



**Alternatives for Florida's Assessment
and Accountability System
Policy Brief**

Sherman Dorn
University of South Florida

Education Policy Research Unit (EPRU)
Education Policy Studies Laboratory
College of Education
Division of Educational Leadership and Policy Studies
Box 872411
Arizona State University
Tempe, AZ 85287-2411

April 2004

EPSL | EDUCATION POLICY STUDIES LABORATORY
Education Policy Research Unit

EPSL-0401-107-EPRU
<http://edpolicylab.org>

Education Policy Studies Laboratory
Division of Educational Leadership and Policy Studies
College of Education, Arizona State University
P.O. Box 872411, Tempe, AZ 85287-2411
Telephone: (480) 965-1886
Fax: (480) 965-0303
E-mail: eps1@asu.edu
<http://edpolicylab.org>

**Alternatives for Florida's Assessment
and Accountability System**

Sherman Dorn

University of South Florida

Alternatives for Florida's Assessment and Accountability System

Sherman Dorn
University of South Florida

Executive Summary

There is broad agreement that public education must be accountable. Florida's current accountability system is, however, not the only model available. Accountability does not have to mean tax money provided or withheld on the basis of test scores. This brief describes the current federal mandates for state accountability, professional standards for testing and accountability, and testing and accountability options that currently exist or have existed in practice outside Florida.

The No Child Left Behind Act does not bind states to their current plans indefinitely. Florida, therefore, has considerable freedom to change its current assessment and accountability system. The state is free to create alternatives to those components of its accountability system which are ineffective while retaining those that work well. This brief recommends that Florida continue to track student achievement and provide technical assessment assistance to low-performing schools. In addition, this brief recommends that legislators and education policymakers seeking a more effective educational accountability system in Florida enact the following recommendations:

1. *Institute a moratorium on monetary rewards and then reform the rewards system.* An effective accountability system that meets professional standards for test use and is credible to educators across the state *requires a moratorium on the monetary rewards attached to single letter grades assigned to schools.*

There may well be a method of monetizing accountability without violating professional testing standards or undermining the system's credibility.

Developing such a method involves considerable consultation with teachers across the state, as well as with testing experts and the general public.

2. *Break the tie between a single letter grade and recognition of merit in schools.* Provide different avenues for recognition:
 - a. Recognition that can be earned through test scores in one year.
 - b. Recognition that can be earned through improvement across multiple years.
 - c. Recognition attached to other measures of school performance, including measures of school violence and suspensions.
 - d. Recognition based on the use of assessment data to guide instruction—an option that is particularly important to encourage appropriate instruction for some students with disabilities and other very difficult-to-teach students, where data-driven instructional decisions may not have measurable performance improvements.
3. *Restrict the spending of any monetary rewards, especially the payment of individual staff.* The following options are less likely to cause the problems that currently exist:
 - a. Sharing the school's expertise with other schools.
 - b. Permanent salary increases for staff members, teachers, and administrators when they voluntarily transfer to low-performing schools for at least three years.

- c. One-time bonuses for staff members, teachers, and administrators when they have significant direct contact with students attending low-performing schools.
- 4. *Reduce the categories used for school accountability from five to three.* The only categories needed under any of the options above and the No Child Left Behind Act are failure, passing, and passing with distinction.
- 5. *Use testing primarily to screen for early intervention in schools.* Meeting professional standards for test use requires either an accountability system that has lower stakes or a system that accounts for measurement error and standard errors. The simpler option is to lower the stakes moderately and to use failure in the accountability system as a screening device, to select low-performing schools for intervention. Thus a failing mark in statewide testing would trigger intervention, not sanctions.
- 6. *Reform the Assistance Plus program.*
 - a. Switch from a consultant-based model to a model of on-site educational auditing. Such educational auditing teams need to be led by former classroom teachers with significant experience in instructing difficult-to-teach students and have a staff comprised of a majority of current or former teachers and specialists.
 - b. Continue development and support of assessment used throughout a school year, including support for curriculum-based measurement such as Dynamic Indicators of Basic Early Literacy Skills (DIBELS).

7. Develop and implement a pilot county-level accountability process. Making county school systems more accountable for equal educational opportunities and for student outcomes requires some process to hold county administrators responsible. Given Florida's history of racial inequality and the distrust many African-American and Latino residents feel toward county school systems, that process must extend beyond test scores and must be independent of school systems. An appropriate mechanism would be the use of grand juries to examine county school systems. California's grand juries investigate the effectiveness of local governments, and many states used to give grand juries that authority. Expanding the role of the grand jury in Florida thus has current as well as historical precedents. Granting authority to a grand jury to investigate local government would not be the first expansion of grand-jury authority in Florida: the grand jury system has in fact been used in the past to serve special needs.

Alternatives for Florida's Assessment and Accountability System

Sherman Dorn
University of South Florida

Section 1: The Issue

This brief provides policy options for K-12 assessment and accountability in Florida. A review of options is timely because of continuing concerns in the state about the package of testing and accountability provisions in Florida. Since 1999, Florida has operated its local public schools under an experiment in intensive high-stakes testing and monetized accountability. Children are tested annually in grades 3-10, and schools can receive \$100 per enrolled student depending on test results. Polls in 2002 indicated that the general public in Florida is ambivalent about the current system of high-stakes testing and accountability. Two surveys of teachers since the policy's establishment indicate that Florida educators are also ambivalent.¹ Yet the education policy debate in Florida is devoid of substantive options in testing and accountability. This brief describes the current federal mandates for state accountability, professional standards for testing and accountability, and testing and accountability options that currently exist or have existed in practice outside Florida.

