

A Comprehensive Assessment of the A-Plus Plan's Impact on Public Education

March 2007



Sherman Dorn



<http://www.civicconcern.org>

Copyright © 2007 by CIVIC Concern, Inc., and Sherman Dorn.

All rights reserved.

CIVIC Concern
104 South Monroe Street
Tallahassee, Florida 32301

CIVIC Concern is dedicated to providing Floridians with plain-language, quickly digestible discussions of the public policy that affects their daily lives and to empowering them to make their voices heard.

CIVIC Concern's mission is to bring renewed recognition to the Florida Constitution's overarching principle that all power is inherent in the people. CIVIC engages in the political process as an advocate of ordinary Floridians. CIVIC serves as an incubator of reasoned public policy and work to ensure government accountability. CIVIC provides Florida voters with valuable information and access, empowering them to demand government action.

Contents

| | |
|--|----|
| Executive Summary..... | 1 |
| <i>Eleven Questions</i> | 1 |
| <i>Findings</i> | 3 |
| A Comprehensive Assessment of the A-Plus Plan’s Impact on Public Education | 6 |
| <i>Independent Evidence of Academic Achievement</i> | 6 |
| <i>What Is Responsible for Achievement Trends</i> | 10 |
| <i>What the FCAT Can and Cannot Measure</i> | 17 |
| <i>Extending the Grading System</i> | 21 |
| <i>Minimizing Growth</i> | 24 |
| <i>Classroom Side Effects of the A-Plus Plan</i> | 26 |
| <i>The Best Research on Vouchers</i> | 28 |
| <i>Other Concerns Losing Attention</i> | 29 |
| <i>Helping Students</i> | 32 |
| <i>Third-Grade Retention</i> | 34 |
| <i>An Adequate Education in Florida</i> | 36 |
| References | 39 |
| About the author | 47 |

[This page intentionally left blank.]

A Comprehensive Assessment of the A-Plus Plan's Impact on Public Education

Executive Summary

In 1999, the Florida legislature passed the A-Plus Plan (Fla. Statutes ch. 99-398), which mandated annual standardized tests for all grades 3–10, established a system to assign a single letter-grade to each local public and charter school based on those tests, created monetary rewards based on grading schools, established a public-choice program for students in schools labeled **F**, and established a voucher program for students in schools labeled **F**. The former governor's administration is recognized by both supporters and opponents for an assertive and innovative education reform program, one that was groundbreaking in several ways for the state and for the country.

Eight years later, it is time to assess this package of reforms. A complete assessment would take several volumes (and several books *have* been written on Florida's education policy since 1998), but this report summarizes evidence that currently exists.

Eleven Questions

This report answers eleven key questions evaluating the current state of public education in Florida and looking forward to future policy changes.

1. What independent evidence exists about the status and trends of academic achievement as well as the achievement gap between minority students and white students?
2. What might explain the difference in Florida Comprehensive Assessment Test (FCAT) trends and independent data between elementary and secondary grades?

3. What are the strengths and limits of the FCAT in describing student and school performance?
4. Are indicators such as Florida's public high school graduation rate, SAT and ACT scores, and National Assessment of Educational Progress (NAEP) scores tied directly to state education policy?
5. Does the A-Plus Plan currently focus on a narrow range of achievement or recognize and reward substantial improvement in the achievement of individual students?
6. What evidence is available regarding schools' narrowing the curriculum and engaging in inappropriate test preparation in response to state education policy?
7. Do we know if vouchers have helped poor children and those with disabilities improve their academic performance? What research describes the overall impact of vouchers, opportunity scholarships, McKay scholarships, and corporate income tax scholarships on public education?
8. Has the A-Plus Plan's grading system's reliance on so few factors drawn attention away from important concerns in non-FCAT areas?
9. How is Florida providing remediation and other services to help low-performing students? How does that intervention compare with recommended practices based on research?
10. Is there clear evidence that retaining children in third grade who fail the reading portion of the FCAT is helping or hurting students?
11. What are the key questions to ask about the adequacy of Florida's student funding?

Findings

The following conclusions are drawn from a broad range of evidence and the research literature.

Academic achievement trends are mixed. The independent National Assessment of Educational Progress (NAEP) scores show mixed results for Florida students in reading and math between 1998 and 2005. Scores have not improved for eighth-grade students in reading, fourth-grade reading scores have stagnated during the last few years, and few changes in the achievement gaps among different population subgroups are statistically significant.

It is impossible to explain academic achievement trends with the A-Plus program. After accounting for improvement trends that began before the 1999 A-Plus education reform package, only elementary-grade achievement appears to have improved. But factors other than accountability may be responsible for these changes.

The FCAT is a valuable assessment tool, but it is limited. The FCAT can describe achievement in reading, writing, math, and science. The FCAT

- Does not test in other subjects and cannot assess skills beyond paper-and-pencil demonstrations.
- Has limited scope within the covered subjects.
- Is less accurate than the Florida Department of Education often implies.
- Cannot distinguish academic knowledge from test-wisness
- Cannot responsibly link individual teachers to student achievement.

Florida's grading system could extend beyond the FCAT, but several commonly-mentioned indicators should not be included in such an extension. The state should not tie the accountability system to the current flawed official high school graduation measure, SAT and ACT scores, or NAEP scores, but there are other ways to expand accountability to include non-FCAT measures. These could include independent and rigorous surveys of school climates and measures of how schools challenge students. In addition, if the grading system included more than one year of data, other longitudinal and lagging measures (such as data on violence in schools) could be included.

Florida's grading system does not reward large improvements in achievement. The mechanisms of school grading within the A-Plus Plan focus on students' meeting minimum thresholds. While there is a growth component, students who barely meet the growth threshold count as much as students who far exceed the growth threshold.

Florida's grading system has led to widespread emphasis on test preparation and other inappropriate responses by schools. Florida schools are likely to respond to test pressures by reducing instruction in untested subjects and by focusing on test-wiseness rather than general student achievement.

There is no trustworthy research basis to support vouchers as a method to encourage improvement in Florida public schools. The best research in peer-reviewed economic journals refutes claims that the existence of voucher programs has improved Florida's schools.

Florida's accountability system draws attention away from other issues. The grading of schools dominates news coverage, making some topics more difficult to raise.

Few state-directed remediation efforts for students with low academic achievement have a basis in research. Of various state efforts, only the activities of the Florida Center for Reading Research have significant support in the research literature.

Florida's third-grade retention policy has little support in research. There is no clear evidence that retaining children in third grade helps Florida students, despite a fourfold increase in retention between the 2001–2002 and 2002–2003 school years.

Florida needs a definition of a high-quality education and an analysis of the resources and systems necessary to provide that high-quality education. While voters approved new constitutional language raising the adequate provision of education to “a paramount duty” of the state of Florida, the meaning of the new language has yet to be interpreted in a substantive manner. Currently, a private commission’s report in 2005 has been the only effort to define the extent of the state’s new obligations.

The publication costs of this brief have been supported by CIVIC Concern, a non-partisan policy incubator, legislative watchdog, and advocate for ordinary Floridians. Andrea Dale copyedited the manuscript, and William Piecuch and Bernard Campbell provided support in shaping this for a public audience.

A Comprehensive Assessment of the A-Plus Plan's Impact on Public Education

Independent Evidence of Academic Achievement

What independent evidence exists about the status and trends of academic achievement as well as the achievement gap between minority students and white students?

