THE FLORIDA STATE UNIVERSITY

K-12 School Accountability

An Analysis of Options to Measuring Accountability in Florida’s School System

AN ACTION REPORT SUBMITTED TO THE FACULTY OF THE COLLEGE
OF SOCIAL SCIENCE IN CANDIDACY FOR THE DEGREE OF
MASTER OF PUBLIC ADMINISTRATION

REUBIN O’D. ASKEW
SCHOOL OF PUBLIC ADMINISTRATION AND POLICY

BY

TONY D. JOHNSON

Tallahassee, FL
May 2003
Dear Mr. Horne:

It gives me great honor to submit to you *K-12 School Accountability: An Analysis of Options to Measuring Accountability in Florida’s School System*. The report is the product of extensive research and analysis of the fall of 2002 and the spring of 2003. School accountability is a very important policy area for students, parents, professional groups, and legislators. Accountability in Florida has yielded criticism over the design and implementation of a desirable accountability program. Florida has a heterogeneous population of students, teachers, and parents; because of this diversity it is important that Florida has a comprehensive assessment tool in place to gauge student progress as they matriculate through our public schools.

My recommendation is that Florida should consider concentrating on and implementing a more equitable means of assessing public schools. The FCAT is a good measure of student achievement, although it should not be used as a “high-stakes” exam. The morale of teachers and students has decreased since the introduction of this assessment tool. Although the FCAT has positive attributes there are also areas of discontinuity. This policy is recommended based on the use of three evaluative criteria: political desirability, administrative feasibility, and program costs. It is my recommendation that this report be used to change the current implementation of the FCAT exam from a “high-stakes” exam to one that demonstrates educational progress.

Respectfully,

*Tony D. Johnson*

Tony D. Johnson
Tallahassee Community College Instructor
Florida State University
Candidate for the Master’s in Public Administration & Policy
TABLE OF CONTENTS

LETTER OF TRANSMITTAL ................................................................. ii
LIST OF TABLES ........................................................................... iii
EXECUTIVE SUMMARY ................................................................. iv

CHAPTER

I. PROBLEM STATEMENT ............................................................... 1

II. BACKGROUND AND LITERATURE REVIEW ......................... 4

   Background ............................................................................. 4
   Literature Review .................................................................... 10

III. METHODOLOGY AND EVALUATION CRITERIA ..................... 15

   Methodology .......................................................................... 15
   Evaluation Criteria ............................................................... 15

IV. MANAGEMENT POLICY OPTIONS ....................................... 17

   Performance-Based Assessment ............................................ 17
   FCAT Exam Implementation .................................................. 24
   Standards Assessment .......................................................... 29

V. CONCLUSIONS ......................................................................... 35

REFERENCES ............................................................................... 38

ABOUT THE AUTHOR .................................................................. 42
List of Tables

Figure 1: The Development of the Florida Comprehensive Assessment Test……………8

Figure 2: Descriptions of Alternative Assessment Models……………………………..11

Figure 3: The 2002 Florida Statutes that support or prohibit performance-based Assessment……………………………………………………………………21

Figure 4: Hypothetical Program Cost Correlation Between a Performance-Based Assessment and a Standardized Assessment…………………………………..23

Figure 5: The 2002 Florida Statutes that support or prohibit the alternative use of the FCAT……………………………………………………………………..28

Figure 6: The 2002 Florida Statutes that support or prohibit standards-based assessment……………………………………………………………..33

Figure 7: Different school assessments that are administered in Florida’s public schools……………………………………………………………………..34

Figure 8: Summary of Alternatives & Evaluation Criteria……………………………………..35
Executive Summary

January 8, 2002 serves as the pinnacle of American pedagogical education. This day marked the inception of the “No Child Left Behind” legislation, which would soon serve as a blueprint for the adoption of statewide assessment methods. This legislation challenged each state to develop a rigorous curriculum and assessment tool that will demonstrate increased: accountability; choices for parents and students, flexibility for states, school districts, schools; and priority on reading. The Florida Department of Education developed a norm-referenced assessment that would be later coined the Florida Comprehensive Assessment Test (FCAT). This action research problem addresses the design, implementation, and alternatives available in K-12 assessment.

Met with support and opposition, the FCAT is the new assessment tool that is used to gauge the progress of Florida’s schools. Public opinion polls suggest that Floridians demand an accountability system that will improve the quality of K-12 education. Since Ronald Reagan’s report, A Nation at Risk of 1983, many reports have outlined the problems in American education. Since that time, and the inception of the FCAT exam, there is a new pressure in public education: this new pressure has added an increased accountability for schools. The result is a decreased morale of teachers and students, and political pressures from other states’ successful assessment models may cause Florida to consider alternative assessment models.

Information for this report was gathered by three methods of inquiry. Initially, popular media reports, public opinion polls, unstructured telephone interviews, and surveys were collected and analyzed. Secondly, federal and state reports were analyzed. Lastly, academic literature was consulted and analyzed.
This report presents three policy options: performance-based assessment, the FCAT exam’s implementation, and standards-based assessment. Each policy option was evaluated against three criteria: political desirability, administrative feasibility, and program costs.

The alternatives as measured against the criteria yielded a recommendation of the continued implementation of the FCAT, but with modifications. The FCAT is the most viable means of student assessment because it is written into statutes, norm-referenced, statistically sound, and focuses the efforts of teachers on particular content material.

The performance-based assessment faired very low on each of the criteria’s evaluation instruments. Performance-based assessment is generally not supported because it lacks both quantifiable data elements and implementation by many states. The standards-based assessment faired well as a viable option on each evaluation stratum. There is a higher level of administrative feasibility and lower program cost of the FCAT in comparison to standards-based assessment model.

Public opinion polls have shown that the “high-stakes” component of the FCAT divides districts, faculty, and students. This negative outcome of standardized assessment may open a political window that may allow for an alternative means of assessment. The FCAT serves as a viable option for student assessment and its continued implementation will serve as an adequate measure of student achievement.
I. Problem Statement

In 1999 a teacher had this to say about the Texas Standardized Examination:

“...I think the TAAS (Texas Assessment of Academic Skills) takes up pretty much the day, and I think as teachers we get bogged down with those worksheets and don’t come up with other creative ways to implement the objectives that they test on in TAAS. So I think we’re very worksheet oriented because I think when children do get that test booklet, it won’t be in the form of a game, it won’t be in the form of a project. But it would be in paper/pencil test” (Berends, 2002, p.102).

As evidenced by this Texan educator, accountability in American public school education is a highly debated public policy issue. All stakeholders in public education typically embrace the need for a system of accountability. The problem arises from the lack of a shared vision of the design and implementation of assessment systems. Teachers and administrators alike agree that high-stakes testing shifts the priority of what substantive material that will be taught in classrooms. High-stakes testing, as school accountability systems are often characterized, are linked to financial, employment, and promotion and retention backlashes. Yet the justification that undergirds accountability systems is aimed to improve teacher quality and student achievement.

