Evaluation of the Florida Tax Credit Scholarship Program
Participation, Compliance, Test Scores and Parental Satisfaction in 2008-09

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

June 2010

Executive summary

This is the third in a series of reports evaluating Florida's Corporate Tax Credit (FTC) Scholarship Program, as required by the Florida Statutes, s. 220.187(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. For convenience, this report refers to the program by its current name, even though the data were collected when the program was still called the Corporate Tax Credit Scholarship Program.

During the 2008-09 academic year, David Figlio, the Project Director, collected test score data from private schools participating in the FTC Scholarship Program in real time. This is the third year for which program participants' test score data were collected, but the second year in which this data collection occurred in real time. This is the first year for which test score gains for private school participants in the program can be interpreted with confidence.

Compliance with program testing requirements, 2008-09:

● Compliance with program testing requirements remained very high in 2008-09. Private schools provided usable test scores for 89.8 percent of program participants in grades 3-10. Another 6.2 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 1.9 percent rate of reported illness/absence is somewhat higher than in 2007-08, but remains comparable to the public school illness/absence rate. Test administration compliance errors by participating schools continue to decline, with reporting problems involving only 1.3 percent of participants in 2008-09.

● The vast majority (68.8 percent) of test-takers took the Stanford Achievement Test. Other popular tests were the Iowa Test of Basic Skills (22.2 percent) and the TerraNova (4.0 percent).

● Scholarship students whose test scores were received are modestly more advantaged than are those scholarship students whose scores were not received. The same is true for program participants with two consecutive years of test scores (thereby facilitating the calculation of test score gains) as compared with those without two consecutive years of test scores.

Selection into the FTC Scholarship Program:

● Program participants tend to come from less advantaged families than other students receiving free or reduced-price lunches.

● Program participants are more likely to come from lower-performing public schools prior to entering the program. In addition, they tend to be among the lowest-performing students in their prior school, regardless of the performance level of their public school.

Test scores of program participants, 2008-09:

● The typical student in the program scored at the 45.3rd national percentile in reading and the 46.2nd percentile in mathematics. 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 2009. 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 -0.1 national percentile ranking points in reading and -1.1 national percentile ranking points in mathematics. These mean gains are indistinguishable from zero. In other words, the typical student participating in the program tended to maintain his or her relative position in comparison with others nationwide. It is important to note that these national comparisons pertain to all students nationally, and not just low-income students.

● Because families can choose whether to participate in the program, it is inappropriate to consider the differences in test score gains between FTC Scholarship Program participants and their public school counterparts to be *caused* by program participation. Credible comparisons of program participants and non-participants must take into account this *selection problem.* This report makes use of the best available statistical tools for determining the causal effect of program participation.

● The best statistical estimates (using a tool called regression discontinuity design) of the effects of program participation indicate that participation is associated with no differences in reading gains and possibly small improvements in mathematics, relative to public school students who applied for participation in the program, though these differences are not statistically significant. The results are most consistent with a finding of small differences between program participants and non-participants.

● Recent statistical research has shown that the FTC Scholarship Program has improved the performance of the 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 Scholarship Program.

Parental satisfaction, 2008-09:

● In a survey of parental satisfaction conducted in spring of the 2008-09 school year, parents of participating students were more likely to consider their child's school to be "good" or "excellent" than were parents of non-participating students who applied to the program. These results are suggestive, but should not be interpreted as causal.

I. Background

This is the third in a series of reports evaluating the Florida Tax Credit Scholarship Program, as required by the Florida Statutes, s. 220.187(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 awarded a contract to the University of Florida at 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 6, 2008. A second report, dated June 16, 2009, presented data from the 2007-08 academic year, the first year in which real-time testing data were collected for program participants.

This report presents the results of the real-time test score collection in 2008-09. This report details key information about program participation and test scores, and

because score reporting was high in the two consecutive years, provides the first reliable calculation of student test score gains for program participants.

II. Test score collection in 2008-09

Data collection protocol

As required by s. 330.287(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 statewide assessments at a public school in accordance with s. 220.187(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 2008-09 academic year was the second year in which it was possible to collect participant test score data in real time. Pursuant to s. 220.187(8)(c)(2), in Winter 2009 the Independent Research Organization contacted the 951 private schools that had participating students in grades three through ten during the 2008-09 school year, as reported on the October roster of program participants. The Florida Department of Education provided the Project Director with a list of all participating students in 2008-09, as of the October participant roster; of these, 11,508 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