Answer

National Assessment of Educational Progress (NAEP) scores show mixed results for Florida students in reading and math between 1998 and 2005. While scores improved for fourth-grade students in reading and math and eighth-grade students in math, scores have not improved for eighth-grade students in reading. Only fourth-grade math scores have continued to rise in the last few years and few changes in the achievement gaps among different population subgroups are statistically significant (or are somewhat likely to have happened by chance).

The sole independent, broad-based exam that Florida schoolchildren have participated in for the past decade shows mixed results for both status and trends of academic achievement and the achievement gap between minority students and white students. The National Assessment of Education Progress (NAEP) scores show mixed results for Florida students over the past decade, with some record of improvement for fourth-grade students but not as much for eighth-grade students. In addition, for most achievement-gap comparison pairs, the evidence does not support a claim in most comparisons that any narrowing of the achievement gap is statistically significant (or unlikely to have happened by chance).

... mixed results for Florida students over the past decade, with some record of improvement for fourth-grade students but not as much for eighth-grade students.

Generally, most researchers consider the National Assessment of Educational Progress (NAEP) as the best independent measure of student achievement. There are some concerns with the uses of NAEP scores, which do not follow students through time and whose composition for state-level tests has changed over the years. Nonetheless, NAEP is the only independent, broad-based assessment one can use to assess trends and the status of academic achievement.

Chatterji (2004) documented the mixed record through 2003, and the 2005 administration of NAEP does not change the substantive conclusions of Chatterji. In fourth-grade math and reading and in eighth-grade math, Florida students' average scale score on NAEP increased between the late 1990s and 2003, but the changes in average scale scores were not statistically significant in eighth-grade reading. In fourth-grade math, Florida students' average scale score on NAEP increased between 2003 and 2005, but in fourth-grade reading, eighth-grade math, and eighth-grade reading, scores were stagnant.

Florida's record of closing the achievement gap is also mixed. Table 1 shows the changes in achievement gaps by gender, race and ethnicity, participation in the federal free and reduced-price lunch program, and the difference between the 25th and 75th percentiles (or the interquartile gap). Between the late 1990s and 2005, the achievement gap *may have* shrunk in both math and reading in fourth grade in any comparison. But only for three comparisons is the change statistically significant, or unlikely to have occurred by chance. No change in eighth-grade gaps is statistically significant.

Comparison of Data Sources

There are three types of exams that children in Florida have taken during the past eight years. One set of exams is the FCAT Sunshine State Standards tests, which by law should align with the state curriculum standards in reading, writing, math, and science. In theory, the tests should measure how well students meet the state's standards. The disadvantage of using the state standards tests for assessing achievement is that it is not independent; if the stakes attached to the FCAT SSS exams distort instruction or encourage teaching to the test (e.g., Nichols & Berliner, 2007; Popham, 2004), then it is not appropriate to use the same test for both high-stakes accountability and also to evaluate the same accountability system.

State law also requires norm-referenced tests, which since 2001 have consisted of one commercial subtest in reading (the Reading Comprehension subtest of the Stanford Achievement Test) and one commercial subtest in math (the Problem Solving subtest of the Stanford Achievement Test). The advantage of using norm-reference tests is that they should theoretically be independent of any distortions caused by teaching to the test with high stakes for students and schools. There are two weaknesses of using the FCAT norm-referenced scores. One is that the form of the exam changed in 2005, with the adoption of the Stanford 10 subtests, based on different norms and with different exam qualities from the Stanford 9 subtests. While test publishers try to equate different forms of an exam, it is generally unwise to assume that scores in 2005 are equivalent to scores in 2004, and the average percentile rank jumped in most grades (some jumps up and some jumps down) in 2005. A second concern with using the FCAT norm-referenced results is that the scope of the subtests are narrow and are unlikely to represent the broader curriculum in either reading or math.

Readers should be cautious in interpreting Table 1 and other evidence in isolation. While NAEP scores are an important indicator of achievement within states, these scores are not definitive evidence regarding the effects of education policy.

Table 1. Trends in NAEP Achievement Gaps in Florida, 1996–2005

| Subject and grade | Male-female | White-Black | White-Hispanic | Nonpoor-Poor* | 25th-75th Percentile Gap |
|--------------------------------|--------------------|--------------------|-----------------------|----------------------|---|
| 4th-grade math | | | | | |
| 1996 | -2.6 | 33.1 | 18.6 | 23.5 | 43.8 |
| 2005 | 2.0 | 23.0 | 14.1 | 20.1 | 37.3 |
| <i>Significant difference?</i> | No | Yes | No | No | No |
| 8th-grade math | | | | | |
| 1996 | 3.5 | 42.2 | 23.4 | 27.2 | 48.9 |
| 2005 | 4.0 | 35.3 | 21.4 | 25.3 | 48.9 |
| <i>Significant difference?</i> | No | No | No | No | No |
| 4th-grade reading | | | | | |
| 1998 | -8.9 | 31.2 | 19.7 | 29.4 | 53.3 |
| 2005 | -4.8 | 25.6 | 12.9 | 20.9 | 45.3 |
| <i>Significant difference?</i> | No | No | No | Yes | Yes |
| 8th-grade reading | | | | | |
| 1998 | -13.5 | 28.1 | 16.8 | 24.1 | 45.8 |
| 2005 | -12.8 | 26.4 | 13.0 | 17.9 | 48.2 |
| <i>Significant difference?</i> | No | No | No | No | No |

* Poverty indicator is participation in the federal lunch program. Negative signs in gender are for test administrations where girls' average scale score is higher than that of boys. The 1996 NAEP administration tested math at the state level but not reading; the 1998 NAEP administration tested reading at the state level but not math. Statistical significance in NAEP reporting is set at $p < .05$. Source: National Assessment of Educational Progress (<http://nces.ed.gov/nationsreportcard/nde/>).

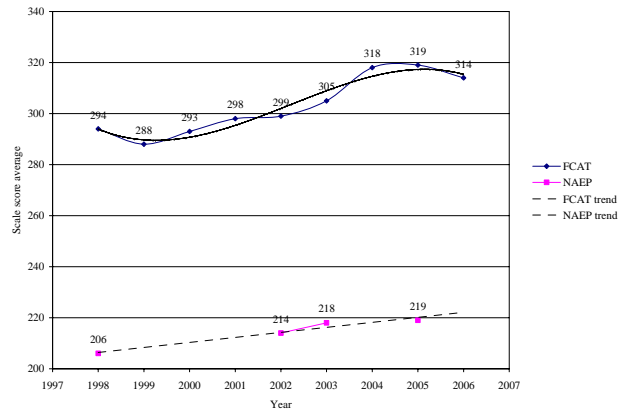
What Is Responsible for Achievement Trends

What might explain the difference in FCAT trends and independent data between elementary and secondary grades?

