Review of Florida Formula for Student Achievement: Lessons for the Nation

Reviewed By

William Mathis
University of Colorado at Boulder
June 2011

Summary of Review

Former Florida Governor Jeb Bush and the Foundation for Excellence in Education have embarked on a well-funded campaign to spread selected Florida education reforms to other states. These reforms include assigning letter grades to schools, high-stakes testing, promotion and graduation requirements, bonus pay, a wide variety of alternative teacher credentialing policies, and various types of school choice mechanisms. This policy potpourri was recently presented by Gov. Bush in Michigan, and the documents used allow for a concrete consideration and review. Regrettably, Bush’s Michigan speech relies on a selective misrepresentation of test score data. Further, he offers no evidence that the purported test score gains were caused by the recommended reforms. Other viable explanations, such as a major investment in class-size reduction and a statewide reading program, receive no or little attention. Moreover, the presentation ignores less favorable findings, while evidence showing limited or negative effects of the proposed strategies is omitted. Considering the overwhelming evidence that retention is ineffective (if not harmful), it is troubling to see Mr. Bush endorse such an approach. Finally, Florida’s real problems of inequitable and inadequate education remain unaddressed.
Introduction

The mission of the Foundation for Excellence in Education is to ignite a state-by-state transformation agenda to “support reform, primarily based on the success of the Florida Formula on Student Achievement” model. The foundation’s president is former Florida Governor Jeb Bush.

As part of the foundation’s national agenda, Bush met with Michigan legislators on June 15, 2011. The linchpin of his PowerPoint presentation is an analysis of Florida fourth-grade reading scores on the National Assessment of Educational Progress (NAEP), where substantial gains are presented as being the result of six sets of reforms: assigning letter grades to schools, high-stakes testing, promotion and graduation requirements, bonus pay, a wide variety of alternative teacher credentialing policies, and various types of school choice mechanisms.

Bush’s presentation is remarkably similar to (and uses some of the same graphics as) a Heritage Foundation report that has been repeatedly presented and re-packaged for a number of states. The lead author of the Heritage report, Dr. Matthew Ladner, was formerly at Arizona’s Goldwater Institute and is now listed as a staff member with Bush’s foundation. Typically, the National Education Policy Center would not review a presentation or PowerPoint. However, as this presentation has been replicated in several different states by a prominent national leader, it invites external and independent review. Accordingly, this review is merited, and to make allowances for the format of the reviewed document, the focus here is on the evidence underlying Mr. Bush’s proposal.

Fundamentally, Mr. Bush’s presentation is based on the fallacious claim that this selected set of loosely coupled reforms introduced in Florida between 1992 and 2011 caused fourth-grade reading score gains. As discussed below, there is reason to doubt the extent of these reported gains. Additionally, no evidence is provided or available to support the causal linkage between the test scores and the reforms. Further, this claim ignores the fact that some of the favored reforms were implemented as late as 2010, and some are not yet implemented.
Moreover, Florida’s concentration on early education starting in 1999⁵ and the introduction of a statewide reading program in 2001⁶ are downplayed, and the 2002 constitutional amendment in Florida limiting class size⁷ is not mentioned, even though the state has appropriated $18.7 billion for this program since 2003-2004.⁸ Direct reforms of this type can reasonably be expected to have more to do with fourth-grade reading scores than more distant factors such as high school graduation requirements.

In the introductory section of his PowerPoint presentation, Bush’s claim of success is staked on increases in NAEP scores—and more specifically, on five slides showing fourth-grade reading gains. Other than one later passing reference, no data are presented for other grade levels in reading or for mathematics. (However, the presentation closes with Florida state, or FCAT, reading scores.⁹)

The presentation offers trend lines beginning with the 1992 and 1994 NAEP reading tests. However, the accuracy and appropriateness of a trend line beginning with these early results is questionable because testing accommodations were not permitted back in 1992 and 1994. That is, the degree of participation in the NAEP examinations by students with disabilities and students who could not speak English is uncertain. At the other end of the same trend line, a mandatory third-grade retention policy initially eliminated low-scoring children from the fourth-grade test and then delayed those children’s exam dates by a year, thereby exaggerating gains attributed to the reforms. Fourth-grade math and eighth-grade reading and math scores, although readily available, are not presented. This is perhaps because these scores do not show as favorable a pattern either in relative gains or in national comparisons.¹⁰

**The Six Reform Clusters**

Listed below are the six reforms promoted by Mr. Bush and an assessment of what the Florida and national research indicates about each of them. I give the greatest emphasis to research specifically drawn from Florida.

**Reform #1: Grading Schools**

Several of Bush’s slides explain the evolution of and changes to the calculation of the “A-F” school grading system. The Foundation for Excellence in Education claims that this system...
has improved school pride, increased community support and focused attention on instruction. How publicly labeling a school with a “C,” “D,” or “F” improves school pride and causes the community to “rally around” these schools is not explained.

