

**PART TWO**

**The Contested Role of the State  
in Educational Policy**



## **CHAPTER 3**

### **ASSESSMENT IN FLORIDA SCHOOLS**

**Willa Wolcott  
School of Teaching and Learning  
University of Florida**

During the last three decades, assessment has been a driving force behind Florida's efforts to improve its schools. Through such examinations as the Florida Teacher Certification Exam (FTCE) and the Florida Educational Leadership Exam (FELE), the state has sought to ensure competencies for K-12 teachers and administrators. It has sought to demand competency from college students through the College Level Academic Skills Test (CLAST) and minimum competency from high school students through the High School Competency Test (HSCT). Most recently, Florida Writes! and the Florida Comprehensive Assessment Test (FCAT) in reading and mathematics, tests developed specifically for Florida schools, are an effort to obtain accountability for student learning.

Florida Writes! is a direct writing test that measures student writing achievement by asking students in the fourth, eighth, and tenth grades to compose an essay on an assigned topic during a 45-minute period. Topic prompts for fourth graders call for either narrative or expository writing. Topic prompts for eighth and tenth graders call for either expository writing or persuasive writing. The Florida Comprehensive Assessment Test (FCAT), developed in the mid-'90s to assess students' performance in reading and in mathematics across three grade levels, is designed to be closely linked to Florida's Sunshine State Standards. The FCAT differs from some earlier Florida tests in terms of the challenging questions asked and the open-response items included for either short or extended answers. These three tests, which have recently been merged under the FCAT testing umbrella, represent challenging assessments in writing, mathematics, and reading for students at three

---

different levels of K-12 schooling. (An additional FCAT test in science has been authorized, and one in social studies may be forthcoming.)

By promoting writing through the direct writing assessment and by including performance items on the FCAT in reading and mathematics, state legislators have sought to use testing as a means of improving overall instruction and of aligning it more closely with current educational theories. Florida is certainly not alone in using assessment as its major approach for upgrading the quality of education. A number of other states have also developed stringent assessments, many in response to the 1994 Congressional reauthorization of Title I that required states, as a condition of participating in the Title I program, to develop rigorous standards and define advanced skills (D'Agostino, 2000, p. 330). However, despite the increasing prevalence of assessments nationwide—and despite the attendant publicity that accompanies test results—research suggests that assessment has serious limitations, in addition to advantages, as a tool for enhancing educational quality. This chapter will focus on the advantages and disadvantages of this strong emphasis on assessment as specifically related to *Florida Writes!* and the FCAT. The tests will be discussed separately despite their recent merging under the one FCAT umbrella.

### **Benefits of the Testing Mandate**

Educators have long recognized the significance of testing. As Scott Paris et al. (1999) emphasize, “Assessment is fundamental to the improvement of education; it provides measures of success for students’ learning, for educators’ leadership, and for continuous evaluations of instructional programs” (p. 30). First, through assessments, students themselves know whether they have learned the material; second, teachers and administrators know whether their teaching has been effective; third, parents know that their children’s diplomas will have value; and finally, the public knows that teachers are, in fact, teaching certain concepts or material deemed important to know. Assessment can provide guidance for schools, ensuring that, despite their diversity, they are all working toward the same goals set forth by state standards. If programs or performances are found wanting, then resources and efforts can be directed toward helping those schools improve.

In Florida, the FCAT and Florida Writes! have unified schools throughout the state in a common purpose—to improve the writing, reading, and mathematics education of their students. For example, by requiring students to compose original essays, Florida Writes! has given new emphasis statewide to the importance of writing. Through the inclusion of different topic types, the test has encouraged schools to teach a variety of writing forms. Similarly, teachers’ understanding of what good writing means has been broadened through workshops, and some teachers have improved their teaching through the holistic scoring sessions they have attended.

The FCAT exams in reading and in mathematics have also promoted a new view of mathematics and reading and have broadened educators’ understanding of what needs to be accomplished in the schools. While administered by a professional testing company, the tests are “homegrown” in that they have been developed under the auspices of the Florida Department of Education with the involvement of numerous in-state educators, and the tests have been linked directly to the Sunshine State Standards. The challenging, higher-order thinking questions that require students not only to analyze, synthesize, and evaluate but also to respond to selected items in writing have served to promote new instructional practices in both subject areas. Furthermore, the effort to ground a few of the mathematics problems in other subject areas has illustrated the underlying importance of mathematics in many subjects, and the desired goal of emphasizing problem solving rather than rote memorization has been underscored. Because of the challenging nature of the tests and their link to the state standards, many schools are now making more efficient use of the school day.

Finally, the Florida Department of Education has made efforts to help schools statewide prepare for the tests. The department has held workshops to explain the tests, distributed sample materials, prepared informational materials on a website, and sponsored the publication of teacher-made practice tests and instructional guides. Thus, along with demanding accountability, the tests have spurred a multi-pronged effort to help schools reach the new goals.

### **Negative Effects of the Testing Mandate**

While the tests have provided several benefits for Florida, there are limitations—those which are attributable to external testing in general and those which relate specifically to the Florida tests.

The broader problems encompass “teaching to the test” and the “constriction of the curriculum” that generally result from external tests. Given that assessment drives instruction, it is not surprising that much time in schools is devoted to preparing for tests, administering tests, and then recovering from tests. The time, as Mary Lee Smith (1991) warns, comes out of instructional time. The testing time for the FCAT and for Florida Writes! is extensive when we consider that each of the FCAT exams in reading and mathematics takes 160 minutes and the writing samples another 45 minutes or over 6 hours total. Typically, a slump follows in which students and the teacher regroup from the testing pressure. (Paris et al., 1999, took note of a “canyon effect” of low motivation that followed the administration of some standardized tests, p. 32.) These time factors do not include the time spent on preparing for the tests.

Certainly, good instruction and test preparation are not mutually exclusive—especially if the tests are, as is the case with the Florida tests, linked to state standards and especially, too, if they are challenging. However, when a central focus of a classroom or an entire goal for the school year becomes that of preparing for the test, then the overall quality of instruction and the overall value of education are diminished.

In a survey of elementary teachers on the impact of recent high-stakes testing on North Carolina teachers, Gail Jones et al. (1999) found that teachers spent more time teaching the mandated subjects with less time devoted to those subjects not tested and that more than 20% of instructional time was spent preparing for tests. (In some cases, students spent 60% of instructional time practicing.) Similarly, in exploring the effect that external testing has on Arizona teachers, Mary Lee Smith (1991) found that entire subjects, such as social studies, art, and health instruction, disappeared from some elementary schools in the weeks

leading up to the test, and that concepts not expected to be tested were excluded from instruction. She warns: “The focus on material that the test covers results in a narrowing of possible curriculum and a reduction of teachers’ ability to adapt, create, or diverge. The received curriculum is increasingly viewed as not amenable to criticism or revision” (p. 10).

Martin Brooks and Jacqueline Brooks (1999) see this restriction of the curriculum to what appears on the tests as an unfortunate narrowing and a reduction of the “holistic endeavor” that the educational experience should represent; it signifies, in their view, a denial of the complexity of the learning process. They caution, for example, that the accountability segment of standards-based reform efforts is “eclipsing the intent of standards and sound educational practice;” that is, so much time and effort are spent on preparing for tests that meaningful instructional approaches and subjects not readily assessed are being eliminated from the curriculum (p. 23).

Such restrictions on the curriculum are unfortunate for all students but may be especially so for disadvantaged students whose understanding of the world may be disproportionately derived from their school experiences. In a study in which upper elementary school teachers were sampled throughout the nation, Joan Herman and Shari Golan (1989) note that even though teachers said other non-tested subjects were still being taught, schools with low socioeconomic (SES) populations spent more time on practice and preparation for tests than did schools in more advantaged areas. Pamela Moss (1994) points to that and similar evidence as indicating that “the narrowing of the curriculum associated with high stakes standardized assessment may be falling disproportionately on certain groups of students for whom concerns about equality of education have been most salient” (p. 113). For these students in particular, what is NOT taught, what is NOT included in the curriculum, might never be learned or experienced. The time spent directly and tangentially on test preparation and administration, with the consequent narrowing of the curriculum, remains a major limitation of external tests.

## **UNDERSTANDING WHAT THE FLORIDA TESTS MEASURE**

In addition to general testing problems, it is important to understand the limitations, as well as the strengths, of the tests themselves.

### **Strengths and Limitations of Florida Writes!**

Florida Writes!, like other direct writing assessments, requires students to respond immediately with an essay on a topic prompt that is given them. Depending on what the topic is, students may be required to write a persuasive, narrative, or expository essay in a limited time with no real opportunities—or resources available—for much revision. Although this type of writing is indeed a valid form, it represents only a small sample of students' total writing experiences. In fact, critics of direct writing assessments argue that such impromptu writing (also called timed writing or demand writing) runs counter to what research has emphasized as being fundamental to the teaching of writing in classrooms. That is, writing is seen as a goal-oriented, meaningful, recursive, or back-and-forth process of prewriting, drafting, revision, and final editing (Flower and Hayes, 1981). The solitary demand of the writing test also denies the importance of the social community to writing, as shown through such collaborative processes as peer editing (Bruffee, 1986).

Furthermore, by giving students specific topics they must address, timed essays such as those used in Florida Writes! overlook the importance of a writer's "owning" the topic. Such "stimulus-response" formats, in the words of educator Garth Boomer (1985), deny a "basic condition for good writing: having something to say and wanting to say it" (p. 63). As is always the case with assigned topics, the accessibility of the topics used for all students remains an important factor, and this is why topics are field tested and carefully reviewed to eliminate any potential bias.