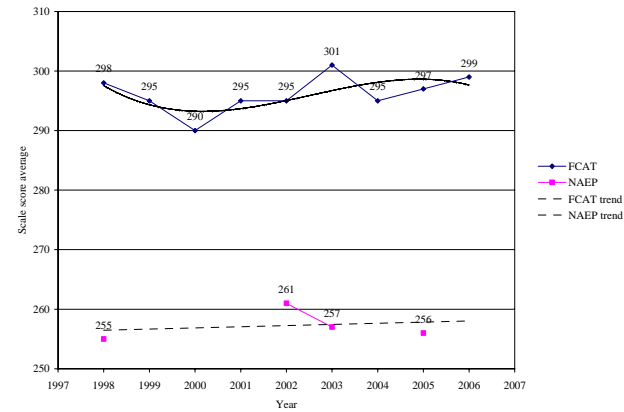
Answer

NAEP and FCAT average scale score trends are different by grade level and subject. After accounting for improvement trends that began before the 1999 A-Plus education reform package, only elementary-grade achievement appears to have seen improvements corresponding to the A-Plus program. But factors other than accountability may be responsible for these changes. These factors include the concentration of Title I programs in elementary schools, the state support of elementary instruction through the Florida Center for Reading Research, the importance of nontested subjects in secondary schools, the cumulative effects of grade retention, and a cohort effect of the Healthy Start program.

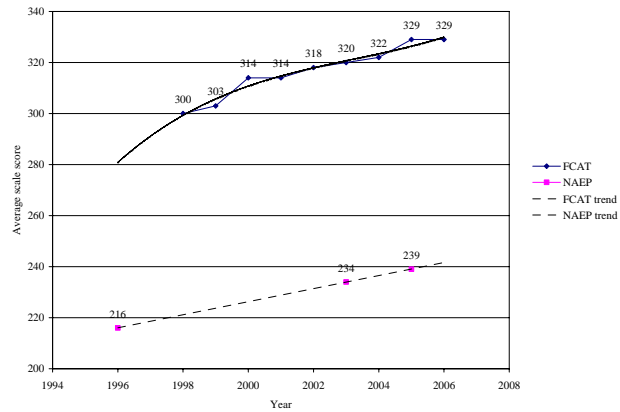
The trends in standardized test scores for Florida's public-school students in both the state standards-based tests and the federal NAEP exams are not identical. Figure 1 illustrates the trends in elementary and secondary math and reading. Trends in both the FCAT and NAEP show greater improvement in math scores than in reading, with a plateau in fourth-grade reading in recent years and largely flat scores in eighth-grade reading.



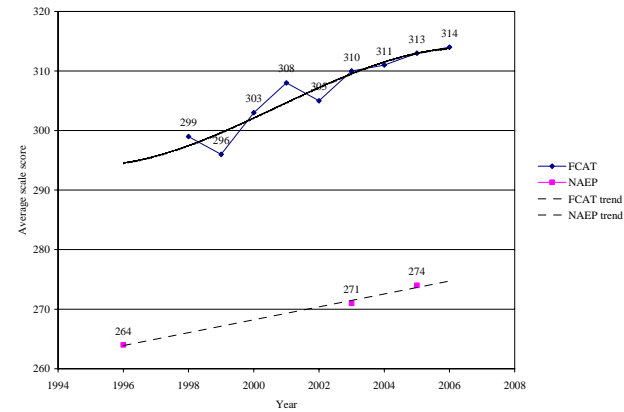
1a. Fourth-grade reading, 1998–2006



1b. Eighth-grade reading, 1998–2006



1c. Fourth- and fifth-grade math, 1996–2006



1d. Eighth-grade math, 1996–2006

Figure 1. Average scale scores of Florida students by grade level and subject, 1996–2006, with trend lines. Trend lines for FCAT scores are third-degree polynomials; trend lines for NAEP scores are linear. The elementary math time series uses fifth-grade math FCAT scores and fourth-grade NAEP scores. Sources: Florida Department of Education and National Center for Education Statistics.

Separating out the effects of different education policies in Florida is difficult for several reasons. First, in most cases the data is only available from the late 1990s to the present, making distinctions between pre-1999 and post-1999 trends less secure. In the subject and grade with the most data points from NAEP, eighth-grade math, the upward trend in scores began well before the passage of the A-Plus accountability plan in 1999, as Figure 2 illustrates. In at least eighth-grade math, one should be skeptical of attributing improving trends to the 1999 education reform package.

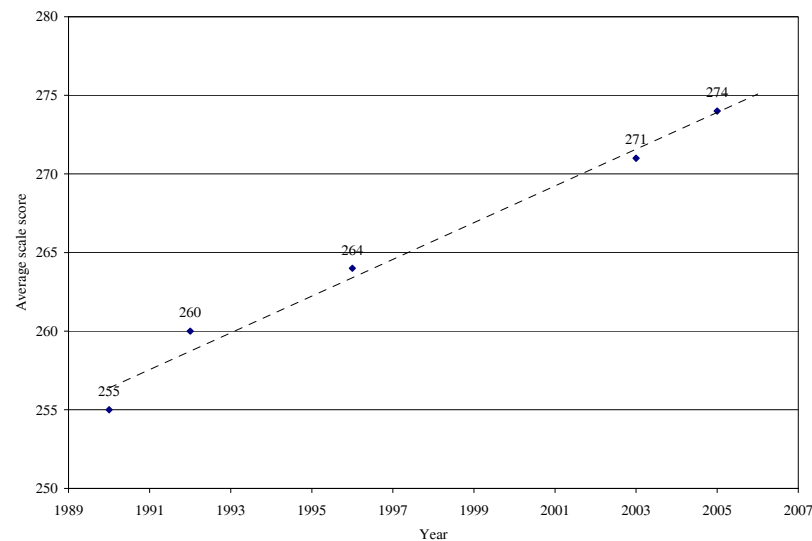


Figure 2. Florida public-school students' scores on eighth-grade NAEP math exams, 1990–2005. Source: National Center for Education Statistics.

After considering the limits of time-series data, one is left with flat scores in eighth-grade reading through Governor Bush's administration and an eighth-grade math trend continuing earlier progress. Both facts suggest that policy changes in 1999 did not improve student achievement in secondary education. That tentative conclusion is supported by trends in passing rates on the 10th grade FCAT reading and math tests, required for graduating with a standard diploma (Table 2). While the passing rates in math improved slightly between 2002 and 2006, passing rates in reading declined.

Policy changes in 1999 did not improve student achievement in secondary education.

Table 2. Florida 10th Grade Students Passing Math and Reading Graduation Exams, 2002–2006

| Year | % Passing math | % Passing reading |
|-------------|-----------------------|--------------------------|
| 2002 | 73 | 59 |
| 2003 | 73 | 58 |
| 2004 | 76 | 54 |
| 2005 | 77 | 52 |
| 2006 | 77 | 52 |

Source: Florida Department of Education.

How can one explain the differences between achievement trends in elementary and secondary grades? One must acknowledge that education policy changes commonly appear in combination as larger packages. While high-stakes accountability was the most public part of the 1999 A-Plus Plan, the 1999 education reforms in Florida included a variety of instructional support programs focused on elementary grades. Isolating claims about the various parts of the package—especially claims concerning the high-stakes accountability system in Florida—requires that one consider alternative explanations of test score trends.