The evidence presented in favor of this grading system is a chart showing an increasing number of schools assigned grades of “A or B.” There is a sharp increase through 2003, which then continues to advance at a more moderate pace through 2010. What is telling, however, is the far flatter trend of “D and F” schools. The number of such schools fluctuates from 143 to 308 between 2001 and 2010. A plain reading of the table suggests a growing gap and greater inequities between higher- and lower-rated schools. It is troubling that the poorly performing schools appear to be being left behind.

While an Urban Institute study found that Florida schools with lower grades did, in fact, show improvements, the effects could have been caused by test preparation schemes. In reviewing this study, Dr. Damian Betebenner concluded that the fear of getting bad letter grades could not be demonstrated to have a causal relation to the improved test scores. That is, the Bush claims regarding the effectiveness of the grading system are substantially overstated.

Reform #2: Rewards and Consequences for Results

This reform provides funds to schools for improving their letter grade or maintaining an “A.” Most of the available funds were converted to bonuses for teachers. In his review of the Florida bonus system, Professor Sherman Dorn found “there is persistent evidence that giving . . . reward money or bonuses [to] teachers and non-teaching staff has led to infighting and disruption in some schools.”

Given the high correlation between socio-economic level and test scores (school grades are based “solely” on test scores), it is not surprising that the outcome of the “Rewards and Consequences for Results” policy is that schools serving more affluent communities receive the most bonus money. This accentuates and sharpens funding disparities, to the detriment of low-income students.

Also mentioned in the Bush presentation is a voucher program that existed until ruled unconstitutional in 2006. This choice program allowed students in failing schools to transfer to schools with better scores, and the Bush presentation suggests that such moves resulted in improved student achievement and test scores. However, since only 753 students out of a population of 2.7 million students were actually exercising this option in 2006, attributing any of the reported statewide learning effects to this program would not be warranted.

But the bonus program masks a far more important financial issue—one not addressed by Mr. Bush in his presentation materials. Florida’s school funding system is inequitable, according to an earlier study by Professor Douglas Harris and a recent study by Baker, Sciarra and Farrie.
Reform #3: Promotion and Graduation Requirements

The imposition of promotion and graduation requirements is also claimed to increase academic achievement and graduation rates.

Mr. Bush asserts that the improved fourth-grade NAEP scores were due largely to learning improvements caused by a test-based retention policy in the third grade. As noted earlier, there are several key problems with this assertion. Initially, it should be noted that Mr. Bush pays much less attention to a massive state reading initiative that included coaches, class-size reduction, and early literacy screening. These are all evidence-based programs (that is, programs supported by substantial outside research) with the clear potential to have a positive influence on fourth-grade scores.

Moreover, the Florida third-grade retention policy was shown by Columbia University Professor Madhabi Chatterji to very likely be the cause for much or most of the NAEP gains—but not in the positive learning sense that Mr. Bush is arguing. Chatterji demonstrates that by screening for low reading scores and then holding these students back a year, the state is able to initially exclude low-scoring students from the fourth-grade NAEP. Then, once these students are promoted to the next grade, the state is able to give the fourth-grade test to a group of students who would otherwise be fifth-graders. That is, these students have another year of learning under their belts. Further, these retained students are disproportionately from minority groups, meaning that the retention policy simultaneously falsely inflates overall scores while creating a misleading impression that the achievement gap is closing.

A troubling aspect of Mr. Bush’s proposal is his assertion that test-based “Retention Works.” In support of this claim, he quotes a Manhattan Institute article by Jay Greene. However, the massive body of research literature on the effects of retention lead to the opposite conclusion. In summarizing this research for the peer-reviewed \textit{Educational Researcher}, Professor Randall Penfield states,

\begin{quote}
The lack of any beneficial effect of retention on later achievement has also been strongly supported by the results of a series of meta-analyses. . . . Without exception, the results of the meta-analyses indicated that retention generates no positive impact on student achievement.\end{quote}

Penfield goes on to observe that in 169 analyses the retained group scored 39% of a standard deviation below the matched promoted group. The clarity of these findings demonstrating the adverse impact of retention (particularly on minority groups) leads the
author to question the fairness of retention as an educational policy, whether such a policy violates professional standards and whether such use is ethical. 18

Penfield uses two references from the Florida Department of Education “demonstrating the disproportionately high retention rates for (minority) students,” 19 but those two links are not operative.

