

3

The A+ Plan

Paul E. Peterson

Florida's bold, innovative A+ Accountability Plan (hereinafter referred to as A+) has been a pace-setter for the nation. Since 1999, when A+ was first put in place, student achievement on the Florida Comprehensive Achievement Test (FCAT) has risen significantly, especially in the elementary grades. These gains are confirmed by other signs that show that Florida is making progress at a considerably better than average rate. Although the Florida program can be enhanced, its record of spurring student achievement, if less than ideal, surpasses that of most states. The federal accountability system established by No Child Left Behind (NCLB) would profit by modeling itself along the lines of the Florida A+.

Principal Findings and Recommendations

Certain features of Florida's A+ deserve special commendation, all of which should be taken into account by Congress when it considers the re-authorization of NCLB.

1. The A+ sets up an intuitive grading system, ranking schools on a five-grade scale—"A," "B," "C," "D," and "F"—that is readily understood by any parent, taxpayer or news reporter. NCLB, by comparison, only identifies schools as making or not making "adequate yearly progress" a simple dichotomy expressed by a misleading circumlocution.
2. The A+ is supported by a comprehensive warehouse of data that enables educators to track each individual student's progress from one year to the next. NCLB should require other states to establish a similar data collection and retrieval system.
3. Making use of its information warehouse, the A+ has a scoring system that evaluates schools on the basis of the students' educational growth as well as on the students' overall level of accomplishment. The less sophisticated NCLB grading scheme does not follow individual students but instead traces a trajectory of cohorts of students toward a targeted level of proficiency to be reached by 2014.
4. The grading system under A+ does a satisfactory job of identifying higher quality schools and an even better job of identifying those that are the least effective.¹ The pass/fail grading system employed by NCLB does a less effective job of detecting school effectiveness.
5. A+ sets up clear positive and negative consequences for schools, depending on the grade they receive. By comparison, NCLB's

1. A higher quality school is one in which students are learning at a more rapid rate, as measured by the growth in student test-score performance.

negative consequences for schools not making “adequate yearly progress” are minimal, and NCLB does not reward outstanding growth or overall achievement.

6. A+ holds students accountable, requiring passage of an examination prior to graduation and expecting students in third grade to reach a certain level of proficiency before being promoted to fourth grade. NCLB does not hold students directly accountable.

Although A+ has these positive features, the state is correct in deciding to undertake important changes in the coming year. Most importantly, its decision to raise the level of proficiency students are expected to reach is to be commended. In addition, the accountability system can be further enhanced by distributing grades among schools less generously, and by giving greater emphasis to student growth in test score performance in its grading scheme. In other respects, however, it is an accountability system worthy of emulation, especially as Congress considers NCLB re-authorization. The remainder of this chapter reviews the evidence that provide the basis for these findings and recommendations.

Student Achievement in Florida

Except for 10th grade reading scores, students have been performing increasingly well on the FCAT over the past eight years (see Chapter 2). But some critics have suggested those gains are artificially produced, not accurate reflections of the learning that is occurring.² While

2. It has been argued that student performance on high-stakes tests such as the FCAT are inaccurate measures of student achievement, because teachers are “teaching to the test” by focusing the curriculum narrowly on test-related material, or are spending an undue amount of time explaining to students how to take the test so as to become test-savvy, or are, in some cases, actually cheating, by assisting students with the answers, either during the examination or afterwards. Schools are also said to be issuing suspensions to low-performing students just before test day, classifying students as disabled so as to excuse them from the test, and giving students better lunches on test day on the theory that well-fed children do better. See, for example, Daniel