- *Concentration of Title I programs in elementary schools.* In Florida, school districts target federal funds for the education of poor children in elementary schools. In 2006–2007, for example, \$222 million of all \$290 million in Title I funds are allocated to elementary schools (Florida Department of Education, 2006a). One plausible explanation for the disproportionate rise in achievement in elementary grades since 1999 is the extra funding of elementary programs through the allocation of Title I funds.
- *State support of elementary reading instruction.* During Governor Bush’s administration, he supported the investment of state funds to support elementary instruction through dedicated funds for reading coaches and the creation and support of the Florida Center for Reading Research. The dramatic jump in eighth-grade scores on the FCAT reading exam in 2006 for sixth- and seventh-grade students suggests that the support of reading coaches in 2005–2006 may be responsible for one-year increases in reading test scores. Former Governor Bush’s support for elementary reading instruction is thus a plausible explanation for the rise in elementary reading achievement but not for the rise in elementary math achievement.
- *Less scheduling flexibility in secondary schools.* Beginning in sixth or seventh grade, most Florida students spend their school days in classes focused on specific

The support of reading coaches in 2005-2006 may be responsible for one-year increases in reading test scores [in sixth and seventh grade].

subjects, changing classrooms and teachers between periods. For the vast majority of secondary schools, the schedule is divided by subject with little flexibility apart from the substitution of extra academic time for electives. In contrast, elementary instructional time is more flexible. While 71% of school principals across the country have reported that their schools responded to high-stakes testing by reducing time devoted to nontested subjects such as history and science (Center on Education Policy, 2006), in

Florida only elementary schools can shift time in a dramatic fashion. One plausible explanation for the rise in elementary achievement since 1999 might be the greater capacity in elementary grades to shift time from untested to tested subjects.

- *The cumulative effects of grade retention.* Beginning with the 2003 spring tests, third-grade students in Florida have been subject to increasing retention (to repeat third grade) based on scores on the FCAT reading test. The number of students retained in third grade jumped from 6,435 at the end of 2001–2002 to 27,713 at the end of 2002–2003 (Florida Department of Education, 2006b). Haney (2006) argues that increasing third-grade retention since 2002 pollutes test scores since 2003, because an increasing proportion of third-grade students in 2004 and subsequent years were older and thus had more exposure to the curriculum. This might be a plausible explanation for part of the rise in FCAT fourth-grade scores, because half of the 26-point jump in fourth-grade reading scale scores between 1999 and 1996 can be attributed to a single year's gain between 2003 and 2004 (see Figure 1a). However, a parallel jump in fifth-grade math FCAT scores between 2004 and 2005 (Figure 1b) does not explain as much of the rise in fifth-grade math scores during

Governor Bush's years in office. The potential confounding effects of grade retention thus explains part but not all of the elementary achievement gains.

- *The Healthy Start program.* Beginning in 1991, Florida's Healthy Start program has supported prenatal care, needs assessment, smoking-cessation counseling, nutritional counseling, and vaccinations for poor women and their children. The first cohorts of children affected by Healthy Start entered Florida's public schools in the late 1990s. Those cohorts would not have reached eighth grade until 2005 at the earliest. One plausible explanation for the increase in elementary achievement and *only* elementary achievement during most of Governor Bush's administration is the improved health of the birth cohorts of 1992 and later years.

Given the different plausible explanations, one cannot reasonably conclude that there is a single reason why elementary achievement in reading and math rose after the enactment of the A-Plus Plan while secondary achievement in reading and math either remained stagnant or rose at the same rate as in pre-1999 years.

What the FCAT Can and Cannot Measure

What are the strengths and limits of the FCAT in describing student and school performance?

Answer

The FCAT can describe achievement in some ways in reading, writing, math, and science for intermediate and some high-school grades. The FCAT cannot describe achievement in untested subject areas, cannot describe achievement in all areas of tested subjects, cannot pinpoint test scale scores of individual students down to single digits, cannot sort out the effects of knowledge and skills from the effects of test-wiseness, cannot identify which teachers are responsible for the knowledge of students in subjects, and cannot assess student and school achievement for which paper-and-pencil tests do not provide an accurate proxy.

In the early 1990s, Florida legislators and state officials responded to the standards movement with a new set of state tests that they hoped would be closely tied to newly-created curriculum standards. In creating this new system, the Florida Comprehensive Assessment Test (or FCAT), state officials created a hybrid test, partly multiple-choice and partly based on performance items, with testing of writing, reading, and math at three points in K-12 grades. (The new 10th grade exam also became the graduation test during a phase-in period in the late 1990s and early 2000s.) In the mid-90s, the state rolled out its new program, beginning with the writing exam, in which fourth-, eighth-, and tenth-grade students had to respond to a prompt with an essay (or a story in fourth grade, depending on the prompt). With the 1999 A-Plus program, the new law expanded the FCAT to all grades between third and tenth grade. The hybrid exams have remained at the original subjects and grade levels, though the additional tests are multiple-choice only.

The Florida Comprehensive Assessment Test (FCAT) describes the generalized achievement of Florida students in grades 3–10 (and eleventh grade in science) in sampled curriculum areas within reading, writing, math, and science. For three grades in all subjects, the test includes performance items. For its original purposes, and within the limits of established validity, the FCAT is an important source of information about student achievement. On the other hand, FCAT test scores should not be used without understanding the limitations of any standardized test (Popham, 2004).

- *FCAT scores do not measure achievement in untested subjects.* The FCAT includes tests in reading, writing, math, and science. The FCAT does not measure achievement in untested subjects such as U.S. or world history, foreign languages, civics or government, health, geography, or the visual and performing arts.
- *FCAT scores do not measure skills and knowledge that do not match up well to paper-and-pencil tasks.* The FCAT include activities that can be completed in a specified time, such as individual computation problems or word problems in math, a five-paragraph essay in writing, or answering comprehension questions about short passages in reading. The FCAT cannot measure achievement in long-term assignments such as a research paper. There are also important expectations we hold for students that cannot be measured in paper-and pencil tasks: performances in music or public speaking, for example, or laboratory work in science classes. Finally, the FCAT cannot measure achievement in collaborative efforts and the ability to work with peers during complex tasks.

There are important expectations we hold for students that cannot be measured in paper-and pencil tasks: performances in music or public speaking, for example, or laboratory work in science classes.

- *FCAT scores do not measure achievement in all areas of tested subjects.* An hour-long test is necessarily selective and cannot cover all of a subject that students learn throughout a year. In the high-school science exam, for example, the test specifications identify 50 important objectives in the Sunshine State Standards for high-school science (Florida Department of Education, 2002). Only 15 of those objectives are covered in the science exam every year. The others are sampled from year to year.

A report to parents might [misleadingly] identify a student's developmental score as 2037 in large type, while a footnote in much smaller lettering notes that the score could range from 1971 to 2103.

- *FCAT scores cannot accurately measure student achievement to single digits of the single-grade or developmental scales.* Because of measurement error, individual scores represent a range of plausible values, but FCAT reports to parents typically imply that the scores are far more accurate. For example, a report to parents might identify a student's developmental score as 2037 in large type, while a footnote in much smaller lettering notes that the implied score could range from 1971 to 2103. In most cases, the developmental score cannot accurately identify scores down to the nearest 10 points, let alone the nearest point, yet the Florida Department of

Education continues to imply a false level of precision.

- *FCAT scores cannot sort out the knowledge and skills of students from efforts to improve test-wiseness.* Florida schools have responded to the pressures of the A-Plus Plan in a variety of ways. In some ways, schools respond productively by improving instruction. In other ways, schools respond inappropriately by narrowing the curriculum or by engaging in short-term tactics to raise test scores rather than

improve instruction. Florida's schools are no different in this way from national responses to high-stakes accountability (Nichols & Berliner, 2007). One result is that Florida's citizens can never be certain how much FCAT scores reflect long-term, meaningful student achievement and how much the scores reflect short-term, tactical test preparation.