A few common-sense definitions are in order. *Assessment* refers to any collection of information on student performance or skills. While the most commonly known assessments are annual standardized tests, there are many types of useful assessments in schools. A *criterion-referenced* test compares the scores of students to predetermined criteria in a subject; in Florida, these criteria are supposed to be drawn from the Sunshine

State Standards curriculum framework. A *curriculum-based measurement* test is a type of criterion-referenced test drawn from the official curriculum. These tests are given frequently in the year, and the scores should be comparable across the year. In a *norm-referenced* test, the scores of the students are compared to the scores of an original sample of students who were given an early version of the test. (This field-test sample is commonly represented as a national norm, even if it was not a random sample of the country's students.) *Accountability* refers to any method of comparing what happens in schools to the accepted obligations of schools. Accountability can include the results of annual standardized tests, but it also refers to schools' responsibilities to be financially prudent, to look after the safety of students, and to be part of a democratic society.

Section 2: Background

Why Accountability?

There is broad agreement that public education should be accountable, for a number of reasons. A school is one of the local faces of government, and while Americans are willing to spend large sums on education, they want to make sure that schools fulfill their obligations.

Education is a much bigger political issue at the state and national levels than it was before World War II. States and the federal government have funded a much greater share of elementary and secondary education in the last 40 years than at any other time in the nation's history. In return, legislators and policymakers want some control over what schools do with the money.

As state and federal politicians have paid more attention to education, legislation has looked to the schools to support political efforts such as the Cold War, the War on

Poverty, the fight against racism, and the ability of the American workforce to compete in the world economy. Each time schools were called upon in this way, the pressure on them to perform increased.²

In light of the history of education since World War II, no one should be surprised at the call for accountability in education. Florida's current accountability system is not the only model available, however. Accountability has not always been thought of as tax money given or withheld on the basis of test scores.

Testing as Accountability in Florida

Over the past 35 years, the meaning of educational accountability has gradually focused more on test results. In the early 1970s, legislators in Florida and elsewhere knew that they wanted some accountability for money spent, but they weren't sure what they wanted, and so they engaged in some experimentation, requiring state testing but without anything like the consequences in Florida today. In the late 1970s, the most notable change was the requirement that a high-school student pass a minimum competency test before graduating. In the first half of the 1990s, Florida responded to the national education reform strategy (America 2000) by creating a state equivalent (Blueprint 2000), a set of statewide curriculum standards (the Sunshine State Standards), and a new set of state tests (the Florida Comprehensive Assessment Tests, or FCAT).³

In 1999, Governor Jeb Bush signed the *A+ Plan for Education*. This law mandated two sets of grade-by-grade tests, the labeling of each local public school and charter school with a single letter grade, rewards to schools for grades of A and for grade improvements from year to year, the creation of a voucher program for students assigned to individual schools with multiple F grades, and the mandate of one-time bonuses to

teachers based on test scores.⁴ Since 2002, state intervention for specific schools—those assigned D or F grades—has been provided by the *Assistance Plus* program. The Assistance Plus plan for schools assigned D or F grades has included a mandatory summer meeting for key educators (principals and district staff), a commitment to intensive improvement plans, the requirement that districts redirect funds to support these schools, and a state-assigned Office of School Improvement set of consultants. Governor Bush has proposed a \$1.5 million budget for Assistance Plus in 2004-05 that would cover state expenses. (Districts are required to pay for local school costs.)⁵

Florida and the No Child Left Behind Act

Florida's state government created its extensive set of annual tests with high-stakes consequences several years before the passage of the federal No Child Left Behind Act (NCLB), the 2002 reauthorization of the Elementary and Secondary Education Act.⁶ This reauthorization of the largest federal school-aid program established mandates for testing and accountability policies in each state. When President Bush signed the law, Florida's existing policies were closer than most states to the new federal assessment and accountability mandates: testing children annually in math and reading in grades 3-8, testing children at least once in high school, evaluating and labeling schools based on performance in math and reading, and tying some sanctions and rewards to statewide accountability. Florida's plan relies on the Florida Comprehensive Assessment Test (FCAT) results.⁷

Professional Testing Standards

Multiple professional organizations of experts in educational and psychological testing and research have crafted standards for the fair use of testing and have made

formal recommendations regarding high-stakes testing. These documents include the *Standards for Educational and Psychological Testing*, the *Code of Fair Testing Practices in Education*, as well as position statements by the American Psychological Association, the American Educational Research Association, the National Association of School Psychologists, and the Florida Association of School Psychologists.⁸ Together these documents establish several professional standards for ethical, fair use of large-scale tests in accountability systems. These standards include the following:

1. Tests must not be used in a way that violates their technical limits; test users (including a state) must accommodate the limits of test reliability and must validate each separate use of a test.
2. Tests should not be the sole determinant of important educational decisions.
3. Test users (including a state) must guard against perverse outcomes such as teaching to the test or higher referrals to special education.

Section 3: Data

Florida's assessment and accountability system benefits the state's children and their education in three significant ways:

1. *It affirms children's rights to a high-quality education.* Official policy sets positive academic expectations for children.⁹
2. *It requires state-supported intervention for schools, including training in curriculum-based measurement at the earliest grades.* The Florida Center for Reading Research provides training in the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) to schools who have been assigned D or F grades as well as schools receiving federal reading grants. DIBELS is a form of

curriculum-based measurement for pre- and early-literacy skills. Thus far, the Florida Center for Reading Research has trained teachers primarily in using this curriculum-based measurement four times a year, but they may be used more frequently to help teachers make low-stakes classroom decisions.¹⁰

3. *It provides a state testing structure that is aligned with state standards and that includes performance tasks.* Official Florida policy dictates that the FCATs are aligned with Sunshine State Standards, the state curriculum framework. The FCATs include written performance tasks at four grade levels.

