



# Policy Brief

---

Matthew Di Carlo  
June 2015

## The Evidence on the “Florida Formula” for Education Reform

### Executive Summary

The education reforms implemented in Florida throughout the late 1990s and 2000s, commonly known as the “Florida Formula,” have received a great deal of attention in recent years. The policies included in this package are focused on test-based accountability, competition, and choice. Its supporters often rely on crude, speculative forms of policy evidence, such as aggregate, unadjusted test score changes coinciding with the period of the reforms. In reality, while not all of the policies constituting the “Formula” have been subject to empirical scrutiny, there is a relatively large body of high quality research available on a number of its key elements. Several of these policies have had a positive estimated impact, gauged mostly in terms of testing outcomes, whereas others have not. And there is virtually no evidence of any negative impacts. Overall, however, most of the evidence on the “Florida Formula” is likely still to come, and the research that does exist supports nuanced, cautious policy conclusions.

## Introduction

During the late 1990s and 2000s, the State of Florida enacted a set of education reforms spearheaded by Governor Jeb Bush. These policies, which emphasize test-based accountability, competition, and choice, have since become known as the “Florida Formula for education success,” or, simply, the “Florida Formula.” In recent years, there has been a coordinated, aggressive effort to advocate for its implementation in other states.

The “Formula” is a multifaceted package that might be summarized as a set of concepts or goals, which are manifested in specific policy interventions. A brief summary of these concepts, along with the primary policies that embody them, is as follows:

1. **Hold schools accountable** – “A-F” school grading system, attached to rewards and consequences;
2. **School choice** – charter schools and different forms of private school choice programs;
3. **High expectations** – retention/remediation of low-scoring third graders, higher graduation standards;
4. **Funding for school and student success** – tying funding to performance and more flexibility in how districts can spend money;
5. **Quality educators** – alternative teacher certification and new teacher evaluations.

In arguing that the Formula has been a success, many of its proponents employ as their evidence changes in aggregate testing results, most commonly unadjusted increases in statewide proficiency rates on Florida’s state assessment (FCAT), and/or increases in average fourth and eighth grade scores on the National Assessment of Educational Progress, or NAEP (e.g., Foundation for Excellence in Education, 2010).

The basic argument among these supporters is that the FCAT rates and NAEP scores increased substantially during roughly the time period the policies were put in place, and that this improvement was due to the enactment of the policies themselves.

This approach, while common and intuitively appealing, violates basic tenets of causal inference and policy evaluation. For one thing, the increases in question compare the performance of different groups of students – NAEP changes, for example, might compare fourth graders in one year with fourth graders from previous years.

Compositional differences between samples can influence trends in average scores (Kane and Staiger, 2002), and such differences cannot be addressed fully by disaggregating trends by student subgroup. There is a big difference between, on the one hand, testing whether students in one year perform better than did previous cohorts, and, on the other hand, assessing whether either group of students *improved* during their time attending a given school or district.

On a related, perhaps more important note, one cannot simply assume that changes in *student* performance are due to changes in *schools'* effectiveness (Glazerman and Potamites, 2011), to say nothing of assuming they are attributable to a specific policy or set of policies.

The latter is particularly salient in Florida, where a series of prior and concurrent reforms were also in place, including not only state-level policies such as class size reduction and an investment in reading coaches, but also the sweeping changes of the federal No Child Left Behind law, which took effect right near the start of the time period in question. And several of the Florida Formula's constituent policies, as discussed below, were targeted directly at relatively small subgroups of students – for example, students in F-rated schools and third graders who were retained due to low reading scores.

It also bears mentioning that, while most of the Formula's core components were enacted in the late 1990s and early 2000s, some were added and/or altered substantially over the course of more than a decade, and even those in place since the beginning might have taken years to unfold (e.g., the proliferation of charter schools or the hiring of alternatively-certified teachers). This means that the aggregate change in testing results over the roughly 15 years usually presented by advocates reflects varying configurations of the Formula and different “extents of its implementation.”