- *FCAT scores cannot identify which teachers are responsible for the knowledge of students in subjects.* The Special Teachers Are Rewarded (STAR) program mandates that teachers' pay bonuses be tied to test scores, but that policy assumes a single teacher's efforts can be tied unambiguously to FCAT scores. There are three reasons why this assumption is incorrect. First, the FCAT testing window is in late winter, and improvement in FCAT scores from year to year includes any learning that occurs after the FCAT exams are administered at the end of one year

and any learning that occurs before the FCAT the next year. Second, more than one teacher can help a student with specific skills—history teachers can give feedback on writing, and science teachers can help students with math. Third, students learn outside the classroom. Not only can parents, friends, and tutors help students learn, but the No Child Left Behind Act mandates that hundreds of Florida schools provide outside tutors when parents request it. There is no statistical software program in the world that can address each of these confounding factors.

The Special Teachers Are Rewarded (STAR) program mandates that teachers' pay bonuses be tied to test scores, but that policy assumes a single teacher's efforts can be tied to FCAT scores.

Extending the Grading System

Are indicators such as Florida's high-school graduation rate, SAT and ACT scores, and NAEP scores tied directly to state education policy?

Answer

There are two ways in which state policy could be tied to the listed indicators: as a goal of public education for the state, to meet constitutional standards for a high-quality education; or as a specific trigger mechanism for state responses, such as the state's labeling districts and schools with a single letter grade. Currently, there is no legal definition of what the constitutional standard for a high-quality education requires. The state's definition of adequate yearly progress for the purposes of the No Child Left Behind Act does include the official high-school graduation measure, but that measure currently inflates the true rate at which students earn standard diplomas. One should be cautious in tying state education policy to indicators such as Florida's high school graduation rate, SAT and ACT scores, and NAEP scores, but there are other ways to expand accountability to include non-FCAT measures.

"Tied directly to state education policy" has two potential meanings:

- Defining the language in Article 9, Section 1 of the Florida constitution
- Establishing triggers for specific consequences in accountability systems

The answer to this question is divided to address both meanings of the phrase.

Defining Constitutional Language

In 1998, voters in Florida approved new constitutional language for Article 9, Section 1. Yet, thus far, there has been no explicit, official definition of what constitutes a high-quality education in Florida. A longer discussion of the new constitutional language lies later, in the answer to the last question, but today, there is no explicit tie between the constitutional language guaranteeing “a high quality education” and the indicators listed in the question.

As Part of Accountability Formulas

Currently, only one of the indicators is used in Florida school accountability formulas. The labeling of individual districts and schools with single letter grades does not use any of these indicators, but the state’s definition of adequate yearly progress for the No Child Left Behind Act *does* include the official Florida high school graduation rate as one potential trigger for placing a high school in the “needs improvement” category. The official Florida high school graduation measure inflates true graduation rates in two ways: removing dropouts from school responsibility when they transfer into GED programs, and including GED certificates in the official count of who graduates (Dorn, 2006).

While the FCAT is limited in the information it provides about student achievement (see the answer to an earlier question), one should be very cautious about adding the listed indicators to any accountability formula triggering consequences:

- The state’s official graduation rate is not accurate, as explained earlier.
- SAT and ACT scores are used for college admissions purposes. These tests are not connected to the Sunshine State Standards, and they have never been validated for any purpose in elementary and secondary schools.

- The National Assessment of Educational Progress (or NAEP) provides scores at the state level, not at the district or individual school level.

On the one hand, the state needs a revised definition of the graduation rate for it to be accurate, and the other two indicators are inappropriate for accountability purposes. On the other hand, the state's school-grading accountability formulas *can* be extended beyond FCAT scores, in the following ways:

The state's school-grading accountability formulas *can* be extended beyond FCAT scores.

- If the state's labeling of schools includes more than one year's worth of data, the state could feasibly and fairly include data reflecting more than one year's worth of school effort (such as a graduation rate, once the formula is improved) or measures that are checked and become official months after the end of a school year (such as data on school-related violence).
- The state could include measures of school climate in accountability if it invested in independent and rigorous surveys of teachers, students, and parents. Such surveys could also help the state guard against inappropriate responses to the pressures of high-stakes accountability.
- The state could include measures of how schools challenge students, such as the proportion of students taking advanced or honors classes or the proportion of students participating in challenging extracurricular activities such as academic tournaments or juried competitions in speech and the performing arts, as well as the proportion of students meeting challenging standards in such extracurricular activities. Such measures should be disaggregated by gender, lunch-program status, race/ethnicity, and disability status to guard against unequal access to challenging opportunities.

Minimizing Growth

Does the A-Plus Plan currently focus on a narrow range of achievement or recognize and reward substantial improvement in the achievement of individual students?

Answer

In Florida, the evaluation of public schools asks narrow questions of student achievement. The mechanisms of school grading within the A-Plus Plan focus on students' meeting minimum thresholds. Students who barely meet the growth threshold count as much as students who far exceed the growth threshold. While the grading rule for Florida schools includes a growth component, the formula is based on whether individual students meet a threshold of growth. In addition, the formula is indifferent to improving performance for students who are in achievement levels 3, 4, and 5.

The state rules defining the criteria for labeling public schools are based largely on achievement status, not growth. Administrative Rule 6A-1.09981 describes a point system that includes the following elements:

- The percentage of students in a school scoring in or above achievement level 3 in the FCAT reading test.
- The percentage of students in a school scoring in or above achievement level 3 in the FCAT math test.
- The percentage of students in a school scoring in or above achievement level 3.5 in the FCAT writing test.
- The percentage of students in a school who make satisfactory annual progress between two successive FCAT administrations in reading, as defined in the rule.

- The percentage of students in a school who make satisfactory annual progress between two successive FCAT administrations in math, as defined in the rule.
- The percentage of students in the lowest quarter of the school's distribution of reading scores the prior year who make satisfactory annual progress between two successive FCAT administrations in math, as defined in the rule.

While three of those elements include a measure of growth, the formula asks whether the test-score growth was *sufficient* according to the rule, not *how much* growth in scores a student demonstrated. Students who barely meet the growth threshold count as much as students who far exceed the growth threshold. Some other states such as North Carolina have accountability systems that incorporate a measure of growth into the accountability formula. There is no written analysis currently available comparing the use of growth thresholds, such as Florida's, with growth measures, such as North Carolina's. But there are alternatives to Florida's minimal use of growth measures.

Students who barely meet the growth threshold count as much as students who far exceed the growth threshold.

Classroom Side Effects of the A-Plus Plan

What is the evidence available about schools' narrowing the curriculum and engaging in inappropriate test preparation in response to state education policy?

Answer

Florida schools are likely to respond to test pressures by reducing instruction in untested subjects and by focusing on test-wiseness rather than general student achievement.

Smith (1991) described a broad range of possible responses to high-stakes testing, including “ordinary curriculum with no special preparation, teaching test-taking skills, exhortation, teaching content known to be covered by the test, teaching to the test in format and content, stress inoculation, practicing test or parallel test items, and cheating” (p. 521). The improvements of instruction or focusing on the state standards would probably be considered appropriate by most Florida citizens, and a limited amount of time spent in encouragement and teaching test-taking skills. But many would be very concerned by extensive time or other resources devoted to test-wiseness, narrowing the curriculum to match test specifications rather than the state standards, or attempting to mimic test conditions as a substitute for instruction.

