

Competition passes the test

Still more evidence from Florida that public schools improve when threatened with the loss of students and money

by JAY P. GREENE & MARCUS A. WINTERS

Do public schools respond to competition from private schools by improving the quality of instruction? This is one of the key questions in the voucher debate. Advocates of vouchers believe that public schools facing the threat of losing students and funding to private schools will take the measures necessary to raise student performance. Opponents worry that vouchers will actually leave public schools worse off by draining them of funds and encouraging the best students and the most involved parents to flee a failing school.

Florida's A+ program affords a unique opportunity to test these competing predictions. The A+ program offers all the students in schools that chronically fail the Florida Comprehensive Assessment Test (FCAT) the opportunity to use a voucher to transfer to a private school. Schools face the threat of vouchers only if they are failing. They can remove the threat by improving their test scores. Comparing the performance of schools that were threatened with vouchers and the performance of those that faced no such threat gives a measure of how public schools respond to competition.

The A+ Program

All public school students in Florida enrolled in grades 3 through 10 take FCAT exams in math, reading, and writing. Test results have consequences for both students and schools.

Students must pass the reading portion of the FCAT in order to be promoted to 4th grade, and they must pass the 10th-grade test to graduate. In addition, all Florida schools are graded from A to F based on the share of their student bodies that scores at high levels on the FCAT and experiences gains in their test scores from year to year.

A school's grade is lowered a level if less than half of its worst students (those in the bottom 25 percent at the school) make a year's worth of learning gains. In order to receive a grade, schools must test at least 90 percent of their students; otherwise, they receive an Incomplete and, after an investigation, the state commissioner of education assigns a grade to the school.

Schools that receive a grade of F twice during any four-year period are deemed chronically failing. Their students then become eligible to receive vouchers, called opportunity scholarships, which they can use at another public school or at a private school. The vouchers are worth the lesser of per-pupil spending in the public schools or the cost of attending the chosen private school.

Schools can take themselves off the chronically failing list by earning higher grades in future years. However, students who use vouchers to attend private schools can keep their vouchers until either they return to a public school or the grade levels offered by the private school run out. For example, if a student uses a voucher to attend 6th grade at a K-8 private school and the failing public school manages to turn things around the next year, the student may keep his voucher until he completes the 8th grade. Thereafter, if his family wants to



Florida's system of school grades and sanctions gives schools differing incentives. We thus separated schools into five categories based on their grades and the degree of actual or potential competition they faced from vouchers. We then compared the performance of the schools in these categories with the performance of the rest of Florida's public schools, looking at each category's change in FCAT and Stanford-9 scores from the 2001–02 school year to 2002–03. The five categories are:

♦ *Schools Eligible for Vouchers.* These schools have received at least two Fs since grades were first given in 1998–99 and have been deemed chronically failing by the state. Students at these schools have already been offered vouchers to attend private schools. Thus

keep him in private school it must do so at its own expense.

Entering the 2002–03 administration of the FCAT, the focus of this study, 129 schools had received at least one F. Students in ten schools had become eligible for vouchers since the grading of schools began during the 1998–99 school year.

Florida does offer failing schools special funding that may temper any financial loss they suffer from students' choosing to transfer into private schools. The lowest-performing schools are given priority when applying for certain grants, and the state has earmarked funds to recruit teachers to work in schools that received D and F grades. However, since such funds are temporary solutions, they do not dramatically reduce the financial incentive for failing schools to remove themselves from voucher competition by improving their performance on the FCAT.

Five Categories of Schools

To analyze the program's impact on public schools, we collected school-level test scores on the 2001–02 and 2002–03 administrations of the FCAT and the Stanford-9, a national norm-referenced test that is given to all Florida public school students around the same time as the FCAT. The results from the Stanford-9 are particularly useful for our analysis. Schools are not held accountable for their students' performance on the Stanford-9. As a result, they have little incentive to manipulate the results by "teaching to the test" or through outright cheating. Thus, if gains are witnessed on both the FCAT and the Stanford-9, we can be reasonably confident that the gains reflect genuine improvements in student learning.

voucher-eligible schools are currently competing against private schools in the market for students. This is the group with the greatest incentive to improve and also the greatest likelihood of being harmed by vouchers if vouchers are in fact harmful.