The Bush presentation also suggests that increasing the “rigor” of graduation requirements has been a vital factor in Florida’s decrease in dropouts. However, the recent National Academies review of the literature states that high-stakes accountability systems do not improve instruction or retention and may, in fact, have the opposite effect. 20 Additionally, the international PISA results conclude that nations retaining students have greater socio-economic gaps and that retention policies negatively affect students from lower socio-economic groups. 21

Interestingly, Mr. Bush’s PowerPoint presentation gives no credit to Florida’s dropout prevention program for the better graduation figures. 22 Also, the effect of tying driver’s licenses to staying in school may have a significant, but unexamined, effect. 23

Reform #4: Funding for School Success

This category summarizes grants for “D and F” schools and summer schools. The majority of the section focuses on showing that college entrance examination scores and participation rates increase as a result of the state paying for the PSAT and for bonuses paid to “incentivized teachers” whose students pass advanced placement tests. As previously noted, teacher bonuses lead to infighting and low morale, 24 and more generally the research on incentive pay for teachers strongly suggests that this is a dead-end approach for teacher improvement. 25 Amrein and Berliner have also documented the perverse actions that occur when too much emphasis is placed on test score gains. 26

In a slide titled “Stop Funding Failure,” Mr. Bush makes a point of saying “$700 million per year” (emphasis in the original) is appropriated in Florida for various types of needy students. While the presentation suggests a major investment, this must be considered in light of a $17 billion annual state department of education budget. 27 That is about 4% of the state department’s budget. A conservative review of the literature suggests an average figure of 27.5% additional support is needed for such students. 28 Thus, what is cast as a large sum of money can be viewed as grossly inadequate.

More importantly, Florida’s funding problem cannot be resolved by such weak measures. In 1999-2000, Florida ranked 38th out of the 50 states on per-pupil spending. Harris found that the system not only is inadequately funded but also distributes funds inequitably. 29 This was confirmed in 2010 by Baker, Sciarra and Farrie, who gave Florida a grade of “D” for funding distribution and a rank of 46th in funding coverage, a measure of fairness across income level. 30 For 2011, Education Week gives Florida a more generous “C,” but the state still comes in below the national average, with a score of 71.9 in school finance. 31
Reform #5: Quality Educators

In contrast to the other reforms that call for “higher standards,” the reforms grouped under “quality educators” propose a set of policy actions that appear to lower standards for entry into teaching. Routes into teaching would include eight additional paths. Among the more non-traditional proposals for teacher licensing are state community college programs for career changers, on-the-job training, persons who have taught in higher education for two years, education minors, and Teach for America recruits.

The proposed solutions may help relieve teacher shortages in a system like Florida’s that is afflicted with low salaries, regional shortages and subject matter shortages. Yet the linkage to improved teacher quality is not evident. In fact, although these initiatives were implemented in Florida in 2002-2003, Mr. Bush presents no evidence that any of them has been successful in achieving the stated goal of improving teacher quality.

Another set of reforms includes changes to the seniority and tenure systems and a new test-based teacher evaluation system, which is to be implemented by 2013-2014. In the Florida context, Professor Gene Glass warned back in 2004 that “Value-added teacher evaluation methods, which attempt to evaluate teachers in terms of the standardized achievement test score gains of their students, are of uncertain validity, have drawn heavy criticism from measurement experts, and raise serious concerns about fairness.”

The issue has since been the subject of considerable research. In the national context, ten of the nation’s most prominent educators and psychometric authorities jointly issued a statement in 2010 declaring that such test-based teacher evaluation is unwise. In his New York City study, Professor Sean Corcoran found that the variation in teacher value-added scores provided an unacceptable uncertainty rate of 34% when three years of data were used. The figure increased to 61% when only a single year’s test scores were employed. Similarly the re-analysis Briggs and Domingue conducted of the Los Angeles Times rating of teachers found that teacher success categories changed in 54% of the cases when an equally (or more) appropriate model was used. The error rates and fundamental invalidity of such systems argue that they cannot be used to promote student achievement or improve teacher quality.

Further, Florida’s test-based teacher evaluation system is yet to be implemented. Mr. Bush’s apparent suggestion that past success can be ascribed to reforms that have not yet been implemented seems far-fetched.

Reform #6: Choices, “Lots and Lots of Choices”

A potpourri of school choice mechanisms is presented as part of Florida’s “success” formula. Vouchers for special education students (McKay scholarships), tax credits (so-called neovouchers) to fund private school for low-income children, charter schools, pre-K vouchers and a newly expanded virtual schooling law are all part of the mix.
While the Bush presentation does allude to three un-named papers extolling the advantages of competition, the presentation includes no evidence of how these choice schemes may have influenced test scores. In fact, the proportion of students involved (10% or less cumulatively) suggests that even if these options have been producing positive results, they are unlikely to have substantially contributed to the touted statewide gains. But even the underlying assumption of positive results is highly questionable. There exists, in fact, a very strong research base concerning school choice policies. Notwithstanding their political popularity, they have failed to produce achievement scores exceeding those of traditional public schools (some studies show negative associations with test scores), and they segregate the student population by race, class, and status as students with special needs and students learning English. Further, charter schools tend to re-direct funds from instruction into administration and overhead.