In 2001 a public opinion poll asked several questions about accountability systems in public education. This poll found the following outcomes for parents and teachers 53% and 55%, respectively, agreed that the U.S. Department of Education (DOE) should continue accountability measures; 42% and 55% believed the U.S. DOE’s efforts of school accountability should continue with adjustments; 1% and 2% believed that accountability efforts should go back to the way they were designed previously [eliminate high-stakes testing] (USDOE, 2001). This era of whole-school reform and increased global competition led the federal government to enact legislation aimed at demanding schools to be more accountable.
On January 2, 2002, President George W. Bush signed into law the *No Child Left Behind Act of 2001 (NCLB)*. This legislation fostered four emergent themes which include: stronger accountability for results, record of flexibility for states and communities, concentrating resources on proven education methods, and school choice for parents. Florida’s call to meet this new legislative requirement was enacted in the form of an assessment plan commonly known as the A-Plus Plan. This plan assigns school grades based on the performance of students on the Florida Comprehensive Assessment Test (FCAT).

This exam tests students in the areas of reading, writing and mathematics. The FCAT is administered in grades 4th, 8th, and 10th. The grade that each school receives is based upon the percentage of students scoring above the thresholds established by these five categories or levels (Greene, 2001). If a school receives two F grades in a four-year period, students are offered vouchers to attend a private school or a better performing public school (Greene, 2001). This exam has been debated in many circles of academia as it was introduced as the new assessment tool for K-12 school accountability.

Why is school accountability so important and who are the major stakeholders? Hamilton et al. (2002) provide a plausible answer to who the major stakeholders are and why school accountability is such an important concern:

…It is increasingly clear to business leaders, however, that the public education system will not respond to such calls for reform in the absence of pressure to do so… In public schools, educators and students have faced few consequences for their failures and even fewer rewards for their successes. The National Alliance of Business believes that introducing such consequences and rewards into public education is essential to raising student achievement and spurring schools to continually improve (Hamilton, 2002, p.7).

As evidenced here the business community is extremely concerned with the accountability of public education. In Florida, this accountability is measured via school
grading. School grading is the primary method that Florida has adopted to gauge the progress of students and the overall progress of schools. Students are then grouped in broader categories-based on their achievement level. Recent studies in Florida, North Carolina, and Texas suggest that schools in the lowest category on the state’s performance scale (and therefore subject to sanctions) tend to show more improvement than those in the second-to-lowest category (Hamilton, 2002). School grading does not show incremental improvement in student learning, but instead gives a very broad, inadequate view of student learning.

Proponents of school grading agree that it is a very viable plan because it allows all stakeholders an opportunity to see verifiable evidence of the school/student’s records of achievement. Moreover, proponents suggest, “standardized assessments cause students to devote a greater effort to their schooling and that they also orient schools to the goals of education that was adopted at the state level” (Orfield & Kornhaber, 2001, p. 9). Conversely, opponents argue that school grading does not demonstrate an accurate level of student or school achievement. Moreover, critics of high-stakes testing suggest that there is a practice legacy that exists (Dorn, 1998). This practice legacy suggests that high-stakes testing creates inappropriate incentives against good teaching. Additionally, opponents argue that such policy and the overall standards movement is based upon faculty assumptions (Orfield & Kornhaber, 2001).

Are there other alternatives to school grading? If so why has Florida adopted the A-Plus Accountability school-grading model? The purpose of this Action Report is to examine alternative solutions to school grading that will provide an accurate, effective, and efficient measure of student achievement in Florida.
II. Background & Literature Review

**Background**

There are four historical developments that should be examined to understand the background of school accountability: school reform, the National Assessment for Educational Progress, the George W. Bush’s school accountability policy agenda, and history of student assessment in Florida. Each development offers substantive background about K-12 school accountability.

There is an abundance of literature that discusses the role of accountability in public school systems. First, in the 1970’s America experienced a growth in the emergence of the minimum competency requirements in schools (Orfield & Kornhaber, 2001). Yet, the problem with school accountability was not prioritized in public policy initiatives until 1983 when the Ronald Reagan Administration published *A Nation at Risk*. This publication attacked unqualified teachers, broken family structure, social promotion, and general academic permissiveness (Meir, 2000).

Moreover, this publication asserted recommendations for school reform, a more rigorous curriculum, and more direction from experts who know the demands of our culture and the new economy (Meir, 2000). Now with the focus on standards, states began to re-write or develop standards for every major subject. This new focus on standards was an attempt to solve the low level of academic achievement of young people (Stotsky, 2000). The standards wars continued to escalate from the 1990’s until the present. States are competing to develop curricula and assessments that demonstrate a steady longitudinal improvement in student achievement. How can states compare student progress on a state-by-state basis?
Second, the National Assessment for Educational Progress (NAEP) is the only national exam that assesses student educational improvement and can be used to compare student achievement levels on a state-by-state basis. NAEP has changed over time since its developmental inception by the 1963 Commissioner of Education Francis Keppel. As the standards movement progressed some twenty-years later; it would be this psychometric exam that would assess educational progress (National Research Council, 1999). NAEP changed its design as new policies were introduced in secondary schools. In 1984 an attempt was made to design the NAEP exam so that it could assess and monitor learners from disadvantaged backgrounds, linguistically challenged, and students with disabilities.

Third, this national wave of improved assessment models continued in the George W. Bush Administration. This administration continued the school accountability initiatives for higher standards and increased use of assessment, which was initially introduced by the Clinton-Gore agenda. This accountability measure was coined as the Goals 2000 Act; it was premised on three guiding principles: high standards for all children, new accountability measures linked to high standards, and new investments to improve the quality of education for all children (Riley, 2002).

As the administration changed and President George W. Bush began his term, he initiated what he called a World-Class U.S. Education System. Each new administration in the past decade has initiated an educational reform effort to bolster the global competitiveness of American children. Henceforth, the political issues that are involved in assessment further complicates the problem of the design, implementation, and deciding amongst alternatives (Riley, 2002; National Research Council, 1999). The
Standard Movement will continue because of the public outcry for performance results from public schools. School accountability is not an isolated geographical policy priority, but instead it is a national concern. Currently, 48 states test their students, 36 publish annual report cards on individual schools, 19 states provide public rating of their schools, 14 provide monetary incentives for good performance and 49 states have set standards for students, and 15 states test students every year in grades 3 through 8 (Olson 1999; Fletcher 2001).

Fourth, national standards were designed and implemented to improve student achievement in the 1990’s. Surprisingly, Florida began student assessment two decades before many other states. In 1969 the Florida Commissioner of Education recommended to the Legislature that it, by statute, authorize him to develop a plan for evaluating the effectiveness of educational programs. In response the Florida Legislature enacted Chapter 70-399 in 1970 (Florida Department of Education, 2003). The Commissioner of Education was directed to design evaluation procedures. The goal was to provide each school district with the relevant comparative data and to the extent possible, be compatible with the National Assessment of Educational Progress (NAEP). The 1971 Legislature adopted the Commissioner's Plan for Educational Assessment in Florida, enacting the Educational Accountability Act; Section 229.57,of the Florida Statutes (Florida Department of Education, 2003, Educational Accountability Act, para.4).

As a result of the Act, Florida's first statewide assessment was administered in reading from 1971-72. The Research and Development Section (RDS) of the Florida Department of Education contracted with the Center for the Study of Education (CSE), University of California at Los Angeles, to supply reading objectives and test items for
second and forth grades (Florida Department of Education, 2003, Florida’s First Statewide Assessment, para.1). Committees, totaling 112 reading specialists and 236 classroom teachers from all school districts, were asked to confirm the grade level at which each objective should be achieved. Since there were too many objectives to assess, a reading consultant reduced the list to a more practical size (Florida Department of Education, 2003).

Then from 1973-1975 Florida administered its third statewide assessment program that focused on reading, writing, and mathematics objectives from the previous year. Later in the winter of 1975 all students in grades three, six, and nine were assessed in reading, writing, and mathematics with the exception of students with disabilities. That year's assessment was unique in that it marked the first time that Florida tested all students instead of using a random sample (Florida Department of Education, 2003).

“In 1976 an amendment was made to the Florida Accountability Act. While the 1971 and 1974 versions of the Act were concerned primarily with the Statewide Assessment Program, the 1976 version consolidated and expanded existing statutes on all aspects of accountability. Sections of the 1976 law related to the management information systems, evaluation procedures, the student testing program, basic skills development, pupil progression, district planning, and cost accounting” (Florida Department of Education, 2003).

“The Florida Comprehensive Assessment Test (FCAT) is the latest version of Florida’s statewide assessment program. The assessment program was initiated about 1972 and has gone through many changes over the years. Originally, the assessment program was based on measuring only a sample of students, but this was quickly changed
to include all students in selected grade levels. The first series of tests measured students’
acquisition of certain minimum competency skills and the program, generally, was called
a “minimum competency testing program” (Florida Department of Education, 2003, A
Chronology of Events, para.1).

Figure 1 describes the creation and the implementation of the FCAT exam from
the 1970’s wave of assessment models. Figure 1 illustrates that the FCAT is not a new
assessment model that was created for political reasons. But instead, it is an assessment
that has been in its infancy stages for three decades.

Figure 1: The Development of the Florida Comprehensive Assessment Test (FCAT)

Curriculum, Instruction & Assessment

Assessment & Evaluation

The Florida Commission on Education Reform and Accountability began conceptualizing the
FCAT well before the first test was administered in 1998. In 1995, the Commission
recommended procedures for assessing student learning in Florida that would raise educational
expectations for students and help them compete for jobs in the global marketplace. The State
Board of Education adopted the recommendations, called the Comprehensive Assessment
Design, in June 1995. The Design specified the development of new statewide assessments
to address four broad areas described in the first four standards of Goal 3 of Blueprint 2000. These
four areas have been generally referred to as reading, writing, mathematics, and creative and
critical thinking.

In addition, the Design required that educational content standards be developed and adopted. Subsequently, the Florida curriculum frameworks, also called the Sunshine State Standards,
were developed and adopted by the State Board of Education. The frameworks and standards
established guidelines for a statewide system that incorporated assessment, accountability, and
inservice training components of the Sunshine State Standards and does so in the context of
real-world applications. Initially, the FCAT was designed to assess reading, writing, and
mathematics at four grade levels so that each subject was assessed at all levels of schooling:
elementary, middle, and high. With legislative approval of Governor Bush’s A+ Plan in 1999,
the FCAT was expanded to include grades 3-10. In 2001, achievement for all grade levels will
be reported for the first time. The FCAT will become the test required for high school
graduation for the class of 2003.

Note. Adapted from the Florida Department of Education (2003). Retrieved March 16, 2003 from
In summary, these four historical developments and it provides a foundation for analyzing school accountability and its linkage to the Florida school accountability plan. Firstly, the national school reform effort promotes the national concern for teacher quality and student achievement. Secondly, the NAEP examination also provides a mechanism for comparing student achievement on a state-by-state basis. Thirdly, the policy agenda of the Bush Administration demonstrates the federal role in encouraging higher performing schools and accountability from states. Fourthly, Florida has kept pace with the national trends in developing K-12 school assessment tools and re-evaluating multiple methods of improving student assessment. Florida’s current assessment program was created because of a larger national accountability development. It is therefore important that it adopt a reliable and valid assessment to measure student achievement.
Literature Review

The body of literature addresses the problems that can arise when states move towards more stringent methods of imposing accountability. Systems may become corrupt, and morale can decrease as the state shifts from teaching and learning to inputs and outputs of standardized assessments (Dorn, 1998; Orfield & Kornhaber, 2001; Stotsky, 2000). All states have adopted a method of assessing student achievement. The pertinent literature on this topic addresses three primary methods that are used nationally to assess student achievement: performance-based assessment, standardized assessment, and standards-based assessment.

First, Khattri, Reeve & Kane (1998) posit that performance assessment is used for five primary reasons: (1) to inform and influence instruction and curriculum (2) to monitor student progress (3) for accountability (4) to align assessment with instruction and curriculum (5) to certify student achievement; graduation requirement). Moreover, proponents of performance assessment assert that this form of assessment serves three functions:

- Helps teachers and other educators conduct a comprehensive evaluation of students’ achievement, including student’s strengths and weaknesses.
- Support instruction and curriculum aimed at teaching for understanding by providing good pedagogical templates.
- Help teachers and other educators better assess students’ understanding of procedural knowledge, which is not so easily judged through traditional assessment methods (Khattri, Reeve & Kane, 1998; Federal OERI, 1993; Arter, 1995).
Linda Mabry (1999) describes the alternative performance assessment models that are available to K-12 schools. Figure 2 provides a brief synopsis of these alternative assessment models.

**Figure 2: Descriptions of Alternative Assessment Methods**

<table>
<thead>
<tr>
<th>Alternative Assessment</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio Assessment</td>
<td>A collection of information by and about a student to give a broad view of his or her achievement. It contains samples of the student’s work, grades, evaluations, and narrative descriptions.</td>
</tr>
<tr>
<td>Profile</td>
<td>A collection of ratings, descriptions, and summary judgments by teachers and sometimes by the student and others to give a broad view of his her achievement. A profile contains all contents of a portfolio except that of the student’s work.</td>
</tr>
<tr>
<td>Project</td>
<td>A specialized, often interdisciplinary inquiry devised and undertaken by a student or group of students.</td>
</tr>
<tr>
<td>Demonstration or Exhibition of Mastery</td>
<td>Often a formal, more or less public performance of student competence and skill that provides an opportunity for summative or final assessment.</td>
</tr>
<tr>
<td>Discourse Assessment</td>
<td>Evaluation of what a student tells about what he or she knows. Typically talking to with the assessor, the student indicates what he or she has learned, offering evidence of critical thinking or problems solving by producing narratives, arguments, explanations, original summaries, interpretations, analyses, or evaluations.</td>
</tr>
<tr>
<td>Simulation</td>
<td>A task designed to incorporate problem-solving features similar to those found in practical or professional contexts.</td>
</tr>
<tr>
<td>Performance Task</td>
<td>A task problem, or question that requires students to construct responses and may also require them to devise and revise strategies, organize data, identify patterns, formulate models and generalizations, evaluate partial and tentative solutions, and justify their answers.</td>
</tr>
</tbody>
</table>

*Note.* Adapted from Mabry, 1999, p.17

Qualitative performance assessment measures are generally scrutinized more often than quantitative measures. Both assessment methods are evaluated by several criteria prior to implementation. One such criterion is reliability; it is important in portfolio assessment because it relies on the consensus of the assessors to apply the same criteria and standards as uniformly as possible (Khattri, Reeve & Kane, 1998; Mabry,
1999; The Federal OERI, 1993). It seems evident that reliability may seem flawed by assessors because portfolio assessment allows for more subjectivity than standardized assessment. It is this subjectivity that may allow students to develop cognitive critical thinking patterns that cannot emerge via standardized assessment (Gitomer, 1993; Raven, 1992). Furthermore, authors suggest that qualitatively-oriented assessors view portfolios as a reliable assessment tool because students and evaluators are able to capture more detailed information than an objective standardized assessment (Arter, 1995; Gitomer, 1993; Mabry, 1999; Roberts, 2003).