Our study includes nine voucher-eligible schools. During the 2001–02 administration of the Stanford-9 (which was administered at about the same time as the FCAT test used to assign these schools' last grade), the students in these schools scored in the 25th percentile nationwide in reading and the 32nd percentile nationwide in math. The schools serve largely poor and minority student populations; 88 percent of their students are enrolled in the federal lunch program, 18 percent speak limited English, and only 1 percent are white.

♦ *Schools Facing the Threat of Vouchers.* These schools received one F during the three school years before the 2002–03 administration of the FCAT; one more F during the 2002–03 administration and their students would have been offered vouchers. They therefore had an incentive to improve in order to ward off the voucher threat. Our study includes 50 voucher-threatened schools, whose test scores and demographics closely resembled those of the voucher-eligible schools.

♦ *Always "D" Schools.* These schools have never received any grade other than D. Thus always-D schools are not voucher threatened, but they face the prospect of becoming so. Here it is important to note again that a school's grades are based not on its overall average scale score but rather on the percentage

of students meeting levels of proficiency and the percentage of students making adequate gains on the tests. As a result, many always-D schools have similar or even lower test scores than F schools but have still managed to avoid receiving a failing grade.

The relatively low initial test scores and disadvantaged student populations of the 63 always-D schools in our analysis make them an attractive group to compare with the voucher-eligible and voucher-threatened schools. Since the three groups of schools are similar in their observable characteristics, such as the student body's ethnic makeup, and most likely in other characteristics as well, the only major difference between the always-D schools and the other two groups is the competition they face from vouchers. Comparing these three groups thus provides one way of isolating the influence of voucher competition.

♦ *Sometimes "D" Schools.* These schools have received at least one grade higher than D and have never received an F. The 507 sometimes-D schools do not face the imminent prospect of having to compete for students. Like always-D schools, many sometimes-D schools have test scores similar to or even lower than F schools though they have never received a failing grade. Because they face no competition from vouchers and have lower chances of receiving an F grade, sometimes-D schools are expected to make less improvement than the voucher-eligible, voucher-threatened, and always-D schools.

♦ *Formerly Threatened Schools.* These schools earned an F in the first year of grading, 1998–99, but have not received an F since. Thus they once faced the prospect of vouchers but no longer do because they have survived the four-year time period without receiving another F. Analyzing this group clarifies whether schools continue to improve relative to the rest of the public schools in Florida once the threat of vouchers disappears. It also tests whether the stigma of receiving an F, rather than the threat of vouchers, is what motivates schools to improve.

We compared the change in test-score performance for each of these groups relative to the rest of Florida public schools between the 2001–02 and 2002–03 administrations of the FCAT and the Stanford-9. Our method was to follow cohorts of students in grades 3 through 10 and calculate the schoolwide change in test scores. For example, we subtracted a school's 3rd-grade reading score on the 2001–02 FCAT from its 4th-grade reading score on the 2002–03 FCAT to get the change in scores for 4th graders at that school. Following cohorts measures the performance of roughly the same students on the test over time. We then averaged the change in test scores for each cohort in the school on each test and subject. This gave us a single cohort change for each school in Florida.

The changes in performance reported below for each group of schools have all been adjusted to take into account

any changes between 2001–02 and 2002–03 in schools' demographic characteristics, such as the share of students participating in the federal school lunch program and the ethnic breakdown of the student body. Unfortunately, we were not able to control for changes in the number of students who spoke limited English or in the school's operating cost per pupil, because at the time of the study such information was available only up to the 2001–02 school year. We instead controlled only for the percentage of students who spoke limited English and the level of spending per pupil in 2001–02.

The inability to control for changes in spending may seem particularly troublesome. However, a similar analysis of the A+ program in a previous year found that taking into account changes in spending had no effect on the results. Furthermore, if any relative improvements made by schools competing with vouchers were the result of school districts' diverting funds to these schools, this could be seen as part of the voucher effect.

