

Evaluation of the Florida Tax Credit Scholarship Program
Participation, Compliance and Test Scores in 2010-11

David N. Figlio
University of Florida
Northwestern University
and
National Bureau of Economic Research

August 2012

Executive summary

This is the fifth in a series of reports evaluating the Florida Tax Credit Scholarship (FTC) Program, as required by the Florida Statutes, s. 1002.395(9)(j). This report provides information on private school compliance with program rules regarding required testing, describes the attributes of eligible students who participate in the program, and presents data on student test score levels and gains in the program, as well as compared with the eligible population of non-participating students.

During the 2010-11 academic year, David Figlio, the Project Director, collected test score data from private schools participating in the FTC Program in real time. This is the fifth year for which program participants' test score data were collected, and the fourth year in which this data collection occurred in real time.

Compliance with program testing requirements, 2010-11:

- Compliance with program testing requirements in 2010-11 was at its highest level to date, and private school reporting errors continue to decline. Private schools provided usable test scores for 93.5 percent of program participants in grades 3-10. Another 3.9 percent of participants were ineligible for testing or were not enrolled in the school at the time of testing; this is largely driven by the fact that some students arrived in schools after fall testing (for schools that test in the fall, principally those that administer the Iowa Test of Basic Skills) and some students who began the year in a school left the school prior to the more typical spring testing. The 0.8 percent rate of reported illness/absence is the lowest it has been since the beginning of data collection. Test administration compliance errors by participating schools are at the same level as 2009-10, and well below earlier years, with reporting problems involving only 0.3 percent of participants in 2010-11.
- The vast majority (69.2 percent) of test-takers took the Stanford Achievement Test. Other popular tests were the Iowa Test of Basic Skills (20.5 percent) and the TerraNova (3.4 percent).
- Scholarship students whose test scores were received are modestly more advantaged than are those scholarship students whose scores were not received. It is not known whether the gains of those without score reports would have been higher or lower than those with score reports.

Differential program participation rates for different groups of students and families:

- Program participants tend to come from less advantaged families than other students receiving free or reduced-price lunches.
- As in most prior years, program participants tend to come from lower-performing public schools prior to entering the program. Likewise, as in prior years, they tend to be among the lowest-performing students in their prior school, regardless of the performance

level of their public school. The tendency for the weakest prior performers on standardized tests to choose to participate in the FTC Program is becoming stronger over time.

Test scores of program participants, 2010-11:

- The typical student in the program scored at the 45th national percentile in reading and the 46th percentile in mathematics, about the same as in 2008-09 and 2009-10. The distribution of test scores is similar whether one considers the entire program population or only those who took the Stanford Achievement Test in the spring of 2010. The Stanford Achievement Test is the most commonly administered test and is the test most directly comparable to the FCAT.
- The mean reading gain for program participants is exactly 0 national percentile ranking points in reading and -0.9 national percentile ranking points in mathematics. These mean gains are indistinguishable from zero. In other words, the typical student participating in the program gained a year's worth of learning in a year's worth of time. It is important to note that these national comparisons pertain to all students nationally, and not just low-income students.
- Test score gains for program participants are virtually identical to those of income-eligible non-participants remaining in Florida public schools. Participating students gained slightly relative to comparable public school students in 2010-11, though this difference is not statistically significant. It is important to recall that the participating students differ from the income-eligible public school students in important ways – their incomes are substantially lower and their previous test performance in public school tended to be substantially lower. These differences make direct comparison of gain scores more problematic. Because families can choose whether to participate in the program, it is inappropriate to consider the differences in test score gains between FTC Program participants and their public school counterparts to be *caused* by program participation. It is, therefore, best to consider the fact that test score gains are extremely similar between the public and private sector to be suggestive evidence of little difference in average performance across the sectors, rather than causal evidence of differential performance. That said, in past cohorts for whom there existed sufficient data to estimate the causal consequences of program participation, there was evidence of positive effects of participation in the FTC program, especially for math. Little has changed in terms of test scores or factors influencing program participation across cohorts, indicating that one might infer, albeit with caution, that positive effects found in prior cohorts continued to the most recent application cohort.
- Recent statistical research has shown that the FTC Program has improved the performance of Florida public schools to a modest degree. Therefore, the correct interpretation of the findings in this report are that students participating in the program have kept pace with the improvements in the public schools associated with the FTC Program.

- This report marks the first time that individual schools' test score gains have been reported for schools with 30 or more students with gain scores. Seventy schools met this criterion in 2010-11.

I. Background

This is the fifth in a series of reports evaluating the Florida Tax Credit Scholarship Program, as required by the Florida Statutes, s. 1002.395(9)(j). This report provides information on private school compliance with program rules regarding required testing, describes the attributes of eligible students who participate in the program, and presents data on student test score levels and gains in the program, as well as compared with the eligible population of non-participating students.

The Florida Department of Education first awarded a contract to the University of Florida as the Independent Research Group and Professor David Figlio as the Project Director in October 2007 to collect program participants' test scores directly from the private schools. Therefore, the first year in which test score data collection could take place in real time was the 2007-08 academic year; data from the 2006-07 academic year, the first year in which testing was required, could only be collected retrospectively from private schools. It was unclear at the time the degree to which the 2006-07 academic year would make an acceptable baseline for evaluation, but it was decided that to accelerate the possibility of providing concrete information regarding testing and compliance amongst participating schools an attempt would be made to retrospectively collect as complete information from 2006-07 test scores as possible. The results of that effort were presented in the program report dated March 2008. Later reports, released in June 2009, June 2010, and August 2011, presented data from the 2007-08, 2008-09, and 2009-10 academic years, with the 2010 report being the first to present gain scores for program participants where all test scores were collected in real time.

This report presents the results of the real-time test score collection in 2010-11. This report details key information about program participation and test scores, and compares test score gains for program students to comparable students in Florida public schools. By Florida Statute, this report also presents information on test score gains disaggregated to the individual private school level for all schools with a sufficiently large number -- defined statutorily as 30 or more -- gain scores collected.

II. Test score collection in 2010-11

Data collection protocol

As required by s. 1002.395(8)(c)(2), participating schools administered to students an approved nationally norm-referenced test as identified by the Florida Department of Education, including the Stanford Achievement Test, Basic Achievement Skills Inventory, Metropolitan Achievement Test, Iowa Test of Basic Skills, Terra Nova, or the Preliminary Scholastic Aptitude Test and ACT/PLAN (for students in high school grades) or made provisions for participating students to take the FCAT at a public school in accordance with s. 1002.395(7)(e). This testing was first required in the 2006-07 academic year, and the Independent Research Organization attempted to collect retroactively as many of these test scores as possible.

The 2010-11 academic year was the fourth year in which it was possible to collect participant test score data in real time. Pursuant to s. 1002.395(8)(c)(2), in Fall 2010 and again throughout Winter and Spring 2011 the Independent Research Organization contacted the 970 private schools that had participating students in grades three through

ten during the 2010-11 school year, as reported on the October roster of program participants. The Florida Department of Education and Step Up for Students provided the Project Director with a list of all participating students in 2010-11, as of the October participant roster, and refreshed and cross-checked against the January participant roster; of these, 17,724 were in the relevant grades, according to the state records. Schools were provided lists of the relevant students and were instructed to submit test scores to the Independent Research Organization. Schools were also informed that they must provide explanations for any missing or invalid student test scores.

Private school compliance

In over 99 percent of cases, schools submitted photocopies of official score sheets provided to them by the relevant testing company (e.g., Pearson Assessments). In a small number of schools, the schools scored the tests themselves and forwarded to the Project Director detailed information regarding the nature of test administration and scoring. The Independent Research Organization followed up with schools that had provided partial or incomplete data, or that did not provide data regarding students who had attended school in the relevant grades but for whom no valid test score was received. Upon receipt of the test scores, the Project Director and his staff double-entered, audited and reconciled the scores, and once the scores were confirmed, the original score sheets were destroyed and the resulting electronic databases stored in accordance with s. 1002.22(3)(d)(5) of the Florida Statutes. These data were then matched with student FCAT, public schooling, subsidized lunch and disability history, when available, from the Education Data Warehouse, and with information from student scholarship applications provided by the