this may well be true to some extent in some places, any suggestion that performance on the FCAT is unrelated to more general learning is belied by the fact that student FCAT scores were highly correlated with student performance on the Stanford 9, a standardized, norm-referenced test given by Florida's schools concurrently with the FCAT.³ The correlation of performance among individual test takers between the FCAT and the Stanford 9 (in grades 3, 4, and 5 in math and reading in 2002, 2003, and 2004) ranged between 0.79 and 0.84, generally thought to be high correlations. In other words, students who did well on the FCAT did well on another test that was not part of the high-stakes testing system.⁴ Not only are correlations high, but FCAT gains are also echoed by parallel gains on the Stanford 9. Between 2001 and 2004, Florida student performance on the Stanford 9 rose significantly for nearly every grade, both in reading and mathematics (see Figure 1). In 2005 Florida introduced a revised norm-referenced test, the Stanford 10, making comparisons between 2004 and later years inappropriate and, as a result, one must look separately at the growth rate between 2005 and 2006. That is done in Figure 2, which shows improvements in student performance on the Stanford

M. Koretz and Sheila Barron, "The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)," Rand Report, 1998; Gregory J. Cizek, "Cheating to the Test," *Education Matters*, vol. 1, no. 1 (Spring 2001): 40–47; Brian A. Jacob and Steven D. Levitt, "To Catch a Cheat: How to Stop Testing Fraud," *Education Next*, vol. 4, no. 1 (Winter 2004): 68–75; Lawrence A. Baines and Gregory Kent Stanley, "High-Stakes Hustle: Public Schools and the New Billion Dollar Accountability," *The Educational Forum*, vol. 69 (Fall 2004); David N. Figlio, and Joshua Winicki, "Food for Thought: The Effects of School Accountability Plans on School Nutrition," NBER Working Paper No. 9319, November 2002; David N. Figlio, "Testing, Crime and Punishment," NBER Working Paper No. 11194, National Bureau of Economic Research, March 2005; and David N. Figlio, and Lawrence S. Getzler, "Accountability, Ability and Disability: Gaming the System," NBER Working Paper No. 9307, October 2002.

3. See Jay P. Greene, Marcus A. Winters, and Greg Forster, "Testing High-Stakes Tests: Can We Believe the Results of Accountability Tests?" *Teachers College Record*, vol. 106, no. 6 (June 2004): 1124–1144.

4. The specifics are available, upon request, from the Program on Education Policy and Governance, Harvard University.

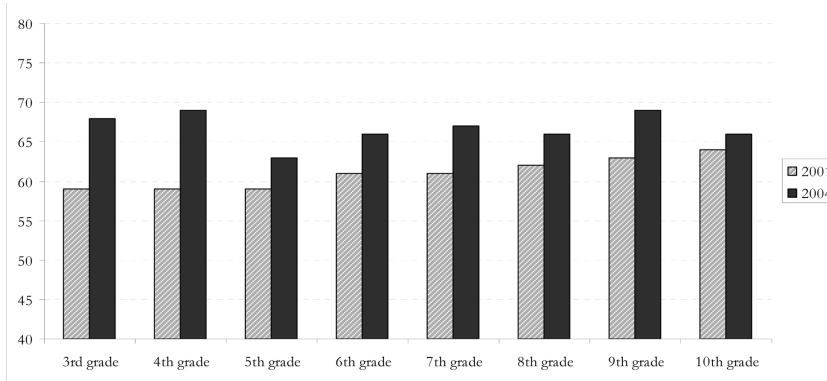


Figure 1a. Stanford 9 Mathematics, Florida's Median National Percentile Rank in 2001 and 2004

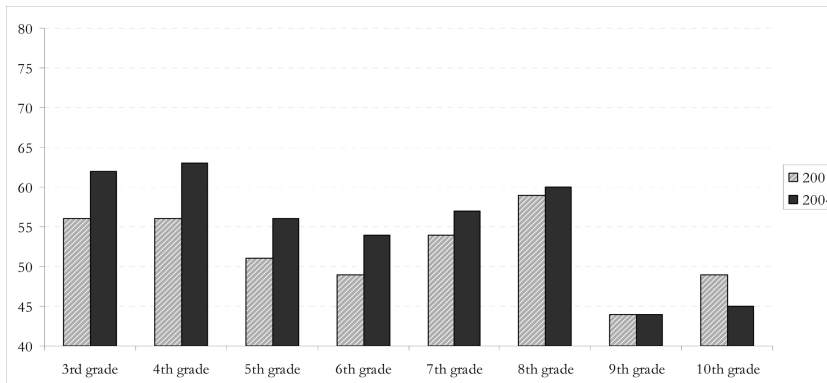


Figure 1b. Stanford 9 Reading, Florida's Median National Percentile Rank in 2001 and 2004

Source: Florida Department of Education, Statewide Comparisons of Norm Reference Test Scores.