Second, other authors suggest that standardized assessment is a viable option to assess student achievement (Baker, Linn, Dunbar 1991-1992; Greene, 2001; Hamilton, Stecher & Klein, 2002). This option can yield applicable results depending upon the context the results will be applied to. The quality of exams can be evaluated based upon three criteria which include reliability, validity, and interpretation.

Reliability refers to the degree to which a test’s scores provide consistent information and are free from various types of chance effects. Validity refers to the extent to which the scores on a test provide accurate information for the decisions that will be based on those scores. Interpretation refers to yielding the same scores for students who are of the same proficiency level regardless of gender, ethnicity, and physical disability (Hamilton, Stecher & Klein, 2002, p.52).

Assessment falls on a continuum in which normative scoring is at one end and criterion-referenced/standards based scoring on the other end. Some authors support normative scoring because it compares students’ performances to those of others, formally or informally ranking them, and awarding high scores to those with top
performances and low scores to those with low performances (Mabry, 1999; The Florida Department of Education, 2003; Dorn, 1998). Criterion referenced/standards-based scoring compares students’ performances to criteria or standards, perhaps ranking them, and awarding high scores to those whose performances meet or exceed criteria or standards and low scores to those who performances do not (Mabry, 1999).

The positive attribute of standardized assessment is, “the potential to devote more instructional time to learning, educators working harder to cover more material in a given amount of time, and teachers working more effectively by adopting better curricula or more-effective teaching methods” (Hamilton, also see Stecher & Klein, 2002, p.77; Strickland & Strickland, 1998). The negative aspect of standardized assessment is the reallocation of instructional time to focus on tested aspects of the standards and the exclusion of untested aspects of the standards. Consequently, some students are coached and some students, teachers, and administrators cheat on high-stakes testing (Hamilton, Stecher & Klein, 2002).

Third, standards-based assessment differs from standardized assessment because local school districts develop their own assessment methods and they are amenable to frequent change (Dorn, 1998; Meir, 2000). The Assessment Movement surfaced in 1989 as the federal government charged states to raise student achievement levels by reforming standards in each state. Standards are used as a coherent and a verifiable means of assessment because, firstly, it provides parents, teachers, and students in a state with clear expectations of what all students should learn. Secondly, standards contribute to a coherent educational practice when teachers align their instructional methods and materials with assessments based on these standards. Thirdly, standards establish relevant
guidelines for teacher preparation, professional development, and certification. Finally, standards reflect high expectations for all students in a state and contribute strongly to the goal of equity (Stotsky, 2000; Orfield & Kornhaber, 2001; Riley, 2002).

Is Florida’s school grading system designed and implemented in such a way that it accurately assesses and monitors student achievement? In the next section, comparable assessment tools performance-based assessment, FCAT implemented as a educational progress instrument, and standards-based assessment will be evaluated. The answers that emerge will allow an assertion of which model more accurately gauges student achievement.

In summary, the literature provides three themes that have implications for assessing teaching and learning in the United States and in specific in Florida. These themes are correlated to the three assessment models that are available for elementary and secondary schools: performance-based assessment, standardized assessment, and standards assessment. The limitation of the literature is that it does not provide an evaluative critique of each assessment model. This report will build upon the literature by comparing the three leading assessment methods against specific criteria: political desirability, administrative feasibility, and program costs. The outcome of this research is an evaluation among policy options and an assertion of the most viable and appropriate alternative.
III. Methodology & Evaluation Criteria

The information and data for this report will be collected using the following methods:

- Five unstructured telephone interviews for a duration of 30 minutes with personnel from the Bay District Schools, Port St. Joe Guidance Counselor, College of Education faculty from the Florida State University, and teachers from Leon County.

- Analysis of state and federal reports on school assessment and school accountability.


- Examination of public opinion polls on school accountability.

Unstructured telephone interviews provided insight into the application of standardized assessment. Teachers and administrators offered a practitioner’s view of assessment and its implication on curriculum and student learning.

State reports provided documentation of the FCAT’s effectiveness in measuring student achievement. These reports will be reviewed for content and political desirability for policy options. The academic literature will serve to support the theories behind each assessment model. Public opinions polls will support or refute the acceptability of accountability measures by the public.

**Evaluation Criteria**

Three criteria will be used to evaluate the proposed policy options. The criteria are: political desirability, program costs, and administrative feasibility.

- Political desirability relates to the support or lack thereof of the accountability mechanism by the public and the Jeb Bush Administration. Recent media coverage (i.e., newsprint) is analyzed to capture the level of support of assessment models by legislators, educators in K-12 schools, and the public. Interview data shows the support or lack thereof by
practioners in K-12 education and higher education on school assessment and accountability (Patton & Sawicki, 1993).

- Administrative feasibility refers to the ease or barriers in implementing an assessment model. The state criterion to implement an accountability mechanism is compared to states that currently possess a system. The alternative assessment model proposed is compared to the legislative mandates that already exist for a feasible method of implementation. Three sub-groups will be considered in this criterion:

  1. institutional commitment- top administrators, office and field staff as well must be in agreement of the policy implementation.
  2. capability- both staff should have the skills necessary and financial capacity to implement the policy.
  3. organizational support- sufficient equipment, physical facilities, and other support services (Patton & Sawicki, 1993).

- Program costs assesses the financial costs associated with the administration, scoring, and evaluation of standardized assessments (Patton & Sawicki, 1993). In comparison, alternative assessment model cost is analyzed. States who already have an alternative assessment model will be compared to the cost of the administration of the FCAT.

These criteria were selected to evaluate the alternatives because they most adequately assess school accountability methods. Other criteria such as testing equity and long-term assessment models could not be evaluated because of the complex study design necessary to perform an accurate assessment. Two limitations are the lack of current interviews of legislators (because of the difficulty in contacting them during and soon after the legislative session) and current opinion poll data. Despite these constraints, the most important criteria were evaluated and it is plausible to assert that most evaluators who used the same methodology would conclude similar recommendations.
IV. Management Policy Options

Section IV details three of the most promising alternatives to school accountability in Florida: Portfolio-based assessment, the use of the FCAT to demonstrate educational progress but not as a high-stakes assessment, and standards-based assessment. Each alternative will be evaluated using four criteria: political desirability, administrative feasibility and program costs.

**Option #1: Performance-based assessment**: Portfolio-based assessment serves as a functional short and long-term tool to measure educational progress. The shortcoming of this assessment tool is that it has few psychometric properties. The advantages of a portfolio-based system are that it is generalizable, allows a calculation of variance because of the subjectivity of the judge, the student, and the work demonstration. This assessment also allows for “high stakes” predictions based on the number of judges and the number of samples (Roberts, 2003). Portfolio-based assessment is centered around referenced assessments. Evidence of student’s work is compiled to demonstrate the student’s learning. This provides a profile of each student’s style of learning and it is readily applied to daily instruction. This alternative uses performance-based assessment to measure student achievement in K-12 schools.

**Political desirability**

Florida does not have a state adopted performance-based assessment system. Although other states have implemented a portfolio-based system, California has had success in its implementation. Many states are skeptical since portfolio-based assessment lacks quantifiable data which is often less refutable. The assessment was accepted and authorized in 1991 by Senate Bill 662 (California Department of Education 2002 Fact
Book). Educationally this performance-based assessment tool yielded much favor among educators, yet it failed to weather the political climate because, in 1994, the Governor, under Senate Bill 1273, struck down the reauthorization of the CLAS system. What is the utility of a performance-based assessment and can political system support their purpose and implementation?