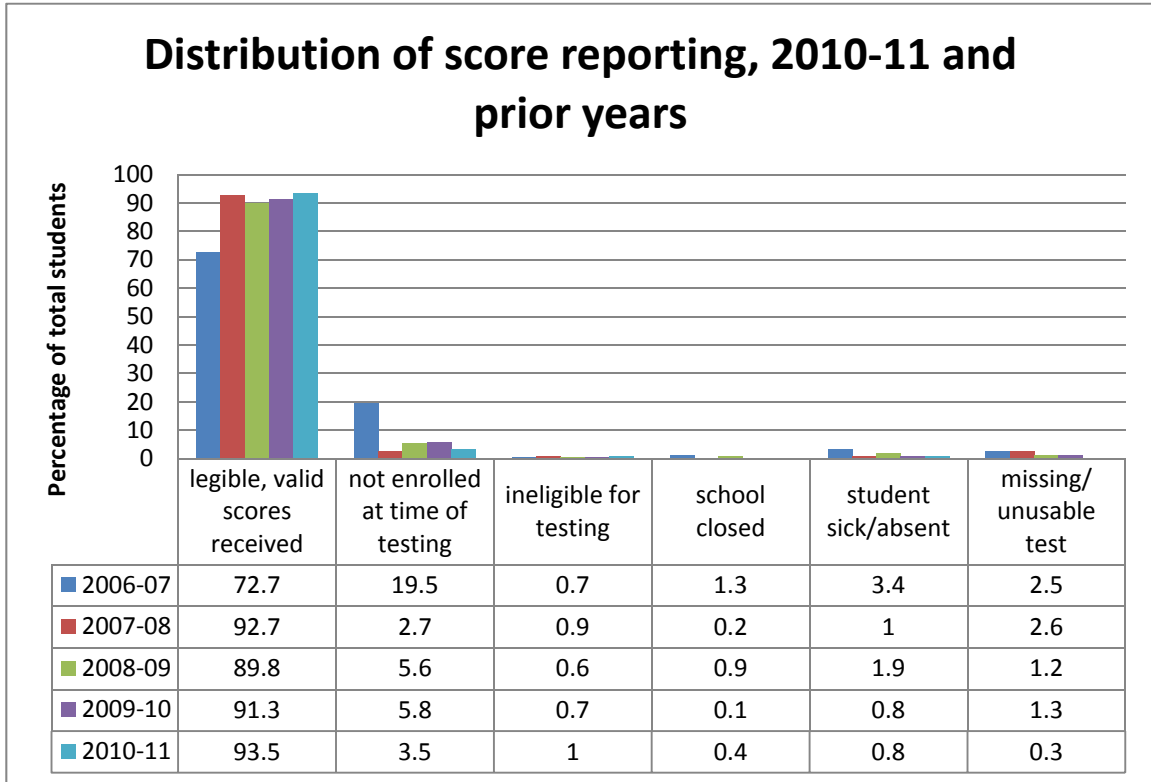
Scholarship Funding Organizations, and then were stripped of individual identifiers such as names, social security numbers or birthdates, for the purposes of analysis.

Of the 970 schools with students in the relevant grades in 2010-11, the overwhelming majority provided evidence of test administration according to the specifications of the program. A small fraction of participating schools closed following the 2010-11 school year and did not provide test scores to the Project Director. In a handful of other cases, the schools administered unapproved tests or neglected to administer tests to participating students; in the case of the small number of non-compliant schools, the Project Director reported the schools to the Florida Department of Education for disciplinary action.

Of the 17,724 students in relevant grades participating in the program in 2010-11, the Independent Research Organization received valid, legible test scores for 16,575 students, or 93.5 percent of all expected students;¹ virtually all of these scores were from tests administered by the private schools themselves. This is the highest rate received to date; we suspect that this is due to a more comprehensive roster reporting system put into place by Step Up for Students as well as ever-improving private school understanding of testing and reporting requirements.

¹ We received 9 additional test scores following the January 21, 2011 date in which we merged score records with school records. This report excludes these 9 test scores, because they cannot be merged with the state records for the purposes of analysis.

Distribution of score reporting, 2010-11 and prior years



Augmenting the 93.5 percent of students for whom we received legible, valid scores in 2010-11 are the 3.5 percent of students who were not enrolled at the time of testing -- this typically happens because over one-fifth of schools administer the Iowa Test of Basic Skills in the early fall and many students begin their time in a school after this test administration took place -- and the 1.0 percent of students on the official roster who were either deemed ineligible for test score reporting pursuant to s. 1002.395(8)(c)(2) or were not enrolled in the school identified on the official rosters. As in prior years, a small number of schools (representing 0.4 percent of potentially eligible students) closed before reporting their scores in 2010-11. Taken together, the percentage of students in 2010-11 with either legible, valid score reporting or one of these other explanations was 98.4 percent, above the 97.9 percent in 2009-10, the 96.9 percent in

2008-09, and the 96.5 percent in 2007-08. Only 1.1 percent of students had either a missing or unusable test or were reported to be sick or absent -- the lowest rate ever in these categories. The category of "missing or unusable tests" includes the school providing test scores that were illegible, not providing scores that could be compared with national norms, testing students using an unapproved test, or failing to test students at all. The percentage of schools falling into these categories continues to fall with each successive round of testing, implying that private school compliance with the testing requirements continues to improve. The small number of remaining expected scores not accounted for in any of these categories are from schools that the Florida Department of Education removed from the program due to non-compliance in testing or other reasons.

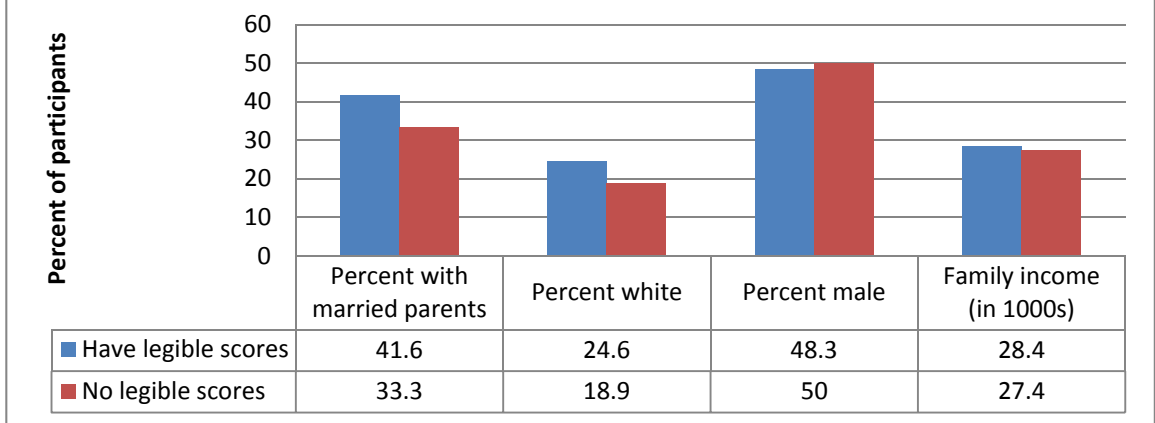
Of the students who have taken tests that were reported to the Independent Research Organization, virtually 100 percent took a test approved by the Florida Department of Education. The vast majority of the students (69.2 percent) took the Stanford Achievement Test, the nationally norm-referenced test administered to all public school students in the relevant grades in Florida through 2007-08, while another 20.5 percent took the Iowa Test of Basic Skills and 3.4 percent took the Terra Nova test. The other students took a number of other tests, most notably the PSAT/NMSQT, taken by 1.8 percent of students, the Basic Achievement Skills Inventory, taken by 1.5 percent of students, the ACT/PLAN, taken by 1.3 percent, and the Metropolitan Achievement Test, taken by 0.4 percent. 1.9 percent took other approved tests. Only two students (fewer than one tenth of one percent) took a test that was not approved by the Florida Department of Education.

Schools have flexibility as to when they administer their exams, and just over 21 percent of participating students took their exam in the fall months. These scores are less likely to be directly comparable to public school students' tests than are those taken during the time immediately surrounding the public schools' test administration. The tests most typically taken in the fall months are the PSAT/NMSQT and the Iowa Test of Basic Skills. The latter case is driven strongly by Florida Catholic schools' uniform assessment of students in October using the Iowa Test of Basic Skills. It is likely to be inappropriate to directly compare status scores of tests administered in March to tests administered in October, as they likely have very different purposes. This speaks to the importance of measuring student learning gains rather than levels comparisons, and also indicates that it would be useful to conduct a fall-spring concordance study if at all possible.

Similarity of students with received legible tests to the overall scholarship population

In 2010-11, the rate of successful test reporting remained at the high levels of previous years. However, as mentioned above, around 6.5 percent of the potentially-tested population of students was not tested (due in large part to students arriving at school after testing or leaving a school before testing, or to students being sick or absent during the testing period), so it is important to gauge whether the students whose test scores were successfully reported are comparable to the overall population of students enrolled in the scholarship program at any time during 2010-11.

Comparison of students with legible scores to other program students, regardless of reason, 2010-11

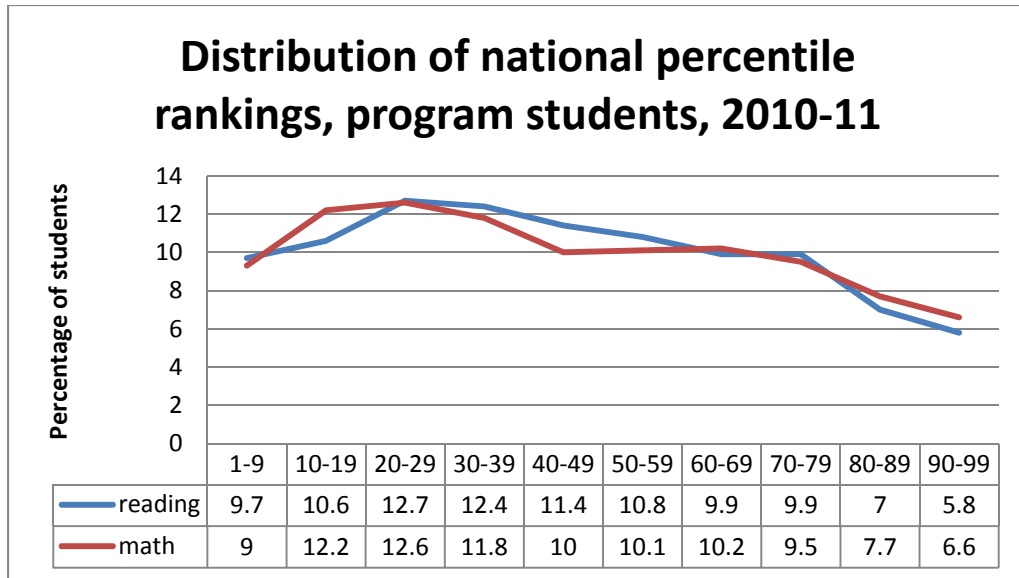


As can be seen from the accompanying figure, there is evidence that students whose test scores were successfully reported are modestly more advantaged than other program participants whose scores were not successfully reported, based on data from the families' scholarship applications. Students whose scores were successfully reported come from families with somewhat higher incomes, with parents considerably more likely to be married, and are more likely to be white, than are students whose scores were not successfully reported, for whatever reason. These differences may have been expected, as highly transient students are likely to be less advantaged, and are more likely to have not been tested because they changed schools. However, even among students who were still in the school at the time of testing, those missing score reports tend to be less advantaged (with family incomes five percent lower), with unmarried parents (33 percent married versus 42 percent married), and nonwhite (15 percent white versus 25 percent white). (Unlike prior years, there was no difference in the percentage male between these groups in 2010-11.) These differences, therefore, underscore the