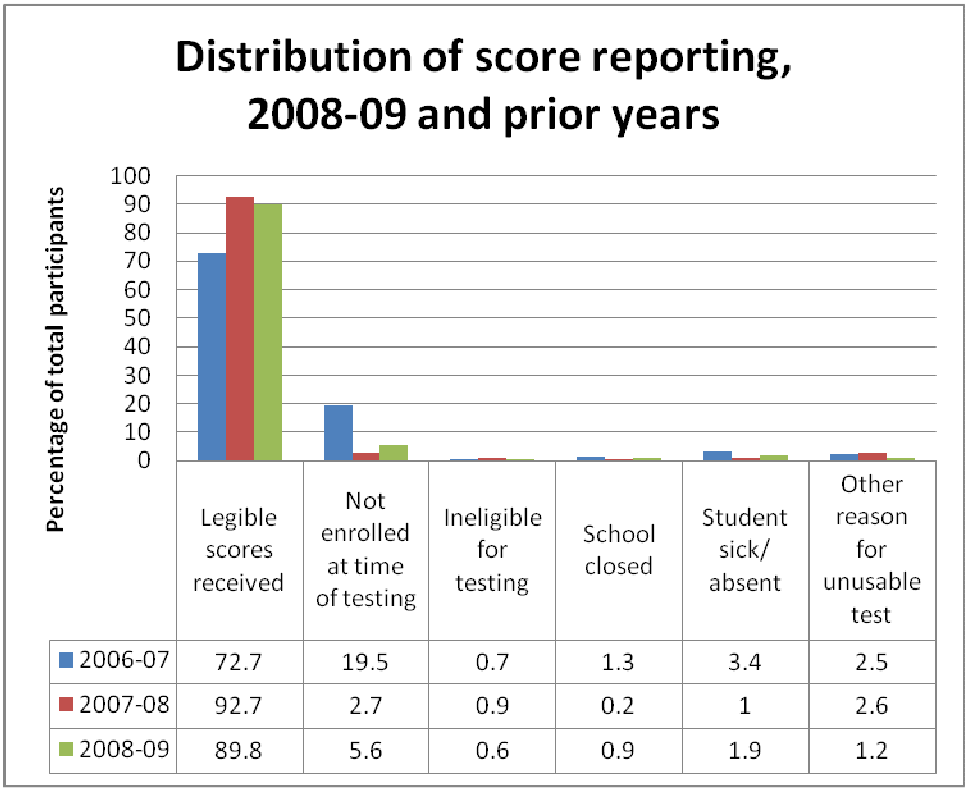
In over 99 percent of cases, schools submitted photocopies of official score sheets provided to them by the relevant testing company (e.g., Harcourt). In a small number of schools, the schools scored the tests themselves and forwarded to the Project Director detailed information regarding the nature 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 951 schools with students in the relevant grades in 2008-09, the overwhelming majority provided evidence of test administration according to the specifications of the program. A small fraction of participating schools closed following

the 2008-09 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 11,508 students in relevant grades participating in the program in 2008-09, the Independent Research Organization received valid, legible test scores for 10,333 students, or 89.8 percent of all expected students;[1] virtually all of these scores were from tests administered by the private schools themselves. This is modestly lower than the 92.7 percent figure for 2007-08, though still in the same vicinity and easily explainable for reasons described below, and represents maintenance of the dramatic improvement in score reporting rates over the retrospective 2006-07 score reporting, in which the comparable figure was 72.7 percent. The difference between the retrospective score reporting in 2006-07 and the real-time score reporting in 2007-08 and 2008-09 underscores the importance of collecting test score data in real time, and demonstrates why the 2008-09 score data present the first credible opportunity to measure the distribution of the student test score gains for program participants.

[1] We received six additional test scores following the January 8, 2010 date in which we merged score records with school records. This report excludes these six test scores, because they cannot be merged with the state records for the purposes of analysis.