10 in 2006 that were, in several grades, larger than any registered in any single year previously. Particularly striking are the middle school improvements and, especially, the dramatic gains in 10th grade test scores, the one grade which had registered few, if any, gains previously.

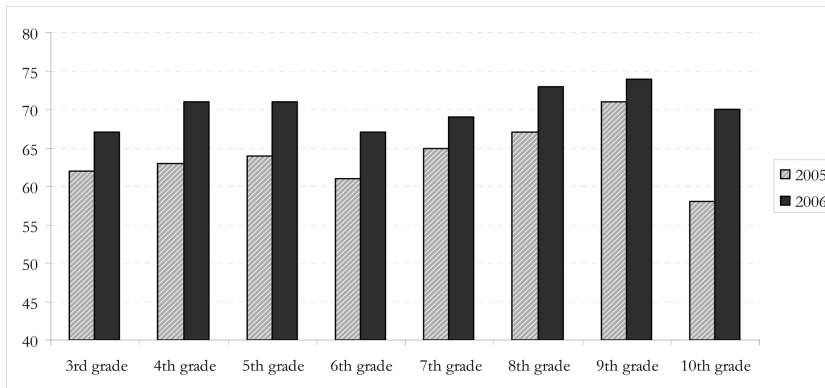


Figure 2a. Stanford 10 Mathematics, Florida's Median National Percentile Rank in 2005 and 2006

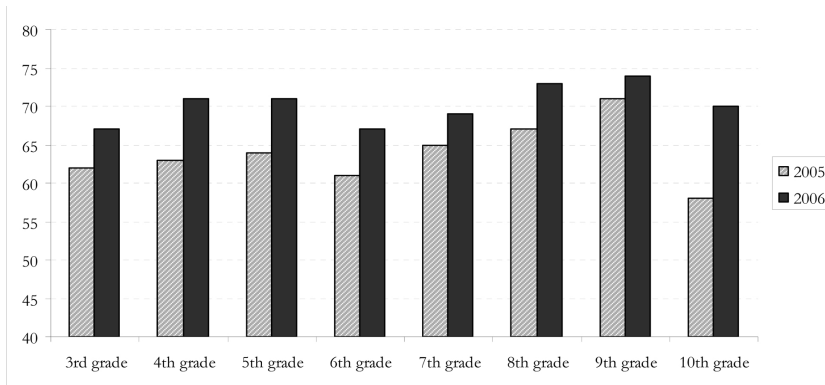


Figure 2b. Stanford 10 Reading, Florida's Median National Percentile Rank in 2005 and 2006

Source: Florida Department of Education, Statewide Comparisons of Norm Reference Test Scores.

Performance on NAEP

Even more convincing evidence concerning educational progress in Florida comes from the National Assessment of Educational Progress (NAEP), generally known as the nation's report card. The test has for

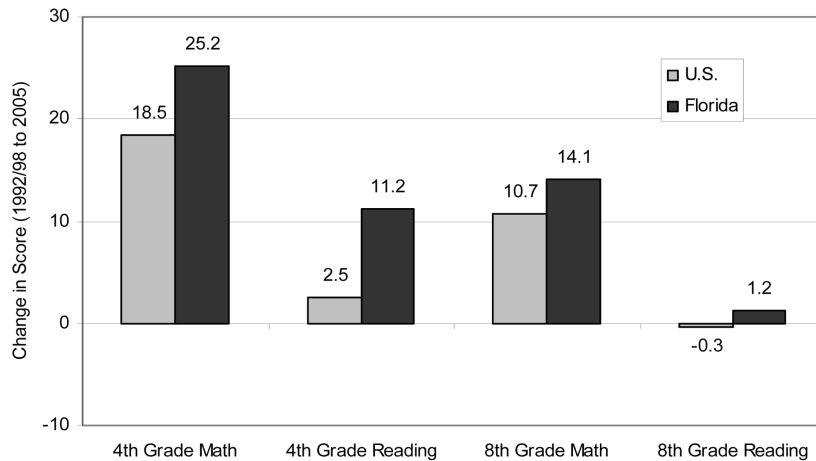


Figure 3. Change in NAEP Scores in the U.S. and Florida (1992/98–2005)