Because topics must enable all test takers to respond immediately, it is assumed that topics eliciting personal experiences are likely to engage the student writers' interest and elicit fluent writing. Some, though certainly not all, of the topic prompts used by Florida Writes! draw on personal experiences. For example, the persuasive prompt for the 1999

Grade 8 assessment asked students to convince a local television station to hire them as student reporters (Florida Department of Education, 1999b, p. 28). However, some research has suggested that personal writing is not necessarily easier for all students (White, 1985) and that, in fact, some students may be uncomfortable at revealing themselves (Ruth and Murphy, 1988). Roxanne Mountford (1999) warns that personal essays are influenced by cultural and gender factors, and Sandra Stotsky (1995) cautions that writing assessments should not overemphasize personal writing.

Still another factor that can affect student achievement is the time constraints. While the 45 minutes currently provided for the Florida writing test is certainly longer than the time provided in some writing assessments, the scope of the task facing each student is significant. Lee Odell (1981) notes, for example, that a typical writing prompt requires students to “contemplate a topic to *which they have likely given little previous thought*; identify their audience and purpose; decide upon the rhetorical strategies they will need in order to achieve their purpose with their intended audience; write a first draft; reconsider and, where necessary, revise that draft; edit their draft to make sure it corresponds to the conventions of standards written English” (pp. 108-109). If, as is contemplated, future forms of Florida Writes! ask students to write on two 20-minute prompts or possibly three 15-minute prompts (so that topic prompts can then be equated), the formulaic, “stimulus-response” nature of demand writing will be intensified even further.

Hence, it must be recognized that the topic assigned, the type of writing required, and the time constraints may not always be conducive to producing the best writing possible on any given student’s part. In fact, it is precisely because of these limitations that researchers have stressed the need for having multiple samples of students’ writing in order to assess their writing ability accurately (Breland et al., 1987). In acknowledgment of these very limitations also, the Florida Department of Education cautions that “the Florida Writes! results do not represent a comprehensive evaluation of writing programs” and that, furthermore, “Florida Writes! does not measure all important aspects of writing” (Florida Department of

Education, 2000). Indeed, many forms of creative writing and the thoughtful research writing required by some subjects are key parts of successful writing programs that are not assessed.

### **Strengths and Limitations of the FCAT in Reading and in Mathematics**

The FCAT in reading, which contains informational and literary passages drawn from a variety of sources, tests two strands, “Constructs Meaning from Literature” and “Constructs Meaning from Information.” These strands, in turn, consist of nine benchmarks or key categories linked to the Sunshine State Standards. A real strength of the FCAT is its inclusion of higher order thinking questions and its use of some open-ended questions to which students must write their responses. A drawback of the test is the time constraints under which students must work and the defined space into which students must write their open-ended responses. A more serious limitation is that the test represents but a small part of what the actual act of reading involves.

Critics of reading assessments point to the complexity of literacy development and to current reading theories that emphasize the meaning-making nature of the reading process in which any subskills, far from being separate as they are customarily treated on tests, are intertwined with one another. Roger Farr points out, for example, that according to analyses of reading tests, “The reading required on most tests is not much like the reading behavior that our new understanding describes. How valid is the content of a reading test in terms of *reader purpose, interests, and background* [italics added], which we now believe are primary influences on reading behavior?” (p. 49)

Indeed, the lower reading test scores that students from disadvantaged areas sometimes obtain may be partially explained by reading theory. The importance of prior knowledge or background experiences is recognized by many theorists as one of the major factors in reading comprehension. That is, networks of related knowledge function almost as a “coat rack” of old knowledge onto which readers can “hang” their new learning to construct a broader and deeper understanding. In

classrooms, teachers can tap into such prior knowledge or enhance it through questions and discussion, thereby enabling students to integrate the new reading material with the old and consequently improve their comprehension. In a formal assessment, students must ferret out the meaning of the new passage and, on their own, integrate the new knowledge with the old. Students who have not had sufficient experience in practicing these strategies are likely to do poorly.

The FCAT in mathematics, like its counterpart in reading, challenges students with problems to analyze, synthesize, apply, and evaluate. Questions for the mathematics problems, which are sometimes based on subject areas such as social studies or science, address five major strands of skills: “Number Sense,” “Measurement,” “Geometry,” “Algebraic Thinking,” and “Data Analysis.” In addition to a few open-ended items, there are some gridded response items, which require students to solve a problem and bubble their answer on a grid rather than choose an answer from a series of options given. The test has drawbacks as well. Although it requires students to explain some of their work in open-ended responses, typically only two of the performance items are extended response, and the multiple choice and gridded response items comprise a much larger percentage. Time constraints are, as in reading, an issue in that students must solve numerous problems within a specified time. The gridded response items, while requiring students to generate their own answers instead of choosing from multiple options, also present a challenge for students in terms of format. Without sufficient practice with this type of response format, students can have difficulty in recording their answers. Additionally, the gridded response items preclude any use of mixed numbers, so all answers must be converted to decimals. Still another issue is that of the variables involved. Students halfway through the tenth grade may not have enough experience with algebra to deal with all the variables required to solve problems.

Thus, as can be seen, all three tests—Florida Writes! and the FCAT in reading and in mathematics—measure *a valid but necessarily limited part of what students study and learn in all these subjects*. Moreover, the tests place additional constraints on students in terms of

time limitations and the testing formats themselves, which often run counter to the learning processes entailed in these subject areas.

### **PROS AND CONS OF HIGH STAKES ASSESSMENTS**

Notwithstanding the strengths and limitations of both Florida Writes! and the FCAT as assessment instruments, we must pay attention, ultimately, to the purposes served by these assessments—whether for information, diagnosis, or accountability. Increasingly, accountability is, as Richard Stiggins (1999) explains, the driving force behind educational reform.

#### **Endorsement of High Stakes**

Recently, Public Agenda (a non-profit, non-partisan research group) conducted, in conjunction with *Education Week*, “Reality Check 2000,” an annual survey on the progress of academic standards. Public Agenda reports that over 80% of employers, professors, teachers, and parents surveyed endorse the importance of guidelines in helping students to learn. Over 60% of this group also support high stakes tests, and no more than 24% of any one segment of the group believe strongly that the high stakes tests “will result in schools’ being overwhelmed by many failing students” (p. S3). It is important to note, however, as Public Agenda also cautions, that **“parents and the public generally think of standards as guaranteeing minimum basic skills, rather than what they may see as more advanced or esoteric knowledge”** (2000, p. S3). Public Agenda notes that parents and the public expect extensive support to be given to weak students; few support the idea of a student’s being held back on the basis of one test.

Some leaders in the business community endorse the national trend toward enforcing higher standards through more demanding tests. For example, a 1999 issue of *WorkAmerica* declares, “Raising standards, with consequences, typically results in the following changes: elimination of bottom-track, low-level courses; provision of extra support for struggling students; and new ways of extending learning time. . . .” (p. 1). Writing in *Education Week*, Edward Rust (2000), the chief executive officer of State Farm Insurance Company and the chairman of the Business Roundtable

Education Task Force, argues on behalf of “having genuinely strong and clear standards, aligned closely to a challenging test and to incentives that demonstrably make students, teachers, and others work harder and smarter” (January 19, p. 40).

Several studies seem to support the idea that high stakes can result in better test scores. In a study that explored the effect of minimum competency tests on learning, Norman Frederiksen (1994) compared the 1978 results obtained by 9-, 13-, and 17- year-olds on the mathematics portion of the National Assessment for Educational Progress to the scores achieved on the same items by students of the same age in 1986—after minimum competency tests were in place. He found that students — especially the 9-year-olds — performed better in 1986 and especially in those states which had high stakes (such as graduation) attached to the tests.

In another study, David Grissmer and Ann Flanagan (1998) conducted case studies in Texas and North Carolina to determine why 9- and 13-year olds in those states “showed the largest average gains among states” in the mathematics and reading portions of the National Assessment of Educational Progress (1992-1996) and also in their own state achievement tests (pp.1, 2). After close analysis, they rejected as possible explanations for the gains such factors as per pupil expenditures, class size, and length of teacher experience (pp. 12-13). They attributed the gains to the implementation of similar policies that emphasized assessment and accountability as part of “systemic reform” (pp. 19-21). Because they examined only two states, Grissmer and Flanagan caution that they can not be definitive about factors that may account for gains, but they speculate that a joint, public/private infrastructure established to support educational improvement may be a substantial factor.

John Bishop (1998) explored the effect on student achievement of curriculum-based external exit exam systems (CBEEES) such as those used internationally and in New York State with its Regents’ Exam. After studying data from four major tests (including the Third International Mathematics and Science Study (TIMSS) and the science and

mathematics scores of 13-year-olds on the International Assessment of Educational Progress), Bishop concludes that “the claims by advocates of standards-based reform that curriculum-based external exit exams significantly increase student achievement are probably correct. . . .When student demography is held constant, New York State, the only state having a CBEEE system in the early 1990s, does significantly better than other states on the SAT test and the NAEP mathematics assessments without experiencing a reduction in high school graduation rates” (p. 21). In Bishop’s view, although exit exams are not the only factor contributing to achievement, the exam counteracts “peer pressure against studying” and supports teachers and administrators in their demand for more “academic rigor” (p. 4, 6).