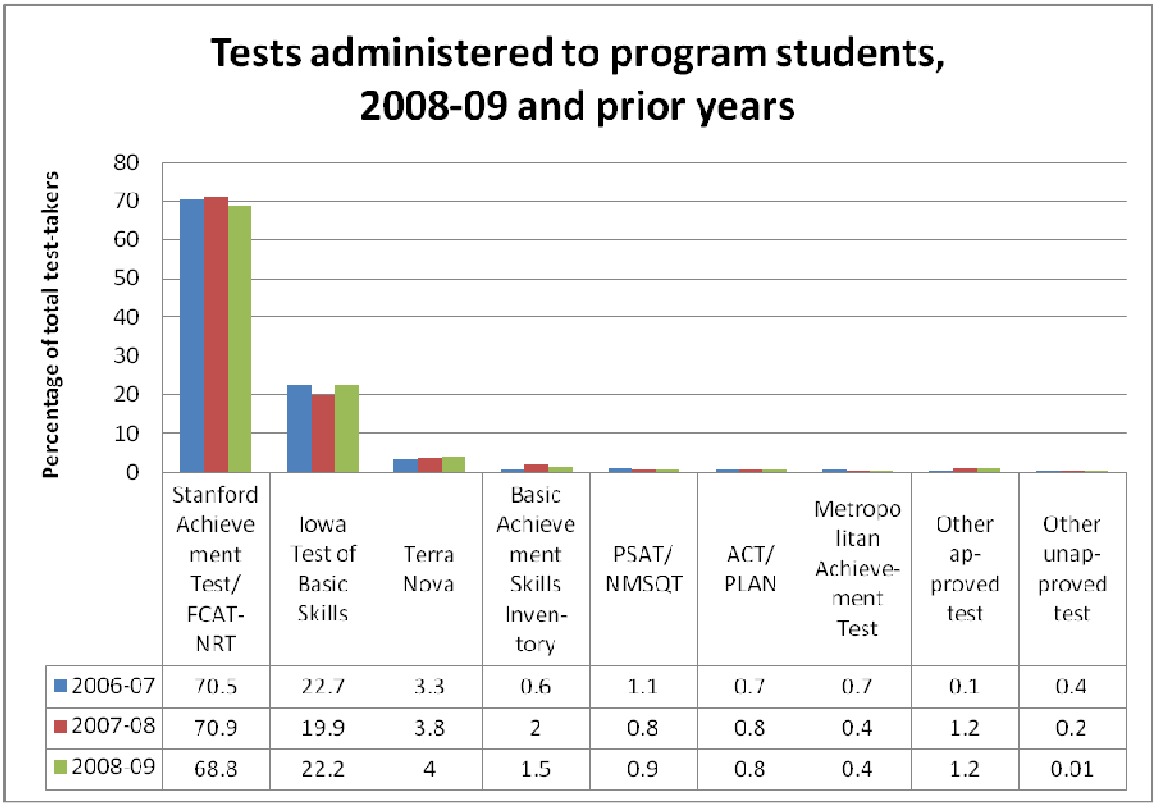
	Legible scores received	Not enrolled at time of testing	Ineligible for testing	School closed	Student sick/ absent	Other reason for unusable test
2006-07	72.7	19.5	0.7	1.3	3.4	2.5
2007-08	92.7	2.7	0.9	0.2	1	2.6
2008-09	89.8	5.6	0.6	0.9	1.9	1.2

The difference between the 2007-08 and 2008-09 percentage of program participants with valid test score gains can be explained by an uptick in the percentage of students who either arrived in the private school after the testing took place -- there is a larger fraction of students attending schools that administered the Iowa Test of Basic Skills in the fall in 2008-09 as opposed to what occurred during 2007-08 -- or left the school before the time in the academic year in which the school administered testing. In 2008-09, the percentage of students falling into one of these two categories increased to 5.6 percent, as opposed to the comparable figure of 2.7 percent in 2007-08. In addition, 0.6 percent of 2008-09 program participants listed on the official roster were deemed ineligible for test score reporting pursuant to s. 330.287(8)(c)(2) -- slightly lower than the 0.9 percent in 2007-08. In 0.9 percent of the cases, the private school closed before reporting its scores, as compared to 0.2 percent in 2007-08. Taken together, the

percentage of students in 2008-09 with either legible, valid score reporting or one of these other explanations was 96.9 percent, highly comparable to the 96.5 percent for 2007-08.

In the remaining cases, the private school either reported the student was absent (1.9 percent, as compared with 1.0 percent in 2007-08) or had some problem with test reporting (1.3 percent, as compared with 2.6 percent.) This last category 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 next table reports the distribution of tests taken by participating students. Of the students who have taken tests that were reported to the Independent Research Organization, virtually 100 percent (all but three students) took a test approved by the Florida Department of Education. The vast majority of the students (68.8 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 22.2 percent took the Iowa Test of Basic Skills and 4.0 percent took the Terra Nova test. The other students took a number of other tests, most notably the Basic Achievement Skills Inventory, taken by 1.5 percent of students.