Overall, then, the unadjusted NAEP changes so often emphasized by the Formula's supporters permit only the most tentative conclusions about the test-based impact of these specific policies. Suggesting otherwise represents a potentially dangerous form of *ad hoc* policy analysis, one by which virtually any policy can be shown to have worked or failed.<sup>1</sup>

Maryland, for example, exhibited the largest NAEP increases in the nation between 1992 and 2011, larger than Florida's (Hanushek et al., 2012), and did so with a completely different, in many respects “ideologically opposite,” set of policies.

Such speculation, fortunately, is not necessary, as there is already a fair amount of strong research evaluating many of the core reforms comprising the “Florida Formula” (see Chatterji [2010] and Mathis [2011] for previous reviews of research on the Formula).

This body of work, which continues to grow, provides a basis for initial conclusions as to the short- and medium term impacts of several of the Formula's components, at least on test scores and related student performance outcomes.

The purpose of this policy brief is to review this evidence in a manner that is fair and useful to policymakers and the public.

---

<sup>1</sup> Some of the Formula's supporters also posit as evidence of the system's success that the distribution of grades has improved over time. This approach suffers from the same basic limitations as the use of simple test score/rate changes, but there is an additional twist in this particular instance: Although it is true that the “average grade” has increased over time, these increases coincide with, and are due largely to, changes in the grades' calculation. For example, a sweeping change in the calculation took effect in 2002, and an analysis of grades using both the old and new systems found that over half of schools (52 percent) would have received the same grade under both systems, but that almost two in five received a higher grade under the new system than they would have received using the previous formula (Rouse et al., 2013).

## Review of the Evidence

Although there is some overlap between the five core areas of the “Florida Formula,” the easiest way to approach this review is to discuss them one at a time.

### Hold schools accountable

In 1995, Florida became one of the first states to adopt its own school grading system, which are now ubiquitous throughout the nation. In Florida, schools receive a grade on the “A-F” scale familiar to students (though, of course, students are not judged by a single grade). The system has changed considerably over the years, but it has always been based largely on standardized testing results for elementary and middle schools, while the current high school grades also incorporate other measures, such as graduation rates.

The main purposes of Florida’s rating system, as well as those in other states, are to inform parents and other stakeholders about “school performance,” as well as to incentivize improvement and innovation by attaching consequences and rewards to the results. Starting in 1999, with the passage of Florida’s “A+ Plan” for school accountability, the grades in Florida were high-stakes. Students who attended schools that received an F in two of the previous four years, including the most recent year, were eligible to switch to a higher-rated public school or to receive private school vouchers, called “Opportunity Scholarships” (this program was shut down in 2006, after being ruled unconstitutional by the state’s Supreme Court).

In addition to the voucher/transfer threat, low-rated schools received targeted assistance, such as reading coaches and assessment teams, while high-rated schools were eligible for bonuses (discussed below). In this sense, the grading system is a cornerstone of the Formula’s school accountability apparatus.

Perhaps the best analysis of the effect of the incentives underlying the school grading system, at least in its earlier years, is presented in a paper by Rouse, et al. (2013). Using multiple student tests, as well as surveys of principals over a five-year period during the early-2000s, the researchers sought to assess both the test-based impact of the grades (presumably, the competitive effect on public schools of losing students to higher rated public schools or private schools), as well as, importantly, how low-rated schools responded to the accountability pressure in terms of changes in concrete policy and practice.

They conclude that test-based performance did indeed improve among the small subset of schools that had received F grades during the system’s first few years, relative to similar schools that had received a higher grade. The difference is sizeable in magnitude, and it persists in the few subsequent years included in the analysis. As with any examination of the impact of a test-based accountability policy, the key question is how schools achieved these improvements, as attaching stakes to tests can encourage undesirable behaviors, such as focusing on tested at the expense of non-tested subjects.