importance both (1) of obtaining as full a collection of test score data as possible, and (2) of measuring student test score gains. It is not obvious that students with missing test scores would have had higher or lower gain scores than those with test scores available. It is also important to note that while public school records do not include data on family income or parental marital status, we observe that those missing public school test scores are also more likely to be nonwhite and eligible for free or reduced price lunches.

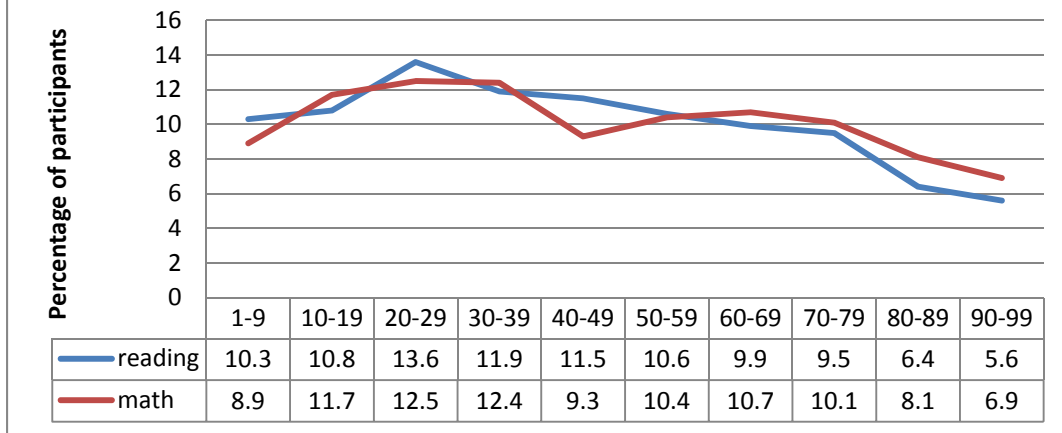
III. Test scores of 2010-11 program participants

Because program participants may take any number of nationally norm-referenced tests and because private schools have some flexibility in the form in which these test scores are reported and the time of year the test is administered, the only way to ensure reasonable comparability across schools and program participants is to report national percentile rankings. National percentile rankings are desirable because they are compared against a nationally-representative group of students; so long as the national norms for one test (such as the Stanford Achievement Test) are comparable to the national norms for another test (such as the Iowa Test of Basic Skills) then there is no inherent bias associated with comparing the national percentile rankings of one student taking a certain test to those of another student taking a different test.

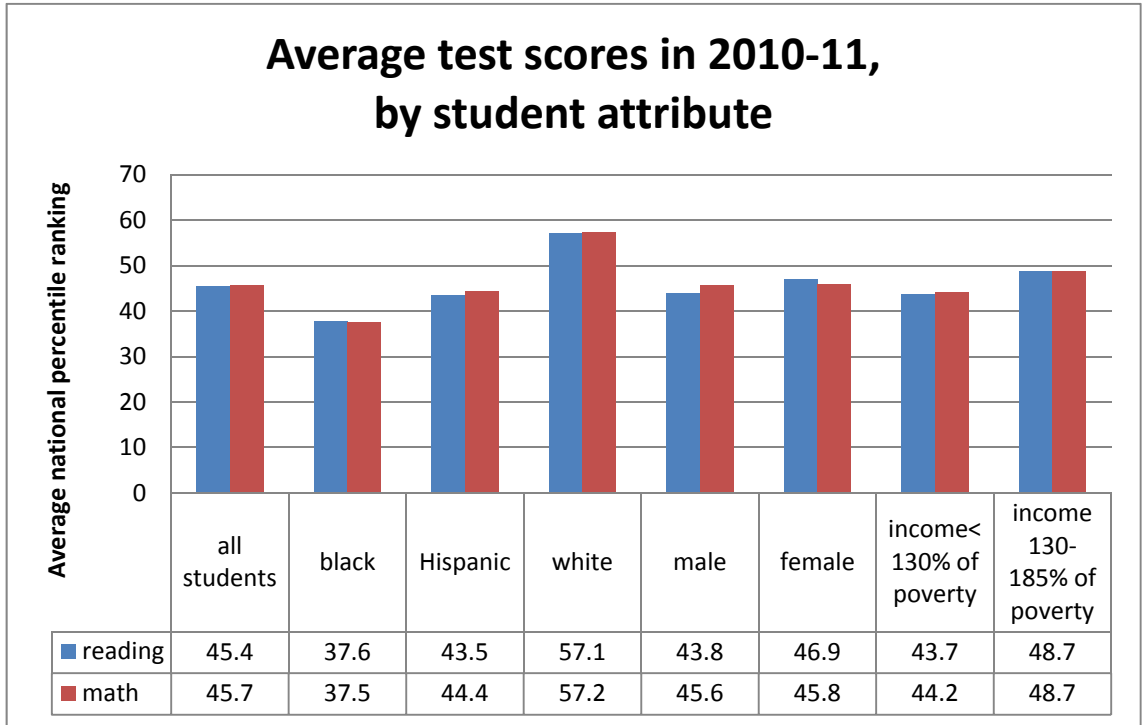


The chart above presents the basic distribution of national percentile rankings among FTC students participating in the program in 2010-11. The typical student in the program scored at the 45th percentile in reading and the 46th percentile in mathematics. This is unchanged from 2007-08, 2008-09, or 2009-10 -- the mean national percentile rankings have varied by less than one percentile point in every year since real-time test score collection began. Were the distributions to be limited to those taking the Stanford Achievement Test in the spring -- the most comparable to the students in the public schools -- the typical student would have scored at the 44th percentile in reading and the 47th percentile in mathematics, the same as in prior years. Given that the distributions of test scores are sufficiently similar for those taking the Stanford Achievement Test in the spring versus the full set of scholarship recipients, and since these differences have been invariant over the years, this report will focus on the full set of students for whom data are available, regardless of test administered.

Distribution of national percentile rankings, participants taking Stanford test in spring 2011



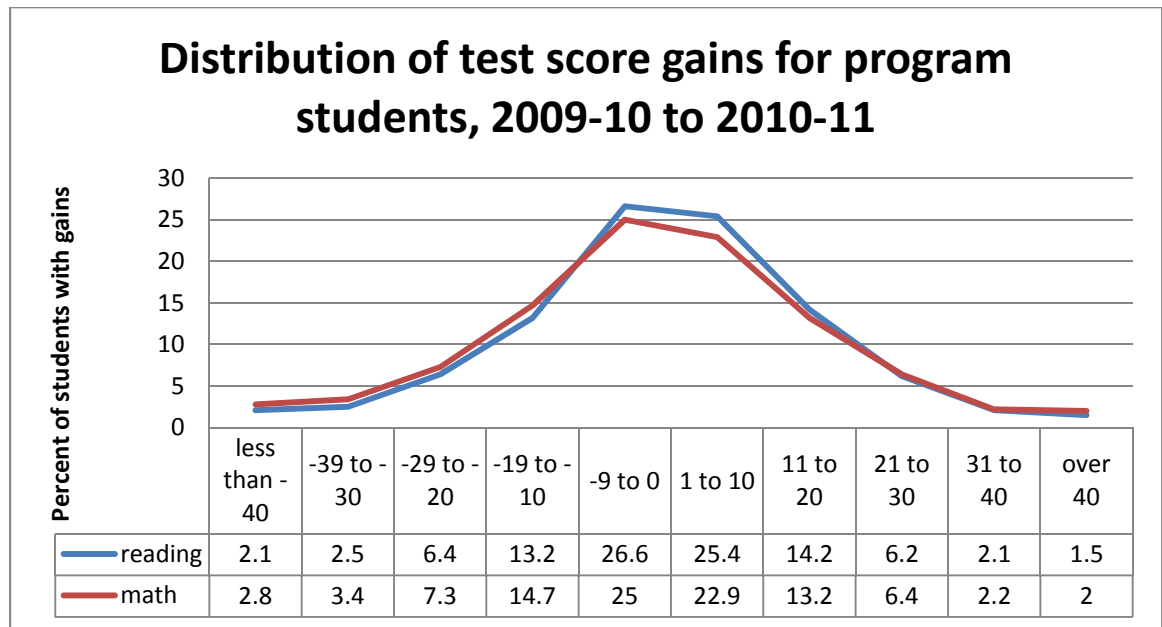
The chart above presents average norm referenced test scores, expressed in terms of national percentile rankings, for various subsets of the FTC Scholarship recipient population, stratified by race, sex, income, and parental marital status. Income is expressed in terms of fraction of the poverty line, to reflect the fact that families of different sizes have different official measures for poverty; those with family incomes below 130 percent of the federal poverty line are eligible for free school meals, while those with incomes between 130 and 185 percent of the poverty line are eligible for reduced-price meals. As can be observed in the next table, white participants tend to score better than do minority participants, females tend to perform better than do males, students with married parents tend to score better than do students with unmarried parents, and relatively high-income families tend to score better than do relatively low-income families. These averages closely mirror the figures presented in previous years' reports.