Results

Between the 2001–02 and 2002–03 administrations of the FCAT, voucher-eligible schools made the largest gain among the five categories of schools. In mathematics they improved by 15.1 scale-score points more than the rest of Florida's public schools (see Figure 1). (Results on the FCAT are reported as the cohort change in mean scale score on a scale from 100 to 500. The median school in Florida had a mean scale score of 291 on the reading test and 300 on the math test. Schools at the 5th percentile of schools in Florida had a reading scale score of 243 and a math scale score of 247, while the 95th percentile school had a reading score of 327 and a math score of 328.) On the Stanford-9 math test, voucher-eligible schools achieved gains that were 5.9 percentile points greater than the year-to-year gains achieved by other Florida public schools (see Figure 2). Results on the Stanford-9 are reported as the cohort change in national percentile rank.

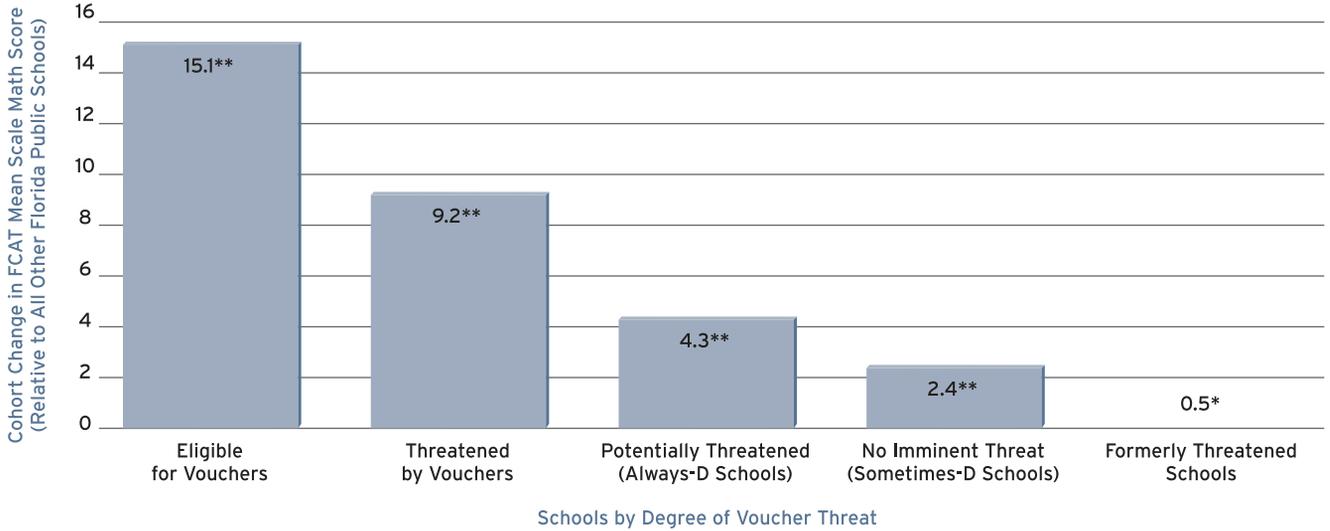
Voucher-threatened schools made the next highest relative gains: 9.2 scale-score points on the math FCAT and 3.5 percentile points on the Stanford-9 in math. Each of these results is statistically significant at a very high level, meaning that we can be highly confident that the test-score gains made by schools facing the actuality or prospect of voucher competition were larger than the gains made by other public schools. As hypothesized, actual voucher competition produced the largest improvements in test scores, while the prospect of facing voucher competition produced somewhat smaller gains.

The results for the always-D and sometimes-D schools were also consistent with our hypotheses. Always-D schools, which, faced with the real danger of receiving their first F, had some incentive to improve, made a relative gain of 4.3 scale-score points on the math FCAT and 1.3 percentile points on the Stanford-9 math test. The sometimes-D

Students Win When Schools Compete (Figure 1)

Gains in test scores were 15 points higher among those schools whose students were eligible for vouchers than the gains among the rest of Florida’s public schools. Schools whose students were on the verge of becoming eligible also made greater gains.

Annual Gain in Test Scores by Degree of Voucher Threat



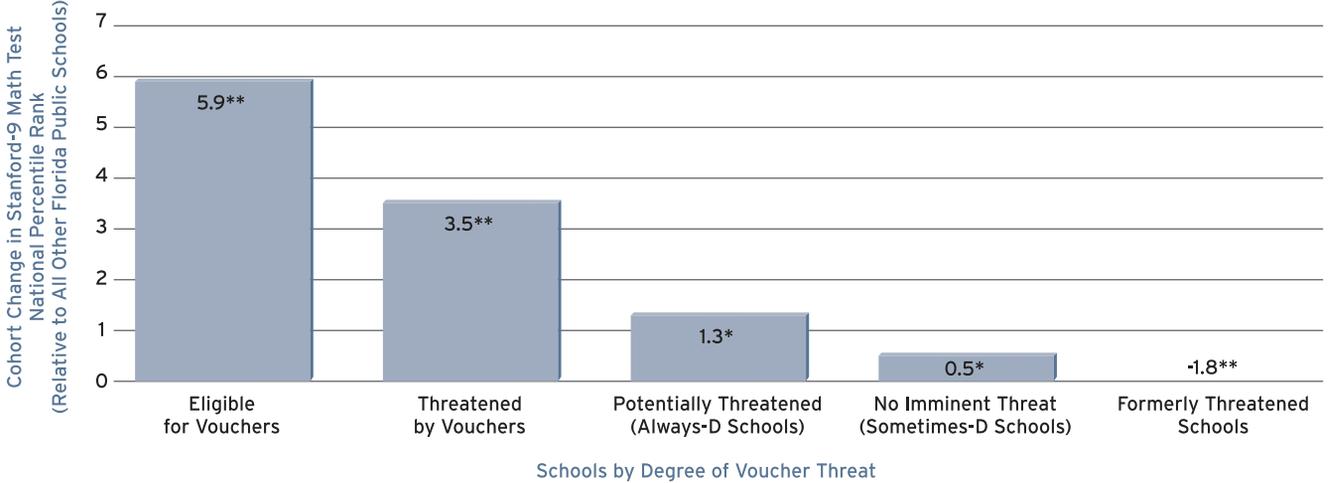
* Result statistically insignificant.
 ** Result significant at the .01 level.

SOURCE: Authors

Genuine Learning (Figure 2)

The same pattern—of greater gains among schools facing competition or the threat thereof—was witnessed on the national Stanford-9 exam, confirming that the gains reflect genuine improvements in learning rather than “teaching to the test” or cheating. The gains among schools whose students were eligible for vouchers were enough to erase almost one-fifth of the gap between their average score in the 2001–02 school year and the average score of all other Florida public schools.

Annual Gain in Test Scores by Degree of Voucher Threat



* Result significant at the .05 level
 ** Result significant at the .01 level

Note: Gains between 2001-02 and 2002-03 school years. Results account for the level of school spending, the proportion of students speaking limited English, and for changes in students’ other demographic characteristics.

SOURCE: Authors

Closing the Gap

Further confirmation that the threat of vouchers caused public schools to improve their performance on Florida's accountability test by **RAJASHRI CHAKRABARTI**

Editor's note: In an independent analysis employing different techniques that allowed her to track trends over time, Rajashri Chakrabarti examined the performance of schools facing the threat of vouchers during the three years after Florida introduced its A+ program. Her findings, reported below, mirror those of Jay Greene and Marcus Winters, who analyzed the effect of the voucher threat for the 2002–03 school year.

Florida's A+ program provides a unique opportunity to evaluate the effect of vouchers on public school performance. Public schools that received an F grade during the 1998–99 school year were directly exposed to the threat of vouchers if they did not improve their test scores. Schools that received a D grade that same year faced no such direct threat. To analyze the effect of the voucher threat, I compared changes in the performance of F schools with the change among D schools from the 1998–99 school year through the 2001–02 school year.