Although their results require cautious causal interpretation (Betebrenner, 2008), Rouse and colleagues, using their principal survey data, find that some of the improvement in low-rated Florida schools appear to be associated with specific steps that the schools reported having taken, such as identifying and increasing instructional focus on lower-performing students and lengthening instruction time (also see Feng, et al. [2010]). This, along with the inclusion of low-stakes exam data in the analysis, serves as suggestive evidence of desirable behavioral/policy changes playing a role in the testing outcomes.

Chakrabarti (2013), using school-level data from the mid-1990s to early-2000s, reaches the same overall conclusion – i.e., that F-rated schools responded to the pressure and were able to generate modest improvements in their performance (also see Chakrabarti [2008]). In this analysis, however, there is more evidence of “gaming” responses, such as reclassifying students, as well as redirecting instruction toward subjects, writing in particular, in which score improvements were perceived as easier to achieve (also see Figlio and Getzler [2006]; Goldhaber and Hannaway [2004]).

Greene (2001) also examines the earlier impact of the grade-based “voucher threat,” using school-level data from the 1998-99 and 1999-2000 school years. He finds that schools receiving F grades (and thus facing the threat) exhibited larger test score improvements than schools receiving higher grades.

West and Peterson (2006), finally, employ student-level data between the 2001-02 and 2003-04 school years, including data from multiple tests, to evaluate the grading system’s incentives. They conclude that F-rated schools improved relative to those receiving D grades, and that this impact is concentrated among African-American students, those eligible for free lunch, and those with low prior test scores. They also provide some evidence that the handful of schools receiving D grades improved a little bit (relative to those receiving C’s), but that schools receiving A-C grades did not vary in their performance. This does not necessarily mean that the latter did not improve, only that, according to this analysis, there was no discernible variation between them in terms of this impact. Note that the findings for D-rated schools, which did not face the voucher/transfer threat, suggest that the impact of the grading system may also stem from the stigma of receiving a poor grade, and not only from the direct threat of voucher eligibility faced by schools receiving F grades (also see Figlio and Rouse [2006]; Goldhaber and Hannaway [2004]; Greene [2001]).

Based on these analyses, which focus on the earlier years of the grading system, prior to the end of the “Opportunity Scholarship” program in 2006, it is fair to say that the small group of Florida schools that received low ratings, when faced with the threat of punishment for and/or stigma attached to the grades, responded in a variety of strategic ways, at least some of them desirable, and that this response was associated with test-based improvement. The latter is consistent with research on state and district grading systems elsewhere (e.g., Winters and Cowen, 2012).

## **School choice**

Florida’s school choice-related policies have been subject to a healthy amount of empirical scrutiny. In a sense, some of the papers discussed in the previous sub-section represent examples of such work, since eligibility for the “Opportunity Scholarships” was, prior to 2006, tied directly to the A-F grading system.

Figlio and Hart (2014) examine the competitive impact of a different choice program, Florida's "Tax Credit Scholarship Program," by which low-income students, starting in 2001, are able to attend private schools, with tuition paid by corporations that receive tax credits in return.

The researchers evaluate whether the introduction of this program spurred improvement among the public schools facing the threat of losing students to nearby private schools. The results suggest that the extent of competition -- in this case, the ease of access to and stock of private schools in the area -- was associated with increased performance of public schools, particularly in Dade County, home to Miami. They also find a more pronounced estimated effect among schools that had more to lose from shedding low-income students, as they were close to the threshold for Title I eligibility.

Florida's other major voucher program, the "McKay Scholarship for Students with Disabilities," which started in 1999, provides students with disabilities the opportunity to switch schools intra-district, attend a public school in an adjacent district, or attend a participating private school. Greene and Winters (2008), analyzing data from the 2000-01 to 2004-05 school years, find that the number of McKay-participating private schools in the area is associated with higher performance among public school students diagnosed with mild disabilities, but not among those with more severe disabilities. They also report that schools responded to the threat of losing students to private schools, at least during the first few years after the introduction of this pressure.<sup>2</sup>

These analyses indicate that Florida's private school choice programs may have had a positive and at least modest effect on public schools facing the competition. The previous literature on competitive effects, however, is not entirely consistent. Some analyses, like those discussed above, find a small positive impact, (e.g., Dee, 1999), while others do not (e.g., Sander, 1999).