**Bush’s Conclusions**

The concluding section of the PowerPoint presentation turns to state test scores (FCAT rather than NAEP). This part of the presentation is limited to six charts, all of which are reading scores between 2001 and 2010. Math and other subjects are not addressed.

The FCAT scores are most interesting when compared to the NAEP scores over the same years. The NAEP reading gains were strongest from 1998 to 2003, while the FCAT gains were most prominent from 2003 to 2007. This suggests that the two sets of tests may be measuring different things, the measures themselves may have changed, or other factors are at play. The presentation of FCAT scores lumps different grade levels together, ignores the shift in grades tested, and assumes that mastery standards across grade levels are comparable.

Finally, the report concludes with various ratings of Florida’s education system, each apparently selected to demonstrate the gains the raters attribute to the foundation’s favored reforms. Noteworthy is the high *Education Week* ranking, which was primarily based on whether the state adopted the reforms the newspaper endorsed. The source or definitions of other high rankings are not identified. Notably, less favorable rankings concerning matters such as spending equity and adequacy (as well as rankings concerning scores on other tests of Florida’s students) are not included, suggesting that the ratings used in the Bush presentation were selected to create the appearance of successful reforms.

**Conclusions**

The Foundation for Educational Excellence’s PowerPoint presentation is an advocacy tool designed for advancing a particular set of reform proposals. Significant flaws invalidate this work:
• **False Ascription of Reform Effects.** Throughout the presentation, the reported test score gains are attributed to the six sets of reforms. Yet no evidence is presented that demonstrates that any of the reforms was responsible for the purported positive outcomes.

• **Inadequate Effectiveness Criteria.** The premise that the six favored reforms have produced achievement gains is based on fourth-grade NAEP Reading scores. Fourth-grade Math results and eighth-grade Reading and Math scores do not show comparable gains and are omitted from the presentation. Furthermore, the early years of the growth chart reflect a time when non-native English speakers and students with disabilities did not receive testing accommodations—a condition that changed in later years. Also, fourth-grade test scores are artificially inflated by Florida’s grade-retention policy, first by the removal of low-scoring children and later by the inclusion of more mature students. Thus, the growth pattern on the NAEP is suspect. The concluding section on FCAT scores is broader, yet it too suffers from being limited to reading, the inability to validly ascribe causality, and the limitations normally associated with using a state’s high-stakes assessment—particularly for long-term comparisons.

• **Cherry-picked Conclusions.** The “gains” are ascribed only to the selected reforms. The effects of other possible causes, such as Florida’s constitutionally mandated class-size reductions, enhanced early education, and the implementation of a statewide reading program, are not seriously considered.

• **Unsupported conclusions.** The proposals ignore significant bodies of both Florida-specific and national research literature that document the limitations—or the outright failure—of many of the proposed reforms. These omissions are most notable in the sections dealing with social promotion, school choice, teacher preparation, funding inadequacy and high-stakes assessment. To be sure, a PowerPoint presentation would not generally be held to the same criteria as a peer reviewed article. Nevertheless, the sweeping and unsupported nature of the reform claims call for substantiation.

• **Real problems are ignored.** The most troubling aspect of the presentation is the avoidance of and failure to address real problems while, instead, embracing false solutions. This avoidance increases the likelihood that these real educational problems will intensify. Unfortunately, if research is our guide, the effect of the Florida reforms will likely prove to be a more inequitable and inadequate educational system.
Notes and References

1 For background, see the Foundation for Excellence in Education at http://www.excelined.org/.

2 Florida Formula for Student Achievement: Lessons for the Nation. (June 15, 2010) Tallahassee, FL: Foundation for Excellence in Education. (69 PowerPoint Slides.) These slides are no longer available on the website of the Foundation for Excellence in Education. They are available upon request from the review's author.


This Heritage report is the top resource listed by the Foundation for Excellence in Education’s “Research Round-up:” http://www.excelined.org/Pages/Reform_Library/Research_Round_Up.aspx.


9 NAEP scores are the generally accepted measures as they are considered less susceptible to teaching to the test, changes in test content and changes in the standards themselves.


See also


See also


See also


DOCUMENT REVIEWED: Florida Formula for Student Achievement: Lessons for the Nation

PRESENTER: Jeb Bush

PUBLISHER/THINK TANK: Foundation for Excellence in Education

DOCUMENT RELEASE DATE: June 15, 2011

REVIEW DATE: June 30, 2011

REVIEWER: William Mathis, University of Colorado at Boulder

E-MAIL ADDRESS: William.Mathis@colorado.edu

PHONE NUMBER: (802) 383-0058

SUGGESTED CITATION: