Report*. Retrieved August 9, 2003, from <http://info.doe.state.fl.us/fsir/>
- Gordon, S. P., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership, 7*, 345-368.
- Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school. *Journal of Educational Psychology, 77*, 631-645.

- Grant, S. G. (2000). Teachers and tests: Exploring teachers' perceptions of changes in the New York state testing program. *Education Policy Analysis Archives*, 8(14). Retrieved July 24, 2002 from <http://epaa.asu.edu/epaa/v8n14.html>
- 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.
- 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.
- Holman, L. J. (1997). Meeting the needs of Hispanic immigrants. *Educational Leadership*, 54(7), 37-38.
- Jones, G. M., Jones, B. D., Hardin, B. H., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81, 199-203.
- Jones, G. M., Jones, B. D., & Hargrove, T. (2003). *The Unintended Consequences of High-Stakes Testing*. Lanham, MD: Rowman and Littlefield Publishers, Inc.
- Kaplan, L. S., & Owings, W. A. (2001). How principals can help teachers with high-stakes testing: One survey's findings with national implications. *NASSP Bulletin*, 85(622), 15-23.
- Knapp, M. S., & Shields, P. M. (1990). Reconceiving academic instruction for the children of poverty. *Phi Delta Kappan*, 71, 753-758.
- Kohn, A. (2000, September 27). Standardized testing and its victims. *Education Week*. Retrieved January 21, 2002, from [http://www.edweek.org/ew/ew\\_printstory.cfm?slug=04kohn.h20](http://www.edweek.org/ew/ew_printstory.cfm?slug=04kohn.h20)
- Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1993). *The reliability of scores from the 1992 Vermont portfolio assessment program* (CSE Technical Report 355). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.
- Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996) *Final Report: Perceived Effects of the Maryland School Performance Assessment Program* (CSE Tech. Rep. No. 409). National Center for Research on Evaluation, Standards, and Student Testing (CRESST): University of California, Los Angeles.
- Kyriacou, C. (1989). Teacher Stress and Burnout: An International Review. In *Human Resource Management in Education*, C. Riches and C. Morgan (Ed.), 60-68. Milton Keynes, U.K.: Open University Press.
- Matthews, L. J., & Crow, G. M. (2003). *Being and Becoming a Principal: Role Conceptions for Contemporary Principals and Assistant Principals*. Boston, MA: Allyn & Bacon.
- Messick, S. A. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment*, 10, 1-9.
- National Research Council (2000). *How People Learn*. Washington, D. C.: National Academy Press.
- Ornstein, A. C., & Levine, D. U. (2000). *Foundations of Education*. New York: Houghton Mifflin Company.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers*. National Board on Educational Testing and Public Policy. Retrieved April 1, 2003, from <http://www.nyssba.org/download/perceivedeffects.pdf>
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education: Theory, Research, and Applications*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56(6), 8-15.

- Popham, W. J. (2000a). *Teaching to the test: High crime, misdemeanor, or just plain good instruction*. Paper presented during AERA, New Orleans, LA, April 24-28, 2000.
- Popham, W. J. (2000b). *Modern Educational Measurement*. Boston: Allyn and Bacon.
- Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in non-directed text learning. *Learning and Individual Differences, 2*, 1-17.
- Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom: self-report and projective assessment of individual differences in children's perceptions. *Journal of Personality and Social Psychology, 50*, 550-558.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.
- State of Florida (2000). *Technical Assistance Paper: Standards for Implementation of Student Assessment Programs in Florida School Districts*. Florida Department of Education. Retrieved April 2, 2003 from <http://www.firn.edu/doe/sas/sasshome.htm>
- Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: SAGE Publications, Inc.

### About the Authors

#### **Brett D. Jones**

University of South Florida St. Petersburg  
College of Education  
140 Seventh Ave. South, COQ 201  
St. Petersburg, FL 33701  
(727) 553-4053  
bjones@stpt.usf.edu  
Homepage: <http://www.stpt.usf.edu/bjones/>

Brett D. Jones is an Assistant Professor of Educational Psychology in the College of Education at the University of South Florida St. Petersburg where he teaches courses in learning and development. Dr. Jones studies the effects of test-based accountability programs on teachers and principals and is a co-author of a book entitled *The Unintended Consequences of High-Stakes Testing* (Rowman & Littlefield Publishers, Inc., 2003).

#### **Robert J. Egley**

Robert J. Egley is an Assistant Professor in the Educational Leadership and Curriculum Studies Program at the University of South Florida St. Petersburg. He has 17 years' experience as an administrator at the elementary, middle, and high school levels. Dr. Egley has several lines of research, including high-stakes testing, invitational education, and instructional leadership. Email: [regley@stpt.usf.edu](mailto:regley@stpt.usf.edu)

***Education Policy Analysis Archives***

<http://epaa.asu.edu>

**Editor: Gene V Glass, Arizona State University**

**Production Assistant: Chris Murrell, Arizona State University**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, [glass@asu.edu](mailto:glass@asu.edu) or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: [casey.cobb@uconn.edu](mailto:casey.cobb@uconn.edu).

