This discussion of high stakes tests explores criticisms commonly made of the use of these tests and suggests some additional reasons educators and the public should not rely on high stakes tests as the single source of data about student achievement. The United States has no national educational policy on standards of knowledge. Each state is left to develop an individual curriculum, yet at the same time each state has to be accountable and show positive results to politicians at state and national levels to keep revenue flowing. National goals and standards have been proposed, but revenue has not come forth to make these goals a reality. It is easy to think that the numbers produced by standardized tests can answer accountability questions, but there are many reasons standardized tests should not be the single basis for educational decisions. This paper suggests four reasons beyond the commonly mentioned why high-stakes tests should not be used for a single purpose: (1) high-stakes tests reduce a child's rich and complex life to a collection of scores, percentiles, and grades; (2) such tests judge children without providing suggestions for improvement; (3) answers on high-stakes tests are final, without opportunities for revision; and (4) high-stakes tests discriminate against some students because of cultural backgrounds and individual learning styles. (Contains 37 endnotes.) (SLD)
I open this paper on “high-stakes testing” by first asking the question “What legally constitutes an adequate public education?” The nation’s state legislatures, through explanation in each of their state constitutions, must be able to satisfactorily answer that question before we set standards, set curriculum guidelines or construct achievement tests to check for competency of skills or learning for understanding.

A review of the fifty state constitutions regarding how state legislators have defined “an adequate public education” led to some interesting results. In all but four of our fifty state constitutions, words describing the educational system as “thorough, efficient, general, uniform, stable and adequate are the norm.” Only in four states (Florida, Illinois, Minnesota and Virginia) do the words “high quality” appear in describing the educational system required by the state constitution. Florida voters added the words “high quality” in a 1998 amendment to the state constitution. My final comment about the adequacy of public education is made regarding the language of the Alabama State Constitution. It states “It is the policy...to further promote the education of its citizens in a manner consistent with its available resources...but nothing [herein] shall be construed as creating or recognizing any right to an education....” This indicates to me that if Alabama determines it has no available resources, then no citizen or student has the right to an education in that state. I truly wonder if the Citizens of Alabama are aware of the language of their own constitution regarding educational policies?

The health of the public education system in the United States continues to be a subject of discourse among all levels within our American Society. Politicians always want to know why children seem to be lacking basic skills at certain measuring points in their educational careers. School
administrators continue to have revenue pulled from their budget, sometimes significantly late in their fiscal or school years. Schoolteachers wonder when their class sizes will stop increasing; where they can put just one more student; and how they can give adequate attention to those students at the lower end of the academic scale? Parents want to know why their child is having difficulty reading, writing coherent paragraphs and solving simple math problems. Students seem to get bored being taught the same type of curriculum year after year with rote memorization and numerous standardized tests being their measure of success. All of these points and questions are valid and every person in America from the President down to the individual student each share a responsibility for ensuring the success of our American Public Education School System. Some can do no more than complain or voice their support for minimal change; others can demand action through organized group efforts; and others can ensure that sufficient revenues are being funneled to support education and all that education is supposed to be in today’s culture. Unfortunately, no consensus exists among states or at the national level as to the content of an “adequate public education.”

Many states have mandated what students should learn in each grade, developed assessments to measure student achievement, rated their schools, publicly identified failing schools and even closed or took over failing schools. Included in the various accountability measures is the “high-stakes testing.” These are tests required by the individual state Departments of Education as mandated by the state legislatures. The tests are to measure and to ensure that each and every student has met the minimum standard of competency in one or more academic areas in order to be promoted to the next grade or in a number of states, to exit high school with a general high school diploma. Test scores are used in this manner to determine a child’s educational future as well as a child’s sense of well being and worth. “The improper use of tests can cause considerable harm to students affected by test based decisions.”