	Stanford Achievement Test/ FCAT-NRT	Iowa Test of Basic Skills	Terra Nova	Basic Achievement Skills Inventory	PSAT/ NMSQT	ACT/ PLAN	Metropolitan Achievement Test	Other approved test	Other unapproved test
2006-07	70.5	22.7	3.3	0.6	1.1	0.7	0.7	0.1	0.4
2007-08	70.9	19.9	3.8	2	0.8	0.8	0.4	1.2	0.2
2008-09	68.8	22.2	4	1.5	0.9	0.8	0.4	1.2	0.01

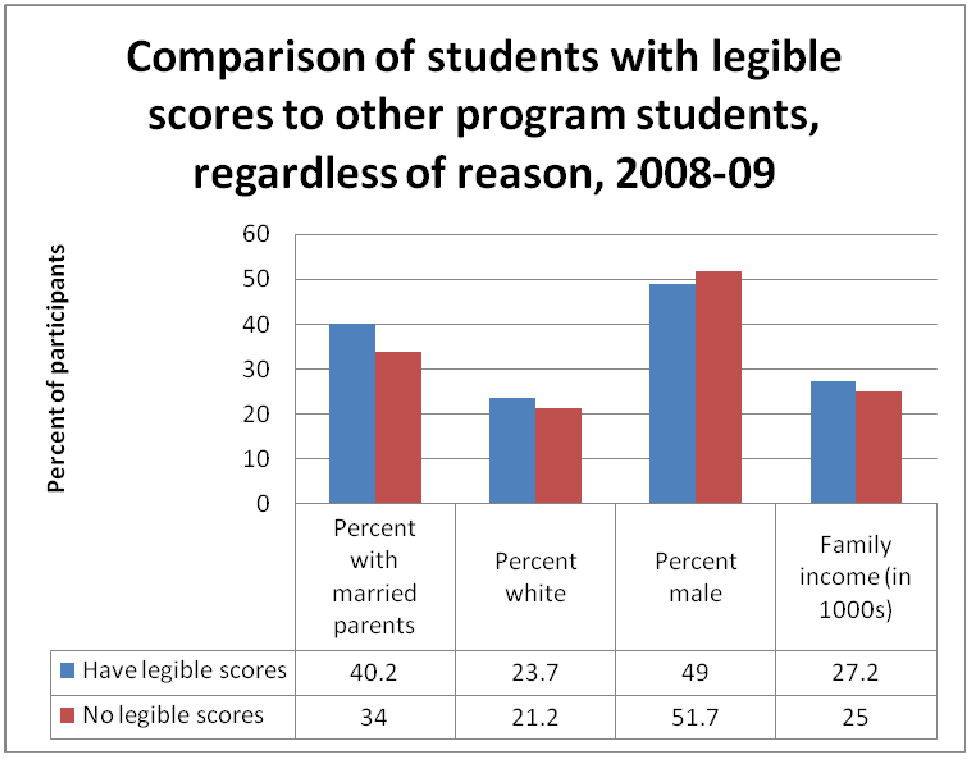
Schools have flexibility as to when they administer their exams, and 21.2 percent of participating students took their exam in the fall months. These scores are less likely to be directly comparable to public school students' tests than are t[2]hose taken during the time immediately surrounding the public 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

[2] The increase from 19.0 percent in 2007-08 to 21.2 percent in 2008-09 in the percentage of students taking tests in the fall months appears to be a major determinant of the modest increase in the fraction of program participants who arrived in their private school after testing took place.

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.

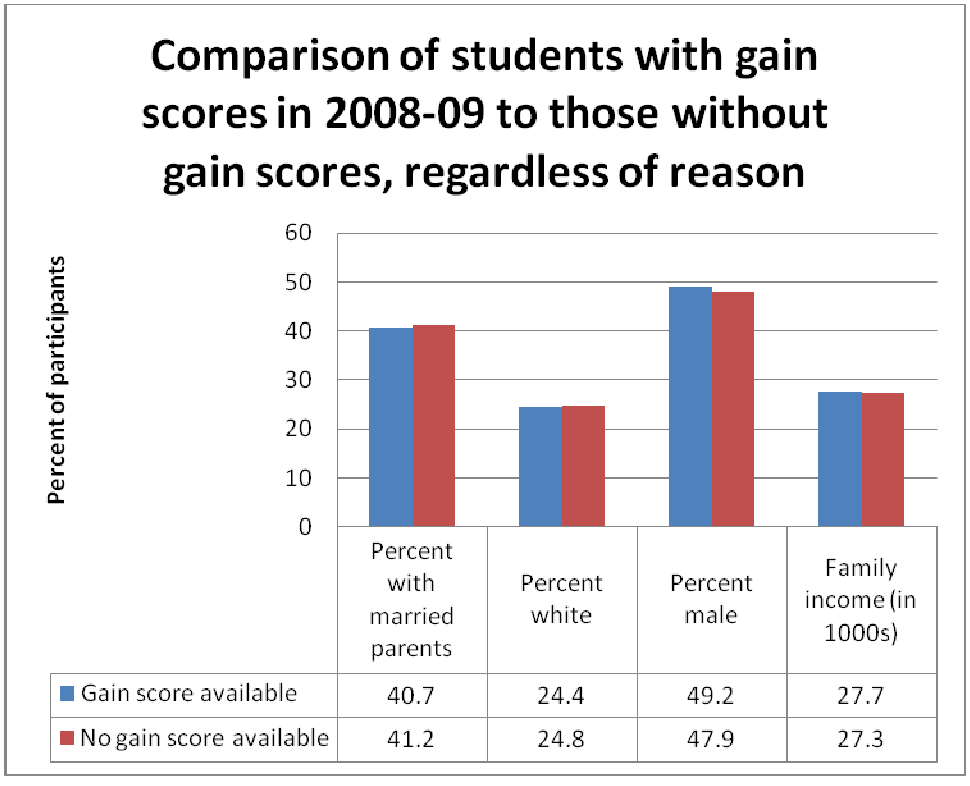
Similarity of students with received legible tests to the overall scholarship population