Many [Florida citizens] would be very concerned by extensive time or other resources devoted to test-wiseness, narrowing the curriculum to match test specifications rather than the state standards, or attempting to mimic test conditions as a substitute for instruction.

The behavior of Florida educators is included in national evidence of both constructive and destructive responses to accountability pressures (Center on Education

Policy, 2006; Nichols & Berliner, 2007). In addition, 23% of teachers participating in a statewide survey described an expansion of inappropriate test preparation directly as a result of Florida's accountability policies (Jones & Egley, 2004). The description of the Florida subset of a national teacher survey came to a similar finding:

The overwhelming majority of teachers reported that the state testing program has led them to teach in ways contrary to their own ideas of sound educational practices. Florida teachers, though, were significantly more likely to hold this view than were teachers in other states with high-stakes policies (90 percent v. 75 percent). Survey results suggest that teachers in high-stakes settings gear the content of instruction to that of the state test, and that they are also modeling their own classroom assessments based on the test's format. Florida teachers reported these changes in greater percentages than did teachers in other high-stakes states. (Abrams, 2004, p. 14)

There is no published professional education research that contradicts these conclusions about the extent of destructive test-preparation practices or the effects in Florida when compared to the rest of the country.

The Best Research on Vouchers

Do we know if vouchers have helped poor children and those with disabilities improve their academic performance? What research describes the overall impact of vouchers, opportunity scholarships, McKay scholarships, and corporate income-tax scholarships on public education?

Answer

The best research in peer-reviewed economic journals refutes claims that the existence of voucher programs has improved Florida's schools.

Figlio and Rouse (2006) is the only refereed, rigorous professional journal article to address the crucial question about Florida's vouchers programs and school competition: Did the threat of vouchers from the Opportunity Scholarship Program provide an incentive for public schools to improve education? They conclude that it did not. The improvement in test scores attributable to the 1999 reforms was from the threat to the reputation of schools and not from the threat of competition posed by the Opportunity Scholarship Program.

State law does not require the reporting of evidence on the (expired) Opportunity Scholarship program. State law also does not require students to take the FCAT when participating in the McKay scholarship voucher program or corporate income-tax scholarship voucher programs. For these two reasons, there is no independent evidence of the quality of such programs or whether they help poor children and those with disabilities improve their academic performance.

Other Concerns Losing Attention

Has the A-Plus Plan's grading system's reliance on so few factors drawn attention away from important concerns in non-FCAT areas?

Answer

The Florida system of assigning each school a single letter grade per year draws attention from other concerns in two ways. First, the consequences of the grading system forces districts and schools to prioritize efforts in the grades and subjects that have high-stakes consequences. Second, the grading of schools dominates news coverage, making some topics more difficult to raise.

Florida's system of assigning single letter grades to individual schools was designed to focus attention on measured student achievement, and it has. First, it has forced school districts and individual schools to pay attention to the specific components of the state's formula to succeed in the grading system. As explained in the answers to earlier questions, this narrow focus has had unintended negative consequences for the curriculum and instruction.

In addition, the grading system has dramatically redefined what is covered in broadcast and print journalism and thus limits what else can be considered important educational news. When the universe of potential problems must compete for limited coverage in daily newspapers and television broadcasts, only a few problems become headline issues (Hilgartner & Bosk, 1988). Examples of policy issues in Florida education that have remained "below the radar" despite explicit efforts to raise concerns publicly include the following:

- *The state of math and science education.* In math and science, only culture wars can compete with the grading system for headlines. In the past few years, more news coverage in Florida has focused on the controversy over teaching evolution (e.g., Blair, 2006b; Blair & Bird, 2005; Date, 2005; Lyons, 2006; Matus, 2005b; Miller, 2005; Winchester, 2006; Winchester & Matus, 2005) than on other issues such as student performance in science, the recruitment and retention of science and math teachers, or the general revision of the state standards (e.g., Blair, 2006a; Dunn, 2005; Flannery, 2002; Matus, 2005a; Matus & Winchester, 2006; Meisels, 2005; Meisels & Potter, 2006; Sedore, 2003; Shah, 2003).

The grading system has dramatically redefined what is covered in broadcast and print journalism and thus limits what else can be considered important educational news.

- *The education of English language learners.* Despite a 1990 federal-court consent decree and a supplemental agreement in 2003, and despite demographic changes in Florida that have dramatically increased the proportion of public-school students whose first language is not English (e.g., MacDonald, 2004), there is scant

coverage of the education of English language learners. Most of the print coverage focuses on the bureaucratic details of teacher certification and endorsements to teach English language learners (e.g., Allen, 2005; Brown, 2006; Dunn, 2004; Flannery, 2003; O'Steen, 2006; Taylor, 2004).

- *The Florida Chamber Foundation's "New Cornerstone" report.* When the Florida Chamber Foundation report on educational investment (or "intellectual infrastructure") was released in 2002, the story attracted almost no news coverage. As much coverage focused on the politics of pre-publication editing (Fineout, 2002;

Hallifax, 2002b; Ulferts, 2002) as on the report itself (Hallifax, 2002a; Salinero, 2002). More than a year later, one economic reporter in the state noted the lack of attention and follow-through after previous Florida Chamber Foundation reports: “This time around, maybe there will be less lip service and more action to strengthen the state's economic status” (Trigaux, 2003).

While the lack of attention to these issues is not solely attributable to the grading system’s reliance, they illustrate the general difficulty of attracting attention to issues that are not closely identified with the grading system.

Helping Students

How is Florida providing remediation and other services to help low performing students?
How does that intervention compare with recommended practices based on research?

Answer

Of various state efforts, only the activities of the Florida Center for Reading Research have significant support in the research literature.

There are several strategies encouraged by state policy to help low performing students, including the following approaches:

- Statewide support for professional development in reading instruction through the Florida Center for Reading Research and through funding for reading specialists. This effort has focused on elementary grades, with some efforts in middle schools. The Florida Center for Reading Research's efforts in promoting phonemic awareness and explicit phonics instruction in early literacy has a research base (National Reading Panel, 2000).
- The statewide policy of requiring retention in third grade for students who neither achieve at level three in the third-grade FCAT nor have a portfolio documenting reading achievement. There is one reputable study arguing that retention can have benefits without long-term side effects (Alexander, Entwisle, & Dauber, 1994), but that study's conclusion has been criticized for going beyond the data (Shepard, Smith,

While legislators and educators understandably want to try new approaches... one danger is the possibility that either policymakers or educators will turn to fads that have little research support.

& Marion, 1996). The bulk of research evidence is that absent other data, retention has poorer long-term outcomes than promotion (Heubert & Hauser, 1999; Shepard & Smith, 1989). (See the next question for more information on Florida's third-grade retention policy.)

- In middle schools and high schools, other strategies are more prominent, such as doubling up reading and math courses for students who have a record of low performance. There is no research literature on this type of intervention.

While legislators and educators understandably want to try new approaches to assisting students with low-academic achievement, one danger is the possibility that either policymakers or educators will turn to fads that have little research support. In the 1990s, for example, President Clinton supported the adoption of school uniforms as an intervention based on anecdotal evidence from a single district, and many Florida schools require school uniforms. There is no research support for claiming that uniforms improve academic or social outcomes, and the spread of uniform policies is more symbolic than substantive (Brunnsma, 2004).