American children nationwide will take over 100 million standardized tests each and every year. Lawmakers have come to view the results of these tests as “the measure” of student achievement and hold schools and school districts highly accountable for their results. A Parents Guide to Standardized Tests in schools was printed in 1998 to help parents improve their child’s chances for success.
High-stakes testing did not come into being until the late 1970's, however some earlier states' legislation regarding accountability led the way. We saw a boom in the 1980's as private enterprise saw opportunities for assisting states in their measurements, for a profit. Florida became the first state in the nation to implement the “High School Competency Test” in 1977 – 1978 and Nevada followed one year later with its’ High School Proficiency Exam. Following their leads in chronological order were North Carolina, New York and Hawaii in the late 1970's and three more states in 1981.6

As of January 2002, forty-nine of our fifty states administer some sort of “standardized” assessment or progress tests to monitor students’ achievement at certain intervals during their academic career.7 Only Arizona, in 2001, suspended through 2004, the use of all standardized testing in their state. The state legislators are questioning the true validity and accuracy of “learning” that these standardized tests are supposed to be measuring. Currently twenty-six states are either using high school exit exams or are in the process of developing such tests.8 All of the states with high school exit exams in place do allow students to take the test multiple times usually beginning in the tenth or eleventh grades. A number of high school graduation tests have been challenged in court with Florida being one of the first to face litigation based on the state’s disproportionate failure rate among Black students. In 1981, the Fifth Circuit Court of Appeals9 affirmed the lower district court’s ruling in favor of the students. The decisions were based on the opinion that the state had not “made any effort to make certain whether the test covered material actually studied in the classrooms of the state.” It also stated that students had a “property interest” in their diplomas since they had attended school during the required years and passed the required courses. When the case was reheard in 1984,10 the court reversed the previous decision. It ruled that the students had, in fact, had an opportunity to learn the material covered in the exit test. The court established the following legal standards for high school exit exams:11

1. The test must measure knowledge and skills that are taught in the state’s school, referred to as “curricular validity.” In Florida, that curricular validity is measured through the Sunshine State Standards.
2. Students must receive adequate notice of the test, requirements for passing the test, and the consequences of not passing the test.

3. The test must not intentionally discriminate against a protected group or class (such as disadvantaged students).

Very recently headlines advising of the “backlash” against standards reform have surfaced, consistently linked to concerns about the high stakes testing practices associated with the standards movement. Some educators, analysts and advocates have concluded that high-stakes testing practices can only perpetuate the discrimination that, at times and in certain places, have plagued our schools. They also claim these tests are doing more harm than good to the very students who need the most help. Many educators and civil rights advocates are suggesting that it has actually added to the problems it had sought to alleviate. These leaders claim that the policies discriminate against minority students, undermine teachers, and reduce opportunities for students to engage in creative and complex learning assignments. Some maintain that these tests are “directing sanctions against the victims, rather than the perpetrators, of educational inequities.” Some have gone so far as to conclude that the practices associated with standards reforms are “dangerous.” All too often these high stakes tests have come to be viewed as the ultimate cure. “Many teachers tend to express themselves as “for” or “against” high-stakes testing.” Unfortunately, the issue is not so simple. “The tests themselves should never be used as a sole criterion when making decisions that have significant consequences.”

In recent years, some critics of American education have proposed establishing a “national” standard achievement test for the nations’ students. So far that has not happened due to the growing opposition among both Democrats and Republicans. Many conservatives in Congress think the testing program intrudes on the rights of the states and local communities to direct their own educational programs. Others fear that this national achievement test/testing program is the first step in establishing a national curriculum. I feel confident that since the “right to an education” is nowhere mentioned in the Constitution, nor in the minutes of the Constitutional Convention kept so meticulously by James
Madison, in the hot summer of 1777 in Philadelphia, that the establishment of a national curriculum would not be upheld by the United States Supreme Court.

In Florida, our state legislature has mandated the use of the Florida Comprehensive Assessment Test (FCAT) to measure benchmark achievements delineated in the Sunshine State Standards promulgated by the Florida Department of Education. In addition, selected FCAT reading and mathematics tests measure student performance against national norms. These norm-referenced tests permit comparison between the performance of Florida students and students throughout the nation on each grade level. Only a very few special public school students in this state are exempt from taking the FCAT in grades 3 through 12. Even students who are speech and language impaired, autistic and traumatically brain damaged are included in the test takers. All standard curriculum and other special categories of students are not excluded. Students attending private and parochial schools are not required by our state to take these tests, however students attending state chartered schools are required to participate in the examinations.

FCAT contains questions that are challenging for all students. Recently what students in Florida are expected to learn (understand) and what they should be able to do (skills) has increased a great deal. The job market of today requires people who are proficient in advanced mathematics and who can read and construct meaning from difficult and technical texts. FCAT is being given to measure achievement of the high level of learning set by the Sunshine State Standards that is being taught in our classrooms and being learned by Florida students.