Portfolios serve a very positive and functional approach for teaching/learning, assessment, evaluation. Portfolios can help student-reflection, critical thinking, responsibility for learning, and content area skills and knowledge (Arter, 1995). Arter (1995) further suggests other uses of this assessment include: (1) the ability to obtain a broader, more in-depth look at what students know and can do (2) assessment is based more “authentic” work (3) have a supplement or alternative to report cards and standardized tests (4) and the ability to have a better way to communicate student progress to parents.

A 2002 Herald/St. Petersburg Times poll found 51% of respondents are not pleased with Governor Jeb Bush’s overall education record. The poll also asked voters to identify the most important issue among nine choices. Forty-three percent responded it was improving public education. Is this dissatisfaction enough to open a political window for school assessment reform?

A performance-based assessment lacks political acceptability in Florida because of the design and implementation of the FCAT. The disconnect between a performance-based assessment and the FCAT is that Governor Bush’s A-Plus Plan rests upon the notion that the FCAT is a true and fair assessment tool. The plan calls for accountability, choices for parents, more resources, rewards for improvement and success, and change in
learning opportunities when students are not progressing (The Office of Governor Jeb Bush, 2003). Schools are assigned performance grades based primarily upon student achievement data from the FCAT. School grades appear to communicate to the public how well a school is performing relative to state standards. Can this system be modified and yet provide the public with the same or similar information that the FCAT provides?

If the accountability system were modified from a norm-referenced assessment to a performance-based assessment, the political desirability of this change would not be favorable. A positive characteristic of a norm-referenced exam is that it lends itself to more generalizability and comparability because of the quantifiable data element. President Bush’s No Child Left Behind legislation requires a rigorous state accountability system that ensures that every child meets the highest possible standards. The Federal Department of Education suggests that they allow for flexibility and variety between states; although politically the federal department has called for comparability between states.

Are portfolios politically desirable? Portfolios are subject-based assessment tools that are deemed unreliable by some education evaluation experts. When evaluated, “Vermont’s statewide system of portfolio assessment showed low reliability (the odds that two different teachers would rate a portfolio the same way). The report recommended improved portfolio training for teachers which in turn will increase assessment reliability” (California Department of Education, 1997, p.4). Interviews with educators showed that the Florida education system must be improved, but many are skeptical that the performance-based assessment is the vehicle for this improvement. D.L. Capes is the Testing & Evaluation Administrator at the Beacon Learning Center (personal
communication, March 8, 2003) says that, “portfolios are another way to evaluate students, probably not the best way but its an option, and I agree we need a new method; our Republican governor would have to change for a new assessment tool to be used. Eventually we will have to follow research and community pressures and come up with something else.”

In short, assessment has always won favor when the measurement tool is recognizable by the public and it is supported by legislation. Districts could implement a pilot performance-based system although they would suffer a negative political backlash. More abstract qualitative measures such as performance-based assessment score low for political desirability.

**Administrative Feasibility**

Portfolio-based assessments are unpopular in some states. Although states such as California, Vermont, Maryland, Kentucky, New York and Oregon have found through implementation that this assessment tool can accurately assess student achievement (Khattri, Reeve, and Kane, 1998). Performance assessment was feasible to implement in these states because it was mandated by the legislature. Each of these states has incorporated portfolio assessment in their legislation. Administratively, performance-based assessment has a few limitations. First, School districts cannot evaluate educational programs who have a specific objective or uniform goal. Second, students cannot be ranking or standardized so that a measurable standard of quality can be set and modified.

Third, students cannot be normed so that comparative data can be used to develop assessment and implement new teaching and assessment strategies. Florida’s current legislation supports the FCAT. Although it is plausible that if the political acceptability
were much higher that administratively a performance-based system could be implemented in statutes. Figure 3 of Florida Statutes 1008.22, 1008.31, and 1008.34 outline assessment mechanisms that must be administered in Florida schools. Moreover the following chart illustrates the statutes that can support or prohibit the implementation of a portfolio-based/performance school accountability model:

Figure 3: The 2002 Florida Statutes that support or prohibit performance-based assessment

<table>
<thead>
<tr>
<th>2002 Florida Statutes that could potentially support a performance-based assessment</th>
<th>2002 Florida Statutes that could potentially prohibit a performance-based assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1008.31(a)(1) the performance accountability system implemented to assess the effectiveness of Florida’s seamless K-20 education delivery system provide answer to the following questions in relation to its mission and goals.</td>
<td>1008.34(2) School performance grade categories- The annual report shall identify schools as being in one of the following grade categories defined according to rules of the State Board of Education—A,B,C,D, F.</td>
</tr>
<tr>
<td>1008.22(1)(a) assess the annual learning gains of each student toward achieving the Sunshine State Standards appropriate for the student’s grade level. 1008.22(1)(c) identify the educational strengths and needs of students and the readiness of students to be promoted to the next grade level or graduate from high school with a standard high school diploma. 1008.22(3)(b) Develop and implement a uniform system of indicators to describe the performance of public school students and the characteristics of the public school districts and the public schools. These indicators must include, without limitation, information gathered by the comprehensive management information system created pursuant to s. 1008.385 and student achievement information obtained pursuant to this section. Also see relevant statute: 1008.22(8)</td>
<td>1008.34(4) School Improvement Ratings- the annual report shall identify each school’s performance as having improved, remained the same or declined. This school improvement rating shall be based on a comparison of the current year’s and previous year’s student and school performance data. Schools that improve at least one performance grade category are eligible for school recognition. Also see relevant statute: 1008.22(3)(c)</td>
</tr>
</tbody>
</table>

Note. Florida Statutes K-20 Education Code, Title XLVIII Ch. 1000-1013 (2002)

One of the primary concerns of portfolios is their inability to be used for administrative purposes. Although contrary to this concern, The Federal Office of Educational Research and Improvement (1993) reports that portfolios are being used successfully for accountability reporting, program evaluation, and a variety of administrative decisions affecting the future of individual students. Although portfolios have value-added utility, there are also shortcomings their use. In summary, most states
are not willing to invest finances and pass laws that are necessary to support and implement a performance-based assessment.

Program Costs

Critics argue that performance assessment is not cost efficient. One of the great appeals of multiple-choice tests is that they are extremely efficient and compared to other alternatives, quite inexpensive. Performance-based assessment requires intensive professional development so that assessors can grade portfolios with minor discrepancies. Therefore, more funding must be expended on the development of efficient data collection designs and scoring procedures for performance assessment (Baker, Linn, and Dunbar, 1991-1992).

Yet several school-based proponents of portfolio assessment argue that performance-assessment tools are a relatively inexpensive investment as compared to psychometric assessment tools. The primary cost involved is the money devoted to professional development opportunities and training for teachers and administrators to assess and evaluate with a higher degree of reliability. Professional development funds are generally already included in a district’s annual budget. Districts can implement a performance-based system by using the financial resources already available. The use of available resources can be achieved by diverting funds used previously for norm-referenced exams to teacher training opportunities.

Program costs are difficult to measure for portfolios assessment for two reasons: (1) schools, districts, and state education departments do not generally record costs for testing and assessment as separate items, but as portions of categories such as personnel, materials, ads, vendors, etc.; therefore separating costs is difficult, and (2) the costs of
machine-scoreable tests and performance assessments are not comparable to professional
development time and money to develop the teacher and student (Kane, Khattri,
Adamson & Pelavin Research Institute, 1997). Although the U.S. Congress’s Office of
Technology Assessment (1992) suggested a hypothetical yet plausible comparison
between performance-based assessment and a norm-referenced assessment as illustrated
in Figure 4:

Figure 4: Hypothetical Program Cost Correlation Between a Performance-Based
Assessment and a Standardized Assessment

<table>
<thead>
<tr>
<th>School Activity</th>
<th>Estimated Outlays</th>
</tr>
</thead>
</table>
| (1) A large urban school district 1990-1991 administers and scores a standardized assessment | ▪ 1.6 million (0.8 million for test administration)  
▪ Approximately $6 per student test administration |
| (2) Test preparation time (3 weeks preparing students and 4 days administering exams) | ▪ $15 million in opportunity costs.  
▪ $110 million per pupil. |

It is not plausible to forecast the direct and opportunity costs of performance
assessment. To attempt to make such an assertion would include assigning a dollar value
to materials, services, and salaries, and indirect costs of time spent on performance
assessment activities. The shortcoming of performance-assessment is the inability to
appropriately define program costs. Districts and state governments are not amenable to
funding programs that may need additional funding to more adequately design and
implement the program. Supporters and critics of performance-based assessment are
sharply divided about the presumed administrative and evaluation cost of such an exam.
In summary, the qualitative nature of performance assessment does not lend itself to
forming an accurate budget forecast to fund this method.
Option #2: FCAT Examination Implementation: FCAT should serve the same role and function of the NAEP exam, and if penalties will be imposed on school districts, they will occur over a course of five years. This five-year time frame will provide schools an opportunity to re-evaluate educational programming and to address problems and concerns before being penalized for low student achievement. FCAT has been used to gauge student achievement and, as such, its implementation is critical for the success of public education. This alternative suggests the continued administration of the exam, although with a modification (i.e., the exam should be implemented as a gauge of educational progress).

Political Desirability

Politically, the FCAT exam was created to serve as the foundation of Governor Bush’s A-Plus plan. This high-stakes serves as the communication mechanism between the public and the quality of Florida’s public schools. The FCAT has demonstrated acceptability by the Governor, legislators, parents, and the business community. This exam suggests that Governor Bush and the Florida Department of Education have a strong accountability mechanism in place and that it is effective. An FSU College of Education Professor says, “The FCAT has a unique implication. It sets schools up to compete, it neglects science and social studies, and it takes away curricula decision-making for teacher. The FCAT is one of the worse things that could have happened to education, and this is the worst I have seen our education system in 50 years” (Anonymous, personal communication, March 8, 2003).

Politically, assessment is an important platform issue, and year after year opinion polls reveal that citizens have concerns about public school education (Garza, 2003; St.
Petersburg Times, 2002; Valle-Greene, 2002). High-stakes testing has forced schools to compete with each other. This has caused schools to reassess their roles and responsibilities as students, teachers, parents, and the legislature. Student retention, bonuses, and jobs are now linked to the grades of schools. Politically, assessment is necessary because it demonstrates to the public that students and schools are being held accountable for performance results. The FCAT has won wide political support in the legislature as a quantifiable means of assessing student achievement and demanding accountability.

Although the FCAT has positive political aspects, it also has three shortcomings. First, there are a high number of students who will not be able to graduate because they have not passed the FCAT. Currently there are 12,772 high school seniors statewide who have not passed the 10th grade FCAT in reading, and there are 9,933 who have not passed the math portion (Garza, 2003). Many of these students are closely approaching their senior year and have taken the FCAT on more than one occasion. This lack of mastery of the FCAT is a problem because dropout rates and the number of General Education Diploma (GED) participants will drastically increase.

Second, the FCAT lacks timeliness. The state administers the FCAT Writing in grades 4, 8, and 10 in February and grades 3-10 are administered in March. In 2002 a Palm Beach resident suggest that this schedule is imposed to give the company who scores the FCAT time to score and release results. Charlie Christ, former Commissioner of the Florida Department of Education affirmed that Florida’s testing timeframe is planned with NCS Pearson’s (the company that scores the FCAT) scoring schedule (Hegarty, 2001, 1B). The timeline is problematic for students and teachers. Students who
are unsuccessful on the FCAT receive their results at the beginning of summer; late June. By this time students who were unsuccessful have missed the summer school or other supplemental instruction registration deadlines. This timeline is difficult for teachers because it does not allow teachers the time to begin remediation with students who are unsuccessful nor does it permit the opportunity for teachers to counsel students about their scores.

Third, The Palm Beach Post suggests that the “FCAT politics” also highlights the necessity of making the scores comparable from one year to the next. Yet this goal has not been achieved because grading formulas constantly change (Palm Beach Post, 2002, 22A). The grading formulas change because the state is raising the expectations of teachers and students on a yearly basis.

In summary, the public and legislators accept the FCAT as a valid and reliable indicator of student performance. There are also long-term negative implications that the FCAT may pose for teachers and students in the next few years. More and more students are feeling overwhelmed and dropping out of school (Garza, 2003). Additionally, the decreased morale of teachers has caused many to change occupations and some projections estimate that Florida will need 40,000 new teachers in the next decade.

Administrative Feasibility

The FCAT is a statutory assessment tool that measures student achievement. In 2001 Level 1 was the lowest performance level, although in 2002 Level 2 students were also included as the lowest performance level. Low FCAT scores and low school performance are the two criteria that are used to retain students. These areas are
important because it demonstrates the high level of autonomy of the Florida Department of Education with its state developed assessment instrument.

The FCAT exam has implications for students, teachers, and districts for success or lack thereof of student achievement on the exam. The FCAT’s implementation can be modified so that it is used to show educational progress. “High-stakes” exams have inherent intentional and unintentional consequences. The advantage is that it encourages schools to perform at higher levels. Similarly, high-stakes assessment plans inadvertently promote schools to compete with each other for recognition and additional funding.

Administratively, the FCAT is an assessment tool that is already in place therefore its continued implementation can occur with relative ease. Schools will likely continue to perform at higher levels with any accountability mechanism in place. Similar to the “Hawthorne effect” schools may improve because they are being monitored and evaluated. The high-stakes component of the FCAT does not necessarily improve learning nor teacher quality.

The Florida Statutes were analyzed to determine how administratively feasible a change in the implementation of FCAT. Figure 5 illustrates the use of the FCAT as an instrument to measure educational growth and not as a “high-stakes” exam:
Figure 5: The 2002 Florida Statutes that support or prohibit the alternative use of the FCAT

### 2002 Florida Statutes that could potentially support FCAT used as a growth measure and not as a high-stakes exam

<table>
<thead>
<tr>
<th>Statute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1008.31(a)(1)</td>
<td>The performance accountability system implemented to assess the effectiveness of Florida’s seamless K-20 education delivery system provide answer to the following questions in relation to its mission and goals.</td>
</tr>
<tr>
<td>1008.34(2)</td>
<td>School performance grade categories- The annual report shall identify schools as being in one of the following grade categories defined according to rules of the State Board of Education—A,B,C,D, F.</td>
</tr>
<tr>
<td>1008.22(1)(a)</td>
<td>Assess the annual learning gains of each student toward achieving the Sunshine State Standards appropriate for the student’s grade level.</td>
</tr>
<tr>
<td>1008.22(1)(c)</td>
<td>Identify the educational strengths and needs of students and the readiness of students to be promoted to the next grade level or graduate from high school with a standard high school diploma. Also see relevant statutes: 1008.22(3)(b), 1008.22(3)(c)</td>
</tr>
<tr>
<td>1008.34(4)</td>
<td>School Improvement Ratings- the annual report shall identify each school’s performance as having improved, remained the same or declined. This school improvement rating shall be based on a comparison of the current year’s and previous year’s student and school performance data. Schools that improve at least one performance grade category are eligible for school recognition.</td>
</tr>
</tbody>
</table>

Note. Florida Statutes K-20 Education Code, Title XLVIII Ch. 1000-1013 (2002).