*Gain in Florida is significantly different from the gain in the U.S.

Note: For 8th Grade Reading, comparison years are 1998 and 2005; for all others, 1992 and 2005.

Source: *Education Week: Quality Counts at 10*, vol. 25, no. 17 (2006).

several decades been given to a representative sample of students nationwide and, under NCLB legislation, is now also being given to a representative sample of students in each state. Consequently, the NAEP provides a calibrating instrument that allows one to determine whether trends on state tests like the FCAT are also to be found by a nationally recognized test given at another time.

Studies have shown that students in states with accountability systems improve at a faster rate than students in states without them.⁵ That pattern holds for Florida as well. Figure 3 shows recent trends in NAEP reading and math performance both for Florida and for all U. S. 4th and 8th graders. In both subjects, at both grade levels, Florida's test gains outpaced the nation—though the gains made by 8th

5. See, for example, Eric A. Hanushek and Margaret E. Raymond, "Lessons about the Design of State Accountability Systems," in Paul E. Peterson and Martin R. West (ed.), *No Child Left Behind?* (Brookings Institution Press, 2003), pp. 127–151.

graders in reading were modest. Among African American students, the improvement by Florida 4th graders was greater than that of the average U. S. 4th grader. For African American 8th graders, Florida gains exceeded the national average in reading, but not in math. For Hispanics, Florida gains exceeded those in the U. S. as a whole in both subjects for both age groups, though the differences were not statically significant (in part because sample sizes were relatively small).

SAT Scores

Confirming evidence that Floridians have been doing better also comes from trends in performance by Florida students on the SAT, a test often used as a criterion for admission to many colleges and universities. Since SAT test participation is voluntary, only a percentage of all seniors actually take this test, making it a less than perfect indicator of how well high school students are doing. Generally speaking, the higher the percentage of test takers in a state, the lower the average score will be (as increasing increments in test takers implies greater participation by marginal students who can be expected to have, on average, lower scores).

In Florida the number of test-takers grew by 61 percent between 1998 and 2005, as compared to a gain of 26 percent nationwide. Some of the growth in Florida can be attributed to the fact that the number of high school graduates in the state grew by 33 percent (compared to 10 percent in the United States as a whole). But overall population growth is only part of the story. In addition, the percentage of high school graduates taking the test increased by 11 percentage points during this period of time, as compared to an increase of only 7 percentage points nationwide. In other words, the composition of the test-taking pool in Florida was changing more rapidly than elsewhere in the United States, as indicated by the fact that the proportion of test takers who were African American, Hispanic, American Indian, or

other non-Asian, non-White categories increased by 7 percentage points (from 32.8 to 39.5 percent) in Florida, but by just 4 percentage points in the United States as a whole (from 24 to 28 percent).⁶

When a state's test-taking population is growing, one expects average test score performance to decline. In Florida, that did happen, but only modestly. The average combined verbal and math SAT scores in 2005 were 996, just below the 1001 combined score in 1998. During that same time period SAT scores across the United States rose from 1017 to 1028. In other words, Florida's average scores fell only modestly short of keeping pace with trends across the United States, despite the fact that its test-taking population was expanding rapidly, itself a sign that schools in the state were encouraging students to seek out a quality institution of higher education.⁷

Evers and Clopton's (in chapter 6) provide a less sanguine interpretation of SAT trends in Florida. They do not think the higher SAT participation rate in Florida can account for the five point drop in test score performance, pointing to the fact that the gain in SAT scores nationally was accompanied by a moderate increase in the percentage of high schoolers taking the test. They recommend a set of curricular reforms that will ensure higher performance in the future. I agree that high school reform is urgently needed and, in this regard, it is encouraging that the Florida legislature has just approved promising reforms in last legislative session. This provides an opportunity to build upon the steady progress the state has already been making toward improving the effectiveness of its public schools. How much of that progress can be attributed to the design of its accountability system?