In 2008-09, the rate of successful test reporting remained at the high levels of 2007-08. However, around ten 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 2008-09.



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 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 there exists strong evidence from national datasets such as the National Education Longitudinal Study that less advantaged families tend to be more transient, and are therefore perhaps more likely to have missed testing because they changed schools. That said, these differences 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.



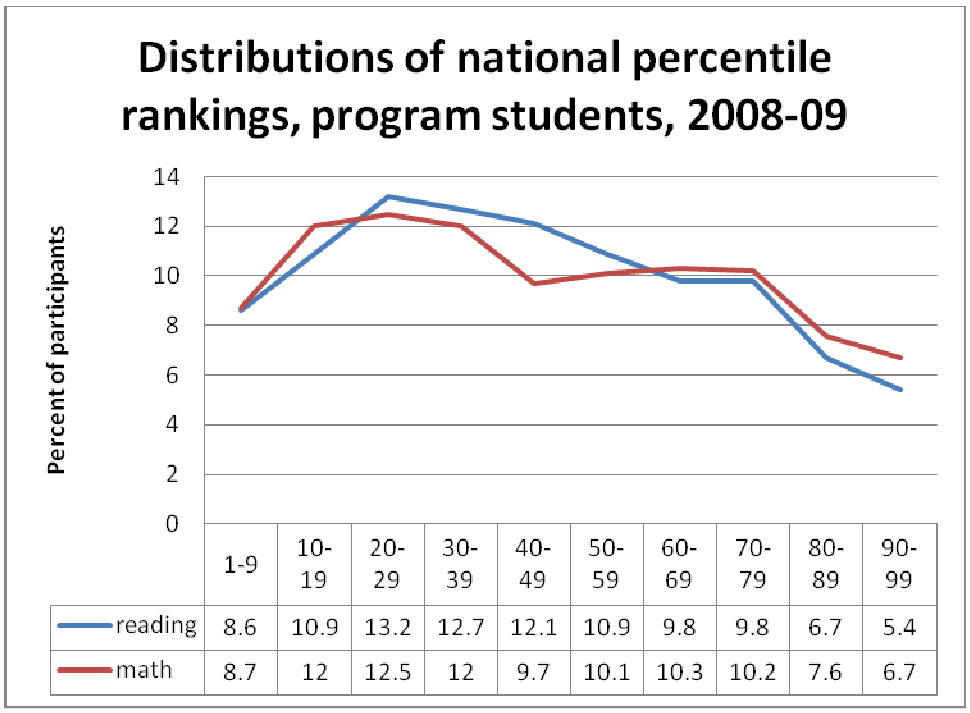
	Percent with married parents	Percent white	Percent male	Family income (in 1000s)
Gain score available	40.7	24.4	49.2	27.7
No gain score available	41.2	24.8	47.9	27.3

As can be seen from the table, the set of students with gain scores from 2007-08 to 2008-09 looks observationally equivalent to the set of students enrolled in the program