Charter schools are also a major part of Florida's reform environment. The state maintains a robust charter sector, and there is some strong research about these schools' impact on student performance.

One somewhat older evaluation of Florida charter schools between the 1999-2000 and 2001-02 school years concludes that they initially (i.e., after opening) produce inferior gains in math and reading, relative to comparable regular public schools, but, by their fifth year, their average performance is modestly higher in reading and statistically indistinguishable in math (Sass, 2006). The results also suggest that nearby regular public schools respond to the pressure from charters, with minor differences showing up in math, but not reading.

An analysis using more recent statewide data from the 2006-07 through 2010-11 school years finds no discernible charter impact in math and a statistically significant negative impact in reading, but the magnitude of this estimated effect is extremely small (CREDO, 2013).

CREDO's (2015) examination of charter performance in 41 urban regions includes data from seven such regions in Florida – Fort Myers, Jacksonville, Miami, Orlando, St. Petersburg, Tampa, and West Palm

---

<sup>2</sup> Note that this paper is the only one in this review that looks at the impact on the performance of students who receive vouchers, rather than their effect on school effectiveness. This is because analyses of the "direct effect" of Florida vouchers and voucher-type programs are scarce. See Rouse (1998) and Peterson et al. (2003) for evaluations of voucher programs outside of Florida.

Beach. The results are decidedly mixed. Estimated charter effects in math are not significant in Jacksonville, Orlando, and St. Petersburg, negative in Fort Myers and West Palm Beach, and positive in magnitude in Miami and Tampa. In reading, CREDO finds no statistically discernible impact in Orlando and Tampa, a negative estimated effect in Fort Myers, Jacksonville, St. Petersburg, and West Palm Beach, and positive results in Miami. Most of the statistically significant coefficients are either small or moderate in magnitude.

And, finally, an analysis by Zimmer and colleagues (2009), which includes charter schools in eight locations, focuses on graduation and higher education outcomes among students attending Florida charter high schools (particularly those who came from charter middle schools). Among these students, they find a large positive impact of charter attendance on the likelihood of graduation and college attendance (also see Booker et al. [2008]).

The evidence therefore suggests that the impact of Florida charters is, at best, rather mixed, and may even vary between urban areas within the state. Where there are significant estimated impacts either way, they tend to be rather small. This squares with the large body of research on charter schools' test-based effects, which have been found to vary substantially, within and between states (Di Carlo, 2011). On the other hand, there is limited evidence that the state's charter high schools may perform better than regular public schools when gauged by high school graduation and college attendance.

## Set high expectations

In 2003, Florida began holding back (and remediating) students with low scores on the state's third grade reading exams – i.e., ending “social promotion” of third graders. It is still a bit too early to get a sense of the longer term effects of this policy, as the first affected cohorts are just recently finishing their K-12 education.

There is, however, some initial work looking at this policy's impact. Winters (2012), in an extension of previous analyses (Greene and Winters, 2007; Greene and Winters, 2009), finds that students who were held back scored substantially higher than their counterparts who just barely made the cut (i.e., were promoted to fourth grade). These (relative) impacts seem to have persisted through seventh grade, the point at which the data end. Schwerdt and West (2013) also find positive but more modest effects of the Florida retention policy on affected students, relative to their similar-but-promoted peers, although their results suggest that the impact fades out over time, and is not statistically discernible after six years.

It is crucial to note that Florida's intervention with these struggling third graders does not consist solely of retention. The retained students are also required to attend summer school and receive ongoing help, which may account for some of the positive estimated effects. In other words, it is plausible that these students would have performed similarly even if they were promoted, but still received the other supports (see Briggs, 2006).

An evaluation of a New York City program with a similar structure, for example, in which retained fifth graders were provided with supports, also finds positive results (Sloan McCombs et al., 2009). Jacob and Lefgren (2004) look at a 1996 Chicago policy tying promotion and summer school decisions to test scores, and conclude that this policy led to higher short term achievement among third but not sixth



graders. Research on grade retention without the additional interventions, however, offers somewhat mixed conclusions (Jimerson, 2001), though many of these studies suffer from data and methodological limitations.

These findings lend some support to the plausible conclusion that an extra year of schooling for low-scoring students, *when accompanied by extensive remediation efforts*, might improve testing results in the short- and perhaps medium term, though there is some indication that the impact fades out by the time the retained students reach high school.

Finally, another element of Florida's "high standards" component is a 2002 policy increasing the standard for passing Florida's high school exit exam. There does not seem to be any high-quality evidence on this policy as yet, and any attempt to evaluate it should probably rely on post-graduation outcomes, such as college attainment. Similarly, the state's recent adoption of the Common Core State Standards is another example of a policy that might fall under the "high expectations" component of the Formula, but implementation of these standards is still in its earlier phases.

## **Funding for school/student success**

This component of the Formula, which is also connected to the grading system discussed above, targets funding toward schools and students based on their measured performance and improvement in that performance. For example, schools receive a certain amount of additional per-pupil funds for earning a grade of A, or for improving their grades between years.

Districts were also given more flexibility in how they distribute their own state funds. For example, money previously allocated for summer school or dropout prevention could be used for interventions during the school year, such as literacy programs. Districts were also required/encouraged to direct funds to schools receiving D and F grades, and/or to subgroups such as struggling third graders and students having trouble meeting FCAT graduation requirements. Dorn (2004) argues that the within-district redistribution may be the most significant policy manifestation of this particular Formula component.

There are no high-quality empirical examinations of the effect of these specific policies on student performance (or their cost-effectiveness), though it is quite possible that part of their impact, if any, is reflected in some of the analyses of the "A+ Plan" discussed above. If, for instance, an F-rated school improved, that outcome may be due in part to the targeted assistance, and not just the accountability pressure. When offering policy implications, one might also have trouble separating the "funding for success" policy itself from the interventions for which it pays – e.g., if the interventions targeted at third graders struggling in reading show results, the takeaway might simply be that it is wise to invest in third grade literacy, rather than interpreting this as evidence in support of funding flexibility.

On a related note, Florida, due to the "A+ Plan," implemented the first statewide teacher performance pay plan in 2001. The specifics of this policy changed quite a bit during the 2000s, but the general idea is that teachers are awarded bonuses based on the testing performance of their students, including that on AP exams (see Buddin et al., 2007).



Once again, there do not seem to be any high-quality evaluations of this particular intervention, but, if the evidence on other U.S. teacher bonus programs is any indication (see Springer et al., 2011; Glazerman and Seifullah, 2012), the incentives are unlikely to have had an effect on *short term* testing results.

On the whole, the “funding for school/student success” component of the Formula *per se* is a difficult target for empirical research, and there is little evidence available. Insofar as it is primarily a resource allocation policy, its evaluation might focus on how it affects the use and distribution of funds, rather than student and school performance outcomes, which are the focus of this review.

## Quality educators

Like many other states, Florida is currently overhauling several of its teacher personnel policies, including the design and implementation of new performance evaluations. The impact of these policies, which might fall under the “quality educators” component of the “Florida Formula,” remains to be seen, as they are still relatively new. Starting in 1999, however, the primary manifestation of the “quality educators” component was expanding alternative certification routes for prospective teachers.

Due to a growing population and a statewide class size mandate, the demand for new teachers was strong throughout most of the 2000s, and a very large proportion of new teachers in Florida are now certified via a variety of alternative pathways. These opportunities were also intended to attract talented individuals into the profession who might have gone elsewhere if they had been required to obtain traditional certification.