Public opinion polls suggest that educational administrators are not accepting of how the FCAT is implemented. The Lakeland Ledger Publishing Company (2002) surveyed 39 principals in Florida: 18% would abolish the battery of tests (measures high-level thinking skills), 41% believe the FCAT has helped Florida’s schools, 38% asserts the FCAT has helped their particular school, 74% believe the A-Plus accountability plan has hurt Florida’s education, and 43% believe the grading system has hurt their particular school.

In summary, statutes are written in Florida’s K-20 Education Code that detail the administration and use of the FCAT exam. This exam is also accepted by the majority of stakeholders and deemed as an important and valid assessment tool. A modification of existing legislation will enable the FCAT to be used for the primary purpose of showing student achievement scores high on the administrative feasibility criterion.

**Program Cost**

There are three direct costs of the FCAT administration/scoring, FCAT informational website design/website maintenance, and the School Recognition Program.
First, the direct cost of the FCAT exam is $105 million to administer and score the exam for three years. This contract equals approximately $35 million annually. The cost of the exam’s administration and scoring is determined based on 2,535,155 students enrolled in 67 public school districts (Florida Dept. of Education, 2003). NCS Pearson has agreed to pay $10,000 for a one-day delay and $30,000 for a two-day delay and after seven days, $250,000 per day. Late scores last year resulted in a $4 million dollar late payment (Hegarty, 2001). This “high-stakes” exam is very promising because there is a potential each year to make additional revenue if Pearson NCS does not fulfill their contractual obligations.

Second, another direct cost is the FCAT information website created by DOE. It assists parents and students with study and support strategies. The site design and construction was $8.2 million and current site maintenance is $1.5 million per year (Palm Beach Post, 2003, 20A).

Third, another direct cost is the School Recognition Program. Schools that earn an A grade from the state or whose school grade goes up a letter grade from the previous year are eligible for $100 per student in the school. The program began in 1998 with a $27.5 million payment in School Recognition dollars, now this current fiscal cycle has paid $122 million to schools (Sarasota Herald Tribune, 2003, 11B). The FCAT exam cost could be absorbed in another high need area in schools. Exceptional education, school construction and maintenance and supplemental academic instruction methods are only a few of these high need areas.
In summary, Florida spends $5 billion on public education. Assessment is important in education although stakeholders have mixed views about its cost. The FCAT exam scored moderately on the program cost criterion.

Option #3: Standards-Based Assessment

Florida may consider implementing standards-based assessments that are aligned with the daily curriculum. Standards-based assessment has been implemented in various forms throughout the U.S. (Meir, 2000). Standards generally exist in the areas of mathematics, reading, language, and science. Standards provide a comparable means of assessing student achievement throughout the state. Nationally, standards have been accepted as a method to assess student achievement. Standards are also a comparative state-by-state assessment tool (Stotsky, 2000).

Political Desirability

Nationally, the Standards Movement has received wide support from legislators, educators, and the public. President George Bush’s Goals 2000: Educate America Act is ambitious legislation that is premised on improving the quality of education for all students. Standards are important politically because they provide a measurable assessment of teaching and learning in public schools.

In the 1990’s as the Systemic Reform Movement took shape, the reform effort was able to align curriculum, standards, and assessments to achieve a powerful impact on the educational system. Politically, the standards-based assessment has had three favorable responses from policymakers: “(1) programs of formal testing controlled by state government agencies allow the level of government with the legal authority for schooling actually occurs, (2) statewide formal testing of
students has the potential to influence the behavior of all major actors in the educational system, and (3) the inherent efficiency of testing as a performance-monitoring process (Orfield & Kornhaber, 2001, p.20).

The shift back to standards-based assessment will not implement a relatively new approach; instead it will revert to the model in place in the early 1990’s. The Sunshine State Standards were developed when the movement was to improve standardization by the national government. The Standards Movement created legislation in Florida that refined the curriculum for each grade and subject area; therefore it is acceptable that politicians would consider the adoption of policies that are centered on standards.

Politically, standards have been viewed favorably by the public view because of students’ equitable treatment with standards. Test-bias does not seem to be applicable to standards-assessment because multiple means of assessment are generally used in standards-based curriculums. Moreover, legally, standards-based assessments have rarely been refuted in courts because of their specificity in design, yet broadness in the implementation of the standards-based assessment models.

In summary, Florida Sunshine State standards serve as the blueprint for all standardized assessments for the state. Therefore these standards score high on the political desirability criterion.

Administrative feasibility

Standardized assessment follows the same convention of the NAEP exam because it sets a measurable objective for policymakers and educational administrators. The Sunshine State Standards as articulated by former Commissioner of Education Frank Brogan:

31
The Sunshine State Standards…they represent rigorous, high stakes academic expectations that will prepare Florida students to compete in one of the toughest marketplaces our nation and world have ever known. The Sunshine State Standards are heavy on academic basics, like reading, writing and mathematics. They reflect the need to set high expectations for our students and to challenge them to meet those expectations. And they reflect a commitment to excellence that is at the heart of every teacher and every parent I have met in the state of Florida (Florida Gulf Coast University, 2003).

Additionally proponents of the Sunshine State standards believe they are well-crafted criteria to promote and sustain student achievement in Florida. A 40-year FSU College of Education Professor agreed with other survey recipients that the morale of teachers and students are significantly declining. This professor believes that this era is the lowest point he has seen in education in 50 years. He says, “the FCAT challenges schools in a negative way and that he thinks we should go back to a curriculum that is driven by standards” (Anonymous, personal communication, March 8, 2003).

Florida’s educational system is based on standards, and proponents of the FCAT suggest that the FCAT exam is derived from the Sunshine State Standards. Although some teachers argue that the standards are measured on the FCAT abstractly and not explicitly and therefore the outcome is a defunct curriculum and the appearance of low achieving students. Teachers generally agree that they negate teaching the standards and that they are now teaching to the test to avoid being fired, being ridiculed by the school administration, or by failing their students. The standards are already in place and they serve as a formidable and viable support for teaching and learning in Florida. Statutorily,
the state government supports standards. Figure 6 shows an analysis of the Florida
Statutes that could support the use of standards-assessment as the primary method to
measure educational achievement:

Figure 6: The 2002 Florida Statutes that support or prohibit standards-based assessment
in Florida

| 2002 Florida Statutes that could potentially support a return to standards-based assessment |
| 1008.22(1)(a) assess the annual learning gains of each student toward achieving the Sunshine State Standards appropriate for the student’s grade level. |
| 1008.22(1)(c) identify the educational strengths and needs of students and the readiness of students to be promoted to the next grade level or graduate from high school with a standard high school diploma. |
| 1008.22(3)(b) Develop and implement a uniform system of indicators to describe the performance of public school students and the characteristics of the public school districts and the public schools. These indicators must include, without limitation, information gathered by the comprehensive management information system created pursuant to s. 1008.385 and student achievement |
| Also see statutes: 1008.22(8), 1008.31(a)(1) |

Note. Florida Statutes K-20 Education Code, Title XLVIII Ch. 1000-1013 (2002).