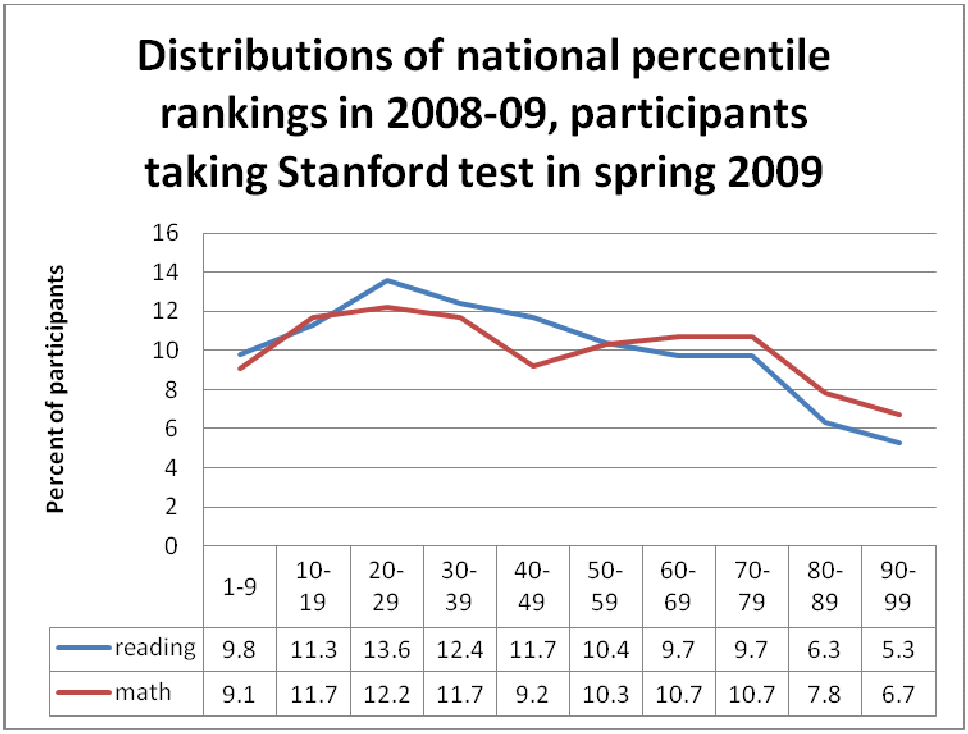
for both years but for whom one or both years of testing are missing. Across all four representative dimensions, the mean attributes of students with gain scores and those without gain scores are highly similar. These results indicate that even though we are missing test scores from a number of students who could have potentially provided test scores, an analysis of gain scores is likely to be highly representative of the overall population of program participants for whom gain scores could possibly have been collected.