In a recent, very extensive analysis of alternative pathways to teaching in Florida, Sass (2007) utilizes detailed data between the 2000-01 and 2006-07 school years. He finds that the pre-service characteristics of alternatively certified teachers are substantially stronger than those of their traditionally certified colleagues, and that the differences in test-based effectiveness between these groups once they reach the classroom varies by the type of alternative pathway.

For example, new teachers certified via the most common alternative option during the period of this analysis (district alternative certification) exhibit test-based performance that is not statistically different from that of their traditionally prepared colleagues. Teachers certified via the “Educator Preparation Institute” alternative option are significantly less effective in boosting student test scores, while those alternatively certified by the “ABCTE Passport” option are substantially more effective.

These results as a whole suggest, again, that the associations vary quite a bit by type of pathway, but that the test-based performance of alternatively certified teachers is generally comparable to that of their traditionally certified colleagues. This squares with evidence elsewhere (Boyd et al., 2008; Kane et al., 2008).

It is, however, almost certain that these additional pathways have helped fill the demand for new teachers in Florida, particularly during the first years of the class size mandate, and it is more than reasonable to assume that they served to attract some qualified individuals into the profession who would have chosen a different career path had they been required to travel the traditional route. But it is

difficult to ballpark how and whether opening up these alternative pathways affected the distribution of teacher quality in Florida.

## Conclusion

Rendering a blanket judgment about a package of reforms as varied as the “Florida Formula” is almost always unwise. It is usually the case that some policies work, while others do not, and that these effects vary by location and subgroups, and are subject to change over time.

In the case of the Formula, there is still no research available for a couple of its core components, including the funding and graduation requirement policies. Moreover, even among those constituent parts that have been subject to empirical scrutiny, the evidence is still building. For example, evaluations of the impact of the third grade retention policy, due to data availability, have not yet followed a cohort beyond 5-6 years, and the data used in the analyses of the grading system and alternative certification pathways are at least ten years old. This is very common – good research takes time.

That said, a few of the evaluated reforms -- most notably, the voucher-triggering school grades during the early 2000s, private school choice programs (gauged mostly in terms of competitive effects), and third grade retention/remediation – seem to have generated discernible and at least modest increases in test-based performance among subgroups of students and schools (i.e., the small group of F-rated schools in the early years of the grades-based voucher program, and low-performing third graders in the short term).

In a couple of other cases, such as charter schools and alternative certification, there seems to have been little discernible aggregate impact on testing outcomes either way, though there is some indication of strong and positive charter high school effects on graduation, and expanding alternative certification pathways may have increased the supply of teachers.

And, finally, there is little if any evidence that these policies had any negative impact on student performance outcomes.

Overall, the (mostly test-based) evidence on these policies, at least those that have been subject to significant empirical scrutiny, might be summarized as mixed but leaning positive. It is, however, most useful to review the evidence about the Formula on a component-by-component basis, rather than putting forth generalizations about their collective impact. And, given the fact that many of these interventions are still relatively young by educational policy standards, it is a safe bet that much important research is still to come.

In the meantime, policy makers, advocates and the public should avoid sweeping conclusions about the impact of the Florida Formula or any combination of its components, whether positive or negative, and regard with extreme skepticism any attempts to attribute raw testing results to these or any other specific reforms.