Test score gains for FTC Scholarship program participants

The relevant statutes call for comparisons of test score gains for FTC Scholarship Program students to similar students in public schools. Because the test scores in both 2009-10 and 2010-11 are measured in terms of national percentile rankings, gain scores can only be interpreted as changes in national percentile rankings, and are, therefore, subject to issues regarding ceiling effects (where students whose scores are already in the high percentiles cannot gain much more) and floor effects (where students whose scores are already in the low percentiles cannot lose much more ground.) Ceiling and floor effect concerns are mitigated for students whose initial national percentile ranking falls in the middle portions of the initial test score distributions, which is the case for the vast

majority of students participating in the FTC Scholarship Program (as well as in the public schools.)



The chart above presents information on the distribution of program participants' test score gains in reading and mathematics for the set of 9,204 students with legible reading scores and 9,241 students with legible mathematics scores in both 2009-10 and 2010-11. The mean gain for program participants is 0.0 national percentile ranking points in reading and -0.9 national percentile ranking points in mathematics, numbers that are numerically slightly better (in math) but statistically indistinguishable from past years' average gains scores. In other words, the typical student participating in the program tended to maintain his or her relative position in comparison with others nationwide. A test score gain of zero, in this context, means that the typical student in the FTC Scholarship Program achieved a year's worth of learning in a year's time. It is important to note that these national comparisons pertain to all students nationally, and not just low-income students -- the students eligible to participate in the FTC Scholarship Program. It

is also important to note that while the typical gain in national percentile rankings compared with the nation as a whole is essentially zero for program participants, this statistic masks considerable variation in individual students' gains. For instance, 11.0 percent of students participating in the program lost 20 or more percentile points in reading relative to the nation as a whole between 2009-10 and 2010-11, while 9.8 percent of program participants gained more than 20 percentile points in reading over this same time period. The corresponding figures for mathematics are 13.5 and 10.8 percent, respectively. Furthermore, these comparisons are very similar when limited to students taking the Stanford Achievement Test during the spring: 0.3 national percentiles in reading and -0.5 national percentiles in mathematics. Put differently, no matter how one aggregates the test score gains, the typical participating student gained a year's worth of learning in a year's time.

IV. Comparisons with public school test-takers

One important purpose of this evaluation is to compare the relative year-to-year gains in the test score of FTC Scholarship Program students to those of comparable public school students. This report compares the distribution of test score gains between 2009-10 and 2010-11 for the two groups of students. It is very important to note, however, that differences in the gains should not be interpreted as causal, for two principal reasons.

One reason to not interpret differences in test score gains between public school students and FTC Scholarship Program students as causal per se involves the fact that students and families choose whether to participate in the program, and these choices

introduce "selection bias" into any comparison of test score gains.² In addition, selection into a public school comparison group is not random. All FTC Scholarship Program students are certified to be low-income, but only three percent of public school free- or reduced-price lunch students' family incomes are audited, so some fraction of the public school comparison population may actually be of higher income than the program allows. The results of these audits strongly suggest that many public school students receiving free or reduced-price lunches are not from families with comparable incomes to those participating in the FTC Scholarship Program. Therefore, it seems to be clear that school meals recipients in the public schools are not a very effective comparison group for FTC Scholarship Program participants, because family incomes of FTC participants are likely to be considerably lower. While it is impossible to measure just how large these differences are, the results of the audits indicate that they may be substantial.

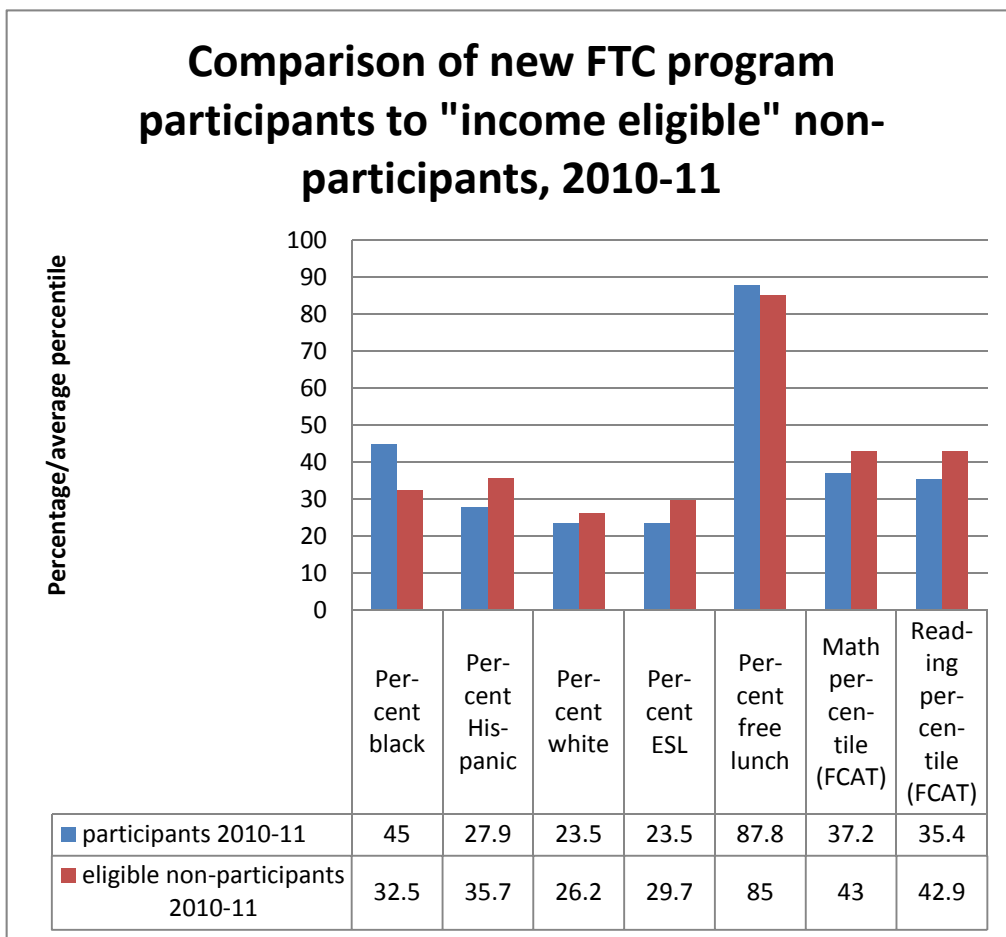
Taken together, these two factors indicate that direct comparisons of average test score gains in the public sector versus FTC Scholarship Program participants, while informative, should not be interpreted as effects of the program on student test score gains.

Summary of key differences between students selecting the FTC Scholarship Program and other income-eligible students

Before directly comparing student test score gains between FTC Scholarship Program participants and others in the public sector, who may or may not be ultimately

² A technical description of selection into the FTC Scholarship Program is provided in David Figlio, Cassandra Hart, and Molly Metzger, "Who Uses a Means-Tested Scholarship, and What Do They Choose?" published in the *Economics of Education Review* in 2009. A brief summary of the key points of that paper is provided in this report.

eligible for program participation, it is important to gauge the degree to which these comparisons are likely to be apples-to-apples comparisons. This report, therefore, begins with a brief summary of some of the key findings of the technical paper mentioned above that describes selection into the program. Any selection findings could reflect either of the two factors -- differential self-selection amongst eligible students; or systematic ineligibility amongst non-participating students who still receive subsidized school meals -- but these findings are highly informative in either case.



The most natural way to make comparisons is to consider a set of students who all spent the prior year in Florida public schools and who received subsidized school meals, making them plausibly eligible to participate in the program. This report employs the

most recent data available at the time of writing -- students who spent the 2009-10 academic year in the Florida public schools, so one can compare the students who entered the FTC Scholarship Program in 2010-11 versus potentially comparable students who did not enter the program in that year but remained free or reduced-price lunch eligible in 2010-11, according to Department of Education records. We exclude students with disabilities who could participate in the McKay Scholarship Program. The chart above presents some basic facts about FTC Scholarship Program participants relative to other potentially income-eligible students. In order to compare similar populations across bars, we restrict analysis to students who had taken either a reading or math test in public school in 2009-10; prior research suggests that this is very similar to the overall population of potential program participants who spent the prior year in a public school. We also limit the analysis to students who would be in grade 10 or below in 2010-11, so that this reflects the set of students for whom a test score is possible. By these standards, there were 3,153 new students in the FTC Scholarship program from this sample and 558,630 students who remained in the public schools and continued on subsidized school lunches in 2010-11.

One observes that FTC Scholarship Program participants differ from non-participants on all of the characteristics easily observed in the administrative record. Scholarship participants are more likely than non-participants to be black, and less likely to be Hispanic or white, and participants are less likely than are non-participants to speak English as a second language. Scholarship participants are more economically disadvantaged than are non-participants on average. While all children in both the participant and non-participant groups were self-reported to be eligible for subsidized

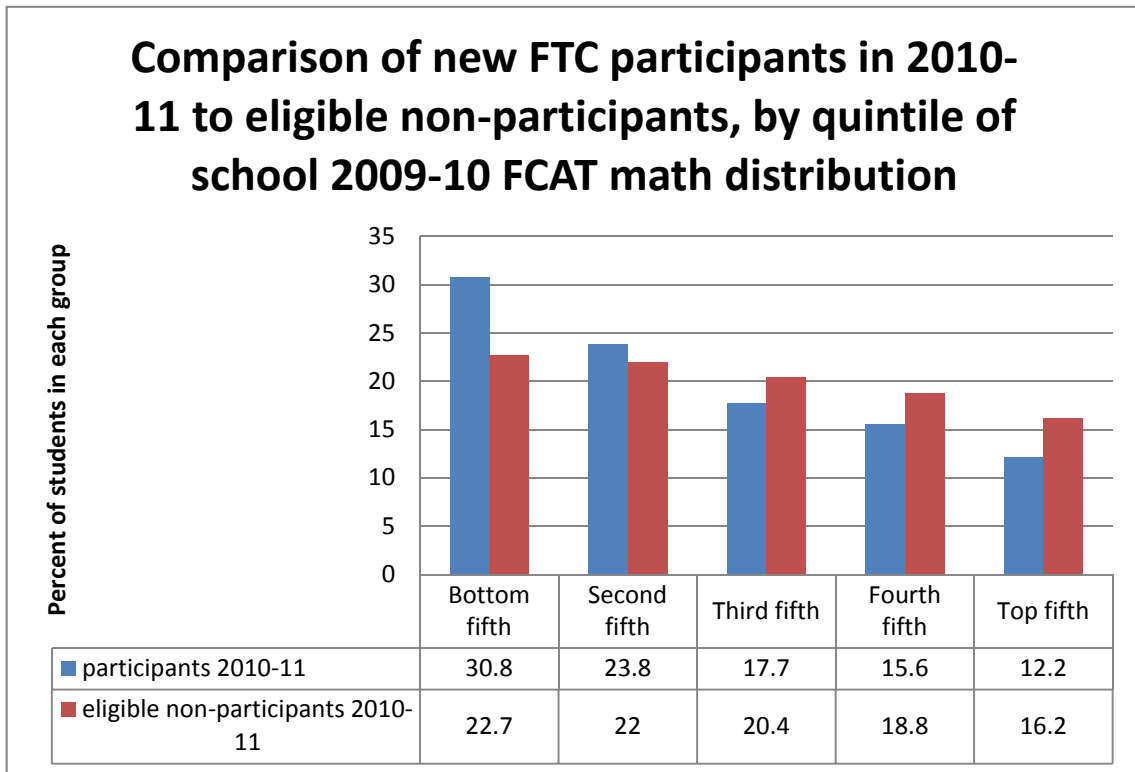
lunch at some point in the 2009-10 school year, participants were more likely to qualify for free lunch as of the last survey taken in 2009-10, while non-participants were more likely to qualify only for reduced-price lunch, indicating that scholarship participants were relatively disadvantaged, even conditional on reported income eligibility. Finally, and perhaps most importantly, scholarship participants have significantly poorer test performance in the year prior to starting the scholarship program than do non-participants. On both the FCAT mathematics and FCAT reading tests, 2010-11 non-participants out-performed 2010-11 scholarship participants in the 2009-10 school year, when both groups were still attending public schools.³ All of these differences are large in magnitude and are statistically significant, and indicate that scholarship participants tend to be considerably more disadvantaged and lower-performing upon entering the program than their non-participating counterparts. These differences are very similar to those observed in years past and reported in prior program reports.

The mean differences in 2009-10 performance between public school students who would ultimately participate in the FTC Scholarship Program in 2010-11 and those who are plausibly income-eligible but who remained in Florida public schools in 2010-11 are compelling, but there are numerous remaining selection questions. For instance, these results are consistent both with the idea that relatively high-performing students from low-performing schools are the ones selecting into the scholarship program, as well as with the idea that relatively low-performing students, regardless of school, are the ones

³ Note that the numbers reported in the test score comparisons are different in this report from those in previous reports. In previous reports, I reported the prior-year norm referenced test national percentile. That is not possible to do in 2008-09 or 2009-10, as students in public schools took only the FCAT. Therefore, in this report, I present information based on the state percentile ranking on the FCAT. All comparisons are qualitatively very similar to those presented in prior years' reports that focused on national norm-referenced percentile ranks.

selecting into the program. It is clear that these two possibilities have very different implications for the interpretation of differential selection into the program.

Consistent with all but one prior year, in 2010-11 FTC Scholarship Program participants came disproportionately from lower-performing schools, according to Florida Department of Education school grades in 2010, as compared to eligible students who did not participate in the program. Amongst the students new to the program in 2010-11, 41.1 percent came from schools graded "A" by the Florida Department of Education in 2010, as compared with 44.8 percent of those public school students eligible for free or reduced-priced lunches who did not participate. At the other extreme, 11.5 percent of new participants came from schools graded "D" or "F" by the Florida Department of Education in 2010, versus 7.0 percent of eligible non-participants.



Also consistent with prior years and in fact strengthening somewhat over time is the fact that regardless of the performance level of the public school that FTC Scholarship Program participants came from, these students tended to be lower-performing before they entered the program. As can be seen in the previous table, 30.8 percent of students who would select into the program were in the bottom fifth of their prior public school's mathematics FCAT test score distribution, while only 22.7 percent of non-participating free- or reduced-price lunch students were in the bottom fifth of the distribution in the prior public school. This gap of 8.1 percentage points is more pronounced than the 6.7 percentage point difference in last year's report and the 4.4 percentage point gap observed in the previous year's report. (Similar differences are present in terms of reading scores.) At the top of the test score distribution, only 12.2 percent of students who would select into the program were in the top fifth of their prior public school's mathematics test score distribution, as compared with 16.2 percent of free- or reduced-price lunch students in the top fifth of the distribution in the prior public school; the 4.0 percentage point gap is in line with last year's 4.3 percentage point gap and modestly larger than the 3.3 point gap reported in the previous year's report. Clearly, public school students who ultimately became program participants are more likely to be the relatively lower-performing students in their schools, a fact that has not changed over time.

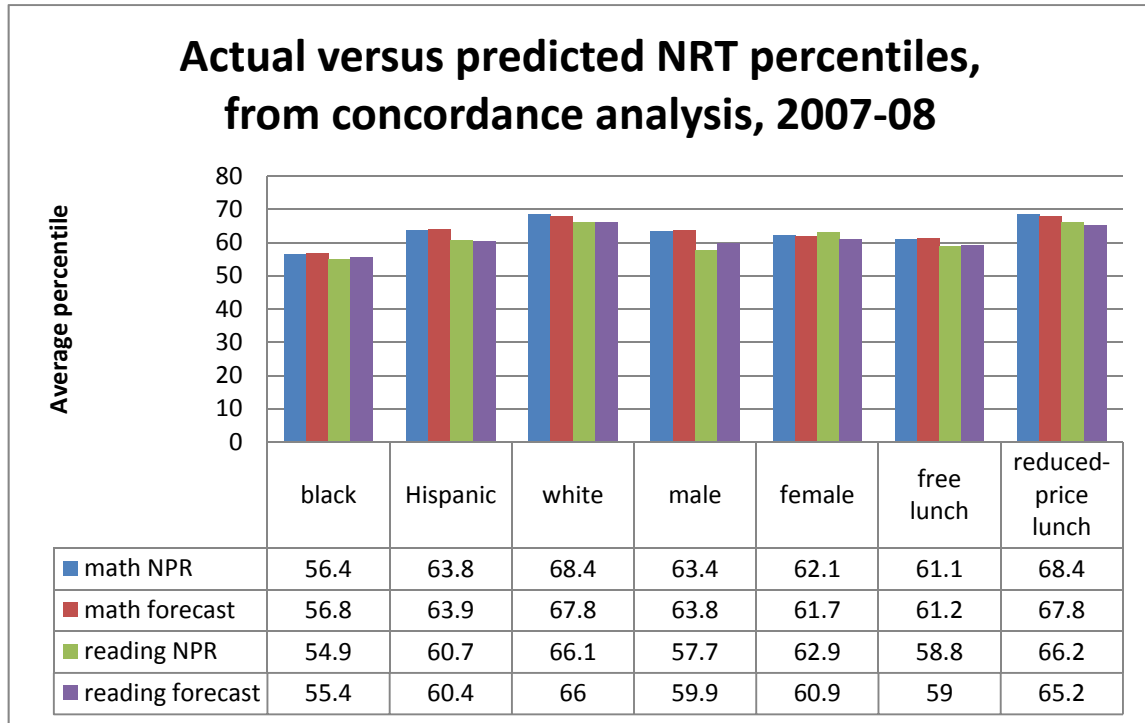
Computing gains of public school students

The fact that program participants are not a random sample of potential students makes clear that direct comparisons of gains of program participants to non-participants

will not yield causal estimates of the effects of the program on participating students. Nonetheless, it is still very worthwhile to benchmark the distribution of measured student learning gains amongst program participants against the distribution of learning gains amongst potentially eligible public school students who elected not to participate in the program.

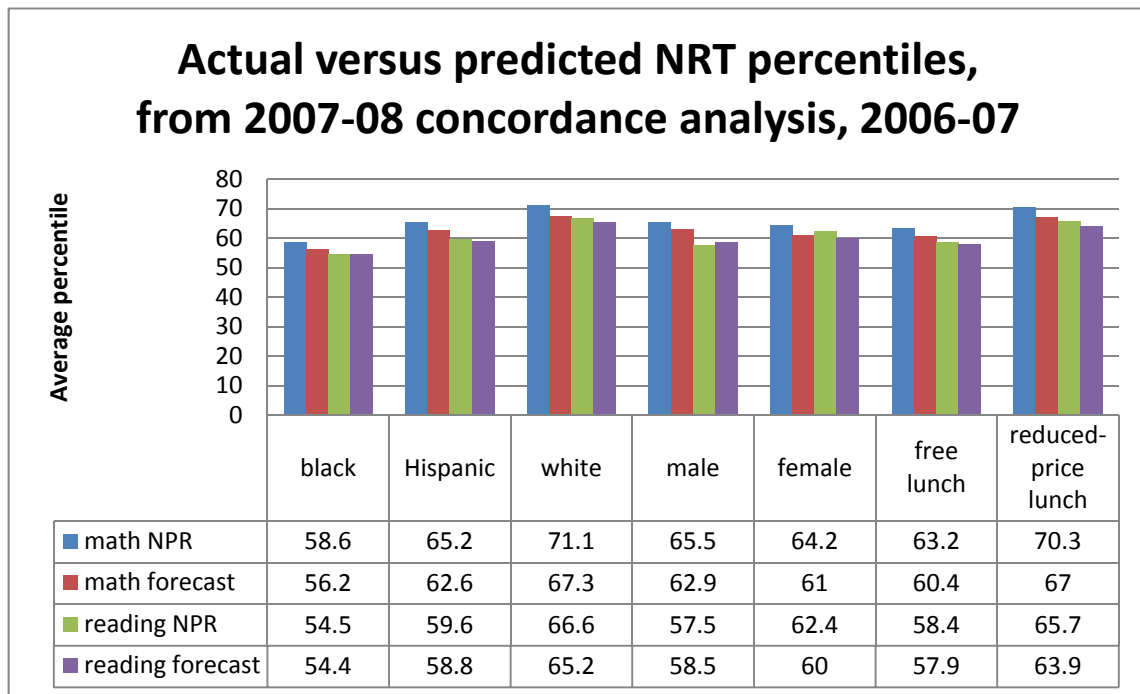
An additional complication is that public school students no longer take a directly comparable nationally norm-referenced test, making comparisons across sectors somewhat more challenging. Through the 2007-08 academic year, public school students took both the criterion-referenced FCAT as well as the norm-referenced Stanford Achievement Test, but the norm-referenced test administration was ended due to budgetary concerns. That said, it is still possible to make comparisons between program participants and non-participants by performing an analysis of the concordance between FCAT scores and Stanford Achievement Test scores. In principle, a concordance analysis predicts what the norm-referenced national percentile would have been, given the level of the FCAT score. This concordance analysis was conducted with the most recent data -- the 2007-08 academic year -- for which the same Florida students took both the FCAT and the norm-referenced test. In practice, for every value of the FCAT developmental scale score in each grade level, I computed the mean NRT national percentile ranking and assigned this mean national percentile ranking as the predicted NRT score to accompany a given FCAT developmental scale score for a given grade level. Because students from different groups might have different concordances between the two tests, the predictions were made using the set of students who were

eligible for subsidized school meals in both 2007-08 and 2008-09. The results of this concordance analysis are highly robust to other population definitions.



The above figure compares mean actual national percentile rankings from the 2007-08 Stanford Achievement Test to predicted national percentile rankings for the same students, based on the concordance analysis conducted in 2007-08, for several subgroups of students. As can be seen in the figure, the actual and predicted scores line up closely across the subgroups. The only place where the match is not as precise involves reading across the genders: The concordance analysis tends to modestly overpredict male reading scores and modestly underpredict female reading scores. However, in general, the concordance analysis using 2007-08 data tends to predict norm-referenced test scores very well. Indeed, the correlation between actual and predicted

math scores in 2007-08 is 0.84 and the correlation between actual and predicted reading scores in 2007-08 is 0.78.

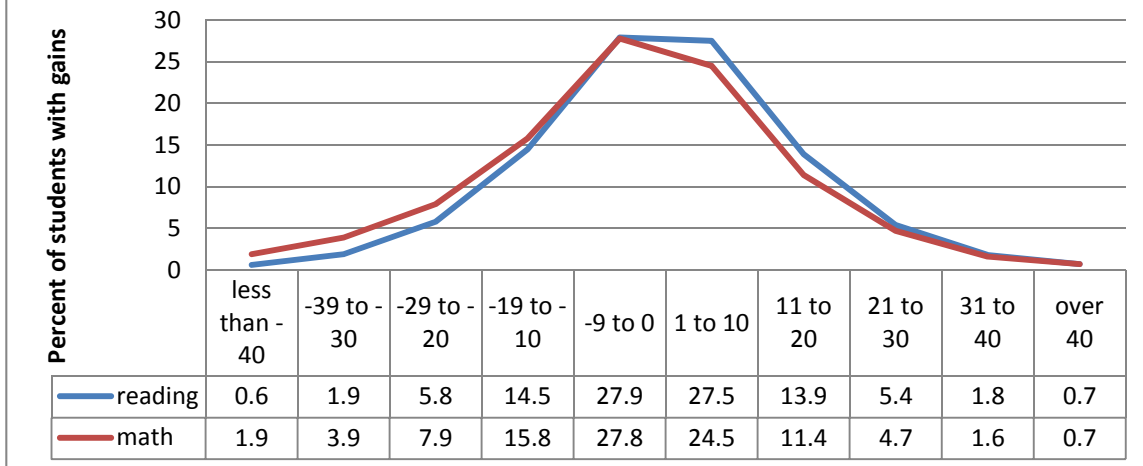


Of course, the purpose of the concordance analysis is to predict norm-referenced test scores in years when there are no norm-referenced scores. To test the potential validity of the concordance analysis, we back the analysis up a year, and predict 2006-07 norm-referenced test scores using 2006-07 FCAT scores, but with the concordance metrics developed using 2007-08 data. As can be seen in the above figure, the relationship between actual NRT scores and predicted NRT scores based on the concordance analysis remains very high: The correlation between 2006-07 predicted scores and 2006-07 actual scores is 0.82 for math and 0.79 for reading. In practice, it appears as if the concordance analysis modestly underpredicts math scores in 2006-07, so the relationship is not perfect, but the correlations are very strong. One can draw similar conclusions when comparing the realized gain scores on the NRT to the forecast gains on the NRT between 2006-07 and 2007-08: In reading, the mean forecast gain based on the

FCAT concordance analysis is 2.0 percentile points while the mean realized NRT gain is a very similar 1.4 percentile points. In mathematics, the difference is greater: The mean forecast gain is 2.1 percentile points while the mean realized gain is -0.6 percentile points. It is not clear whether this implies that the forecasts for the concordance analysis will overstate or understate the true gains between 2007-08 and 2008-09 -- as both are possible, depending on the interpretation of the differences between 2006-07 and 2007-08 -- but the results do indicate that the concordance analysis is perhaps more successful in the case of reading rather than mathematics. The good news, from the point of view of making valid comparisons between gain scores of private school students who take exclusively norm-referenced exams and those of public school students who take exclusively the FCAT, is that it appears possible to make reasonable comparisons across these two sectors even when the examinations taken are different.

With these provisos in mind, one can now turn to measuring test score gains for the public school students who received subsidized school meals in both 2009-10 and 2010-11. This report employs the concordance metrics described above to compute predicted NRT scores in 2009-10 and 2010-11 based on the student's actual FCAT scores in the two years.

Distribution of predicted test score gains (from concordance analysis) for income-eligible public school students, 2009-10 to 2010-11



The distribution of test score gains amongst public school students is very similar to the distribution of gains amongst program participants. The mean gain in the public school comparison group is 0.8 percentile points higher than the mean gain amongst program participants in reading and 1.1 percentile points lower in mathematics, but given the selection issues mentioned earlier in this report, these mean gain differences should not be considered to be meaningful. Participating schools have more students in the tails of the distribution -- those with gains or losses of more than 20 percentile points -- than the public school students, but the differences in the extremes may be due in part to the concordance analysis. In summary, both distributions of test score gains are in the same ballpark, with public schools slightly outperforming private schools on one average and the reverse true on the other average. In prior years it was possible to use a regression discontinuity design to generate causal estimates of the effects of program participation on student test scores, but for the most recent application cohort for which one can

measure gain scores -- the application cohort of 2009-10 -- there exist insufficient data to carry out the regression discontinuity analysis with confidence.⁴ The reason for the change is that the scholarship funding organization implemented additional steps in the online application form to pre-screen families who would be ineligible to participate in the program in order to save ineligible potential applicants time and money; this additional step caused many ineligible potential applicants to no longer decide to apply for the program. Therefore, it is best to think of these test score gain comparisons as descriptive, rather than causal. However, given that the factors leading to participation in the program have been generally stable across cohorts, and since distributions of public and private school test score gains have been stable over time, it stands to reason that the causal effects of program participation seem likely to remain roughly comparable over time. Therefore, since in prior cohorts the estimated effect of program participation were modestly positive, especially for math, one can infer that there is a strong chance that the effects would remain positive in the most recent cohort as well. That said, this last statement is necessarily speculative.

V. Individual school average gain scores, 2009-10 to 2010-11

Beginning with this report, the Florida statutes require that average student gain scores be reported for schools with 30 or more participating students with gain scores.

⁴ Specifically, the regression discontinuity approach employed in previous reports requires a large number of income-ineligible families to apply to participate in the program, and that the families barely ineligible for the program would have similar characteristics to the families that are barely eligible for the program. However, due to improvements in the application process that successfully deterred many ineligible potential applicants from applying, the 2009-10 application cohort (which would provide the gains comparisons parallel to prior years' reports) has only about half the income-ineligible applicants with prior test scores as did the 2008-09 application cohort, and those ineligible families that did apply for the program are not sufficiently similar to the barely-eligible families to carry out the regression discontinuity model with confidence.

Average gain scores are only a single indicator of a school's quality, so should not be interpreted as definitive measures of a school's performance, but rather as one of a large number of ways in which a school could be evaluated. The Appendix Table reports the average gain scores for the 70 schools with sufficiently large numbers of students to qualify them for public reporting. School average gain scores are reported for reading, mathematics, and combined (the average of reading and mathematics.) The combined score is especially informative in cases such as this where average scores are based on a reasonably small number of observations. In addition to presenting the one-year gain scores for 2010-11, the Appendix Table presents the average gain scores over three years, from 2008-09 through 2010-11.

The rationale behind including the three-year moving average of gain scores is that while an average gain score in a single year is one potential indicator of school quality, it is an extremely noisy measure of a school's contribution to student test scores, and the likelihood that noise is dominating the measured gain scores increases the smaller the number of student gains that are being considered. As an example of how average gain scores in a single year can be misleading, consider a school whose students performed idiosyncratically well in 2009-10. That school is likely to experience a negative average gain score in 2010-11 because it is doubtful that the school will have an idiosyncratically positive performance two years in a row. (The same is true, of course, in reverse for schools with students who performed unusually poorly in 2009-10, and for

whom we expect a "bounce back.") This phenomenon is called "regression to the mean," and it is very prevalent in situations such as this.⁵

There are no sure-fire solutions to the faulty inference caused by regression to the mean, but one way to minimize the effects of the phenomenon is to average gain scores across several years. Doing so both adds extra observations -- reducing the potential for a small number of student gain scores to drive the average -- as well as balances out idiosyncratically positive and idiosyncratically negative scores over time. A multi-year moving average, therefore, provides a more accurate measure of a school's contribution to student test scores than a single gain score measure in cases where relatively small number of gains scores are evaluated.⁶ The benefit of presenting both the one-year average gain score and the three-year average of gain scores becomes apparent when one observes that there are occasionally schools with very strong gain scores in 2010-11 that do not reflect the longer-term sustained gain scores of students in the school, as well as schools with very weak gain scores in 2010-11 that are unrepresentative of the longer-term averages. Therefore, one-year average gain scores should be treated extremely cautiously.

Because the three-year moving average is the more reliable measure of a school's average gain scores, the schools are rank-ordered from highest average combined gain in reading and mathematics to lowest average combined gain using the three-year measure. It is important to note that schools near one another in the ranking cannot be statistically

⁵ Regression to the mean is less of a concern in the case of public schools because public schools tend to have many more measured gain scores than do the private schools participating in the FTC Scholarship Program.

⁶ Note that Thumbelina Learning Center is the only school amongst the 70 to not have many gain scores recorded prior to 2010-11. Therefore, its three-year average gain score should be interpreted with more caution than the other three-year average gain scores reported in the Appendix Table.

differentiated from one another. Rather, I identify the schools with average gain scores that are statistically distinguishable from zero (at the 95 percent level of confidence in a two-tailed test), either positively or negatively, by highlighting the cell where the average gain score is reported. Put differently, if a school is reporting having statistically positive estimated gains, that means that one can be at least 95 percent confident that the school's students achieved more than a year's gain in a year's time. (For schools with statistically negative estimated gains, this suggests that one can be at least 95 percent confident that the school's students achieved less than a year's gain in a year's time.) Beside every school's average combined gain score is its average math gain score and its average reading gain score. Recall that an average gain score of zero does not imply that students are not gaining; rather, an average gain score of zero means that students are maintaining their position relative to the national average, or, in other words, achieving a year's gain in a year's time.

VI. Conclusion

This report presents empirical evidence on the compliance and performance of private schools that participate in the Florida Tax Credit Scholarship Program. The report analyzes data from 2010-11, and compares these data to prior years of test score collection and public school data from the Education Data Warehouse of the Florida Department of Education. There is strong evidence of high degrees of compliance with testing requirements for program participants.

Simple comparisons of the distribution of test score gains between FTC Scholarship Program participants and plausibly-eligible non-participants indicate that the test score gains in both populations are comparable in magnitude between program participants and eligible non-participants. But we must recognize that these populations are not equivalent. Program participants enter the program with lower academic performance and with substantially lower household incomes, which makes comparison more problematic. In addition, these are not causal estimates of differences, and the true effect of program participation may be more positive or more negative than the simple means comparisons. There is strong and compelling evidence that relatively low-performing students from relatively low-performing schools tend to be the students to participate in the FTC Scholarship Program, and causal analysis of these differences would need to take this differential selection into account. It is, therefore, wisest to interpret the similar test score gain performance between program participants and eligible non-participants as suggestive, but not definitive, evidence of similar performance across the sectors.

Finally, there exists compelling causal evidence indicating that the FTC Scholarship Program has led to modest and statistically significant improvements in public school performance across the state. Therefore, a cautious read of the weight of the available evidence suggests that the FTC Scholarship Program has boosted student performance in public schools statewide, that the program draws disproportionately low-income, poorly-performing students from the public schools into the private schools, and that the students who moved perform as well or better once they move to the private schools.

Appendix Table: Average gain scores in 2010-11 and three-year moving average of gain scores from 2008-09 to 2010-11 for schools with 30 or more gain scores in 2010-11, ranked by average three-year combined gain score.

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
WORSHIPERS' HOUSE OF PRAYER ACADEMY	MIAMI	38	95	9.5	7.5	11.5	3.6	3.3	5.7
SUNFLOWERS ACADEMY*	MIAMI	72	139	-8.8	-7.3	-10.4	3.3	4.2	2.4
NUR UL-ISLAM ACADEMY	COOPER CITY	69	125	5.1	6.2	4.0	3.0	3.4	2.6
UNIVERSAL ACADEMY OF FLORIDA	TAMPA	48	131	0.8	0.3	1.3	3.0	3.5	2.5
PATHWAYS SCHOOL	ORLANDO	39	89	9.8	9.4	10.1	2.3	2.1	2.5
ESPRIT DE CORPS CENTER FOR LEARNING	JACKSONVILLE	40	103	1.9	1.5	2.7	2.3	2.6	1.2
TALLAVANA CHRISTIAN SCHOOL	HAVANA	30	65	-2.6	-4.1	1.1	2.3	2.7	1.9
THE POTTER'S HOUSE CHRISTIAN ACADEMY ELEM	JACKSONVILLE	82	217	-0.7	-3.5	2.1	2.2	0.4	4.1
ABUNDANT LIFE CHRISTIAN ACADEMY	MARGATE	34	79	4.0	5.0	3.1	2.1	3.2	0.9
SAINT MICHAEL THE ARCHANGEL*	MIAMI	32	72	2.3	-1.6	6.2	2.0	0.4	3.7
BRITO MIAMI PRIVATE SCHOOL	MIAMI	32	89	18.7	19.5	17.8	1.8	1.3	2.1

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
SAINT HELEN CATHOLIC SCHOOL*	FORT LAUDERDALE	31	69	4.9	7.3	2.4	1.8	2.5	1.2
ESCAMBIA CHRISTIAN SCHOOL	PENSACOLA	33	81	-1.8	-4.4	0.8	1.7	0.4	3.0
TREASURE OF KNOWLEDGE CHRISTIAN ACADEMY	ORLANDO	31	77	0.4	-4.4	5.9	1.1	-0.5	2.3
BETESDA CHRISTIAN SCHOOL	OPA-LOCKA	42	83	-0.6	1.5	2.6	1.0	0.7	2.1
VICTORY CHRISTIAN ACADEMY	ORLANDO	53	128	3.3	3.1	3.5	0.8	1.7	-0.1
LIGHTHOUSE CHRISTIAN ACADEMY	DELAND	35	75	3.2	0.7	5.8	0.8	0.4	1.3
AMERICAN YOUTH ACADEMY, INC.**	TAMPA	73	199	-3.2	-4.9	-2.8	0.8	0.9	1.2
TRINITY CHRISTIAN ACADEMY	DELTONA	55	127	-2.2	-4.9	0.5	0.5	-0.1	1.1
WARNER CHRISTIAN ACADEMY	SOUTH DAYTONA BEACH	46	102	0.3	-1.7	2.3	0.4	-0.8	1.5
HERITAGE CHRISTIAN SCHOOL	KISSIMMEE	79	200	4.4	7.8	1.0	0.3	1.6	-1
SAINT JOHN THE APOSTLE SCHOOL*	HIALEAH	54	89	1.5	0.9	2.2	0.2	-0.7	1.2