6. The Digest of Education Statistics published by the National Center of Education Statistics provides data on SAT scores and percentage of graduates taking the test by state. From 1999–2000 until 2003–04, the last year for which data is available, Florida's participation rate experienced the largest percentage point increase and it was the 8th-fastest growing in the country. In five of the seven other states whose SAT participation rates increased at a rapid rate, test scores fell sharply.

7. The difference between the nation's average gain in SAT scores between 1998 and 2005 and that of Florida's was of 11 points in math (or less than one-tenth of a standard deviation) and of 5 points in reading (or .045 standard deviations).

The Design of Florida's Accountability System

Florida inaugurated its first accountability system in 1973 when the legislature called for state assessments to ensure that curriculum standards were being met. Over the years, that accountability system has been strengthened. In 1996, for example, the legislature asked the state to create a five-point, numerical rating system that would rank schools based on student achievement and other factors as well as to identify the schools that were “critically low.”⁸ Although the 1996 program was an important precursor to A+, enacted three years later, the measuring stick introduced by A+ constituted a singular advance. A+ dropped the numerical scoring system in favor of a more intuitive “A” through “F” grading system, enhancing its transparency. After 2002, grades were based exclusively on test-score performance, as other factors such as student attendance rates were dropped from the grading system. Students were tested not just in selected grades but in grades 3 through 10 in reading and math, which gave schools the information they needed to track individual student performance from year to year.

Growth Scores

Little noticed, but ultimately extremely important, the state put into place a comprehensive, statewide system of data collection that gave students identification numbers that allowed each of them to be tracked for as long as that student remained within the Florida educational system. Along with student test performance, the state maintains records as to the school a student attended, whether a student changed schools, the student's teachers, and a host of background information on each student.

Once that warehouse of information had been established, it was

8. Caroline D. Herrington and Christine Johnson, “A+ Plan in Florida: Is it Working?” Paper presented before the Association of Public Policy and Management, Washington, D. C., November, 2006.

possible for Florida to modify A+ in important respects. Originally, the grading system was based solely on the level of student performance at one point in time. Beginning in 2002, the grading system was based in part on student growth, the gains in student's performance from one year to the next. Including learning gains allowed the state to do a better job of identifying more exactly the educational contribution the school was making.⁹

When Florida included student growth within its grading system in 2002, many schools that had previously prided themselves on their performance were shocked to discover that they no longer were "A" schools. No less than a third of the 357 "A" schools serving elementary students lost that ranking in 2002. Meanwhile, some schools originally thought to be of dubious quality received a higher rank. Over

9. Since 2002, approximately one half of the score a school receives depends upon the growth that a child has made from one testing period to the next. The other half of the score is based on the overall level of performance, something that can be influenced by the educational endowment the child brings to school. As West and I explain, "The new grading system gives as much as a 50 percent weight to learning gains on a 600 point scale used to calculate a school's grade. A school can attain a maximum of 200 points on this scale, depending upon the percentage of students making learning gains in reading and math. A gain is defined as improving by one performance level, making more than a full year's learning growth, or by maintaining the same performance level, if it is Level 3 or higher. A school can earn another maximum of 100 points, based on the percentage of its lowest performing students (the bottom 25 percent of the school's test-takers in reading) making learning gains (as defined above) in reading. A school can receive a maximum of 300 points based upon the percentage of its students achieving Level 3 or higher in reading and math and, in writing, the average of the percentage reaching Level 3.0 or higher and the percentage attaining Level 3.5. To receive an 'A,' the school must achieve 410 points; to receive a 'B,' it must receive 380 points; a 'C,' 320 points; 'D,' 280 points; otherwise an 'F.' 'A' schools must also show that at least half of their lowest performing students have made a year's worth of learning gains, and they must test 95 percent of their students. Otherwise, schools, to receive a grade must test 90 percent of their students and have at least thirty students who have been tested in two consecutive years in both reading and math." Martin R. West and Paul E. Peterson, "The Efficacy of Choice Threats within School Accountability Systems: Results from Legislatively Induced Experiments," *The Economic Journal*, vol. 116, issue 510 (March 2006), p. C58.