Program Costs

The program costs that are associated with standards-based assessment are minimal. Florida currently administers the Florida Comprehensive Assessment Test (FCAT), Norm-Referenced Test (NRT), and the National Assessment of Educational Progress (NAEP). Figure 7 provides a description of the current forms of assessment that are currently administered in Florida:
Figure 7: Different school assessments that are administered in Florida’s public school

<table>
<thead>
<tr>
<th>Florida Comprehensive Assessment Test (FCAT).</th>
<th>The FCAT is a state student achievement test that specifically addresses a student's ability to perform on the Sunshine State Standards benchmarks. Initially, students were tested in the areas of reading for grades 4, 8 and 10 and math for grades 5, 8 and 10. The FCAT was expanded in 1999 to assess students in grades 3-10 in reading and mathematics using both criterion referenced and norm-referenced tests. Science will be added in 2003. School grades awarded under the “A+” Plan are based on FCAT results.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm-Referenced Tests.</td>
<td>Florida school districts administer norm-referenced tests to students at various grade levels. Norm-referenced assessments determine how well individuals perform in comparison to others, such as others at the same grade level or age. The Department of Education collects results from the tests administered in grades 4 and 8 and uses the data for educational accountability and public reporting purposes.</td>
</tr>
<tr>
<td>National Assessment of Educational Progress (NAEP).</td>
<td>Congress authorized the National Assessment of Educational Progress as a means of monitoring educational attainments in several subject areas in grades 4, 8, and 12. The subjects include mathematics, science, reading, and writing among others. NAEP collects information on a state level for those states wishing to participate. Section 229.57(2) directs Florida’s participation in measuring national educational goals.</td>
</tr>
</tbody>
</table>


Florida currently spends $57 million on assessment and evaluation; this estimate includes the FCAT and other testing initiatives. In fiscal year 1995-96, the Florida Department of Education expended about $48 million (53% of its total expenditures) on activities related to school improvement and legislative priorities (OPPAGA, 2003).

In summary, Florida’s 2004 budget already includes dollars for assessment. Standards-based assessment preceded the FCAT exam; therefore, reforming the implementation of FCAT is possible. Florida Standards are comparable to many other states cited for having an excellent objectives and goals in its curriculum. Floridians believe strongly in Sunshine State Standards and are willing to invest resources in it (Valle-Greene, 2002). Therefore standards-based assessment scores high on the program cost criterion.
V. Conclusion

This report provided an analysis of three alternative assessment tools. Each policy was evaluated based on three criteria: political desirability, administrative feasibility, and program costs. Figures 8 & 9 summarize the results. Each alternative suggests a potential for assessment measurement tools in Florida.

Figure 8: Summary of Alternatives & Evaluation Criteria

<table>
<thead>
<tr>
<th>Option</th>
<th>Political Desirability</th>
<th>Administrative Feasibility</th>
<th>Program Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Assessment</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>FCAT/Non-High Stakes</td>
<td>10</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Standards Assessment</td>
<td>10</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Ranking Scale: 1 to 10

1 very unfavorable, not considered 6 more than average acceptance
2 unfavorable, likely unconsidered 7 good; modify assessment to meet district needs
3 unfavorable at this time 8 good at this time
4 possibly favorable w/modifications 9 very good alternative available
5 neutral 10 very favorable among any other alternative

First, the performance-based assessment scores low on the political desirability criterion. Legislators believe that the quantifiable data is the best gauge of student achievement level. The current Republican administration is not amenable to assessment models that could hinder the viability of the A-Plus Plan or that may not be generalizable to and comparable to other states’ mainstream assessment methods. Performance assessment also ranked lower on administrative feasibility because legislation, testing, and evaluation would have to change significantly. The option ranked higher on the
program cost criterion because limited materials and supplies are needed to implement this option. The only cost associated with this option is professional development opportunities for teacher to because fair and accurate assessors.

Second, the FCAT as a gauge of educational progress ranked high on the political desirability criterion because both legislators and top-level education administrators view this tool as an accurate gauge of student learning gains. The FCAT also faired high on administrative feasibility because the exam in written in the Florida education administrative code and it is the catalyst for the A-Plus Plan. It did not fair as well on program costs because of alternative norm-referenced assessments that are available at a more inexpensive cost. Additionally students can be tested intermittently and not yearly to save on cost-efficiency.

Third, standards assessment is a likely competitive policy option for the FCAT. The Florida Sunshine State Standards have been in place before the FCAT and have served as a plausible tracking of educational gains. Standards assessment ranks high on the political desirability because it coincides with national standards from professional education associations. It also ranks relatively high on administrative feasibility because assessment has already been created around these standards and teachers, legislators, and other stakeholders are familiar with them. Standards assessment has not scored as high on program costs as FCAT because its implementation is premised on professional development and assessment.

Assessment of the alternatives using the three criteria indicates that the FCAT is a good assessment measure. The shortcoming of FCAT is the high-stakes component the exam imposes on schools. FCAT does gauge the knowledge students are learning in
classrooms, since although it is difficult to measure student-learning gains from year to year. The frequent changes in the scoring of this exam also suggest some political instability. Therefore, FCAT should be aligned with daily instruction that will provide a cohesive curriculum of standards, assessment, and evaluation. Many assessments developed by districts that supported the Sunshine State standards were serving the needs of students, teachers, parents and the community (Florida Gulf Coast University, 2003). These exams should be implemented again in schools. Florida is a textbook adoption state. Therefore, all of the books that are used in K-12 schools are written around Florida’s Sunshine State Standards (Irvin, 2003). The shortcomings of these previous assessments were that they were not generalizable or comparable to other districts’ assessment tools. Standards-based assessment allowed teachers to plan curriculum and assessment from state-accepted standards. Districts also felt more ownership of standards-based assessments. Currently teachers and administrators view FCAT as disconnected from daily instruction. Therefore FCAT should be implemented so that it monitors educational progress and not as a “high-stakes” exam. This will give FCAT the same measurable outcome of the Florida Sunshine State Standards and the NAEP.
REFERENCES


Education is year’s top election issue. (2002, March 31). *St. Petersburg Times,* p. 1A.

End FCAT Politics (2002, November 4). *Palm Beach Post.* p. 22A.


Garza, C.L. (2003). Many see time slipping away on FCAT. If they don’t pass, they don’t graduate. *Florida-Times Union*, Metro p. B-1.


Greene, J.P. (2001). *An Evaluation of the Florida A-Plus Accountability and School Choice Program*. The Manhattan Institute, Center for Civic Innovation, The Florida State University, School of Public Administration & Policy, & The Harvard University Program on Educational Policy and Governance.


State Weaves Tangled Web Over FCAT Site Contract. (2003, February 20). *Palm Beach Post,* p. 20A


About the Author

Tony D. Johnson (B.S. Political Science & Criminal Justice, Florida State University; MPA, Florida State University) has served as a teacher for the Leon County School District and as a grant writer for Tallahassee Community College. Mr. Johnson’s primary research interest includes discrimination and inequality, the history of Black and Latino education and contemporary urban education. He is currently an Adjunct Instructor at Tallahassee Community College and an instructor in the Department of Educational Leadership & Policy Studies with the Florida State University.