Despite these benefits, Florida's assessment and accountability system also prompts a number of concerns:

1. *Significantly more state dollars are spent on rewards for high-performing schools than are spent on intervention in low-performing schools.* The 2003-04 Florida budget includes \$0.5 million for the Assistance Plus intervention program for D and F schools and \$138 million in rewards to schools for being assigned letter grades of A or letter grades that improved between 2002 and 2003. In his budget proposal for 2004-05, Florida Governor Jeb Bush requested \$1.6 million for Assistance Plus and \$140 million for reward money.¹¹
2. *School-recognition reward money has gone disproportionately to schools in wealthier communities.* Two-thirds of the school recognition dollars in Florida have gone to low-poverty schools (schools where fewer than half of the students are eligible for free- or reduced-lunch programs). A changed

grading system in 2002 may have ameliorated that problem. The concern among some in Florida, including civil rights organizations such as the Florida NAACP chapter, is that the accountability system shifts resources away from schools that need assistance to schools that do not need extra assistance.¹²

3. *School-recognition reward money disrupts relationships within some schools.*

At some schools (though the exact number is unknown), there have been disagreements on how to spend school-recognition reward money, especially if the choice of the staff includes bonuses that are not distributed equally within the school.¹³

4. *School-recognition reward money magnifies concerns about the accountability system's inconsistencies.* Since 1999, the distribution of letter grades assigned to Florida's local public schools has changed. In 1999, 8 percent of Florida's public schools received an A; in 2003, 47 percent did. Florida residents in different occupations—from school principal to real-estate agent—have expressed doubts about changing standards. The monetary rewards attached to grades magnify those concerns. Because the reward money goes to schools that improve from one letter grade to a higher grade, a school that received a grade of B for all five years would have received no reward money, in contrast to schools that bounce up and down in assigned letter grades—an outcome some think is problematic.¹⁴

5. *Letter grades assigned to schools ignore important measures of a school's environment.* In 1999, the school-grade criteria included measures of school

suspensions. Because the schedule for calculating such statistics required the use of 1997-98 suspension statistics in the school grade for 1998-99, the state then omitted that measure beginning in 2000. Thus, while both school violence and suspensions are serious topics on many campuses, no measures of either are included in the state's accountability system.¹⁵

6. *The test cutoff scores that trigger rewards or sanctions under Florida's accountability system do not take into account test measurement errors or standard errors of group averages.* Since 1999, individual student scores and school measures have been treated as exact measurements by the state. Assuming that a test score is exact ignores the fact that in all tests, there is both *measurement error* (the uncertainty in a single student's score on a single test) and *standard errors of means* (the uncertainty that an average of test scores for a class, school, or county is truly representative of the group). Thus, two schools in Florida can receive different letter grades while their students have attained test scores that are indistinguishable from a statistical standpoint.¹⁶
7. *Statewide tests can vary in format even when the tests have identical consequences.* The state's accountability system relies almost entirely on results from the Florida Comprehensive Assessment Tests (FCATs). An individual test that is part of the FCATs will require written responses depending on the examination's role within the system, the students' grade level, and the time of year. FCAT reading and math tests sometimes demand written responses and sometimes do not, depending on the grade levels of the

students. In 2001, Florida Education Commissioner Charlie Crist removed the written-response items from the calculation of school grades for 2001 grades and *only* 2001 grades. In addition, FCAT reading and math tests for grade 10—a gatekeeper for high school graduation—do not have a consistent format. Students in the 10th grade initially take tests containing written-response questions, but those given the opportunity to take the test again do so without written-response items. (Students who did not qualify for high-school graduation are retested.)¹⁷

8. *The FCAT is administered early in the second half of the year, but scores typically are returned at the end of the school year.* Students take a writing test—FCAT Writes—in February and the rest of the FCATs in March. But scores are typically issued only in the last few weeks of school (or, in 2000, after the end of the school year).¹⁸ There are two concerns about this timeline: Testing in the winter does not reflect what students learn through an entire academic year; and, late returns on scores do not allow for timely intervention.¹⁹
9. *One set of statewide tests is connected neither to the Sunshine State Standards nor to the accountability system.* In 1999, the *A+ Plan for Education* mandated statewide norm-referenced tests. In practice, these have included one sub-test from reading and one from math in an off-the-shelf commercial norm-referenced test, and they are given in the second week of the state’s March testing schedule. These tests are not aligned with the Sunshine State Standards, and they have no role in the state’s accountability system.

10. The state's intervention program for low-performing schools uses individual consultants as the primary route to maintain contact with and advise schools.

While there are occasional statewide or regional meetings as part of Assistance Plus, most state-sponsored assistance consists of contacts between educators at individual schools and designated state consultants.²⁰ There is some question, however, whether such a model is the best option for intervention at the school level. In Florida, contacts have varied dramatically by individual school, ranging from a low of 14 documented contacts (with the Academic Research Center in Polk County) between mid-November and mid-January to a high of 60 contacts in the same period (with the Eastside Multicultural School). In addition, it is questionable how many of those contacts are part of a coordinated approach or how many focus on the key skills and resources teachers in low-performing schools need. The opportunity to fit in with the school's improvement plan or to coordinate activities with other state consultants is not present, as each advisor visits on different dates.²¹ These concerns parallel the experiences of other programs, such as Chicago's school reforms, which also have relied on consultants.²²

11. No specific accountability provisions exist for counties in terms of rewards, sanctions, or intervention. Counties are required to provide resources to schools with D or F grades, but the state's accountability system does not provide for extensive reports on a county's performance in the same way that the state assigns letter grades to schools, nor does the system provide for intervention in the operations of county school systems.²³

There are two reasons to consider a county-level accountability process. Local public schools and charter schools are the statutory responsibility of county school boards in Florida. Historically, county school systems constituted the level of organization most resistant to desegregation in the 1960s, and at least a significant number of older African-American Floridians distrust the ability of both county and state governments to preserve their interests without some check on the authority of one branch of government.²⁴

Federal accountability mandates

The federal No Child Left Behind Act (NCLB) creates some mandates for state testing and accountability, but few state plans have the same intensity of testing and consequences as Florida's assessment and accountability system: other state plans meet the No Child Left Behind mandate for annual progress in substantially different ways.²⁵ Given the variation in state plans that the federal Department of Education has already approved, it is worth noting what the federal law requires at a minimum—and what it does not:

1. *Annual assessment in several subjects in grades 3-8, with at least one assessment in high school.* The law does not prohibit performance assessments or more frequent assessments, however.²⁶
2. *A state judgment of individual schools.* The law requires that a state decide whether each school is meeting Annual Yearly Progress goals. It does not require more than that single pass/fail decision for a school and does not require Florida's A-through-F grading scale.²⁷
3. *Rewards and sanctions attached to state judgment.* The law requires that a state have *some* rewards and sanctions attached to the Annual Yearly Progress

declaration. The law is more specific about the sanctions (which include a threat to reassign administrators and teachers) than about the rewards, and the law does not require monetary rewards.²⁸

Federally Permissible Operating Alternatives

Based on the foregoing, Florida has the following options for its accountability plan:

1. *Non-monetary recognition of excellent performance.* There are many honors that educators and schools work for other than monetary rewards. Since the 1982-83 school year, the U.S. Department of Education has identified Blue Ribbon Schools for documenting best practices. School districts have sought recognition from a variety of programs such as the Malcolm Baldrige Quality Award, its equivalent in many states (including the Florida Governor's Sterling Award), or the RIT/USA Today Quality Cup.²⁹ Non-monetary systems are appropriate both for a moratorium on financial incentives and also as a permanent alternative to financial incentives.
2. *Rewards and sanctions systems with multiple opportunities for recognition.* North Carolina's accountability system offers multiple opportunities for identification as a school with a distinctive record, with separate opportunities for recognition for student performance in a single year and for growth across several years.³⁰
3. *Fewer categories for school accountability.* The No Child Left Behind Act requires only the identification of schools as passing or failing the state's system.

4. *On-site educational audits.* England, Rhode Island, New York, Maine, and Illinois send teams of experienced educators to conduct intensive examinations of schools. Beyond the paper auditing of high-school accreditation and special-education compliance reviews, these on-site visits focus on what happens inside classrooms, between teachers and students. The audit teams work together and produce reports that include recommendations. In Rhode Island, these reports have been presented publicly to both the school and to the community. In the experience of Rhode Island, each on-site audit visit costs between \$3,000 and \$4,000. Because of the training and time they require, educational audits cannot be conducted for every school every year, but only on a rotating basis.³¹
5. *Regular assessment throughout a school year.* The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) measures of emerging literacy skills are part of a larger set of measures developed in the last 20 years that can be used throughout the school year to guide instruction. The use of such curriculum-based measurement tests is supported by the research of dozens of specialists in special education working in real classrooms. One school system in Minnesota mandated curriculum-based measurement for every child as a tool to identify academic problems early and to prevent spiraling failure.³²
6. *Two-level accountability system.* Maine's accountability system includes data collection and judgments made through a Local Assessment System, and it has included that two-level system in its June 2003 plan to meet No Child Left Behind Act standards.³³

Section 4: Data Quality

Several factors complicate any discussion of assessment and accountability system alternatives. Operating accountability systems typically reflect several simultaneous policy changes, and Florida is no exception. Florida education law changed in several ways in 1999; to the extent that student outcomes have improved or declined, one cannot decisively determine which policy is responsible for what outcomes. Some insight, however, can be gained by comparing states with a range of policies.³⁴

An additional obstacle is the inevitable lag time in the evaluation of major policy initiatives. Yet legislators have a legitimate need to address the difficulty of implementation reforms. Florida's intensive system of assessment and accountability has been in place since 1999, and there are few published evaluations, let alone evaluations published in peer-reviewed journals. A major study of Florida's failing school-voucher program conducted by reputable economists is unpublished as of the writing of this brief (even in working-papers form). Part of the difficulty is the lack of data in some important areas. For example, the Office of School Improvement did not begin tracking Assistance Plus consultant contacts until November 2003, more than a year after the start of the program. In some cases, it is clear that the contact in an individual activity report is by e-mail or telephone conference call. One may also expect that some itinerant consultants have not yet made a habit of reporting activities promptly. Thus, the information about Assistance Plus contacts must be considered tentative.

Section 5: Findings

The No Child Left Behind Act does not bind states to their current plans forever.³⁵ Florida has considerable freedom to change its current assessment and accountability

system to better achieve desired ends while preserving aspects of the system that are working.

Some elements of the state's accountability system are solidly supported by research. They include:

1. *The existence of statewide tests tracking student achievement.* Despite concerns about the form of the Florida Comprehensive Assessment Tests (FCAT) or their multiple uses in Florida's accountability system, there is considerable support for some form of assessment that keeps track of student achievement and that could be used to guide instruction and for accountability purposes. The professional standards for educational testing, described above, affirm the value of assessment where constructed and used appropriately.
2. *Training in the use of curriculum-based measurement in Assistance Plus and reading-grant schools.* The Florida Center for Reading Research has trained dozens of teachers around the state in early elementary reading assessments and has established a Web site to help teachers and principals track the developing literacy skills of younger students. This type of assessment takes little time away from classroom instruction, can be conducted by non-teaching staff members, and can help teachers target individual students for assistance during the year. The Florida Center for Reading Research warns against using the assessments either to evaluate teachers or to make decisions about retaining students in a grade, so the assessments are unlikely to be distorted by efforts to teach to the test. The Florida Center for Reading Research has created versions in multiple languages, including Spanish and Haitian Creole,

so performance is much less likely to be affected by the language of origin.

All of these practices conform to professional standards for test construction and use.

Areas of Concern and Some Working Alternatives

There are several areas of concern in Florida for which the examples of other states provide reasonable alternatives. There are feasible policies in operation in real schools that are alternatives to problematic elements of the state's accountability system.

These include:

1. *The use of a simple monetary reward for single letter grades.* There are several types of alternatives with real-world examples: recognition programs without money attached and monetary incentive programs with multiple opportunities to earn recognition.
2. *The assigning of letter grades as part of accountability.* Several states have accountability systems with fewer categories—some just noting whether or not schools have met NCLB Annual Yearly Progress goals.
3. *Reliance on annual standardized testing for accountability.* Before the No Child Left Behind Act, Rhode Island relied on educational audits for part of its accountability system. The New England Compact states are developing more qualitative tests in response to the No Child Left Behind mandates. At least one school system has relied instead on curriculum-based measurement as its key tool to prevent early reading difficulties.
4. *High-stakes accountability that does not take account of tests' technical limits.* Colorado, Iowa, and Kansas use statistical confidence intervals—a

way of adjusting for standard errors of means—to ensure that chance is unlikely to be responsible for labeling schools as failing.³⁶

5. *Consultant-based intervention programs that may be inconsistent and uncoordinated.* On-site educational audit systems of some form have been used in four states and England.³⁷
6. *Prevention of undesirable consequences.* North Carolina's Testing Code of Ethics forbids "reclassifying students solely for the purpose of avoiding state testing."³⁸ This code thus declares it unethical to assign students to special education or to push them out of school in order to eliminate test scores from the state's accountability system.

Needs With No Clear Solution

In contrast to the areas of concern above, for which feasible alternative policies exist as models, there are some necessary refinements for which a readily available solution or model does not yet exist. These include:

1. *Guidelines for spending reward money in a way that does not undermine educator morale.* There is no example of a state with high-stakes testing and monetary rewards that guides the spending of money in a well-documented fashion. Florida has the largest experiment in such monetary rewards, and there is persistent evidence that giving the choice of spending reward money on bonuses for teachers and non-teaching staff has led to infighting and disruption in some schools.
2. *County-level accountability that is independent of school systems.* The closest example of a state with multiple levels of accountability—Maine's current

system—does not have the history of segregation and deeply-rooted mistrust that exists in Florida.

Section 6: Recommendations

Legislators and education policymakers seeking a more effective educational accountability system in Florida are advised to enact the following recommendations:

1. *Institute a moratorium on monetary rewards and then reform the rewards system.* An effective accountability system that meets professional standards for test use and is credible to educators across the state *requires a moratorium on the monetary rewards attached to single letter grades assigned to schools.* There may well be a method of monetizing accountability without violating professional testing standards or undermining the system's credibility. Developing such a method involves considerable consultation with teachers across the state, as well as with testing experts and the general public.
2. *Break the tie between a single letter grade and recognition of merit in schools.* Provide different avenues for recognition:
 - a. Recognition that can be earned through test scores in one year.
 - b. Recognition that can be earned through improvement across multiple years.
 - c. Recognition attached to other measures of school performance, including measures of school violence and suspensions.
 - d. Recognition based on the use of assessment data to guide instruction—an option that is particularly important to encourage appropriate instruction for some students with disabilities and other very difficult-to-teach

students, where data-driven instructional decisions may not have measurable performance improvements.

3. *Restrict the spending of any monetary rewards, especially the payment of individual staff.* The following options are less likely to cause the problems that currently exist:
 - a. Sharing the school's expertise with other schools.
 - b. Permanent salary increases for staff members, teachers, and administrators when they voluntarily transfer to low-performing schools for at least three years.
 - c. One-time bonuses for staff members, teachers, and administrators when they have significant direct contact with students attending low-performing schools.
4. *Reduce the categories used for school accountability from five to three.* The only categories needed under any of the options above and the No Child Left Behind Act are failure, passing, and passing with distinction.
5. *Use testing primarily to screen for early intervention in schools.* Meeting professional standards for test use requires either an accountability system that has lower stakes or a system that accounts for measurement error and standard errors. The simpler option is to lower the stakes moderately and to use failure in the accountability system as a screening device, to select low-performing schools for intervention. Thus a failing mark in statewide testing would trigger intervention, not sanctions.
6. *Reform the Assistance Plus program.*

- a. Switch from a consultant-based model to a model of on-site educational auditing. Such educational auditing teams need to be led by former classroom teachers with significant experience in instructing difficult-to-teach students and have a staff comprised of a majority of current or former teachers and specialists.
 - b. Continue development and support of assessment used throughout a school year, including support for curriculum-based measurement such as Dynamic Indicators of Basic Early Literacy Skills (DIBELS).
7. *Develop and implement a pilot county-level accountability process.* Making county school systems more accountable for equal educational opportunities and for student outcomes requires some process to hold county administrators responsible. Given Florida's history of racial inequality and the distrust many African-American and Latino residents feel toward county school systems, that process must extend beyond test scores and must be independent of school systems. An appropriate mechanism would be the use of grand juries to examine county school systems. California's grand juries investigate the effectiveness of local governments, and many states used to give grand juries that authority.³⁹ Expanding the role of the grand jury in Florida thus has current as well as historical precedents. Granting authority to a grand jury to investigate local government would not be the first expansion of grand-jury authority in Florida: the grand jury system has in fact been used in the past to serve special needs.⁴⁰

Notes and References

¹ Bridges, T. (2002, June 14). Voters Cool to Bush's A+ Plan, Poll Says. *Miami Herald*.

Jones, B. D. & Egley, R. J. (in press). Voices from the Frontlines: Teachers' Perceptions of High-Stakes Testing. *Education Policy Analysis Archives*, 12.