The above information about the FCAT is considered by some as the "gospel truth" and is viewed by others as only another "political ruse." How does the general public, which includes average citizens, public school parents, students and the Florida educators, know what the true purpose of the test is? Students are provided feedback on whether they passed or failed, but are never allowed to review the test and reprocess their incorrect answers to replace old learning with correct learning for understanding. School administrators are provided with numerous reports and lots of statistics that include school and demographic reports, school district and individual school demographic reports, school, district and state
summary reports, performance reports and holistic rubrics. The “big daddy” of all reports is kept at the Department of Education Headquarters in Tallahassee, Florida, which shows state compliance or non-compliance with the Sunshine State Standards. This report is not available for review by the general public of our state. Is the FCAT really an accurate measurement of achievement of those skills society requires or has “high-stakes testing” such as FCAT become such a “politically dominated issue” so monumental, and our educational system is so broken down that we have to use the “plug the hole in the dike with a finger” approach and hope the issue will solve itself? Is high stakes testing another of the “political ruses” generated by state legislatures to deceive the general public into believing that “all is well” in our public education system?

We have no national educational policy on standards of knowledge. Each state has been left to develop their individual curriculums. Some states have even allowed each school district to write their own standards based on general and broad statements put forth by the political bureaucracy in each state and promulgated in their individual state constitutions. Here we have our great United States of America, governed by a Constitution for 220+ years, in the longest lasting democracy in the history of the civilized world, with no national defining standards for our future leaders. Yet each state has to be accountable and show positive results to the politicians at both the state and national levels in order to keep revenue flowing from the state and national coffers.

In 1989 at Charlottesville, Virginia, President George H.W. Bush, and the nation’s state governors established six National Educational Goals to be achieved by the year 2000. For the purposes of brevity, those goals are not listed here, but can be found at www.ed.gov/pubs/Prog95/pg_6toc.html. The states agreed that they would align their educational policies both vertically and horizontally based on these goals. A summary of the goals for standards based reform were to be 1. high academic standards for all students, 2. accountability for student outcomes, 3. inclusion of all students in the reform initiatives, and 4. flexibility to foster instructional change.

As wonderful as these national goals and standards are, revenue has not come forth from Washington to make these goals a reality. States cannot do it alone with their paucity of revenues and
plethora of needs from its residents in this post 9/11/01 era, and the bursting of the technology bubble on the stock market, that has caused a deep recession costing many states billions of dollars in anticipated revenue. The reality of all of this is that our educational system is broken and we seem to be doing "patch-work" fixes so it does not completely collapse. When schools are charged with the responsibility for both inclusion and sorting of their student populations, and we use a number system that determines how we sort students that is subject to human perception, the consequences for both the student and the school can become devastating. Even at the Graduate Record Examination Level, students who answer 50% of the questions correctly will fall into the upper 70th percentile. Setting cutoffs is a more complex process than most policy makers and educators anticipate, and typically more arbitrary than most would imagine. President Clinton signed the Goals 2000: Educate America Act on March 31, 1994, giving further credence to our need for solving the educational crisis in the United States.

With many reports and quantifiable statistics (numbers), we can report to our superiors "numbers" which carry a certain value. Two is higher than one and forty is higher than thirty. Those are understandable and with these numbers we can show trends, results, achievements, identify schools and districts that are excelling, just mediocre, or failing on a norm-referenced test. Having a number to use such as 1, 5, or 8 says with certainty a value. Why do we need that certainty? Dr. Selma Wasserman, Professor Emerita, Simon Fraiser University, discusses our unending quest for certainty and states in her article "in the area of human judgment, which is at the heart of qualitative assessment, the risks of uncertainty are very great." In her article, she makes reference to a book by John Dewey titled The Quest for Uncertainty. She goes on to quote him stating “judgment and belief regarding actions to be performed can never attain more than a precarious probability.” In the play Copenhagen, (Dr. Wasserman also refers to) character Niels Bohr says “It starts with Einstein...He shows that measurement - measurement on which the whole possibility of science depends...It’s a human act, carried out from one particular viewpoint of one possible observer.” I would also question the true validity of the numbers we assign in the educational settings. If a student knows ½ of an answer or can complete ¾ of a skill, we consider that student to have measured a Zero on most test items. The FCAT does provide a few
mathematical problems where a student has to explain how they arrived at the answer (explain the process) and are given partial credit for each step explained correctly. But there are only a few of these and in math only. We cannot take into account just how much a student really knows. It appears to be an all or nothing scenario.