III. Test scores of 2008-09 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.

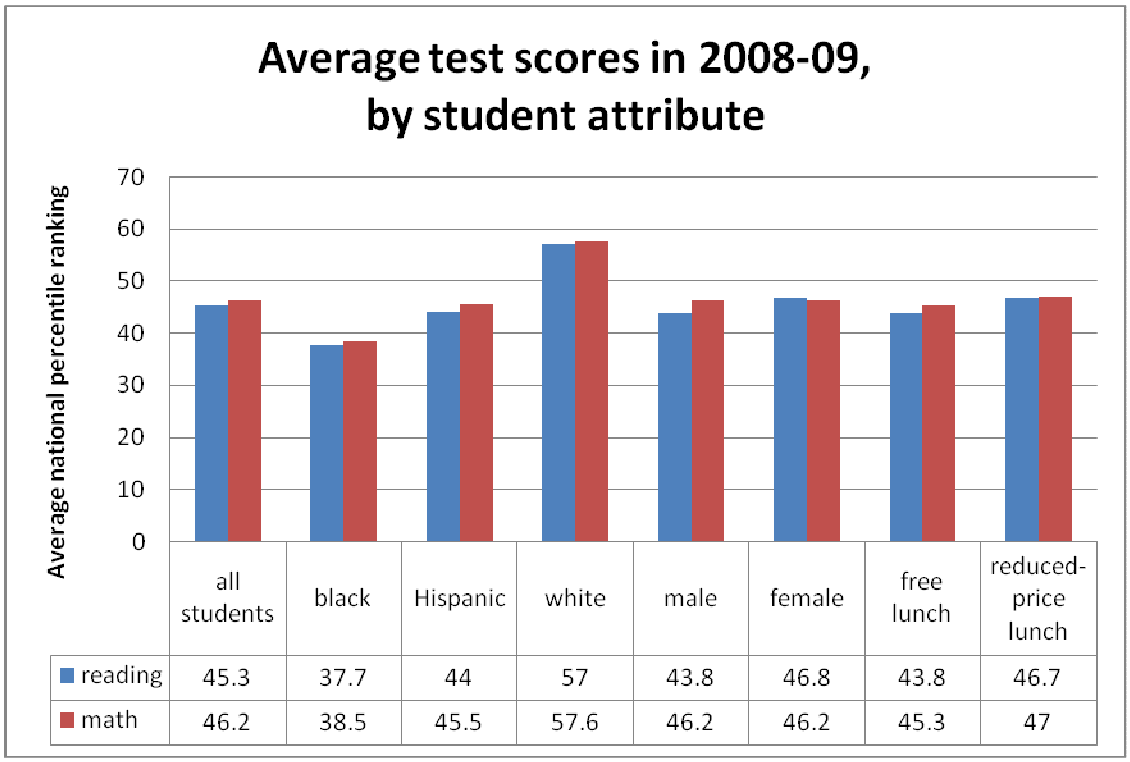


The chart above presents the basic distribution of national percentile rankings among FTC Scholarship students participating in the program in 2008-09. The typical student in the program scored at the 45.3rd percentile in reading and the 46.2nd percentile in mathematics. This is basically unchanged from 2007-08, in which the typical student in the program scored at the 44.8th percentile in reading and the 46.3rd percentile in mathematics. 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 44.4th percentile in reading and the 46.6th percentile in mathematics. Given that the distributions of test scores are so similar for those taking the Stanford Achievement Test in the spring versus the full set of scholarship recipients, this report will focus on the full set of students for whom data are available, regardless of test administered.



	1-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
reading	9.8	11.3	13.6	12.4	11.7	10.4	9.7	9.7	6.3	5.3
math	9.1	11.7	12.2	11.7	9.2	10.3	10.7	10.7	7.8	6.7

The next chart 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 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.

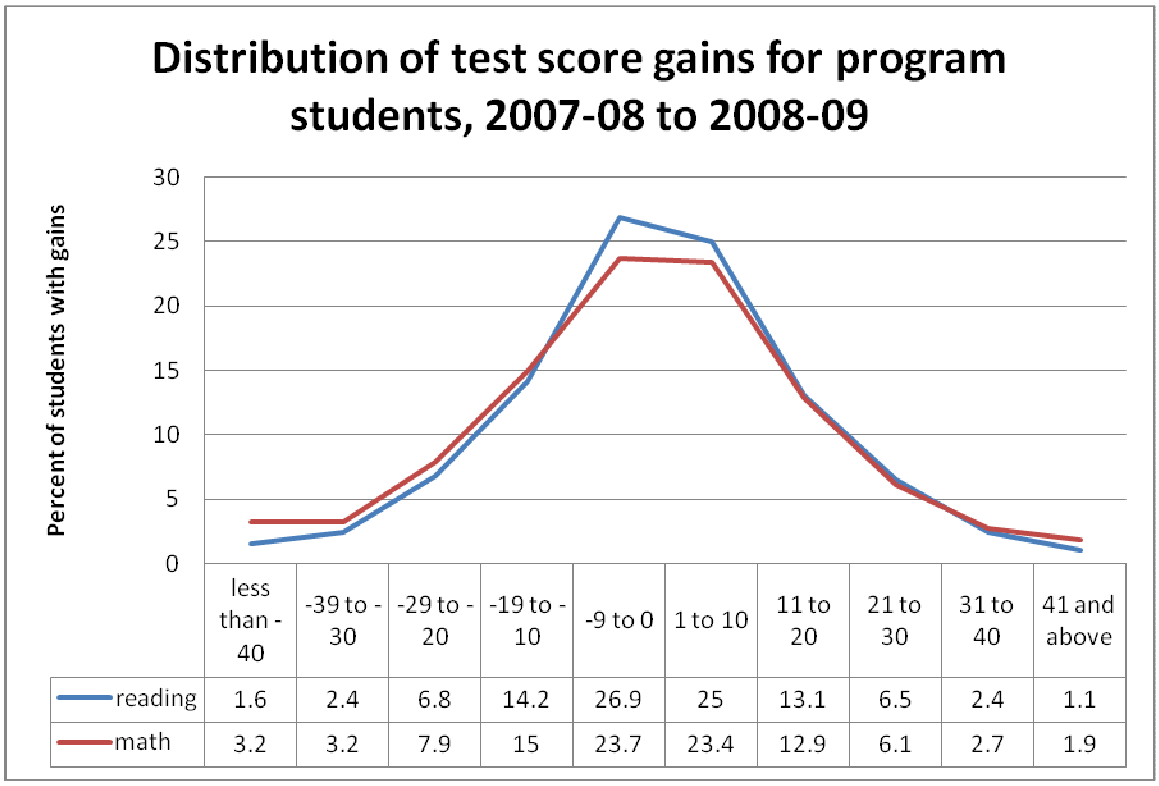


	all students	black	Hispanic	white	male	female	free lunch	reduced-price lunch
reading	45.3	37.7	44	57	43.8	46.8	43.8	46.7
math	46.2	38.5	45.5	57.6	46.2	46.2	45.3	47