Abrams, L. M. (2004, February). Teachers' Views on High-Stakes Testing: Implications for the Classroom. *Educational Policy Research Unit*, Doc. No. EPSL-0401-104-EPRU. Tempe, AZ: Education Policy Studies Laboratory.

² One could say the same about governors, that they are all education governors, Erwin V. Johanningmeier, personal communications. For a more complete discussion of this topic, see:

Dorn, S. (1998, January 1). The Political Legacy of School Accountability Systems. *Education Policy Analysis Archives*, 6, 1. Retrieved December 20, 2003, from <http://epaa.asu.edu/epaa/v6n1.html>

Kaestle, C. F. & Smith, M. S. (1982). The Federal Role in Elementary and Secondary Education, 1940-1980. *Harvard Educational Review*, 42, 384-408.

³ Michael, D. (2003, November 2). Equity vs. Adequacy: A Comparison of Governor Ruben Askew's and Governor Jimmy Carter's Educational Reforms. Paper presented at the Annual Meeting of the History of Education Society, Evanston, IL.

Office of Program Policy Analysis and Government Accountability (2001, April). Justification Review: Kindergarten through Twelfth Grade Public Education Program. Report No. 01-22, p. 12. Retrieved from <http://www.oppaga.state.fl.us/monitor/reports/pdf/0122rpt.pdf>

Florida Association of School Psychologists, (2002, November). Position Paper on the Use of the Florida Comprehensive Assessment Test (FCAT) in High Stakes Decision Making. Retrieved from <http://www.fasp.org/PDFfiles/FASP%20FCAT%20Paper.pdf>

Dorn, S. (2002, February). Reforming the Structure of Florida's Accountability System. *Educational Policy Research Unit*, Doc. No. EPSL-0401-106-EPRU. Tempe, AZ: Education Policy Studies Laboratory.

⁴ Florida Chapter 99-398, available on-line at http://election.dos.state.fl.us/pdf/99laws/ch_99-398.pdf

⁵ For information, see:

Florida Statutes 1008.33

<http://schoolgrades.fl doe.org/assistance-plus.cfm>

Governor Jeb Bush, (2004). Proposed Florida 2004-05 budget. Retrieved from http://www.ebudget.state.fl.us/govpriorities/education/assist_plus.asp

⁶ No Child Left Behind Act of 2001, Public Law 107-110, signed into law January 8, 2002.

⁷ In NCLB terms, annual yearly progress is a measure each state chooses to show progress toward all students' having competence in math, reading, and any other subject the state chooses to add. Florida's plans are available at <http://www.ed.gov/admins/lead/account/stateplans03/index.html>

⁸ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing* (rev. ed.). Hanover, PA: Authors.

Joint Committee on Testing Practices (2004). *Code of Fair Testing Practices in Education*. Draft available on-line at <http://www.apa.org/science/FinalCode.pdf>

American Educational Research Association (2000, July). Position Statement Concerning High-Stakes Testing in PreK-12 Education. Retrieved from <http://www.aera.net/about/policy/stakes.htm>

National Association of School Psychologists (2003). *Position Statement on Using Large Scale Assessment for High Stakes Decisions*. Retrieved from http://www.nasponline.org/information/pospaper_largescale.html

National Association of School Psychologists (2002). *Large scale assessments and high stakes decisions: Facts, cautions and guidelines*. Bethesda, Md.: Author.

American Psychological Association (2001). *Appropriate Use of High-Stakes Testing in Our Nation's Schools*. Retrieved from <http://www.apa.org/pubinfo/testing.html>

Florida Association of School Psychologists (2002, November). Position Paper on the Use of the Florida Comprehensive Assessment Test (FCAT) in High Stakes Decision Making. Retrieved from <http://www.fasp.org/PDFfiles/FASP%20FCAT%20Paper.pdf>

⁹ Dorn, S. (2004, February). Reforming the Structure of Florida's Accountability System. *Educational Policy Research Unit*, Doc. No. EPSL-0401-106-EPRU. Tempe, AZ: Education Policy Studies Laboratory.

¹⁰ The Florida Center for Reading Research's web page for assessment is at <http://www.fcrr.org/assessment/index.htm>

For a description of the research base, see:

Deno, S. L. (1985). Curriculum-Based Measurement: The Emerging Alternative. *Exceptional Children*, 52, 219-32.

Stecker, P. M. & Fuchs, L. S. (2000). Effecting Superior Achievement Using Curriculum-Based Measurement: The Importance of Individual Progress Monitoring. *Learning Disabilities Research and Practice*, 15, 128-34.

Shinn, M. R. (Ed.) (1998). *Advanced Applications of Curriculum-Based Measurement*. New York: Guilford Press.

Hall, T. & Mengel, M. (n.d.). Curriculum-Based Evaluations. Wakefield, MA: National Center on Accessing the General Curriculum. Retrieved from <http://www.cast.org/ncac/classroompractice/cpractice03.pdf>

¹¹ See, for example:

http://www.ebudget.state.fl.us/govpriorities/education/assist_plus.asp

http://www.ebudget.state.fl.us/govpriorities/education/school_recog.asp

¹² Office of Program Policy Analysis and Government Accountability (2000, August). Florida Actions Should Improve Student Performance in High-Poverty Schools. Report No. 00-07. Available on-line at <http://www.opлага.state.fl.us/reports/pdf/0007rpt.pdf>

Davis, C. & Doig, M. (2003, July 20). FCAT Helping 'Rich' Schools Get Richer. *Sarasota Herald-Tribune*.

- Waite, M. (2003, July 6). All FCAT A's Are Not Created Equally. *St. Petersburg Times*.
- Tschinkel, W. R. (2003, February 3). New and Improved A+ Grades: Camouflaged Bias. Unpublished commentary. Available at http://bio.fsu.edu/~tschink/school_performance/unpublished.html
- Davis, B. R. (2003, May 31). District's Best Teachers Are in 'F' Schools (letter to the editor). *Miami Herald*.
- ¹³ Hegarty, S. (2002, August 31). Schools Reap Cash Rewards for Grades. *St. Petersburg Times*
- Davis, C. & Doig, M. (2003, April 28). FCAT Free-for-All. *Sarasota Herald-Tribune*.
- ¹⁴ A school that started with a D in 1999 and bounced up to a C in 2000, down to a D in 2001, up to a C in 2002, and up to a B in 2003 would have earned reward money in three of those years.
- Waite, M. (2003, July 6). All FCAT A's Are Not Created Equally. *St. Petersburg Times*.
- Pinzur, M. I. (2003, October 21). Paradoxes in School Bonuses for FCAT Success. *Miami Herald*.
- ¹⁵ See:
- Eitle, D. & Eitle, T. M. (2003, December). Segregation and School Violence. *Social Forces*, 81.
- Raffaele, L. M. (1999). *An Analysis of Out-of-School Suspensions in Hillsborough County*. Tampa, FL: Hillsborough Constituency for Children.
- ¹⁶ For the brief description of school grade formulae since 1999, see <http://schoolgrades.fl doe.org>
- ¹⁷ These written responses may be short, extended, or be the entire exam, the last only with FCAT Writes (the state's writing test). Generally, the tests with written responses include the following: all FCAT Writes tests (at grades 4, 8, and 10); FCAT criterion-referenced reading tests at grades 4, 8, and 10; FCAT criterion-referenced math tests at grades 5, 8, and 10; and all FCAT criterion-referenced science tests (at grades 5, 8, and 10). The state's norm-referenced tests in reading and math include only multiple-choice options, as do criterion-referenced tests for reading for grades 3, 5, 6, 7, and 9; and for math in grades 3, 4, 6, 7, and 9. For additional information about FCAT, see <http://www.firn.edu/doe/sas/fcat.htm>.
- For the change in 2001, see:
- Brown, M. (2001, March 6). Late FCAT Changes Criticized. *Tampa Tribune*.
- Hegarty, S. (2001, February 11). Psst: Some Test Questions Don't Count. *St. Petersburg Times*.
- ¹⁸ Florida Department of Education (n.d.). History of Statewide Assessment Program. Tallahassee, FL: Author. Retrieved from <http://www.firn.edu/doe/sas/hsap/hsap9000.htm#2000>.
- Scores did not come back on the 2000 tests until June.
- ¹⁹ Hegarty, S. (2000, May 28). Patience Tested by FCAT Delays. *St. Petersburg Times*.
- ²⁰ The Office of School Improvement web site is (<http://osi.fsu.edu>). The database of individual consultant contacts is (<http://osi.fsu.edu/AplusPla.nsf>). The Office of School Improvement has 38 regular employees who support the office's multiple missions (that include statewide reporting of school improvement plans in general and reports of school-recognition funds as well as Assistance Plus operations).
- ²¹ In one site visit in late 2003, a state consultant modeled the use of Apple iBooks for teachers at a Hillsborough County charter school that had been assigned an F letter grade for 2002-03. The in-service demonstration did not align with the school improvement plan, which made no mention of laptops as a

strategy in its key goals. The consultant's job (part-time at a state-funded university center) is to help teachers integrate technology into classroom use.

2003 School Improvement Plan (2003, October 9). Tampa United Methodist Charter School. Retrieved from http://146.201.17.41/s_i_p_n/print_html.asp?sessid=596406977

Worley, G. (2003, December 15). Report on Tampa United Methodist Charter activities. Retrieved from <http://osi.fsu.edu/AplusPla.nsf/75000b4571d081ca85256df800556de8/74ec282964e8379285256e020011c4ef?OpenDocument> Worley is an Educational Technology Integrator at the Florida Center for Instructional Technology. It should be stressed that many consultants are providing time to the Assistance Plus program. The concern expressed here is not about their willingness to help schools but about the overall structure of Assistance Plus.

²² Finnegan, K. & O'Day, J. (2003). *External Support to Schools on Probation: Getting a Leg Up?* Chicago: Consortium on Chicago School Research. Retrieved from <http://www.consortium-chicago.org/publications/pdfs/p63.pdf>

Newmann, F. M. & Sconzert, K. (2000). *School Improvement with External Partners*. Chicago: Consortium on Chicago School Research. Retrieved from <http://www.consortium-chicago.org/publications/pdfs/p0c01.pdf>

Duffrin, E. (2002, June). Seen as a Model, Manley Plan Falls Short. *Catalyst*. Retrieved from <http://www.catalyst-chicago.org/06-02/0602main1.htm>

Duffrin, E. (2002, October). Accountability Impact both Positive, Negative. *Catalyst*. Retrieved from <http://www.catalyst-chicago.org/10-00/1000accountability.htm>

Williams, D. (1998, December). A Third of Schools Switch Partners. *Catalyst*. Retrieved from <http://www.catalyst-chicago.org/12-98/128switch.htm>

Williams, D. (1998, December). But Are They Any Good? *Catalyst*. Retrieved from <http://www.catalyst-chicago.org/12-98/128good.htm>

Daniels, T. (2002, March). Partisan bickerings continue to disrupt monthly Chicago Teachers Union meetings. *Substance*. Retrieved from <http://www.substancenews.com/March02/ctumeeting1.html>

End of year energy' marks June House meeting (2003, June). *Chicago Union Teacher*, 6.

²³ Florida Statutes 1001.42(16)(a).

²⁴ County school boards were not alone. Florida Governor Claude Kirk (1967-71) was the most staunchly anti-integration governor in Florida (at some points in opposition to county school board wishes).