Schools that were originally given a grade of F in 1999 made greater performance gains than the D schools on each of the Florida Comprehensive Assessment Tests (in math, reading, and writing) and in each of the three school years. (See accompanying figure for the results in math.) I also found that F schools made greater gains than a larger group of

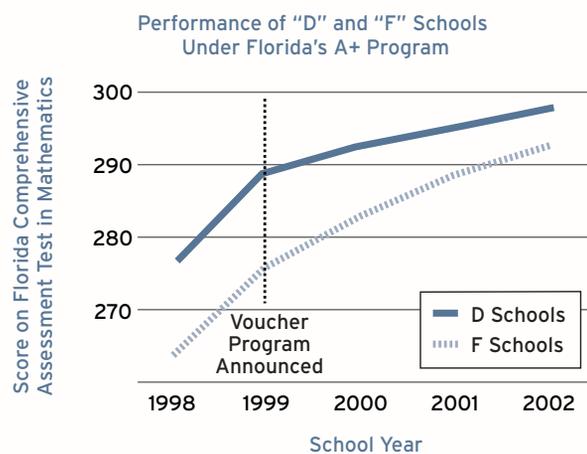
schools that were given a grade of C in 1998–99. D schools also made greater gains than C schools, perhaps reflecting the fact that D schools were not far from earning a grade of F and thereby facing the threat of vouchers.

These improvements did not reflect changes in the observable characteristics of the schools' student bodies or in the schools' levels of spending. Nor can the differences in gains be attributed to performance trends in the different groups of schools before the establishment of the program. Nor do the differences reflect the fact that F schools, being low performers, had more room for improvement.

Could these improvements simply reflect the stigma of being identified publicly as a low-performing school? Tellingly, I did not observe similar improvements among low-performing schools under the state's old accountability system, which rated schools based on their performance but did not impose the threat of vouchers. Beginning in 1997, Florida schools were assigned a

rating of 1 to 4 on the basis of their performance. Schools placed in group 1 (the lowest-performing set) did not improve relative to schools in group 2 or group 3. In short, there is strong evidence that F schools in Florida responded to the threat of vouchers.

—Rajashri Chakrabarti is a doctoral candidate in economics at Cornell University. These results are from her paper "Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs."



Note: Scores represent averages for all schools in each group after adjusting for changes in student characteristics and per-pupil spending and weighted by the number of test-takers.

SOURCE: Author

schools experienced year-to-year changes in FCAT math scores that were only 2.4 points higher than all other Florida public schools, significantly less than the gains in both voucher-eligible and voucher-threatened schools. Their improvement relative to all public schools on the Stanford-9 was less than a percentile point. Formerly threatened schools saw no improvement in their math scores relative to all public schools.

The patterns were similar in reading, though the relative gains made by schools facing voucher competition were smaller and sometimes statistically insignificant. Overall on the FCAT reading test, voucher-eligible schools gained 5.2 points more than other schools gained. However, this gain fell barely short of a conventional standard for statistical significance, likely due to the very small number of schools in this category (only nine). Voucher-eligible schools also made a statistically insignificant relative gain of 2.2 percentile points on the Stanford-9.

Voucher-threatened schools actually made the greatest

gains on the FCAT reading test: 6.1 points. Their relative gain on the Stanford-9 was a statistically significant 1.7 percentile points.

Always-D schools made no statistically significant gains on the FCAT or Stanford-9 reading tests, while sometimes-D schools experienced a decrease of 1.1 points on the FCAT and no significant change on the Stanford-9 reading test. We also found a relative loss of 3.8 points for formerly threatened schools on the FCAT and a relative loss of 1.6 percentile points on the Stanford-9 (both results were statistically significant).

Overall, the schools facing either the prospect or the reality of vouchers made substantial gains compared with the results achieved by the rest of Florida's public schools. They also made strong gains relative to those earned by schools serving similar student populations, which had nonetheless avoided receiving an F.

The smaller gains achieved by always-D and sometimes-

D schools compared with the performance of voucher-eligible and voucher-threatened schools, despite the similar characteristics of all these schools, strengthen our confidence that voucher competition is the cause of the improvements. Always-D schools, in particular, are very similar to voucher-eligible and voucher-threatened schools in their initial test scores, student populations, and resources, as well as other unobserved factors for which we could not adjust the data. Since it is essentially by chance that always-D schools do not receive an F, the comparison approximates a randomized experiment. Yet the schools that faced voucher competition experienced much larger increases in test scores.