Third-Grade Retention

Is there clear evidence that retaining children in third grade who fail the reading portion of the FCAT is helping or hurting students?

Answer

There is no clear evidence that retaining children in third grade helps Florida students, despite a fourfold increase in retention between the 2001–2002 and 2002–2003 school years.

When Florida law began requiring the retention of third-grade students who failed to score well on the FCAT reading exam in 2002–2003, the retention of third-graders jumped from 6,435 to 27,713 (Florida Department of Education, 2004b). While those absolute numbers decreased to 20,121 two years later, third-grade retention is still considerably higher today than before the policy change. The empirical question is whether greater retention has improved the education of retained students.

There is insufficient evidence to answer this question.

The Florida Department of Education (2004b) notes that in grades 3–10, retained students are far less likely to have FCAT scores in the lowest achievement level than nonretained students. This evidence compares *younger*, unretained students *exposed to material for one year* with *older*, retained students who *have been taught the same material two years in a row*. Unfortunately, this evidence does not answer the relevant research question: how would the *same* retained students have performed if not retained (or how would nonretained students perform if retained).

Greene and Winters (2006) argue that their comparison of the third-grade cohort immediately before the 2002–2003 policy and the third graders in 2002–2003 address this

research question and show an advantage for students under the retention policy. Wiley's (2006) review of the Greene and Winters research found five technical and presentation flaws in this research, making it an inappropriate basis for making policy decisions.

One gap in most of the research on grade retention (including Greene and Winters') is the framing of the issue as either retention or promotion, each option having weak or no academic intervention. Holmes' (1989) meta-analysis of grade-retention policies suggests that the only studies finding positive benefits of retention involved early, intensive interventions. The retention policy in Florida hinges on third grade, which is probably too late to intervene effectively to stop most children's early reading problems, and there is no documentation of what intervention the thousands of Florida third-graders who repeat one (or more) years in third grade receive apart from that extra year or a different teacher. From a lay perspective, the research says that if one is faced with a choice between retaining a child and praying that an academic miracle will occur, on the one hand, and promoting a child and praying that an academic miracle will occur, the better choice is to promote the child and pray that an academic miracle will occur. But neither approach is an appropriate instructional or policy choice.

There is no clear evidence that retaining children in third grade helps Florida students.

An Adequate Education in Florida

What are the key questions to ask about the adequacy of Florida's student funding?

Answer

While voters approved new constitutional language raising the adequate provision of education to “a paramount duty” of the state of Florida, the meaning of the new language has yet to be interpreted in a substantive manner. The new constitutional language opens up opportunities for suing the state for failing to provide an adequate and “high-quality system of free public schools,” a change that the Supreme Court recognized in nonbinding language in a 2006 decision. Critical questions for such interpretation include the definition of “a high quality education” and what such an education would require from the state.

For decades, Florida's courts rebuffed a broad range of lawsuits intended to force change in the state's system of funding public education (Staros, 1994). In 1998, voters in Florida approved new constitutional language for Article 9, Section 1, which today begins as follows:

The education of children is a fundamental value of the people of the State of Florida. It is, therefore, a paramount duty of the state to make adequate provision for the education of all children residing within its borders. Adequate provision shall be made by law for a uniform, efficient, safe, secure, and high quality system of free public schools that allows students to obtain a high quality education...

This language shifts the state's obligations in terms of providing an adequate education. In 2006, the Florida Supreme Court's decision in the *Bush v. Holmes* case included a

discussion of the amendment, and the majority opinion implied that the 1998 amendment substantially increased the state's obligations to its children.

After the 1998 revision restoring the [historical] “paramount duty” language, Florida’s education article is ... imposing a maximum duty on the state to provide for public education that is uniform and of high quality. (*Bush v. Holmes*, 2006, p. 16)

The Supreme Court’s language was a side remark, or obiter dictum, not directly relevant to the voucher case before the court. Yet the majority was clearly signaling reinterpretation of the state’s educational obligations based on the amended language without committing itself to a definition of a high-quality education.

The majority opinion implied that the 1998 amendment substantially increased the state’s obligations to its children.

Currently, there is no active lawsuit that involves Article 9, Section 1, a claim tied to the guarantee of an adequate education, and the definition of a high-quality education. Yet there may well be in the future, given the history of funding litigation in Florida’s past and the trends towards adequacy litigation in other states. In 2004, Harris noted,

The interpretation of the adequacy language may ultimately be decided in the courts.... [The legislature should initiate] a cost study for educational adequacy to provide preliminary estimates of potential additional funding requirements. (Harris, 2004, p. 13)

In 2005, a private Constitutional Accountability Commission concluded that the new language imposed requirements both on the *system* of education and also on the

outcomes. In terms of the system, the 2005 commission relied on the language referring to “uniform, efficient, safe, secure, and high quality.” For the outcomes, the commissioned borrowed the language requiring a system that “allows students to obtain a high quality education.” The commission recommended that the legislature adopt fourteen measures to compare Florida’s education system to the nation’s as a whole and define a high-quality system so that the state would not rank in the lowest quarter of states on any measure and would rank in the top half in at least six of the fourteen measures.

The Constitutional Accountability Commission (2005) report discussed above made its recommendations about the two key questions that must be answered about the 1998 amendment:

1. What defines a high-quality education?
2. What resources and other system traits are necessary to reach those standards?

There is a literature on methods to define an adequate education and estimate costs, and the techniques are problematic. As Harris wrote, “Estimating these [high-quality education] costs is quite difficult” (Harris, 2004, p. 13). However, until the legislature or courts act, the Constitutional Accountability Commission report is the only document proposing substantive answers to these questions.

References

- Abrams, L. M. (2004). *Teachers views on high-stakes testing: Implications for the classroom*. EPSL-0401-104-EPRU. Tempe, AZ: Education Policy Studies Laboratory. Retrieved December 26, 2006, from <http://epsl.asu.edu/epru/documents/EPSL-0401-104-EPRU.pdf>.
- Alexander, K., Entwisle, D. R., & Dauber, S. (1994). *On the success of failure: A reassessment of the effects of retention in the primary grades*. New York: Cambridge University Press.
- Allen, J. L. (2005, September 30). Teachers learn how to break the language barrier in schools. *Sarasota Herald-Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Blair, D. (2006b, January 29). Teachers opt for texts without intelligent design. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Blair, R. (2006a, July 2). Bill + governor's signature = more high school math. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Blair, R., & Bird, A. (2005, October 12). Intelligent design debate brews. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Brown, M. (2006, August 31). Students learning English put to test. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

Brunnsma, D. L. (2004). *The School Uniform Movement and What It Tells Us about American Education: A Symbolic Crusade*. Lanham, MD: Scarecrow Press.

Bush v. Holmes, Fla. Supreme Court SC04-2324 (2006). Retrieved January 9, 2006, from <http://www.floridasupremecourt.org/decisions/2006/sc04-2323.pdf>.

Center on Education Policy. (2006). *From the capitol to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author.

Chatterji, M. (2004, April). *Good and bad news about Florida student achievement: Performance trends on multiple indicators since passage of the A+ legislation*. EPSL-0401-105-EPRU. Tempe, AZ: Education Policy Studies Laboratory. Retrieved February 25, 2005, from http://www.asu.edu/educ/epsl/EPRU/epru_2004_Research_Writing.htm.