Several cautions are in order as we interpret the results of these three studies. It is important to note that, at the time of Bishop’s study, only **moderate** stakes for individual New York students were involved; their exam performance comprised only part of students’ course grades, and students could opt not to take Regents’ courses altogether. In the Grissmer and Flanagan study of North Carolina and Texas, the students examined were 9- to 13-year-olds, and high school exams were not considered. In the Frederiksen study, the 17-year-olds, even in high stakes states, did not make the gains the 9-year-olds did. Finally, most of these studies focus on mathematics results; the studies do not address what happens to subjects NOT covered on the exams. As Pamela Moss (1994) notes, “At best, test scores can reflect only a small subset of valued education goals” (p. 114). She cites research showing that even though scores often rise in a high stakes program, the increase does not mean the quality of education has necessarily improved.

### **Criticisms of High Stakes**

In fact, several educators question whether the standards-based reform efforts will ultimately succeed. For example, after conducting a five-site study on the effect of minimal competency testing on educational reform, Mary Catherine Ellwein, Gene Glass, and Mary Lee Smith (1988) found that not only was too much emphasis placed on “the standard’s image” and on *initial* failure rates, but there also was a sharp distinction

between the early and late phases of the reform movement. Thus, they propose that “competency tests and standards function as symbolic and political gestures, not as instrumental reforms” (p. 8), and they stress the need for empirical data. James Catterall (1990) studied students’ awareness of the minimal competency test as a graduation requirement in four states with very different graduation rates. He found that only about half the students knew that the test was required, despite the fact that many had already taken it. He suggests that “the visibly debated competency test has retreated as a school policy to the point of immateriality to educators and students” (p. 6).

Teachers themselves, as the Public Agenda’s (2000) “Reality Check” indicates, hold very different views from those of other sampled groups (parents, employers, and college professors) about numerous issues connected with standards. Most teachers feel public schools are doing as good a job as possible given the lack of parental involvement, and many disagree that “high-stakes testing makes teachers and students more accountable” (p. S7). The complexity of teaching and learning is stressed by Susan Ohanian (2000), who, in deploring the use of business analogies with which to envision and compare teaching, asks, “How many minutes does a fledgling teacher have to be in a real classroom before she realizes that students don’t pass by her desk like goods on a conveyor belt?” (p. 346).

Martin Brooks and Jacqueline Brooks (1999) also take issue with the thinking behind standards-based reform, which they summarize as follows: “Develop high standards for all students; align curriculum and instruction to these standards; construct assessments to measure whether all students are meeting the standards; equate test results with student learning; and reward schools whose students score well on the assessments and sanction schools whose students don’t” (p. 20). They argue that despite the seeming logic of this thinking, this “linear approach” is not only too much like other attempted reforms but also overlooks the complexity of learning.

In a similar vein, Jerome D'Agostino's (2000) comprehensive review of the literature related to the effects of achievement testing has led him to conclude that despite the creation of more sophisticated, more efficient achievement tests, history is repeating itself with tests a century later still being used to "evaluate teachers and schools and to hold them accountable for student learning" (p. 333). He points out that there are "some fundamental limitations in the use of tests to effect change" and states, "Neither the political nor professional reform models use achievement tests effectively" (p. 333). Without teacher ownership in the tests, he argues, teachers fear being held publicly accountable and resent tests. Furthermore, because non-educators frequently have had a large say in the standards created, the actual tests developed for accountability frequently do not help teachers understand problems their students are having (p. 333).

Like D'Agostino, Al Ramirez (1999) argues that the great faith policy makers have placed in assessment overlooks the limitations of assessment (p. 205). Another educator, Thomas Kelly (1999), questions the value of forcing compliance with the standards; he suggests that the higher standards mandated by many states are very difficult and that expecting students to attain high standards in all subjects is not realistic and disregards human diversity (p. 546.) Indeed, researchers have recently commented on the seemingly universal pattern in which test scores "start low, rise quickly for a couple of years, level off for a few more, and then gradually drop over time." Because of this predictability of test outcomes, questions have arisen as to whether tests "are valid measures of progress toward the academic standards set by states" (Hoff, 2000, pp. 1, 12).

### **PROPOSED STAKES FOR FLORIDA TESTS**

In that regard, the projected scope of the Florida tests with serious consequences for all involved—schools, teachers, and students themselves—can only lead to concern. In 1999, for example, results from the FCAT and Florida Writes! were two major criteria, along with attendance, discipline, and drop-out rates, used in grading the schools throughout the state. Schools were graded against a common standard with the result that the lowest performing schools typically were those

containing large numbers of students on free or reduced-price lunches and significant numbers of transient students. More important, the grades were made public so that teachers, the public, and students alike knew whether an individual was part of a “D” or “F” school or a school with a higher rating.

Beginning with the freshman class of 1999, students must pass the FCAT in order to graduate. As of this writing, the passing score has not been set, but there will be six opportunities for students to pass the FCAT. A link between the Florida College Entry-Level Placement Test (CPT) and the FCAT is also envisioned in which students who show proficiency on the FCAT in mathematics or reading may be exempt from the similar section of the college placement test. Additionally, students in the currently non-tested grades (3, 6, 7, and 9) will be given a multiple-choice FCAT to assess the Sunshine State Standards in reading and in mathematics (T. Fisher, personal communication, January 19, 2000). Furthermore, Florida Statute 232.245 requires the retention of fourth graders unable to read at a certain level.

Finally, the score range represented by the achievement levels will, as of 2002, rise. For example, whereas now scaled scores of 100-287 represent the lowest level—Level 1—for fifth-graders in mathematics, scaled scores of 100-300 will represent Level 1 for them in 2002 and afterward. Similar increases will occur at all achievement levels for both reading and mathematics (Florida Department of Education, 1999a). Thus, expectations for, and consequences of, students’ performance on the FCAT are projected to increase.

### **IMPLICATIONS OF HIGH STAKES ASSESSMENT**

The rising stakes attached to the FCAT, (or to any test, for that matter), are troublesome. In *Testing for Learning*, Ruth Mitchell (1992) decries the prevalence of the message “that a single test can determine what students know and can do” (p. vii). Similarly, U.S. Secretary of Education Richard Riley, while advocating real accountability, has recently urged states **not** to “rely on a single test to measure students’ knowledge of the standards” (Sack, 2000, p. 32). Regardless of how strong

the testing instrument, no single test or battery of tests can accurately reflect the full mosaic that students' education represents, nor do state legislators intend for testing to do so. In fact, businessman Edward Rust, cited previously, stresses that no one truly believes that "scores on standardized tests are the last word on academic success" (2000, p. 40). Nevertheless, the emphasis accorded tests through the high stakes attached to them conveys — at least implicitly—that very message. However, in few other segments of public life are either employees or their employers routinely evaluated on the basis of only *two or three days' performance*. (One of the few exceptions is the Olympics, but in that context athletes bring motivation, total focus, and evident talent to the task, in addition to public respect and full resources.) When the results of such tests are then publicly recorded by the media, we can better appreciate the anxiety and concern with which teachers, administrators, and students view assessment.

As Richard Stiggins notes, high stakes testing can have negative, *unintended consequences* for teachers and learners, creating feelings of despair, anger, and anxiety. In a study done on the effect of external mandated testing on elementary schools in Arizona, Mary Lee Smith (1991) found that such tests created great anxiety before, during, and after the tests. Teachers reacted to this anxiety in one of two ways—either by focusing almost exclusively on the subjects to be tested or by passively resisting through ignoring the demands of the testing altogether.

In the survey that sampled North Carolina elementary school teachers throughout the state, 61% of the teachers responding thought their students were more anxious, and over 77% of the teachers believed their own morale was lower (Jones et al., 1999). In fact, many thought the program would not lead to improved education in their schools, perhaps an ironic finding in light of North Carolina's supposed assessment success noted previously in the Grissmer and Flanagan study. Similarly, in the study reported by Joan Herman and Shari Golan (1989), teachers indicated that "testing creates substantial tension for teachers and students" (p. 23).

Although some tension is surely to be expected, too much anxiety can have negative results—on students, as well as on teachers. While acknowledging that pressure to meet high academic standards as shown through test scores is effective with *some* students and with *some* teachers, Richard Stiggins warns, “For students, increasing pressure to score high on tests, combined with a lack of focused opportunities to learn, can lead to a sense of futility—a feeling of hopelessness—that can cause them to stop caring and stop trying. . .” (1999, pp. 192-193). He warns of the undesirable effect that will occur with students ceasing to try and, hence, ceasing to learn. To anyone who has worked closely with students engaged in high stakes test taking, anxiety and distress are, no matter what the students’ age, all too real—and especially so for those hardworking students who do not test well. In fact, Florida teachers at one “D” school spoke of “shielding” their young students from the stress of trying to remove their school from the low-performing list (Carr, 2000).

Given the projected teacher shortage during the next few years, and given, too, Florida’s concern with its substantial drop-out rate, we must worry about effects such high stakes testing will have on students and teachers alike. Will hardworking, dedicated teachers who become associated with failing schools become too demoralized to continue their struggles? Or will embattled teachers who are held accountable for their students’ successes even when many variables are beyond their control choose to remain in the teaching profession? Similarly, will defensive, alienated, and discouraged students decide, if they always seem to be unsuccessful, that dropping out of school is preferable to remaining in class?

#### **ALTERNATIVE ROLES FOR ASSESSMENT**

To express concern about the implications of assessment does not mean that all assessment should be abolished, for, as noted at the outset, assessment serves some very important purposes. At the same time, we must ask whether assessment should become such a dominant force that other, meaningful content or promising instructional strategies *inadvertently* become lost in the curriculum. We must ask for Florida, as Richard Stiggins has done for school systems in general, whether high

stakes tests should be the major vehicle through which we seek to improve our educational system. We must seek to counteract potential ill effects by exploring how assessment might be used to *encourage*, rather than *attempt to enforce*, educational reform.

### **Changing the Stakes**

As we have seen, having the FCAT umbrella carry such heavy stakes for individuals and schools as it is projected to do is fraught with serious implications. One alternative would be an assessment model in which the three FCAT tests would continue to provide district-wide information but high stakes would be eliminated. Matrix sampling, a statistical sampling procedure for obtaining representative data, could be used so that students in targeted grades would each take *some, but not all* of the performance assessments in the specified subjects (Farr, 1999; Moss, 1994). The reading and writing tests could be combined so that students would respond in writing to their readings, thereby creating a more authentic representation of both the reading and writing processes (Claggett, 1999). Results would then be publicly released for *school districts* so that the state and the public would continue to have overall information about how the different school districts are performing statewide. Hence, the public would still be informed as to how well Florida schools, with their diverse populations and extremes of rural and urban settings, are performing against a common measure. While teachers and students would still need to prepare for the tests, using a matrix sampling would enable a broader range of items to be covered and, at the same time, reduce the time taken out of instructional periods for actual testing. Moreover, if high stakes were not attached to the FCAT scores, the exams would not have to be scored soon after their administration, and teachers could be paid during the summer to do the ratings on a regional basis. Scoring the essays and the open-ended responses would provide teachers with invaluable professional development. Thus, the instructional value of the tests would be increased (Bishop, 1998; Mitchell, 1992).

A variation of this approach would be to continue with census testing for the FCAT—that is, giving the same, complete test to all students so that individual scores would be obtained—but *to eliminate the*

*snapshot, single-performance testing as the basis for determining graduation readiness or the need for retention in elementary school.* Rather, individual scores could be considered as a *part* of the final English or mathematics course grades, thereby giving students through moderate stakes an incentive to try their hardest. Of course, with this approach the time taken out of the instructional day for test administration would not be shortened, nor could as broad a spectrum of performance items be tested. While this approach would give students some external reason for applying themselves (which is missing from the matrix approach), census testing would hold students accountable for their own performances before schools were necessarily improved, a potentially troublesome issue (Smith and Levin, 1996). Moreover, the need for getting scores back to the schools soon after the testing would mean that teachers would miss out on the professional development that such scorings can provide.

If neither the suggested matrix nor census alternative is considered feasible, a third alternative would be to delay *the implementation of high stakes until students, teachers, and schools have had a chance to practice the instructional precepts* promoted through these new tests. With all the subjects tested, the learning is often cumulative, and the building of knowledge takes time. As noted with regard to the national mathematics standards, it takes time for students, as well as teachers, to adjust to a new curriculum that requires them to think mathematically, to problem solve, to read, and to write (Bay, Reys, and Reys, 1999). Thus, it is more reasonable to expect today's fourth or fifth graders to have had the appropriate mathematics sequence and mathematics instruction by high school to enable them to pass the FCAT for graduation than it is to expect today's ninth graders to learn the material in a year or two.

As increasingly high stakes become attached to the FCAT, it is important that provisions be made for comprehensively evaluating the overall effectiveness of this movement. Such evaluations could not only include information about those who pass and those who fail, but also explore both the impact on the educational system as a whole and the costs, financial and otherwise, that are incurred (Ellwein, Glass, and Smith, 1988). Jerome D'Agostino (2000) has noted the lack of "few

empirically driven consequential validity studies” on the impact of achievement testing on the classroom (p. 313).

### **Grading of Schools**

Assigning grades to schools is heavily dependent on the criteria used. (For example, one Florida high school, which received a “C” from the state, was recently acclaimed in a national news magazine for its International Baccalaureate and Advanced Placement programs.) Although the practice of grading schools has forced school administrations to confront problems within their districts, in the long term, grading schools seems counterproductive, creating unnecessary competition among schools of vastly different resources and populations and demoralizing teachers, staff, and students at low performing schools.

Because of these differences, it is vital that the improvement made by any school—especially weak performing schools—be considered an important criterion if the assigning of grades is continued. In that regard, the legislative intent to factor improvement into school grades beginning with 2000-2001 is a positive start, as is the intent to give financial resources to schools that improve. Indeed, one of the few potentially positive benefits to stem from grading will occur if significant attention and resources—whether in the form of money, materials such as trade books, or proven programs—are directed toward low performing schools to help them succeed. Because low performing schools cannot be discarded (unlike a low performing vehicle), it is *imperative that causes be identified and remedies undertaken. Resources allocated in this way should be viewed as needs-based, rather than merit-based.*

Schools in lower socioeconomic areas often have larger gaps to close than do schools in more affluent neighborhoods; for example, students enter school at greatly varying levels of language development, which impacts their mastery of all skills, but especially their reading skills. While having high expectations for all students is important, large learning gaps—whether in reading, math, or writing—cannot be closed quickly, regardless of how hard students and teachers work. Such schools need additional resources, more time, and specialized programs to help close

these gaps. Pilot studies can be undertaken to determine which of many possible programs offer the most promise. The Grissmer and Flanagan study (1998) showed that more resources were shifted to schools in disadvantaged areas in North Carolina and Texas. While the proposed pairing of business executives with low performing schools certainly has potential (The Gainesville Sun, 2000, January 26), that practice alone will not suffice, and it, too, will need to be studied to determine its effectiveness.

Furthermore, paying attention to how cohorts— that is, groups of the same students—perform over time is also important. Otherwise, as Darling-Hammond notes, “The use of cross-sectional rather than cohort analyses of performance means that annual scores are as likely to be an artifact of changes in the population of students taking the test as in the quality of teaching they experience” (1997, pp. 250-251). If the FCAT is given in successive years through the majority of students’ schooling, the records will exist to provide in aggregate form an indication of how well students are doing over time.

### **Formative, Classroom Assessments**

In addition to informing the public about school progress through accountability tests such as the FCAT, improving diagnostic classroom testing is also important for helping to meet students’ individual needs. Because such formative, classroom assessments represent the “other 99% of the assessments that happen in a student’s life,” Richard Stiggins (1999) argues that the area of formative, classroom assessment offers the best chance for true educational reform (p. 193).

Formative assessment is a behind-the-scenes, day-to-day, internal process within the classroom that neither attracts publicity nor results in single, newsworthy, high stakes scores allowing for comparisons to the norm to be made. The value of such formative assessment has been stressed by Paul Black and Dylan Wiliam (1998). After reviewing a number of studies that covered students at different levels of schooling and in different subjects, Black and Wiliam (1998) found that improving formative assessment practices contributed to “significant and often

substantial learning gains” (p. 140) and helped low achievers especially by showing them how to correct their specific problems.

Improving formative assessment entails having teachers and administrators understand the principles of assessment and of evaluating student achievement (Ramirez, 1999). It also entails, among other practices, helping teachers understand how their classroom discussion questions and their test items can enhance student learning or helping them see, as Farr (1999) suggests, patterns in their students’ work to which they can then adjust their teaching. It further involves helping teachers try such authentic approaches as portfolio assessment, in which students, who also learn self-assessment through this undertaking, purposefully collect their best work throughout a school term to show growth or improvement over time. It has, as one goal, the involvement of students in their own learning (Stiggins, 1999).

Of course, improving classroom assessment is slow and entails extensive professional development. However, the workshops and training practices initiated when the FCAT was first developed might serve as a point of departure for improving classroom assessment. Expending more attention, more resources, and more time on the in-service training of teachers and administrators in the principles and practices of formative classroom assessment offers the promise of impacting students directly. It represents what Darling-Hammond (1997) calls an investment “in teachers’ abilities to teach more effectively” (p. 251).

### **Dialogues With All Concerned Parties**

Another part of the ongoing process of educational improvement might include convening district-wide roundtable discussions during the summers with representatives chosen from all parties concerned—namely, legislators, parents, teachers, administrators, college professors, employers, and students. Although all share the goal of improved student learning, at times teachers hold quite different views from those of employers or professors about how to achieve it (Public Agenda, 2000, pp. S6-S7). Rather than having the different concerned parties taking stances in the media that often appear adversarial, it might be helpful to have frank

discussions about the reasons behind everyone’s viewpoints. Such dialogues, far broader in scope than those of the current school advisory councils, might help to militate against the impression of “top-down educational reform” and some of the ill effects of high stakes testing.

### CONCLUSION AND RECOMMENDATIONS

We must recognize, as many educators have argued (Moss, 1994; Stiggins, 1999), that the choices we make in assessment not only reflect but also determine the very nature and purpose of education itself. If we want our students’ educational experience to be broader than that measured by single standardized tests such as the FCAT umbrella (regardless of how challenging those tests may be), we must give more than token acknowledgment to the value of that education. We must refrain from the tendency to add ever more assessments with increasingly high stakes to our children’s yearly schedules in the well-intended but misguided perception that single test scores can ever reflect whether or not true learning has taken place. Assessments—both the external, summative kind of accountability and the internal, formative kind of the classroom—have a valuable *place* in education, to be sure, but educational reform should not depend on such high stakes that assessment begins to *REplace*, however inadvertently, the complexity, breadth, and depth of the educational experience itself.

To that end, the following recommendations based on the previous discussions are made:

***1. Continue with the FCAT umbrella but eliminate or minimize the high stakes attached to the score results.***

Matrix sampling could be used so that students would take some, rather than all, of the assessments and scores would be released in terms of school districts. Furthermore, reading and writing tests could be combined in a more integral way to reflect the actual processes. These steps would reduce the time spent taking tests and the dominance of the testing role that high stakes create in the classroom. Alternately, students could continue to take all of the assessments (in a census approach), but the resulting scores could **comprise part of their course grades rather than**

**being the determining factors in graduation or promotion.** At the very least, because of the newness of the FCAT in reading and mathematics, any proposed implementation of high stakes should be delayed until both schools and students have had more years of experience with the FCAT.

2. *Modify the practice of grading schools.*

The practice of grading schools seems counterproductive, but if it must be continued, the improvement of scores or the retention of previously good scores should be substantially factored into overall grades assigned the schools. Most important, resources should be allocated to schools with the greatest need.

3. *Help schools work on improving ongoing classroom assessment.*

In recognition of the importance that formative, classroom assessment plays in student learning, additional attention and resources should be devoted to the long-term improvement of formative assessment within school districts through paired-school mentorings, professional development institutes, and conferences.

4. *Engage representatives of all parties in regular, district-wide discussions about the role of assessment and its implications.*

Dialogues among teachers, business leaders, legislators, parents, and students might enable participants better to understand each other's diverse perspective on education and, hence, to reduce the adversarial nature with which assessment is often viewed—and undertaken—today.

## REFERENCES

- Bay, J. M., Reys, B. J., & Reys, R. E. (1999). The top 10 elements that must be in place to implement standards-based mathematics curricula. *Phi Delta Kappan*, 89 (7), 503-506.
- Bishop, J. (1998). *Do curriculum-based external exit exam systems enhance student achievement?* CPRE Research Report Series Report #40. Philadelphia, PA: Consortium for Policy Research in Education. ERIC Document Reproduction Service No. ED 423 256.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Boomer, G. (1985). The assessment of writing. In P. J. A. Evans (Ed.), *Directions and misdirections in English evaluation* (pp. 63-64). Ottawa, Canada: Canadian Council of Teachers of English.
- Breland, H., Camp, R., Jones, R., Morris, M. & Rock, D. (1987). *Assessing writing skill*. (Research Monograph No. 11). New York: College Board Publications.
- Brooks, M. G., & Brooks, J. G. (1999). The courage to be a constructivist. *Educational Leadership*, 57 (3), 18-24.
- Bruffee, K. A. (1986). Social construction, language arts, and the authority of knowledge: A bibliographical essay. *College English*, 48, 773-790.
- Bush plans to match CEOs with “D” schools. (2000, January 26). *The Gainesville Sun*, p. 1B.
- Carr, C. (2000, February 7). Under pressure: The countdown begins in earnest for students taking the FCAT. *The Gainesville Sun*, pp. 1, 5.

- Catterall, J. S. (1990). *A reform cooled-out: Competency tests required for high school graduation*. CSE Technical Report 320. Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing. ERIC Document Reproduction Service No. ED 338 675.
- Claggett, F. (1999). Integrating reading and writing in large-scale assessment. In C. Cooper & L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (pp. 344-365). Urbana, IL: National Council of Teachers of English.
- D'Agostino, J. (2000). Achievement testing in American schools. In T. L. Good (Ed.), *American education: Yesterday, today, and tomorrow. Ninety-ninth yearbook of the National Society for the Study of Education* (pp. 313-337). Chicago: University of Chicago Press.
- Darling-Hammond, L. (1997). Toward what end? The evaluation of student learning for the improvement of teaching. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 248-263). Thousand Oaks, CA.: Corwin Press.
- Ellwein, M. C., Glass, G. V., & Smith, M. L. (1988). Standards of competence: Propositions on the nature of testing reform. *Educational Researcher*, 17 (8), 4-9.
- Farr, R. (1999). Putting it together: Solving the reading assessment puzzle. In S.J. Barrentine (Ed.), *Reading assessment: Principles and practices for elementary teachers* (pp. 44-56). Newark, DE: International Reading Association.
- Florida Department of Education (1999a). *FCAT owners' manual: Background information for citizens about the Florida Comprehensive Assessment Test (FCAT)*. [On-line]. Available: <http://www.firm.edu/doe/sas/fcathome>.

Florida Department of Education (1999b). *Florida Writes! Report on the 1999 assessment: (Grades 4, 8, and 10)*. Tallahassee, FL: Author.

Florida Department of Education (2000). *Florida Writing Assessment Program (Florida Writes!)*. [On-line]. Available: <http://www.firn.edu/doe/sas/fsapdrct.htm>.

*The 1999 Florida Statutes*. [On-line]. Available: <http://www.leg.state.fl.us.citizen>.

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32, 365-387.

Frederiksen, N. (1994). *The influence of minimum competency tests on teaching and learning*. Princeton, NJ: Educational Testing Service. ERIC Document Reproduction Service No. ED 369 820.

Grissmer, D., & Flanagan, A. (1998, November). *Exploring rapid achievement gains in North Carolina and Texas*. National Education Goals Panel. ERIC Reproduction Document Service No. ED 425 204.

Herman, J., & Golan, S. (1993, Winter). The effect of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12, 20-25, 41-42.

Hoff, D. J. (2000, January 26). Testing's ups and downs predictable. *Education Week*, 19 (20), 1, 12.

Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999) The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81 (3), 199-203.

Kelly, T. (1999). Why state mandates don't work. *Phi Delta Kappan*, 80 (7). 543-546.

- Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools*. New York: Macmillan, The Free Press.
- Mountford, R. (1999). Let them experiment: Accommodating diverse discourse practices in large-scale writing assessment. In C. Cooper & L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (pp. 366 -396). Urbana, IL: National Council of Teachers of English.
- Moss, P. (1994) Validity in high stakes writing assessment. *Assessing Writing, 1* (10), 109-129.
- Odell, L. (1981). Defining and assessing competence in writing. In C. Cooper (Ed.), *The nature and measurement of competency in English* (pp. 95-138). Urbana, IL: National Council of Teachers of English.
- Ohanian, S. (2000). Goals 2000: What's in a name? *Phi Delta Kappan, 81* (5). 344-351.
- Paris, S. G., Calfee, R. C., Fillby, N., Hiebert, E., Pearson, P. D., Valencia, S. W., & Wolf, K. P. (1999). In S. G. Barrentine (Ed.), *Reading assessment: Principles and practices for elementary teachers* (pp. 30-43). Newark, DE: International Reading Association.
- Public Agenda (2000, February 16). Reality Check 2000. *Education Week, 19* (23), S1-S8.
- Ramirez, A. (1999). Assessment-driven reform: The emperor still has no clothes. *Phi Delta Kappan, 81* (3), 204-208.
- Rust, E., Jr. (2000, Jan. 19). Higher standards, stronger tests: There's no turning back. *Education Week, 19* (19), 40, 60.

- Sack, J. (2000, March 1). Riley urges “review” of standards. *Education Week*, 19 (25), 1, 32-33.
- Smith, M. L. (1991) Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20 (5), 8-11.
- Smith, M. S. & Levin, J. (1996). Coherence, assessment, and challenging content. In J. B. Baron & D. P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities*. Ninety-fifth yearbook of the National Society for the Study of Education (pp. 104-124). Chicago: University of Chicago Press.
- Stiggins, R. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81 (3), 191-198.
- Stotsky, S. (1995). The uses and limitations of personal or personalized writing in writing theory, research, and instruction. *Reading Research Quarterly*, 30, 758-776.
- White, E. (1985). *Teaching and assessing writing*. San Francisco: Jossey-Bass.
- Author’s note: I am grateful to Dr. Margaret Early, professor emerita at the University of Florida, and to Dr. Sharon Buckley-Van Hoek of the University of North Florida for reviewing earlier drafts of this manuscript and providing helpful insights. I also appreciate the assistance of Dr. John Gregory of the University of Florida in discussing with me the mathematics section of the FCAT.*



## **CHAPTER 4**

### **EDUCATORS AND THE A+ PLAN: VOICES FROM THE FIELD**

**Deborah Inman**  
**College of Education and Human Services**  
**University of North Florida**

Policy makers in Florida have turned to standards-based reforms and performance-based assessments to improve education by setting high standards for what students should know and be able to do, using tests to measure whether the goals are being met, and encouraging results by rewarding success and penalizing failure. However, current efforts to ensure that all students learn and achieve to high standards have resulted in controversial legislation. From curriculum reform and the development of state standards and assessments to high-stakes school accountability, traditional ways of organizing and delivering education are being challenged.

Educational accountability is not a new reform initiative in Florida (see Herrington and MacDonald in this book). In 1991, the Florida Legislature passed the School Improvement and Education Accountability Act, which was designed to hold the schools accountable for student success. Areas of reform included student readiness to start school, increased graduation rates, improved student performance, improved learning environments, increased school safety, and better trained teachers and staff. State level policy makers continued to increase their leadership role in education accountability and standards-based reform in Florida. In 1996, the Sunshine State Standards were adopted by the State Board of Education (SBE). These standards established specific expectations regarding what public school students in Florida should know and be able to accomplish. The standards were developed by the Florida Department

---

of Education in the subject areas of language arts, mathematics, science, social studies, the arts, health and physical education, and foreign languages. In an effort to align assessments with standards, the Florida Comprehensive Assessment Test (FCAT) was developed to measure student learning in the areas of reading and mathematics based on the Sunshine State Standards. The Florida Writes! (FCAT writes) exam assesses writing proficiency. Statewide administration of the FCAT began in the spring of 1998. Currently, the tests are taken by students in elementary, middle, and high school in grades 4, 5, 8, and 10. In 1999, field testing of the FCAT began in additional grades. By 2001-2002, the FCAT will be administered to students in grades 3 through 10.

Despite these efforts to improve education in the state, the challenges for improving education remained daunting. In proposing his A+ Plan, the Governor (Bush, 1999) indicated that half of Florida's fourth graders were unable to read at the fourth grade level; over one-third of Florida's ninth graders had a "D" or "F" average; and Florida's high school graduation rate was only 52% . Clearly there were still many issues to address even though Florida was moving forward with various reform efforts to strengthen education in the state. It was in this context that Governor Jeb Bush and Lieutenant Governor Frank Brogan proposed the A+ Plan for Education, subsequently adopted by the legislature in 1999. The controversial A+ Plan for Education raised the stakes regarding accountability by providing different consequences for success and failure. The legislation required all schools to be evaluated and then "graded" in an effort to improve school accountability with the ultimate goal of improving student achievement for all children. The A+ Plan currently affects 2.3 million students enrolled in kindergarten through 12<sup>th</sup> grade in 2,877 public schools in 67 public school districts in the state.

The purpose of this chapter is to provide an overview of the current issues related to the implementation of the comprehensive A+ legislation in 1999. Since state policy makers enacted this legislation and initiated assignment of grades of "A" through "F" to schools, critics and proponents of the plan have been lobbying to express their perspectives and make further recommendations. Although it is too early to document

what is working and what is not working regarding the implementation of the A+ Plan, it is important to identify and understand the intent of the legislation as well as the controversies related to the implementation.

This chapter will focus on the general intent of the A+ legislation, the initial impact of the legislation, and the controversies related to the legislation. The chapter will conclude with recommendations to optimize the benefits of the A+ Plan.

### **RESEARCH STRATEGY**

There has been much rhetoric from proponents and critics regarding the A+ Plan since its enactment in 1999. However, there has been little documentation of how educators and their educational programs have been affected by the implementation of this legislation. The intent of this research was to glean a broad understanding of the issues related to the implementation of the A+ Plan from selected individuals.

The author first interviewed twelve staff members within the Florida Department of Education to gather background information regarding the A+ Plan and to obtain their perspectives on the state role in systemic reform. To further assess the intent of the A+ Plan, the author reviewed: (1) the legislation, (2) speeches by Governor Bush to Congress as well as to the Florida Legislature, and (3) Florida Department of Education reports and rules. In order to determine the initial impact of the legislation and emerging controversies, the author: (1) reviewed over 20 news articles from major newspapers as well as local community newspapers around the state; and (2) collected information from educators using purposeful sampling techniques.

### **Procedures**

Since the intent of this research was to gain a broad understanding of the issues related to the implementation of the A+ Plan from selected individuals, the sampling procedure used to identify the respondents was purposeful sampling. The selection of sites where data were collected was deliberate to: (1) utilize professional contacts in educational settings; (2)

ensure a spread of geographic locations, and (3) to include sites that seemed particularly interesting or atypical.

The sample included 2,538 educators (663 school administrators and 1,875 teachers) at the elementary, secondary, and high school levels from across the state. Almost half (1,200) of the educators were from school districts in Northeast and Central Florida.

The data were collected from January 2000 through May 2000, primarily by written surveys. Approximately 10% (300) of the respondents were interviewed, providing them the opportunity to elaborate and to respond to follow-up questions. Professional educators helped to collect the data. The information collected provided insights regarding the initial impact of the legislation, reactions to the legislation, and emerging controversies related to the legislation.

### **Limitations**

It is important to note that the purpose of the interviews and surveys was not to generalize these findings to the state population. The sample was not random and it was not geographically representative of Florida. The intent was to attain a broad understanding from selected individuals from which additional, more rigorous and more in-depth research can be pursued. The percentage of educators surveyed for this initial investigation was very small compared to the total number of educators throughout the state. However, the sample is large enough and varied enough to provide information useful to identify initial issues and areas for further research.

### **THE INTENT OF THE A+ LEGISLATION**

The A+ Plan was developed to provide a process and a set of procedures to ensure that all students achieve to high standards. As a result of Florida's System of School Improvement and Accountability established in the early 1990s, 150 critically low achieving schools were identified and a process was developed to improve education in those schools. Florida raised the expectations in 1999 through the implementation of the A+ Plan with more challenging assessments, higher

standards and a system of grading schools “A” through “F” based on the performance of their students on state assessment tests. Grades are publicized and a system of interventions and assistance is provided for “D” and “F” schools. The components of the A+ Plan (Florida Department of Education, 2000a) include:

**Assess student learning.** Florida students in grades 3 through 10 will be tested annually on their progress. In order to more accurately assess student learning benchmarks set forth in the Sunshine State Standards, the FCAT will be extended so that all 3<sup>rd</sup> through 10<sup>th</sup> graders take the test (not just the 4<sup>th</sup>, 5<sup>th</sup>, 8<sup>th</sup>, and 10<sup>th</sup> graders).

**Raise standards by ending “social promotion.”** Students will not advance to the next grade without mastering the skills of their current grade level. By state law, no student may be assigned to a grade level based solely on age or other factors that constitute social promotion. Specifically, students with substantial reading deficiency in grades 1, 2, or 3 must receive intensive reading instruction. If the deficiency is not remedied, as indicated, by the fourth grade state assessments in reading, the student must be retained, unless a district determines to promote for good cause. High school graduation requirements include earning passing scores on the state’s High School Competency Test or the Florida Comprehensive Assessment Test.

**Grade schools.** The state ranks every public school on an “A” to “F” scale, based primarily on student learning. As part of the A+ Plan, the terminology for grading Florida’s schools changed from a “1” through “5” grading scale to an “A” through “F” grading scale to clarify for parents, teachers, and community leaders how schools are performing. School report cards are sent to parents and posted on the Internet.

**Reward high-performing and improving schools.** For 1999-2000, \$30 million was provided to reward highly successful schools. That amount will increase for 2000-2001. Schools that move up a grade and schools that receive an “A” grade receive a bonus of \$100 per student to spend at school discretion.

**Assist low-performing schools.** For 1999-2000, \$527 million was provided to assist schools with low-achieving students, particularly in low-performing schools, and that amount is expected to increase in the future. Low-performing schools designated by grades of “D” or “F” develop school improvement plans to address the particular needs of the school. The Department of Education’s Office of School Improvement coordinates technical assistance and intervention to low-performing schools. Further, local school districts provide support and assistance based on needs identified by the school improvement plans.

**Provide alternatives for students in failing schools.** If a school receives an “F” grade in two of four consecutive years, children in those schools are given alternative educational choices through the Opportunity Scholarship Program. These scholarships allow students in failing schools to attend another school of their choice—public, private or religious. In order to receive the scholarships, private schools must accept all students who apply, meet health and safety standards, and accept the value of the scholarship as full tuition. However, the vouchers tied to this plan were ruled in violation of the state constitution in March 2000 by Circuit Judge L. Ralph Smith. The case is currently under appeal.

The A+ Plan is a comprehensive effort to improve student achievement for all children by holding schools accountable. The implementation issues as described in the Florida Department of Education Rules are summarized below:

Results from the statewide assessment program form the basis of Florida’s system of school improvement and accountability. Student achievement data from the Florida Comprehensive Assessment Test (FCAT) will be used to establish both proficiency levels and annual progress for individual students, schools, districts, and the state. Results will further be used as the primary criteria to calculate school performance grades, school improvement ratings, school rewards and recognition, and performance-based

funding, and will be reported annually. Ultimately, the statewide assessment program will be used to measure the annual learning gains of each student toward achievement of the Sunshine State Standards appropriate for the student's grade level. This requires an expansion of the assessment program to include grades 3 through 10, a fundamental shift to assess annual student progress, and other comprehensive changes. Full implementation of the statewide accountability program will occur beginning with school year 2001-2002 (paraphrased from Florida Department of Education Rules, 6A-1.09981, available at <http://www.firn.edu/doe/rules/6a-11.htm>).

This rule forms the framework for the implementation of Florida's system of school improvement and accountability.

### **INITIAL IMPACT OF THE A+ PLAN**

From the state to the local level, political and social attitudes are tied to beliefs about educational practices. Policy makers are in a difficult position trying to support their constituents' demands for better schools while at the same time trying to respond to their requests for lower taxes.

Policy makers, teachers, school officials, parents, and the public are constantly faced with dilemmas about what schools should teach students. Regarding the implementation of the A+ Plan, educators are clearly caught in the middle. The teachers are placed in the most difficult position trying to meet the needs of the students yet being held accountable by the legislation. Principals, as school leaders, are faced with providing the support and environment for their teachers to produce results mandated by the A+ legislation without necessarily having the capacity or the resources. Further, they are under direct fire from district superintendents and their local community who want to have successful and publicly acclaimed high performance schools.

Schools and school districts throughout the state have responded to the increased accountability measures of the A+ legislation by

implementing various initiatives to improve student learning. High-performing schools are trying to maintain their status, while low-performing schools are trying to improve their status.

This section summarizes information collected by the Florida Department of Education (2000b) regarding the initial impact of the A+ Plan in selected school districts. The department found that some school districts are using the education funding provided by the A+ Plan to hire more teachers; focus on basic skills like reading, writing and math; increase the involvement of parents through more frequent parent teacher conferences; and increase the opportunities for tutoring and mentoring. For example, when the A+ Plan was first proposed, Broward school district identified 67 schools that needed support. They held workshops for principals, added three district level administrators to help those schools, and offered special training to teachers. Miami-Dade responded to the A+ Plan by hiring 210 additional teachers to provide extra help in the districts' 26 low-performing schools. In Duval County, low-performing elementary schools implemented reforms to improve scores and student learning through summer academic enrichment programs to raise mathematics and reading skills as well as needed professional development to change the way teachers teach. Clay and Volusia school districts hired new teachers and provided resources to support smaller classes and/or tutoring programs. Marion County school district hosted a teaching academy, which offered intensive mini-courses on teaching strategies for teachers. In Escambia County, the two failing schools that qualified for vouchers implemented after-school and Saturday tutorials, brought in additional teachers and paraprofessionals, reduced teacher/student ratios, planned to lengthen the school year from 180 to 210 days, and dedicated time blocks to reading, writing, and mathematics classes. Further, principals at those two schools were given control over the schools and the ability to hire or transfer school personnel.