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
EASTLAND CHRISTIAN SCHOOL	ORLANDO	37	112	2.6	2.9	2.2	0.0	0.0	0.0
CITY OF LIFE CHRISTIAN ACADEMY	KISSIMMEE	59	159	-4.0	-5.2	-2.8	-0.3	-0.8	0.0
ACADEMY PREP CENTER OF TAMPA INC.	TAMPA	53	92	3.3	3.6	3.1	-0.4	0.5	-1.2
BERYL WISDOM ADVENTIST SCHOOL*	ORLANDO	33	75	-3.8	0.9	-7.5	-0.6	0.3	-1.6
LINCOLN-MARTI COMMUNITY AGENCY 01-931	MIAMI	104	200	-3.9	-7.3	-0.4	-0.6	0.9	-2.3
SAINT MARYS CATHEDRAL*	MIAMI	82	138	1.7	-0.3	3.5	-0.7	-2.2	1.2
SAINT ANDREW CATHOLIC SCHOOL*	ORLANDO	37	101	1.1	1.2	1.0	-0.8	-0.8	-0.8
LIFE ASSEMBLY OF GOD LIFE ACADEMY	KISSIMMEE	57	129	-5.3	-7.1	-3.5	-0.8	-1.2	-0.3
SAINT BARTHOLOMEW SCHOOL*	MIRAMAR	31	84	5.5	6.0	5.1	-0.9	-0.8	-1.0
CALVARY CHRISTIAN ACADEMY	ORMOND BEACH	31	66	2.0	2.8	1.2	-0.9	-0.6	-1.2
FAITH CHRISTIAN ACADEMY	ORLANDO	66	156	-2.8	-4.5	-1.2	-0.9	-2.5	0.7

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
MUSLIM ACADEMY OF GREATER ORLANDO	ORLANDO	31	101	4.0	0.3	7.8	-1.0	-1.1	-1.0
LA PROGRESSIVA PRESBYTERIAN SCHOOL INC.	MIAMI	84	210	-0.6	-2.5	0.7	-1.0	-2.4	0.5
OUR LADY OF LOURDES CATHOLIC SCHOOL*	DAYTONA BEACH	34	76	3.7	3.5	4.2	-1.1	-0.9	-1.1
NORTH FLORIDA CHRISTIAN SCHOOL	TALLAHASSEE	38	95	0.4	-2.3	3.1	-1.1	-2.5	0.3
JOSHUA CHRISTIAN ACADEMY	JACKSONVILLE	78	179	-0.5	0.4	1.4	-1.1	-2.5	0.2
FOREST LAKE EDUCATION CENTER*	LONGWOOD	42	117	-0.8	-2.0	0.4	-1.1	-1.8	-0.5
TRINITY CHRISTIAN ACADEMY	JACKSONVILLE	63	165	-2.4	-3.2	1.6	-1.1	-0.5	-1.6
PLEASANT HILL ACADEMY	KISSIMMEE	63	132	2.9	4.3	1.6	-1.2	-2.2	-0.3
LINCOLN-MARTI COMMUNITY AGENCY 17	HIALEAH	70	182	-2.2	-2.4	1.9	-1.3	0.0	-2.9
OCALA CHRISTIAN ACADEMY	OCALA	35	71	0.0	-0.9	0.8	-1.4	-2.7	-0.1
MIAMI UNION ACADEMY*	NORTH MIAMI	86	201	0.6	-0.8	1.5	-1.5	-2.5	-0.3

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
FIRST COAST CHRISTIAN SCHOOL	JACKSONVILLE	33	78	0.1	-3.2	3.3	-1.5	-3.5	0.4
WEST MELBOURNE CHRISTIAN ACADEMY	WEST MELBOURNE	31	68	-3.0	-4.3	1.4	-1.5	-0.8	-2.3
NORTHRIDGE CHRISTIAN ACADEMY	HAINES CITY	30	75	-7.2	-11.8	-2.6	-1.7	-4.0	0.7
NORTHWEST CHRISTIAN ACADEMY	MIAMI	38	83	-1.7	-2.0	1.3	-1.8	-3.9	0.5
SOUTH ORLANDO CHRISTIAN ACADEMY	ORLANDO	46	137	7.8	10.2	5.3	-1.9	-2.3	-1.9
HOLY ROSARY CATHOLIC SCHOOL*	JACKSONVILLE	33	103	1.2	1.1	1.2	-1.9	-2.4	-1.5
EDISON PRIVATE SCHOOL	HIALEAH	54	122	-0.9	-2.2	0.4	-1.9	-1.9	-2.1
GREATER MIAMI ADVENTIST SCHOOL*	MIAMI	40	91	-2.2	-4.2	0.0	-1.9	-3.4	-0.5
MELODY CHRISTIAN ACADEMY	LIVE OAK	44	100	-4.8	-5.8	-3.8	-1.9	-1.8	-2.1
AGAPE CHRISTIAN ACADEMY	ORLANDO	58	182	-1.0	1.0	3.1	-2.0	-3.3	-0.5
LINCOLN-MARTI COMMUNITY AGENCY 10	MIAMI	79	175	3.8	3.1	4.8	-2.2	-2.0	-2.4
S.L. JONES CHRISTIAN ACADEMY	PENSACOLA	46	89	0.9	-0.6	2.5	-2.2	-3.0	-1.5

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
HOLY FAMILY CATHOLIC SCHOOL*	NORTH MIAMI	67	156	2.7	1.7	5.3	-2.6	-4.9	-1.2
CEDAR CREEK CHRISTIAN SCHOOL	JACKSONVILLE	44	105	3.0	0.5	5.5	-2.7	-4.0	-1.8
KINGSWAY CHRISTIAN ACADEMY	ORLANDO	71	164	-1.9	0.0	3.8	-2.7	-1.5	-3.9
SAINT JAMES CATHOLIC SCHOOL*	MIAMI	74	165	-2.0	-4.6	0.5	-2.7	-5.6	0.1
CORNERSTONE CHRISTIAN SCHOOL**	JACKSONVILLE	32	101	-10.2	-16	-4.4	-3.3	-3.4	-3.1
LEADERS PREPARATORY SCHOOL	ORLANDO	37	69	-1.1	1.8	3.9	-3.8	-2.7	-4.8
LANDOW YESHIVA CENTER*	MIAMI	66	102	-3.1	-3.7	2.6	-4.1	-5.2	-3.0
WEST OAKS ACADEMY	ORLANDO	37	82	-3.8	-3.5	4.2	-5.0	-5.4	-4.1
CHAMPAGNAT CATHOLIC SCHOOL OF MIAMI	MIAMI	49	99	-9.8	-10.6	-9.0	-5.2	-7.2	-3.9
HERITAGE PREPARATORY SCHOOL	ORLANDO	55	165	-2.3	-0.2	4.4	-5.7	-5.5	-5.9
MONSIGNOR EDWARD PACE HIGH SCHOOL**	MIAMI GARDENS	54	105	-1.7	2.0	-5.4	-5.8	-6.3	-5.4
CHAMPAGNAT CATHOLIC SCHOOL OF HIALEAH	HIALEAH	58	121	-9.1	-8.5	-9.6	-6.0	-5.2	-6.8

SCHOOL NAME	CITY	NUMBER OF GAIN SCORES OBSERVED		AVERAGE GAIN SCORE IN 2010-11			AVERAGE GAIN SCORE FROM 2008-09 TO 2010-11		
		2010-11 SCHOOL YEAR	BETWEEN 2008-09 AND 2010-11	READING+ MATH COMBINED	MATH	READING	READING+ MATH COMBINED	MATH	READING
ARCHBISHOP CURLEY/NOTRE DAME HIGH SCHOOL**	MIAMI	31	61	-6.8	-7.5	-6.1	-6.3	-9.1	-3.5
THUMBELINA LEARNING CENTER 4 (see note)	HIALEAH	49	52	-7.6	-9.2	-5.6	-6.5	-8.6	-4.7

Notes:

- (1) Cells report average gain scores. Cells that are bolded and highlighted are statistically distinct from the national average at the 95 percent level of confidence.
- (2) Nearly all of Thumbelina Learning Center's gain scores were recorded in 2010-11. Therefore, this school's three-year average gain score should be interpreted with more caution than the reader might interpret the other three-year average gain scores.
- (3) All schools administered the Stanford Achievement Test in 2010-11 except when marked by the school's name. Schools marked with * administered the Iowa Test of Basic Skills (except for Landow Yeshiva Center, which administered the Iowa Test of Basic Skills in grades 3-6 and the Stanford Achievement test in grades 7-9.) Schools marked with ** administered another test besides the Stanford or the Iowa test. American Youth Academy administered the ERB test; Archbishop Curley/Notre Dame High School administered the ACT/PLAN; Cornerstone Christian School administered the Terra Nova test; and Monsignor Edward Pace High School administered the PSAT.