**EPAA Editorial Board**

**Michael W. Apple**

University of Wisconsin

**Greg Camilli**

Rutgers University

**Sherman Dorn**

University of South Florida

**Gustavo E. Fischman**

Arizona State University

**Thomas F. Green**

Syracuse University

**Craig B. Howley**

Appalachia Educational Laboratory

**Patricia Fey Jarvis**

Seattle, Washington

**Benjamin Levin**

University of Manitoba

**Les McLean**

University of Toronto

**Michele Moses**

Arizona State University

**Anthony G. Rud Jr.**

Purdue University

**Michael Scriven**

University of Auckland

**Robert E. Stake**

University of Illinois—UC

**Terrence G. Wiley**

Arizona State University

**David C. Berliner**

Arizona State University

**Linda Darling-Hammond**

Stanford University

**Mark E. Fetler**

California Commission on Teacher  
Credentialing

**Richard Garlikov**

Birmingham, Alabama

**Aimee Howley**

Ohio University

**William Hunter**

University of Ontario Institute of  
Technology

**Daniel Kallós**

Umeå University

**Thomas Mauhs-Pugh**

Green Mountain College

**Heinrich Mintrop**

University of California, Los Angeles

**Gary Orfield**

Harvard University

**Jay Paredes Scribner**

University of Missouri

**Lorrie A. Shepard**

University of Colorado, Boulder

**Kevin Welner**

University of Colorado, Boulder

**John Willinsky**

University of British Columbia

**EPAA Spanish & Portuguese Language Editorial Board**  
**Associate Editors**

**Gustavo E. Fischman**  
**Arizona State University**

**&**

**Pablo Gentili**  
**Laboratório de Políticas Públicas**  
**Universidade do Estado do Rio de Janeiro**

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez  
Universidad Nacional Autónoma de México

**Argentina**

- Alejandra Birgin  
Ministerio de Educación, Argentina
- Mónica Pini  
Universidad Nacional de San Martín, Argentina
- Mariano Narodowski  
Universidad Torcuato Di Tella, Argentina
- Daniel Suarez  
Laboratorio de Políticas Públicas-Universidad de Buenos Aires, Argentina
- Marcela Mollis (1998—2003)  
Universidad de Buenos Aires
- Ana Inés Heras Monner Sans  
Universidad Nacional de Jujuy
- José Luis Bernal Agudo  
Universidad de Zaragoza
- Carlos Mora-Ninci  
Universidad Nacional de Córdoba

**Brasil**

- Gaudêncio Frigotto  
Professor da Faculdade de Educação e do Programa de Pós-Graduação em Educação da  
Universidade Federal Fluminense, Brasil
- Vanilda Paiva
- Lilian do Valle  
Universidade Estadual do Rio de Janeiro, Brasil
- Romualdo Portella do Oliveira  
Universidade de São Paulo, Brasil

- Roberto Leher  
Universidade Estadual do Rio de Janeiro, Brasil
- Dalila Andrade de Oliveira  
Universidade Federal de Minas Gerais, Belo Horizonte, Brasil
- Nilma Limo Gomes  
Universidade Federal de Minas Gerais, Belo Horizonte
- Iolanda de Oliveira  
Faculdade de Educação da Universidade Federal Fluminense, Brasil
- Walter Kohan  
Universidade Estadual do Rio de Janeiro, Brasil
- María Beatriz Luce (1998—2003)  
Universidad Federal de Rio Grande do Sul-UFRGS
- Simon Schwartzman (1998—2003)  
American Institutes for Resesarch–Brazil

#### **Canadá**

- Daniel Schugurensky  
Ontario Institute for Studies in Education, University of Toronto, Canada

#### **Chile**

- Claudio Almonacid Avila  
Universidad Metropolitana de Ciencias de la Educación, Chile
- María Loreto Egaña  
Programa Interdisciplinario de Investigación en Educación (PIIE), Chile

#### **Colombia**

- Jorge Ossa  
Universidad de Antioquia

#### **España**

- José Gimeno Sacristán  
Catedratico en el Departamento de Didáctica y Organización Escolar de la Universidad de Valencia, España
- Mariano Fernández Enguita  
Catedrático de Sociología en la Universidad de Salamanca. España
- Miguel Pereira  
Catedratico Universidad de Granada, España
- Jurjo Torres Santomé  
Universidad de A Coruña

- Angel Ignacio Pérez Gómez  
Universidad de Málaga
- J. Félix Angulo Rasco (1998—2003)  
Universidad de Cádiz
- José Contreras Domingo (1998—2003)  
Universitat de Barcelona

### **México**

- Hugo Aboites  
Universidad Autónoma Metropolitana-Xochimilco, México
- Susan Street  
Centro de Investigaciones y Estudios Superiores en Antropología Social Occidente,  
Guadalajara, México
- Adrián Acosta  
Universidad de Guadalajara
- Teresa Bracho  
Centro de Investigación y Docencia Económica-CIDE
- Alejandro Canales  
Universidad Nacional Autónoma de México
- Rollin Kent  
Universidad Autónoma de Puebla. Puebla, México
- Javier Mendoza Rojas (1998—2003)  
Universidad Nacional Autónoma de México
- Humberto Muñoz García (1998—2003)  
Universidad Nacional Autónoma de México

### **Perú**

- Sigfredo Chiroque  
Instituto de Pedagogía Popular, Perú
- Grover Pango  
Coordinador General del Foro Latinoamericano de Políticas Educativas, Perú

### **Portugal**

- Antonio Teodoro  
Director da Licenciatura de Ciências da Educação e do Mestrado Universidade Lusófona de  
Humanidades e Tecnologias, Lisboa, Portugal

### **USA**

- Pia Lindquist Wong  
California State University, Sacramento, California
- Nelly P. Stromquist  
University of Southern California, Los Angeles, California
- Diana Rhoten  
Social Science Research Council, New York, New York
- Daniel C. Levy  
University at Albany, SUNY, Albany, New York
- Ursula Casanova  
Arizona State University, Tempe, Arizona
- Erwin Epstein  
Loyola University, Chicago, Illinois
- Carlos A. Torres  
University of California, Los Angeles
- Josué González (1998—2003)  
Arizona State University, Tempe, Arizona