Sanders, R. (2002). Rassling a Governor: Defiance, Desegregation, Claude Kirk, and the Politics of Richard Nixon's Southern Strategy. *Florida Historical Quarterly*, 80, 332-59.

For an example of remaining African American skepticism about government authority, see:

Dawson, E. M. (1998, June). The Preservation of Florida's Historically All-Black School Records. *Annotation*, 26. Retrieved from http://www.archives.gov/grants/annotation/june_98/school_records.html

²⁵ Maine incorporates locally determined tests; other states, such as Texas, have different student-number (not test-score) thresholds for the minimum number of children in a demographic group within a grade for the assessments to be broken down for that group. While Florida reports scores for any school where there are at least 30 students in a demographic group (e.g., African American fourth-graders), other states do not

look at a population group unless there are 45 or more students in that group within a grade. Other state plans are available at: <http://www.ed.gov/admins/lead/account/stateplans03/index.html>

The plans of Maine and Nebraska, in particular, are different from that of Florida's in several respects. The Enhanced Assessment Project of the New England Compact is described at <http://www.necomcompact.org/enhanced.htm>. This brief does not analyze the appropriateness of the federal law. There is ongoing debate about the balance between accountability and resources, among other issues, but that is beyond the scope of this discussion.

²⁶ No Child Left Behind Act §1111(b)(3).

²⁷ No Child Left Behind Act §1111(b)(2)(B).

²⁸ No Child Left Behind Act §1111(b)(2)(A)(iii).

²⁹ The Malcolm Baldrige Quality Award web site is <http://www.quality.nist.gov>. The list of all Blue Ribbon Schools from 1982 through 2002 is at <http://www.ed.gov/programs/nclbbrs/list-1982.pdf>. The Quality Cup web site is at <http://www.qualitycup.org/>, maintained as an archive though the program ended in 2001.

³⁰ North Carolina Department of Public Instruction (2004, January 20). Determining Composite Scores in the ABCs Model. Retrieved from <http://www.ncpublicschools.org/Accountability/reporting/DeterminingCompositeScoresABCs04.pdf>

³¹ Wilson, T. A. (1995). *Reaching for a Better Standard: English School Inspection and the Dilemma of Accountability for American Public Schools*. New York: Teachers College Press.

Baker, P., Rau, W., Ashby, D., & Harper, A. (1999). Quality Assurance Improvement Planning: Second Annual Report. Retrieved from <http://www.coe.ilstu.edu/accountability/2nd%20Annual%20Report.htm>

Ancess, J. (1996). *Outside/Inside, Inside/Outside: Developing and Implementing the School Quality Review*. New York: National Center for Restructuring Education, Schools, and Teaching.

E-mail, (2001, January 17). Rick Richards to Roderick Colbert.

Smith, D. R., & Ruff, D. J. (1998). The Southern Maine Partnership's School Quality Review. In David Allen (Ed.). *Assessing Student Learning*. New York: Teachers College Press.

Rhode Island's School Accountability for Learning and Teaching web site is at <http://www.ridoe.net/schoolimprove/salt/default.htm>

Southern Maine Partnership web site is at <http://www.usm.maine.edu/smp/about/mission.htm>

³² Howe, K. B., Scierka, B. J., Gibbons, K. A., & Silbergliitt, B. (n.d.). A Schoolwide Organization System for Raising Reading Achievement Using General Outcome Measures and Evidence-Based Instruction: One Education District's Experience. Unpublished manuscript. Retrieved from <http://www.edformation.com/PDFs/School-wide%20organ%20w%20revise.pdf>

³³ Maine Department of Education (2003, June 6). Consolidated State Application Accountability Workbook (rev. ed.). Retrieved from <http://www.state.me.us/education/nclb/state%20app/Con%20App%20Workbook%20013103.doc>

Hickok, E. W. to Gendron, S. A. (2003, July 1). Retrieved from <http://www.ed.gov/admins/lead/account/letters/me.doc>

³⁴ For examples of multi-state analyses, see:

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving Student Achievement: What State NAEP Test Scores Tell Us*. Publication MR-924-EDU. Santa Monica, CA: RAND.
Retrieved from <http://www.rand.org/publications/MR/MR924/>

Amrein, A. L. & Berliner, D. C. (2002, December). The Impact of High-Stakes Tests on Student Academic Performance: An Analysis of NAEP Results in States with High-Stakes Tests and ACT, SAT, and AP Test Results in States with High School Graduation Exams. *Education Policy Research Unit*. Doc. No. EPSL-0211-126-EPRU. Tempe, AZ: Educational Policy Studies Laboratory. Retrieved from <http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0211-126-EPRU.pdf>

Carnoy, M. & Loeb, S. (2003). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Education Evaluation and Policy Analysis*, 24, 305-31.

³⁵ No Child Left Behind Act §1111(b)(1)(F).

³⁶ See the NCLB state plans at <http://www.ed.gov/admins/lead/account/stateplans03/index.html>

³⁷ Wilson, T. A. (1995). *Reaching for a Better Standard: English School Inspection and the Dilemma of Accountability for American Public Schools*. New York: Teachers College Press.

³⁸ 16 NCAC 6D .0306(g)(6). Retrieved from
<http://www.ncpublicschools.org/accountability/testing/policies/testcode080100.pdf> The Testing Code of Ethics allows and encourages test preparation, a policy at odds with the recommendations of this brief.

³⁹ Jameson, M. (2000). The Grand Jury: An Historical Overview. California Grand Jurors Association.
Retrieved from <http://www.nvo.com/cgja/nss-folder/briefhistoryofgrandjuries/Grdjry.rtf>

⁴⁰ The authority to draw funds independently is in Florida Statutes 125.59, and the description of the statewide grand jury system is in Florida Statutes 905.31-905.40.