## References

- Betebrenner, D. 2008. Review of "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit.
- Booker, K., Sass, T. R., Gill, B., and Zimmer, R. 2008. *Going Beyond Test Scores: Evaluating Charter School Impact on Educational Attainment in Chicago and Florida*. Santa Monica, CA: RAND Corporation.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., and Wyckoff, J. 2009. Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis* 31(4), 416-440.
- Briggs, D. 2006. Review of "Getting Farther Ahead By Staying Behind: A Second-Year Evaluation of Florida's Policy to end Social Promotion." Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit.
- Buddin, R., McCaffrey, D. F., Nataraj Kirby, S., and Xia, N. 2007. *Merit Pay for Florida Teachers: Design and Implementation Issues*. Santa Monica, CA: RAND Corporation.
- Chakrabarti, R. 2013. Accountability with Voucher Threats, Responses, and the Test-Taking Population: Regression Discontinuity Evidence from Florida. *Journal of Education Finance and Policy* 8(2), 121-167.
- Chakrabarti, R. 2008. Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs. *Federal Reserve Bank of New York Staff Report No. 315*.
- Chatterji, M. 2010. Review of "Closing the Racial Achievement Gap: Learning from Florida's Reforms." Boulder, CO: National Education Policy Center.
- Center for Research on Education Outcomes (CREDO). 2015. *National Charter School Study 2013*. Stanford, CA: Center for Research on Education Outcomes.
- Center for Research on Education Outcomes (CREDO). 2015. *Urban Charter School Study: Report on 41 Regions*. Stanford, CA: Center for Research on Education Outcomes.
- Dee, T. 1998. Competition and the quality of public schools. *Economics of Education Review* 17(4), 419-427.
- Di Carlo, M. 2011. *The Evidence on Charter Schools and Test Scores*. Washington, D.C.: Albert Shanker Institute.
- Dorn, S. 2004. *Alternatives for Florida's Assessment and Accountability System*. Tempe, AZ: Education Policy Studies Laboratory, Education Policy Research Unit.
- Feng, L., Figlio, D. N., and Sass, T. 2010. *School Accountability and Teacher Mobility*. CALDER Working Paper 47. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research.
- Figlio, D. N., and Getzler, L. S. 2006. "Accountability, Ability, and Disability: Gaming the System?" In Gronberg, T. J., and Jansen, D. W. (eds.), *Advances in Applied Microeconomics, Volume 14*. Oxford: Elsevier Science Press.
- Figlio, D. N., and Hart, C. M. D. 2014. Competitive Effects of Means-Tested School Vouchers. *American Economic Journal: Applied Economics* 6(1), 133-156.
- Figlio, D., and Rouse, C. 2006. Do Accountability and Voucher Threats Improve Low Performing Schools? *Journal of Public Economics* 90, 239-255.
- Foundation for Excellence in Education. 2010. *Florida Formula for Student Achievement: Lessons for the Nation*. Tallahassee, FL: Foundation for Excellence in Education. (69 Powerpoint slides)
- Glazerman, S. M., and Potamites, L. 2011. *False Performance Gains: A Critique of Successive Cohort Indicators*. Washington, D.C.: Mathematica Policy Research.
- Glazerman, S., & Seifullah, A. 2012. *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) After Four Years*. Princeton, NJ: Mathematica Policy Research, Inc.
- Goldhaber, D., and Hannaway, J. 2004. Accountability with a Kicker: Preliminary Observations on the Florida A+ Accountability Plan. *Phi Delta Kappan* 85(8), 598-605.
- Greene, J. P. 2001. *An Evaluation of the Florida A-Plus Accountability and School Choice Program*. New York: Manhattan Institute for Policy Research.