Constitutional Accountability Commission. (2005, August). *Final report*. Tallahassee, FL: Florida School Boards Association. Retrieved February 16, 2007, from <http://www.fsba.org/documents/finalcacreport.pdf>.

Date, S. V. (2005, October 9). Many private schools taking vouchers teach creationism. *Palm Beach Post*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

Dorn, S. (2006, June). *Inflation of official Florida graduation rates*. Working paper retrieved June 29, 2006, from <http://www.shermandorn.com/mt/archives/Inflated-Florida-Graduation.pdf>.

Dunn, A. (2004, January 27). New rule sought for ESOL teachers. *Lakeland Ledger*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

- Dunn, A. (2005, April 22). Polk schools set math, science goals. *Lakeland Ledger*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Figlio, D. N., & Rouse, C. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, 90(1-2), 239-255.
- Fineout, G. (2002, January 31). Researchers met with members of Bush's office of policy and budget. *Lakeland Ledger*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Flannery, M. E. (2002, October 17). Algebra may be added to 8th-graders' studies. *Palm Beach Post*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Flannery, M. E. (2003, February 9). Spanish-speakers at an FCAT disadvantage. *Palm Beach Post*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Florida Chamber Foundation. (2002). Preparing Florida's intellectual infrastructure for the 21st century economy. Chapter 3 of *New Cornerstone* (pp. 3-1—3-75). Tallahassee, FL: Florida Chamber of Commerce. Retrieved June 21, 2002, from http://www.newcornerstoneonline.com/PDFs/FLCor_Brochure_Chap3.pdf.
- Florida Department of Education. (2002). *Science grade 10 test item and performance task specifications* [draft]. Tallahassee, FL: Author. Retrieved December 24, 2006, from <http://www.firn.edu/doe/sas/fcat/fcatis01.htm>.

Florida Department of Education. (2006a, November 1). 2006–2007 Title I school list. Tallahassee, FL: Author. Retrieved February 16, 2007, from http://www.firn.edu/doe/title1/pdf/2006_2007finaltitlelist.pdf.

Florida Department of Education. (2006b, October). Non-promotions in Florida's public schools, 2004–2005. Tallahassee, FL: Author. Retrieved February 16, 2007, from <http://www.firn.edu/doe/eias/eiaspubs/pdf/nonpromotions.pdf>.

Fla. Statutes ch. 99-398.

Greene, J. & Winters, M. (2006). Getting ahead by staying behind: An evaluation of Florida's program to end social promotion. *Education Next*, 6(2), 65–69.

Hallifax, J. (2002a, April 8). Chamber study concludes Florida's schools falling behind. *Associated Press*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

Hallifax, J. (2002b, January 30). Lawmakers raise questions about Chamber education study. *Associated Press*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

Haney, W. M. (2006, September). *Evidence on education under NCLB (and how Florida boosted NAEP scores and reduced the race gap)*. Boston: National Board on Educational Testing and Public Policy. Paper presented at Hechinger Institute, Broad Seminar for K-12 Reporters, New York City. Retrieved December 29, 2006, from <http://www.bc.edu/research/nbetpp/statements/nbr6.pdf>.

- Harris, D. N. (2004, April). Class size, pre-kindergarten, and educational adequacy: Costs and funding options for Florida. EPSL-0401-108-EPRU. Tempe, AZ: Education Policy Studies Laboratory. Retrieved December 26, 2006, from <http://epsl.asu.edu/epru/documents/EPSL-0401-104-EPRU.pdf>.
- Heubert, J. P., & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Hilgartner, S., & Bosk, C. L. (1988). The rise and fall of social problems: A public arenas model. *American Journal of Sociology*, 94(1), pp. 53–78.
- Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard and M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16–33). London: Falmer Press, 1989.
- Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12(39). Retrieved December 26, 2006, from <http://epaa.asu.edu/epaa/v12n39/>.
- Lyons, T. (2006, January 1). Jeb helps confuse science, religion. *Sarasota Herald-Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Macdonald, V.-M. (2004, April). The status of English language learners in Florida: Trends and prospects. EPSL-0401-113-EPRU. Tempe, AZ: Education Policy Studies Laboratory. Retrieved December 26, 2006, from <http://epsl.asu.edu/epru/documents/EPSL-0401-113-EPRU.pdf>.
- Matus, R. (2005a, December 30). Florida gets an "F" in science. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

- Matus, R. (2005b, August 28). Will Darwin come under fire in Florida? *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Matus, R., & Winchester, D. (2006, May 25). Science scores latest to slip. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Meisels, G. G. (2005, May 6). Florida faces crisis in fields of math and science. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Meisels, G. G., & Potter, R. (2006, March 26). Florida's 'F' in science reflects reviewers' bias. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Miller, K. (2005, October 13). Florida education chief noncommittal on evolution. *Palm Beach Post*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development. Retrieved December 24, 2006, from <http://www.nichd.nih.gov/publications/pubskey.cfm?from=reading>.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

- O'Steen, C. (2006, February 28). ESOL provides benefits. *Lakeland Ledger*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Popham, W. J. (2004). *America's "failing" schools: How parents and teachers can cope with No Child Left Behind*. New York: Routledge.
- Salinero, M. (2002, April 9). Florida education system loses ground in key areas. *Tampa Tribune*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Sedore, D. (2003, August 1). Intensive course fills teaching need. *Palm Beach Post*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Shah, N. (2003, January 13). Math test reveals 'lousy' results. *Palm Beach Post*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention*. London: Falmer Press.
- Shepard, L. A., Smith, M. L., & Marion, S. F. (1996). Failed Evidence on Grade Retention. *Psychology in the Schools*, 33(3), pp. 251-261.
- Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal*, 28(3), 521-542.
- Staros, B. J. (1994). School finance litigation in Florida: A historical analysis. *Stetson Law Review*, 23, 497-519.

- Taylor, V. (2004, September 30). Top of the class. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Trigaux, R. (2003, October 2). Florida's vision looks sadly familiar. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Ulferts, A., (2002, January 31). School report delayed for analysis. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Wiley, E. W. (2006, February 23). *Review of Getting ahead by staying behind: An evaluation of Florida's program to end social promotion*. EPSL-0602-119-EPRU. Tempe, AZ: Education Policy Studies Laboratory. Retrieved February 17, 2006, from <http://epsl.asu.edu/epru/ttreviews/EPSL-0602-119-EPRU.pdf>.
- Winchester, D. (2006, February 15.) Intelligent design no big bang here. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.
- Winchester, D., & Matus, R. (2005, December 30). Darwin's lost ground. *St. Petersburg Times*. Retrieved February 16, 2007, from the LexisNexis Academic Universe database.

About the author

Sherman Dorn is an associate professor of education at the University of South Florida, editor of the professional journal *Education Policy Analysis Archives*, co-editor of *Education Reform in Florida* (State University of New York Press, 2007), and author of *Accountability Frankenstein* (Information Age Publishing, 2007). His research traces how our society defines school problems and how those definitions shape education policy. Since 1996, he has taught social-science and humanities perspectives on education at the University of South Florida. He writes regularly on state and national education policy at <http://www.shermandorn.com>.



<http://www.civicconcern.org>