half the 196 elementary schools that had received a “D” now were given a “C” or better. Five jumped to “A” status.¹⁰ With its powerful new warehouse of data that allowed the tracking of students from year to year, Florida was now able to recognize the difference between a school that got good students and schools that helped them grow.

Incentives to Improve

Not only was the measuring stick improved but A+ also gave schools clear incentives to enhance the performance of their students. If schools improved from one grade level to another, they received an extra \$100 per student that could be used for staff bonuses or for a variety of school improvement measures, at the school’s discretion. Schools initially awarded an “A” also received the \$100 bonus, and they continued to receive the bonus if they retained their “A” level standing.¹¹ Florida schools that received an “F” had the strongest incentives to improve. They bore both the stigma of being (in 2003) among the 2 percent of all schools in Florida given a failing grade as well as the threat that a repeated “F” would give students at the school the opportunity to use a voucher to go elsewhere. In addition, “F” schools were assigned a community assessment team made up of parents, business representatives, educators, and community activists who were to write an intervention plan for the school. Schools that received a “D” were also stigmatized as being (in 2003) among the 10 percent worst performing schools in the state and, like the “F” schools, were assigned an assessment team.

10. West and Peterson, 2006, table 1, p. C49.

11. Apparently, high scores also boosted property values. See David N. Figlio and Maurice E. Lucas, “What’s in a Grade? School Report Cards and the Housing Market,” *The American Economic Review*, vol. 94, no. 3 (June 2004): 591–604. Also, a high score may, at least initially, help in the re-election of school board members. See Christopher R. Berry and William G. Howell, “Democratic Accountability in Public Education,” in William G. Howell (ed.), *Besieged: School Boards and the Future of Education Politics* (Brooking Institution Press, 2005), pp. 150–172.

Identification of School Effectiveness

If one defines effective schools as the places where the average student's test performance improves the most from one year to the next, then A+ discriminates quite well between higher quality and lower quality schools.¹² For students in grades 4 through 10, Martin West and I calculated such improvement or growth for each student between 2002 and 2003, and again for the following school year.¹³ We first estimated the growth made by the average student at that school.¹⁴ Then we calculated the difference in growth among the schools in the five categories, "A" through "F."

A good scoring system will identify stark differences in gains or growth in student test scores between those schools awarded a high grade and those given a low one. The Florida grading system does quite well in this regard. Schools given "A's" are, on average, places where students are learning more than schools receiving lower grades. At the other end of the scale, our results show that schools receiving a "D" or an "F" were clearly very low performing schools, not just schools with disadvantaged students. But even though the A+ yard-

12. Admittedly, this definition of school quality places the greatest weight on student learning, not on other factors such as improvement in student character or self-esteem. But given the fact that the State of Florida has defined as its primary objective the enhancement of a student's performance on the FCAT, a good accountability system should at least indicate clearly which schools are meeting that objective.

13. Paul E. Peterson and Martin West, "Is Your Child's School Effective: Don't Rely on NCLB to Tell You," *Education Next* (Fall 2006): 76–80.