Moreover, the similarity of our findings on the Stanford-9 and FCAT math tests suggests that the gains being made by schools facing voucher competition are the result of real learning and not simply manipulations of the state's high-stakes testing system (see Figure 2). If schools facing voucher competition were only appearing to improve by somehow manipulating Florida's high-stakes testing system, we would not have seen a corresponding improvement on another test that no one had incentives to manipulate.

Other Possible Explanations

Could the gains witnessed among voucher-eligible and voucher-threatened schools actually be the product of some influence other than their being forced to compete against private schools?

Let's first consider the possibility that it was the stigma of being labeled a failure, rather than the competitive incentives introduced by vouchers, that spurred improvement among F schools, as several researchers have suggested. If this were the case, we would expect to see similar gains among formerly threatened schools, which have also received at least one failing grade. Quite the contrary, however: formerly threatened schools made no gains in math and experienced losses in reading. In other words, formerly threatened schools still had the stigma of an F grade, but once the threat of vouchers was removed, they actually lost ground (see Figure 2).

Nonetheless, it is possible that the stigma of the F grade fades over time. In that event, schools that received an F in 1999 might no longer feel the stigma in 2003. But although the voucher-eligible and voucher-threatened school categories include some schools that received their most recent F in 2000, those categories experienced gains. We find it implausible that the stigma effect exists for only three years and then suddenly disappears. The more compelling explanation is that the actuality or prospect of voucher competition provides incentives for schools to improve, an effect that disappears when the four-year window expires.

Another potential explanation for the exceptional gains made by schools facing voucher competition is that their extremely low initial scores are affected by a statistical tendency

called "regression to the mean." Schools that report very high and very low scores may report future scores that come closer to the average for the whole population. This tendency is created by nonrandom error in the test scores, which can be especially troublesome when scores are "bumping" against the top or the bottom of the test-score scale. For instance, if a school earns a score of 2 on a scale from 0 to 100, it is hard for students to do worse by chance but easier for them to do better by chance. Schools that are near the bottom of the scale are likely to improve, even if only by statistical fluke.

To test for this possibility, we compared the gains made by F schools with the performance of an even smaller subset of schools whose 2002 test scores were similar but had never received an F (which we termed low-performing non-F schools). If there were no difference between the gains experienced by F schools and those among low-performing non-F schools, we might worry that regression to the mean was driving our results.

In mathematics, the gains made by voucher-eligible and voucher-threatened schools relative to low-performing non-F schools on both the FCAT and the Stanford-9 were nearly as large as their gains relative to all other schools in the state. Thus in math there seems to be no effect from regression to the mean. In reading, however, we found no difference in the test-score gains achieved by F schools and low-performing non-F schools, suggesting that regression to the mean could be influencing our results in reading.

Even so, it seems unlikely that regression to the mean is the entire story, even in reading. The very fact that fewer schools were included in this section of the analysis made it less likely that significant differences would emerge. Moreover, the low-performing non-F schools actually had average test scores that were *lower* than those among the F schools. These schools also clearly faced pressure to improve in order to avoid the voucher threat, even if that threat was less immediate. Many of the schools in the low-performing category were also in either the always-D or sometimes-D categories, which were shown above to have made gains relative to all Florida public schools—probably due to the likelihood that they would receive an F if they did not improve.

Having largely ruled out these other explanations, we are left with the conclusion that the gains witnessed among low-performing schools are the result of the competitive pressures introduced by school vouchers. Moreover, the similarity of our findings on both the high-stakes FCAT and the low-stakes Stanford-9 indicates that the gains reflect genuine improvements in learning. In the absence of student-level information, results must remain tentative. Nonetheless, this study yields solid evidence that public schools will react positively to being forced to compete with private schools for students and the dollars they carry.

—Jay P. Greene is a senior fellow and Marcus A. Winters a research associate at the Manhattan Institute.