Edward Fiske, editor of the New York Times, criticized standardized high stakes testing for the following reasons:\(^{28}\)

- Tests assume a single correct answer to a problem. There is no allowance for complex answers or multiple ways of achieving them.
- Tests measure how good students are at recognizing information, not generating it. We do not ask them to synthesize, solve problems or to think independently.
- Tests are timed – more emphasis is placed on thinking quickly than on thinking profoundly.
- Most tests focus on basic skills. Higher order thinking is ignored.
- Tests emphasize isolated learning. They do not require integration of facts and ideas. Many students do not see the connection of the isolated facts or information they have learned.

In the September 2001 KAPPAN MAGAZINE POLL,\(^{29}\) 65% of the respondents favored classwork and homework as the best way for measuring student achievement. What a difference a year makes! In the September 2002 KAPPAN/Gallup Poll,\(^{30}\) the use of a test alone dropped from 31% (2001) to 26% (2002) and the use of only class-work and homework dropped from the 65% reported above to 53% (2002). A new category had to be added by the volunteering of this additional information from the poll participants. Twenty percent (20%) of the participants believed that both tests and class work/homework should be combined to assess achievement of a student. This category did not exist on the 2001 or prior polls. The uncertainty category (Don’t Know) dropped from 4% (2001) to only 1% (2002). I consider these to be significant findings and could be the subject of further examination. More people (66%) would withhold money from schools not making progress (depriving students who truly need the additional resources and maybe, technology), and 54% of the respondents’ favor not renewing the principal’s contract. An even split is seen on blaming or helping the teachers. Teacher pay in Florida
ranks about 40th in the nation. How do we ensure that our students learn higher order thinking skills when teachers are paid lower level wages? What kind of message does that send from the state to the educators who have answered the “calling” to become an educator of tomorrow’s leaders and citizens? It is true that some teachers refuse to move from the “pen/pencil/worksheet/book” conditioning response to a true instruction, according to Howard Gardner, the author of many books and articles on “Multiple Intelligences.” Money can be a motivator, but some educators feel “indoctrinated” and being baby sitters because politicians will hold them more accountable for instruction versus districts with broad, general guidelines and a lack of resources to back up those guidelines.

In our educational preparation to become a teacher, we are taught the many different ways to measure achievement and to assess student performance. That is a skill we should acquire over a very short time due to our daily contact in 180+ days in the classroom. This should allow each teacher, if paying attention to their class, to view the students in many different learning environments. High-stakes tests “fail to catch the subtleties of incremental improvement that impact teachers day-to-day curricular and instructional decisions.” The common sense principle that multiple factors can help ensure better educational decisions affecting students is a widely shared view. In fact, all leading test publishers observe that, “if appropriately validated for the particular use, test are valuable instruments to use when making important decisions affecting a student’s education.” But they confirm that test scores are not perfect measures and should not be the sole basis for making such decisions. “Using test scores to make high-stakes decisions denotes a quality of testing perfection that does not exist” If the states and our national government do not believe that we, the teachers, are capable of measuring accurately the students’ achievement in our class, then the remedy lies with us, the teachers, not with high-stakes testing or legislators. These tests routinely cause unnecessary stress, anxiety and other harmful effects on our student test takers. The STAR Reading Score is a high-stakes test based on a one-time “vocabulary usage” administered to students. Even though it is a one-time test, given on any day chosen by the administration, which could be the worst emotional day for the student, it is routinely used and scores passed to other teachers as the accurate measure of the student’s reading level. This “number” stays with
the student for the entire semester and maybe the entire school year. Would it really matter to the overall system if we assessed a student as reading, 1) above the current grade level, 2) at the present grade level or 3) below the current grade level? We do this on our professional observations and do not need another label to put on each student so we can be accountable to politicians. Are we so needy of "numbers" (as previously discussed) that we need to know a specific number, 5.1, 6.7 or others, to give an accurate measurement?