- Greene, J. P., and Winters, M. A. 2007. Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy. *Education Finance and Policy* 2(4), 319-340.
- Greene, J. P., and Winters, M. A. 2008. The Effect of Special Vouchers on Public School Achievement: Evidence from Florida's McKay Scholarship Program. Manhattan Institute Civic Report No. 52. New York: Manhattan Institute for Policy Research.
- Greene, J. P., and Winters, M. A. 2009. The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review* 28(1), 135-142.
- Hanushek, E. A., Peterson, Paul E., and Woessmann, L. 2012. Is the U.S. Catching Up? International and State Trends in Student Achievement. *Education Next* 12(4), 24-41.
- Jacob, B. A., and Lefgren, L. 2004. Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics* 86(1), 226-244.
- Jimerson, S. R. 2001. Meta-analysis of grade retention research: Implications for practice in the 21<sup>st</sup> century. *School Psychology Review* 30(3), 420-437.
- Kane, T. J., Rockoff, J. E., and Staiger, D. O. 2008. What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review* 27(6), 615-631.
- Kane, T. J., and Staiger, D. O. 2002. Volatility in School Test Scores: Implications for Test-Based Accountability Systems. In Ravitch, D. (ed.), *Brookings Papers on Education Policy* 2002. Washington, D.C.: Brookings Institution.
- Mathis, W. 2011. Review of "Florida Formula for Student Achievement: Lessons for the Nation." Boulder, CO: National Education Policy Center.
- Peterson, P., Howell, W., Wolf, P., Campbell, D. 2003. School vouchers: results from randomized experiments. In Hoxby, C. (Ed.), *The Economics of School Choice*. Chicago: University of Chicago Press.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. N. 2013. Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy* 5(2), 251-281.
- Rouse, C. 1998. Private school vouchers and student achievement: an evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* 113(2), 553-602.
- Sander, W. 1999. Private Schools and Public School Achievement. *The Journal of Human Resources* 34(4), 697-709.
- Sass, T. R. 2006. "Charter Schools and Student Achievement in Florida," *Journal of Education Finance and Policy* 1(1), 91-122.
- Sass, T. R. 2011. Certification Requirements and Teacher Quality: A Comparison of Alternative Routes to Teaching. CALDER Working Paper 64. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research.
- Schwerdt, G., and West, M. R. 2013. The Effects of Test-Based Retention on Student Outcomes Over Time: Regression Discontinuity Evidence from Florida. CESifo Working Paper Series N. 42013. Munich: The Center for Economic Studies.
- Sloan McCombs, J.; Nataraj Kirby, S.; and Mariano, L. T. 2009. Ending Social Promotion Without Leaving Students Behind: The Case of New York City. Santa Monica, CA: RAND Corporation.
- Springer, M.G., Ballou, D., Hamilton, L., Le,V., Lockwood, J.R., McCaffrey, D., Pepper, M., and Stecher, B. 2010. Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.
- West, M. R., and Peterson, P. E. 2006. The Efficacy of Choice Threats Within School Accountability Systems: Results from Legislatively Induced Experiments. *The Economic Journal* 116(510), C46-C62.
- Winters, M. A. 2012. The Benefits of Florida's Test-Based Promotion Policy. Manhattan Institute Civic Report No. 68. New York: Manhattan Institute for Policy Research.
- Winters, M. A., & Cowen, J. M. 2012. Grading New York: Accountability and Student Proficiency in America's Largest School District. *Education Evaluation and Policy Analysis* 34(3), 313-327.
- Zimmer, R., B. Gill, K. Booker, S. Lavertu, T. R. Sass, and J. Witte. 2009. Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition. Santa Monica, CA: RAND Corporation.

The Albert Shanker Institute is a nonprofit, nonpartisan organization established in 1997 by the American Federation of Teachers to honor the life and legacy of its late president Albert Shanker. The Institute's mission is to promote democracy by supporting two fundamental principles: the right of every child to quality public education and the right of every worker to a voice on the job.

This document does not necessarily reflect the views of members of the Institute's Board of Directors

#### **BOARD OF DIRECTORS**

Danielle Allen  
Paul Almeida  
Anthony Bryk  
Barbara Byrd-Bennett  
Han Dongfang  
Carl Gershman  
Sara Goldrick-Rab  
Ernest Green  
Sarita Gupta  
Linda Darling Hammond  
E. D. Hirsch, Jr.  
Sol Hurwitz  
John Jackson  
Clifford Janey  
Lorretta Johnson  
Susan Moore Johnson  
Richard Kahlenberg  
Ted Kirsch  
Stanley S. Litow  
Michael Maccoby  
Herb Magidson  
Harold Meyerson  
Daniel Montgomery  
Susan Neuman  
Pedro Noguera  
Mary Cathryn Ricker  
Richard Riley  
William Schmidt  
Randi Weingarten  
Deborah Wince-Smith

#### **EXECUTIVE DIRECTOR**

Leo E. Casey