14. To correct for mean reversion, learning gains are calculated for the students in each performance decile to have an average of zero and a standard deviation of 1.0 within that decile. The adjustment partially corrects for mean reversion by comparing the gains made by students with similar initial performance levels. For additional discussion on decile standardized gains, see Eric A. Hanushek, John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin, "The Market for Teacher Quality," NBER Working Paper No.11154, February 2005; Brian A. Jacob and Lars Lefgren, "Principals as Agents: Subjective Performance Measurement," NBER Working Paper No. 11463, July 2005; and Brian A. Jacob and Lars Lefgren, "When Principals Rate Teachers," *Education Next*, vol. 6, no. 2 (Spring 2006): 59–69.

stick used in Florida is fairly accurate—and, as discussed below, is certainly much better than the NCLB yardstick—it could be improved further. The difference between an “A” school and a “C” school is still fairly modest. This is due in part to the fact the Florida measuring stick does not focus as tightly on student growth as it should but also in part to the fact that grading standards are generous. In 2005 A+ gave 45 percent of the state’s public schools in Florida an “A,” and it gave another 21 percent a “B.” Only 11 percent of the schools were given a “D” or an “F.” In other words, nearly half the schools are given the highest grade on a grading system where the proficiency standard itself is not very high.

Florida is currently undertaking a major reassessment of the educational content of the standards it sets, the need for which is a step strongly endorsed in the curriculum section of this volume. It also is planning to raise the level of proficiency vis a vis these standards that students are expected to reach at any particular grade level. That policy, too, is well worth pursuing, as current proficiency standards, while 14th among all states, is still only modestly better than the national average, earning a “C” on the *Education Next* report card that compares stringency of standards among states.¹⁵ Admittedly, Florida is a state where student achievement has historically been quite low, and setting proficiency levels unduly high might be excessively discouraging. But as Florida continues to make educational progress, it will want to lift its own proficiency standards accordingly.

Student Accountability

A+ holds students accountable by asking them to pass the FCAT at a certain level, if they are to receive a high school diploma. The law builds upon provisions in previous Florida accountability plans, which as early as 1978 required students to achieve a passing score on a

15. Paul E. Peterson and Frederick M. Hess, “A Race to the Bottom? Keeping an Eye on State Standards,” *Education Next*, vol. 6, no.3 (Summer 2006), pp. 28–29.

basic skills examination in order to receive a diploma.¹⁶ A+ replaced the basic skills requirement with a more stringent one that said students must pass the FCAT at an acceptable level. Despite the tougher standard, Florida's graduation rates have continued to rise (see Chapter 2).

Florida's student-accountability provision is even more innovative, because it seeks to end what is known as social promotion. That occurs when students are passed on from one grade level to the next, regardless of their achievement levels. Advocates of social promotion defend this policy on the grounds that holding students back for another year undermines a child's self-esteem and results in higher dropout rates. Those who wish to end it say that requiring students to pass standards motivate higher levels of effort, even among young children.

Florida is the first state to require students to achieve a minimal level of accomplishment before being passed on to the next level. Specific expectations are set at the state level for 3rd graders, while allowing local school boards to set the expectations in all the other grades. Beginning in 2003, 3rd graders, to be promoted, must achieve at a minimally acceptable level on the reading portion of the FCAT. The requirement is somewhat less demanding than first appears. The bar is set at Level 2, one level below proficiency. And roughly 40 percent of the students who did not reach that level in 2003 were nonetheless passed on to the next grade.¹⁷ Still, that was a much lower percentage than before social promotion for 3rd graders was brought

16. Herrington and Johnson, 2006.

17. The law exempts from the "no promotion" rule students who have limited English proficiency status, have a severe disability, have already been held back for two years, or who have otherwise demonstrated competence, such as performing well on the Stanford-9 standardized test or through a performance portfolio. Jay P. Greene and Marcus A. Winters, "Getting Ahead by Staying Behind: An Evaluation of Florida's Program to End Social Promotion," *Education Next*, vol. 6, no. 2 (Spring 2006), p. 66.