The various strategies developed by these school districts in response to the A+ Plan exemplify some of the positive initiatives related to the legislation. It will be important to examine the details of

implementation and the results of those efforts as information becomes available.

Further, it is important to note that all school districts in the state have not been able to implement such extensive activities. Moreover, these examples do not convey the anxiety of educators related to the challenge to overhaul their teaching and curriculum and increase their school grade through the performance of their students on the state assessments within the school year. Many controversies have emerged with the implementation of the A+ Plan. It is important to identify the issues surrounding the major controversies and try to provide guidance and perspectives to support successful implementation of policies to encourage learning for all students.

### **CONTROVERSIES RELATED TO THE A+ PLAN**

The Bush/Brogan A+ Plan for Education is one of the most controversial education initiatives Florida has enacted. Since the implementation of the A+ Plan in 1999, the media has kept citizens throughout the state abreast of various concerns emerging within the schools.

As stated in the methodology section of this chapter, purposeful sampling techniques were used to survey and/or interview 2,538 educators to glean a broad understanding of the issues related to the implementation of the A+ legislation. The basic questions that guided the surveys and interviews were:

- How has the implementation of the A+ Plan affected you and your school?
- What are your major concerns?
- What do you see as the strengths or promise of the A+ Plan?
- What do you see as the weakness or perils?

The information collected from the interviews provided the opportunity to elaborate and follow-up on data from the written surveys.

The information collected confirms that educators have strong opinions regarding this comprehensive reform effort. The respondents related the effects of the A+ Plan directly to their concerns. The following seven issues emerged as the most noteworthy: changing criteria for grading schools; the use of tests for high-stakes accountability; the impact of high-stakes testing on teaching and learning; recognition and rewards; equity and adequacy issues; school leadership; and vouchers. These issues were consistent with many of the issues targeted by the press. Further, many of these issues were consistent with implementation of similar standards-based accountability legislation in other states. It is consequential that these seven issues emerged consistently from open-ended questions. Therefore, these seven issues are the focus of this section.

### **Changing Criteria for Grading Schools**

Over three-fourths of the educators surveyed stated one of the primary and continuing concerns regarding the A+ legislation is the changing criteria that determine the grading of schools. Since the inception of the A+ Plan, the guidelines and expectations for determining school grades have not remained constant (for example, guidelines surrounding criteria related to discipline, suspension, expulsions, special education students, and student mobility).

### **The Use of Tests for High-Stakes Accountability**

One of the major criticisms of the A+ Plan is the current grading of the schools based on student assessment by state tests. Generally most educators surveyed believe standards-based assessments are important. However, the A+ Plan, as currently implemented, grades the schools based on the performance of students in certain grades on the Florida Comprehensive Assessment Test (FCAT). Over three-fourths of the educators surveyed stated they do not believe the FCAT should be the dominant measure of a school's performance, rather they believe it is important to measure student gains on a variety of measures. Research confirms that if schools are to be graded on their contributions to a student's achievement, it is critical to measure the value the school adds if the accountability system is to be useful (Clotfelter and Ladd, 1996).

Although the heart of the A+ Plan is to measure the progress of all students from year to year effective in 2002, at this time the plan does not take into consideration student gain but only student scores on the state test. Therefore, there are many concerns regarding the use of the assessment data.

Over half of the educators surveyed stated the value added will strengthen the validity of the assessment data, but they are apprehensive because they are not certain how the measures of that performance will be determined. Almost half of the educators interviewed were concerned about how the benchmarks for student progress will be established and expressed apprehension as to whether FCAT scores will be the sole source of gains.

Other issues regarding the use of assessments for accountability are related to how much improvement schools can reasonably be expected to achieve in a given time period and how such expectations can be determined. The National Research Council of the National Academy of Sciences suggests that more research is needed to better determine a fair measure of what can be expected to be learned in a given time frame (see Elmore and Rothman, 1999).

Research documents that when appropriately developed and used, content and performance standards can be a basis for high-quality performance assessments that provide a broad range of information about how students and schools are doing (Darling-Hammond and Falk, 1997). The information obtained can be very useful to inform teaching decisions, to identify needed resources for student learning, and to evaluate school practices. While almost all of those interviewed agreed with these positive uses of assessment information, almost half expressed concern that in Florida, the initial emphasis seems to be on the use of the data to determine rewards and sanctions.

More than three-fourths of the educators interviewed stated that the system will remain flawed until the state grades schools on the improvement of individual students from year to year. Moreover, they

stated they were even more concerned that the current assessments may provide misinformation about the quality of schools resulting in inappropriate consequences for schools and students.

### **Impact of High-Stakes Testing on Teaching and Learning**

One of the greatest concerns regarding the A+ Plan is the impact of school grades on teachers and students. The perspectives of teachers and principals surveyed regarding the impact of the A+ Plan on student learning were surprisingly consistent. Generally, most educators interviewed agreed with the intent to provide optimum learning experiences for all children. However, over three-fourths of those same educators stated they believe the legislation as currently implemented will further increase inequities among schools. Over half of the teachers surveyed, especially those with students taking the FCAT, stated they feel as if all they do is focus on FCAT skill preparation. Over one-third of the educators interviewed stated they fear the students' best interests may no longer be at the heart of the education program. They expressed concern that priorities have been redirected to the school grade rather than each child's educational experience.

Over three-fourths of the educators surveyed reported that the legislation enacting the A+ Plan has had a profound effect on the content and delivery of curriculum and on teacher stress and morale. Over one-half of those interviewed stated that many teachers are intimidated by the test and the school grading system. They also reported that most school staff meetings between August and March were focused on the tests and what the results would mean to the school. They further stated that the legislation has resulted in teachers who teach the FCAT as a separate piece of their usual curriculum. While this is not the intent of the A+ Plan, these respondents stated that this occurs because teachers have not been appropriately oriented to the increased expectations, nor have they received the needed professional development targeted to teaching standards-based curriculum. Those teachers perceive the FCAT as one more thing they have to do, rather than integrating it into their total curriculum. In addition, over half of all teachers surveyed indicated a need for additional professional development targeted to teaching curriculum

related to the state standards, so they can be confident their students will do well on the FCAT without having to devote instructional time to teaching to the test, doing FCAT related worksheets, and coaching to the test.

There is increasing consensus that investing in teacher professional development can increase student achievement (National Commission on Teaching and America's Future, 1996; Sanders and Rivers, 1996). Darling-Hammond (1999) reported that states investing heavily in a teacher capacity building strategy over the past decade had substantial gains in students' mathematics scores for grades 4 and 8 on the National Assessment of Educational Progress (NAEP).

One of the most noted concerns regarding the impact of the A+ Plan on teaching and learning was directly related to the effect of school grades on teacher and student morale. More than three-fourths of all educators surveyed reported that when a school receives a grade of "A," teacher spirits and student confidence soar; however, when low-performing schools receive a grade of a "D" or an "F," the teachers and students feel they have failed. Further, over three-fourths of those interviewed stated since many of the students who attend the low-performing schools come from homes in low socioeconomic environments, their schools may be the only success they experience. When those schools are deemed failures, the students' expectations and optimism for their own future may be diminished.

### **Recognition and Rewards**

Further controversies regarding the A+ Plan emerge among teachers when recognition, rewards, and bonuses are attached to these grades. Rewards for high-performing schools in Florida can take the form of public recognition and cash bonuses for the schools or for the teachers and staff within the schools. The pressure of a school grade with rewards based on FCAT scores could tempt some to do whatever it takes to get the grade. For example, teachers in two counties in Florida are facing possible disciplinary action because of charges related to cheating on the March 2000 FCAT. Cheating on state tests has been found in other states as well,

including Kentucky and Texas (Sacks, 1999). Further, the financial bonuses provide good teachers at low-performing schools incentives to transfer to a school where they are more likely to win an award.

When asked about their concerns regarding the A+ Plan, over half of the educators surveyed identified recognition and rewards as an issue. Further, more than half of those interviewed stated they felt the financial rewards created divisiveness among teachers and administrators within as well as among schools. An example cited by many related to how to spend the bonus money. Many teachers wanted to split the money equally among all staff but were divided as to whether it should be last year's staff who earned the grade or the current staff who needed the incentive to keep up the good work. Others did not want new staff members to get anything, while some wanted a partial share for new staff.

Over one-third of those interviewed stated they believed the competitive nature of the grading system has destroyed much of the collegiality that formerly existed among schools. The schools now view one another as competition for the needed funds. While many of these same respondents agree with the concept of the A+ Plan, they do not support the competitive nature of the plan. They believe education is a field that needs *cooperation* among faculty, schools, the districts and the state.

### **Equity and Adequacy Issues Related to the A+ Plan**

When asked about the promise and perils of the A+ Plan, equity and adequacy issues emerged for both the promise and the perils. More than half of the educators who were interviewed agreed the promise of the A+ Plan, over time and with appropriate implementation, is that educational opportunities could be greatly improved. Schools that need help are being identified and are receiving the support they need to improve education in their schools. However, these same respondents stated that the peril of the A+ Plan is that as currently implemented the plan will further increase inequities among the schools.