Research results from the "Civil Rights Project at Harvard University" have concluded four major findings:\[^{34}\]

1. High-stakes tests do not necessarily make teachers and students more motivated in the classroom.

2. High-stakes testing is correlated with drop out rates. Nine of the ten states with the highest drop out rates use high-stakes testing. None of the ten states with the lowest drop out rates use high stakes testing. (This is an area that could use additional research)

3. A student's performance on a high-stakes exam is significantly tied to the level of their teacher's experience. And finally,

4. Test based grade retention does little to improve learning.

I have provided comments from the experts on high-stakes testing and why students, educators and even test makers do not advocate it as the single measure. I would like to add here my four personal reasons for being against the use of high-stakes tests for a single purpose:

1. I believe that is reduces a child's rich and complex life to a collection of scores, percentiles and grades.

2. I believe that it judges the child without providing suggestions for improvement seeming to regard testing and instruction as two separate activities.

3. The answers on the test are final. Students do not receive an opportunity to revise, reflect or redo a testing experience because the test focuses on only "the right answer."
4. I believe that the tests seem to discriminate against some students because of their cultural background and individual learning styles.

These criticisms I have just outlined have become part of my educational belief system influenced by Thomas Armstrong ideas in his many writings. Both he and Howard Gardner impact my teaching life, teaching method and solved for me a problem I have had for many years. The question was "how do I make sure that I can "include" every child in the learning process, capturing and focusing on their individual strengths, instead of teaching to the middle group of students and let the bottom ones fall to the wayside and letting the high end students race ahead?"

Students need to be taught ordered thinking versus scattered knowledge; taught how to integrate ideas and skills generating new knowledge versus conditioning to a test and presented with different ideas on an issue subject to appropriate discussion versus saying "that subject is taboo, we...." Without the above mentioned skills, our students will continue to get bored with our curriculum, drop out at ever increasing rates and will achieve only about 20% of their potential. Will students instead be moved ahead one grade without regard for achievement and become doomed to failure prior to the 12th grade on the high school exit examinations? I personally favor an approach that is very simply and thoroughly explained by Thomas Armstrong in "Multiple Intelligences in the Classroom." Each child comes to school with their engrained strengths and weaknesses. No two children are alike. It is true that we should not be required to develop individual lesson plans for each of those students, but we can expose those students to a variety of our teaching methods and strategies, so that we can capture their interest and help them to motivate themselves by appealing to their strengths on a regular basis. Students would be more likely to take risks in areas where they would not be as smart if we can ensure that they are fully aware of and accept that our classrooms are indeed safe places for them to take the risks of a wrong answer without being labeled or "put-down" by teachers or students for having taken the risk. The only risk they now take is not being willing to take any risk at all. Then we know in certainty that the student will not succeed. I believe that we should allow our teachers to teach, let our teachers measure achievement and do assessments for understanding in non-traditional ways following some of the examples provided by
Thomas Armstrong in his above referenced book. Politicians need to let teachers excel at what they do best, and that is to teach, not to “teach the test.”

An editorial appeared in the Saint Petersburg Times on June 16, 2002, and it was titled “ARROGANCE IN EDUCATION.” Without going into a lot of detail of what our governor had to say about his educational policies and his wish to be remembered as “Florida’s Education Governor”, he has organized the state’s educational policy around the results of one high-stakes standardized test, the FCAT. Then he has set up the state itself as the only gauge of the entirety of our state’s public school performance. Governor Bush has maneuvered himself into the position of being the “judge of our public schools in Florida.” I find that attitude pompous and much beyond arrogance of the highest degree.

There are many issues which politicians could address involving education – more resources, more staffing, adequate buildings and better teacher pay. Politicians do a very good job when it comes to spending money and that is what they should stick to. Spend the tax dollars to help education achieve the standards which are set by the teachers, not the politicians.

ENDNOTES

9 Debra P Versus Turlington, 644 F.2d 397 5th Circuit, 1981.
10 83-3326, 11th Circuit Court of Appeals, 1984.
I. DOCUMENT IDENTIFICATION:

Title: HIGH STAKES TESTING - FLORIDA COMPREHENSIVE ASSESSMENT TEST: A TRUE MEASURE OF ACQUIRED SKILLS OR A POLITICAL RUSE?

Author(s): PAUL WAYNE DEVILLER

Corporate Source: 

Publication Date: Jun 15, 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only.

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: 

Organization/Address: 

Telephone: 

Fax: 

E-mail Address: 

Date: 

@ ERIC: 035150
III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:

Address:

Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)