to an end. Previously, 90 percent of low-scoring students were being promoted.¹⁸

Limiting social promotion in 3rd grade gave strong incentives to 3rd graders—and their teachers—to focus on reading skills. By and large, the policy seems to have had a profound impact. The percentage of very low scoring students in 3rd grade declined from an average of 23 percent in 2003 to an average of 14 percent in 2006, four years after the policy was introduced. In a careful study of the impact of the policy on student performance, Jay Greene and Marcus Winters found that the students who benefited the most were those who were held back, perhaps because the state required that they be given focused, intensive instructional services. Whatever the reason, the retained students in this year did exceptionally well (as compared to a similar group of students not held back in 2002, before the policy took effect).¹⁹ FCAT score performance was roughly 10 percent of a standard deviation higher in reading and 30 percent higher in math for those affected by the policy, as compared to those in the grade the year before the policy took effect. Significantly, similar gains were registered on a separate standardized test, suggesting that improvements could not be attributed to “teaching to the test.” Nor is there any sign that teachers focused only on the lowest-performing students, thereby overlooking the needs of higher achievers. Average FCAT scores for 3rd graders continued to rise and the percentage of students performing at the highest levels continued to inch upward in both reading and math.²⁰

18. In this context, low-scoring students are identified as those who scored below the Level 2 threshold.

19. Greene and Winters, 2006.

20. From 2003 until 2006, third grade average FCAT scale scores rose by 15 points in reading (from 298 to 313) and by 16 points in math (from 308 to 324). By 2006, 38 percent of 3rd graders performed at levels four or five in reading (up from 30 percent in 2003), and 37 percent did so in math (up from 29 percent in 2003).

How NCLB Can Learn from A+

There is much in the A+ Accountability Plan for Congress to consider when it re-authorizes NCLB. For one thing, the federal law should make distinctions similar to the “A” to “F” ones that are the heart of the A+ system. Currently, NCLB simply distinguishes between schools said to be making “adequate yearly progress” and those not so designated. That way of classifying schools is both opaque and legalistic, obscuring as much as it reveals.

But it is not only the rubric that is the problem. In addition, the measuring stick itself is flawed, because it does not directly measure student growth and therefore cannot discriminate clearly between higher quality and lower quality schools. Under NCLB, schools do not make adequate yearly progress, if the percentage of students in the next cohort of students is not achieving at a level of proficiency in reading and math that meets a statewide target necessary to bring all students to full proficiency by 2014. If the previous sentence is incomprehensible, that is the point. NCLB has devised a crude and obtuse way of measuring school effectiveness, so crude that it is also very misleading. Suburban schools with advantaged students can easily reach NCLB proficiency standards even when there is little student growth at the school. And even as NCLB makes some schools look adequate when they are not, it also does not zero in on those schools that are doing the least effective job at securing student growth.²¹

When it comes to the rigor of the grading system, both accountability systems can learn from each other, as Florida’s is too generous, while NCLB’s is too harsh. In 2005 A+ gave 66 percent of its schools either an “A” or a “B,” a generous grading system. NCLB has exactly the opposite problem. According to its way of calculating school effectiveness, no less than 77 percent of Florida’s schools were failing to make Adequate Yearly Progress (AYP) in 2004. The U. S. De-

21. West and Peterson, Fall 2006.

partment of Education modified its policy in 2005 by agreeing to grant “provisional AYP” status to 27 percent of the schools, with 36 percent said to be making AYP and just 37 percent not making AYP. It is not clear how long the “provisional” status will be granted, but its introduction adds new complexity and confusion and certainly underlines the need for further reform. In the meantime, the contrasts between the Florida grading system and the federal one are little short of bizarre, leaving parents hopelessly confused as to just why, for example, a school that receives an “A” from the state of Florida should be said to be not making adequate yearly progress. Clearly, the two systems need to be reconciled, perhaps meeting somewhere in the middle. Florida should follow through on its plans to raise proficiency standards, while NCLB should adopt a five-level grading system like Florida.

NCLB has still other things to learn from Florida’s accountability system. NCLB does not have any graduation requirement or limit on social promotion that would hold students accountable. Schools are not given any rewards or incentives to perform. And if a school is found not to be making adequate yearly progress, the immediate impacts are small. Perhaps the most unfortunate feature of the NCLB accountability system, as it operates in Florida, is its inability to differentiate cleanly between higher and lower quality schools. That would seem to be a minimum expectation for a sound accountability system. In this regard, A+ does significantly better.