Over three-fourths of the educators surveyed stated the A+ Plan is inequitable because it rewards schools located in upper economic areas while penalizing those in lower income areas. For example in 1999, of the elementary schools where more than 80% of students were poor enough to qualify for a free or reduced-price lunch, 93% received either a “D” or “F” grade. None of these schools earned an “A” or “B.” In contrast, in elementary schools where fewer than 20% of students are considered poor, 83% received an “A” or “B.” None received a “D” or “F.” Specifically, of the 61,000 students in “F” schools, 85% were minorities: 63% (38,430) were African-American children, and 20% (12,200) were Hispanic children. Furthermore, of all students attending Florida’s “F” schools, 81% participate in the free and reduced-price school lunch program (Florida Department of Education, 2000b).

It is important to hold all students accountable to the same standards. However, if students and schools are going to be graded and compared on the same standards, then all schools must have at minimum the same resources. The basic argument is that if tests have high stakes or consequences, the education system must ensure that all students have equal opportunity to learn the curriculum (Odden and Odden, 1995). This means that the same curriculum must be taught to all students and taught by teachers with the expertise to teach it well. Resource inequities further complicate these expectations and are especially problematic when schools serving students with greater needs may require additional resources. More than three-fourths of all educators interviewed stated they are concerned that the A+ Plan is not fair in this regard. When all resources are not equitably and adequately available to all students to support their learning, it is difficult to implement high-stakes testing and unfair to compare results related to performance and productivity.

### **School Leadership**

Over three-fourths of the educators interviewed reported that school leadership has evolved as even more critical during the transition to the A+ Plan. Specifically, principals with strong leadership skills, who know how to build capacity, motivate teachers and students, and who can communicate with parents the responsibilities regarding their child’s

achievement and who can increase school spirit and support throughout the community are instrumental to successful implementation of the A+ Plan. When school leaders were asked specifically about teacher and student morale, they expressed concern about the direction school spirit and morale may take if, after dedicated effort on the part of all players, the school still does not receive the expected grade. Further, school leaders interviewed—who believed they had the ability and support to increase teacher skills, build teacher capacity, raise and utilize new resources or re-prioritize existing resources to improve student performance within the school year—stated that they were anxious to see if the criteria used by the state will reflect the efforts and accomplishments of teachers and students. These school leaders were determined that students in their schools would not qualify for vouchers.

### **The A+ Plan and Vouchers**

When the A+ Plan was first introduced by Governor Bush, one of the most controversial aspects of the plan was the legislation that provides vouchers to students at chronically failing schools. While educators, policy makers and citizens throughout the state have voiced concerns and taken positions regarding the use of public funds to pay for private education, their counterparts around the nation have been watching closely to see what happens. Florida's voucher program was the first statewide plan in the nation. Students at schools that earn failing grades from the state two years out of four were eligible for vouchers that could be used at public, private, and religious schools. Immediate concerns emerged from educators and parents regarding the use of public school funds to pay for private education and the unknown quality of non-graded private and religious schools, which were not held accountable to the same standards and measures (school grades) as the public schools.

The biggest issue for state policy makers was whether appropriating public funds that could be used for tuition at religious schools was constitutional. As the legal issues emerged, the plaintiffs argued that the 1999 law violates the state constitution by using tax dollars to pay for private school tuition. On March 14, 2000, Circuit Judge L. Ralph Smith ruled that the plan, in fact, does violate the Florida

Constitution because tax dollars cannot be used to send children to private schools. Governor Bush appealed the ruling. The case will be heard by the First District Court of Appeal in Tallahassee but could automatically be sent to the State Supreme Court.

Florida's decision to link its voucher plan to the performance of public schools shifted the equity argument from one linked to family poverty to one tied to school performance. Florida was one of the first states to look at plans that base students' eligibility for vouchers on how their public schools perform on state accountability measures. Prior to Florida's plan, vouchers had not been viewed as an accountability tool. Vouchers have become less a means of helping low-income students achieve educational choice and more a sanction aimed at turning around failing schools. However, a recent study (Fiske and Ladd, 2000) found that while voucher plans may or may not be good policy for the poor, they should not be viewed as a substitute for programs to assist struggling schools.

### **SUMMARY**

Policy makers in Florida have turned to standards-based reforms and performance-based assessments to improve education by: (1) setting high standards for what students should know and be able to do, (2) using tests to measure whether the goals are being met, and (3) encouraging results by rewarding success and penalizing failure.

As legislation related to standards-based accountability is implemented, controversies arise based on disagreements with the concept of standards-based reform and assessments and/or concerns regarding the criteria related to implementation. While educators generally support standards based reform and assessments, the problems emerge when the support to fully implement and build the capacity needed for successful standards-based reform is not in place before high-stakes accountability tied to those standards is implemented. Further concerns relate to how the assessment scores are used. The current assessments may provide misinformation about the quality of schools resulting in inappropriate consequences for schools and students.

## RECOMMENDATIONS

Depending on how standards and assessments are developed and used, they could support more ambitious teaching and greater levels of success for all students, or they could serve to create higher rates of failure for those who are already least well-served by the educational system. Therefore, it is important to consider carefully both the substance of standards and assessments and their uses. These issues are especially relevant to Florida policy makers and educators at this moment in time. Because of the comprehensive nature of the A+ Plan and the potential if properly developed and implemented to improve education for all children, the careful deliberation of these issues by educators and policy makers is imperative. Further, as with any state initiated reform, it is important to build an infrastructure at the state level to ensure continued support of the reform efforts as state leadership changes.

Because policy makers are so far removed from the classroom, governors, legislators, school boards and other policy makers have always found it difficult to know which schools need help. Although educators across the state do not agree with the idea of grading schools, some policy makers have found the grading system to be useful in identifying the schools that need help. However, many educators are concerned that the current criteria for grading schools and the current implementation of the A+ Plan may result in inaccurate information and inappropriate consequences for schools and students. The following recommendations are offered to provide the optimum level of success in reaping the benefits of the A+ Plan.

1. ***Release grading criteria at the beginning of each school year and keep them constant*** through the end of the school year.
2. ***The FCAT should be administered as late as possible in the school year*** to allow maximum instructional time for students to learn and to provide a better assessment of what is taught during the entire school year.

3. ***Reduce the FCAT's influence on school grades*** as soon as possible. The FCAT should not be the dominant factor in determining the school grade. The plan should focus on student gains on a variety of measures. Students' progress toward reaching the benchmarks should be evaluated continuously with in-class assessments, including documentation and student work samples. Although many of these issues are expected to be addressed and implemented in 2002, the current use of FCAT to determine grades can result in misinformation now about the quality of learning in schools and in inappropriate consequences for schools and students.
4. ***Resolve inequities***. Until the inequities are resolved, the argument remains that high-stakes tests are unfair. In order to implement state curriculum and testing standards and tie performance to consequences, all students must be assured at minimum adequate and equivalent resources with the same high quality curriculum taught by equally skilled teachers.
5. ***Recognize schools that are making progress***. The current grading system does not account for the schools' gains when student improvement does not raise the school grade. Morale drops for students and teachers when low-performing schools can document greater gains in student achievement than high-performing schools but receive no recognition or reward bonuses for having accomplished more.
6. ***Strengthen the teaching profession*** by providing ongoing professional development, competitive salaries, and benefits for continuing education. Specifically, strengthen classroom skills of teachers and build teacher capacity. Provide mentor teachers or teacher leaders to help current teachers in the classroom through professional development with a focus on instruction techniques. Provide incentives for highly qualified teachers to serve in low-performing schools. Align teacher evaluations with standards.

7. ***Ask why outcomes occur and make necessary changes in the conditions that influence teaching and learning.*** Specifically, policy makers need to establish the capacity at the state level to undertake analyses of student outcomes and to use the knowledge gained to develop further policies to improve the conditions that positively influence teaching and learning.
8. ***Conduct in-depth ethnographic and case study research*** to document the implementation of the A+ Plan over the next three to five years.

## REFERENCES

- Bush, J. (1999, September). *Testimony of The Honorable Jeb Bush, Governor of Florida to the U.S. House Committee on the Budget, U.S. House of Representatives, Washington, DC [On-line]*. Available: <http://www.fl.us/eog>.
- Clofelter, C., & Ladd, H. F. (1996). Recognizing and rewarding success in public schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance based reform in education* (pp. 23-64). Washington, DC: The Brookings Institution.
- Darling-Hammond, L., & Falk, B. (1997). Using standards and assessments to support student learning, *Phi Delta Kappan*, 79 (3), 190-199.
- Darling-Hammond, L. (1999). *Supply, demand, and quality in mathematics and science teaching*. Briefing for the National Commission on Mathematics and Science Teaching for the 21<sup>st</sup> Century, Washington, D.C. New York: National Commission on Teaching and America's Future.
- Elmore, R. F., & Rothman, R. (Eds.). (1999). *Testing, teaching and learning: A guide for state and school districts*. Washington, DC: National Academy Press.
- Fiske, E., & Ladd, H. F. (2000). *When schools compete*. Washington, DC: The Brookings Institution.
- Florida Department of Education. (2000a), *Elements of the A+ plan*. Tallahassee, FL: Author. [On-line]. Available: <http://www.flgov.com/eog/aplusplan/index.html>.
- Florida Department of Education. (2000b). *Positive activities and changes in schools statewide*. Tallahassee, FL: Author. [On-line]. Available: <http://www.firn.edu/doe>.

*Florida Department of Education Rules, 6A-1.09981. [On-line].*  
Available: <http://www.firn.edu/doe/rules/6a-11.htm>.

National Commission on Teaching and America's Future. (1996). *Doing what matters most: Investing in quality teaching*. New York: Author.

Odden, A., & Odden, E. (1995). *Educational leadership for America's schools*. New York: McGraw-Hill.

Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Books.

Sanders, W. J., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Research Progress Report. Knoxville: University of Tennessee Value-Added Research and Assessment Center.