Test score gains for FTC Scholarship program participants

While any such analysis is complicated by the fact that the 2006-07 test score collection was conducted during the 2007-08 academic year and therefore the 2007-08 round of test score collection is the first over which there is sufficient control to guarantee a reasonably complete score analysis, it is nonetheless possible to evaluate the distribution of test score gains in the FTC Scholarship Program for the students who participated in both 2006-07 and 2007-08. Because the test scores in both 2006-07 and 2007-08 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.



	less than -40	-39 to -30	-29 to -20	-19 to -10	-9 to 0	1 to 10	11 to 20	21 to 30	31 to 40	41 and above
reading	1.6	2.4	6.8	14.2	26.9	25	13.1	6.5	2.4	1.1
math	3.2	3.2	7.9	15	23.7	23.4	12.9	6.1	2.7	1.9

The chart above presents information on the distribution of program participants' test score gains in reading and mathematics for the set of 5,700 students with legible reading scores and 5,738 students with legible mathematics scores in both 2007-08 and 2008-09. The mean gain for program participants is -0.1 national percentile ranking points in reading and -1.1 national percentile ranking points in mathematics, virtually the same gain distribution as was reported in the 2007-08 report that was based on incomplete score reporting from the baseline 2006-07 year. In other words, the typical student participating in the program tended to maintain his or her relative position in

comparison with others nationwide. 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.

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 2007-08 and 2008-09 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.[3] 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

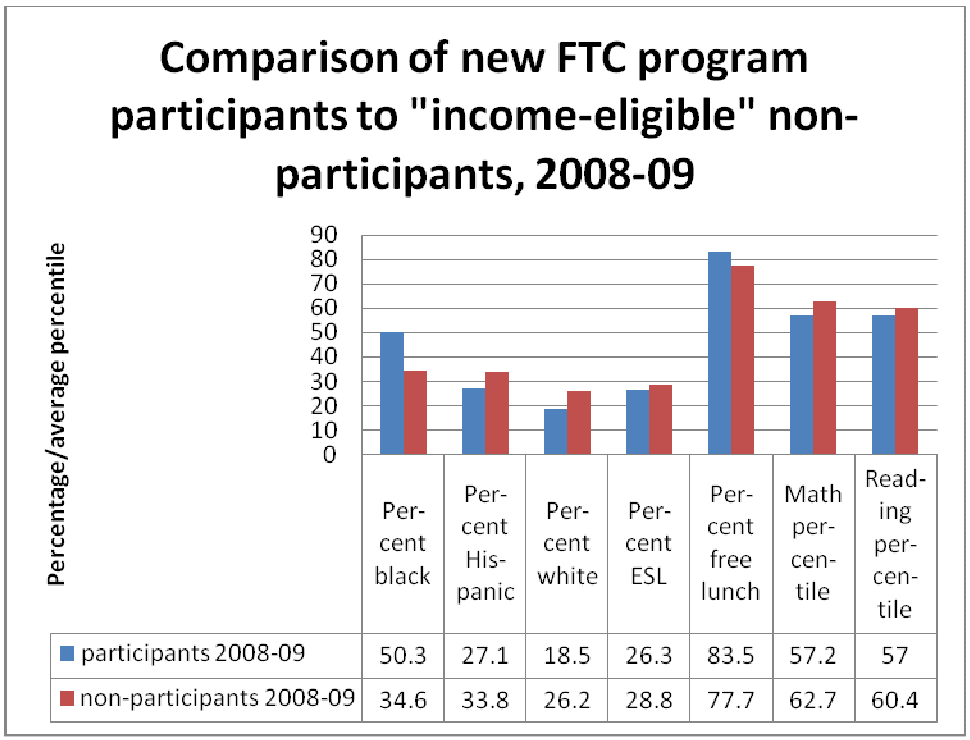
[3] A technical description of selection into the FTC Scholarship Program is provided in David Figlio, Cassandra Hart, and Molly Metzger, "Who Uses a Means-Tested Scholarship, and What Do They Choose?" published in the *Economics of Education Review* in 2009. A brief summary of the key points of that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

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 their family incomes are likely to be considerably different. 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. This report presents these basic comparisons of student test score gains in the public and private sectors, and then presents the results of more sophisticated empirical methods aimed at more compellingly deducing the causal effect of participating in the FTC Scholarship Program.

Summary of key selection findings

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 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 2007-08 academic year in the Florida public schools, so one can compare the students who entered the FTC Scholarship Program in 2008-09 versus potentially comparable students who did not enter the program in that year but remained free or reduced-price lunch eligible in 2008-09, 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 2007-08; prior research suggests that this is very similar to the overall

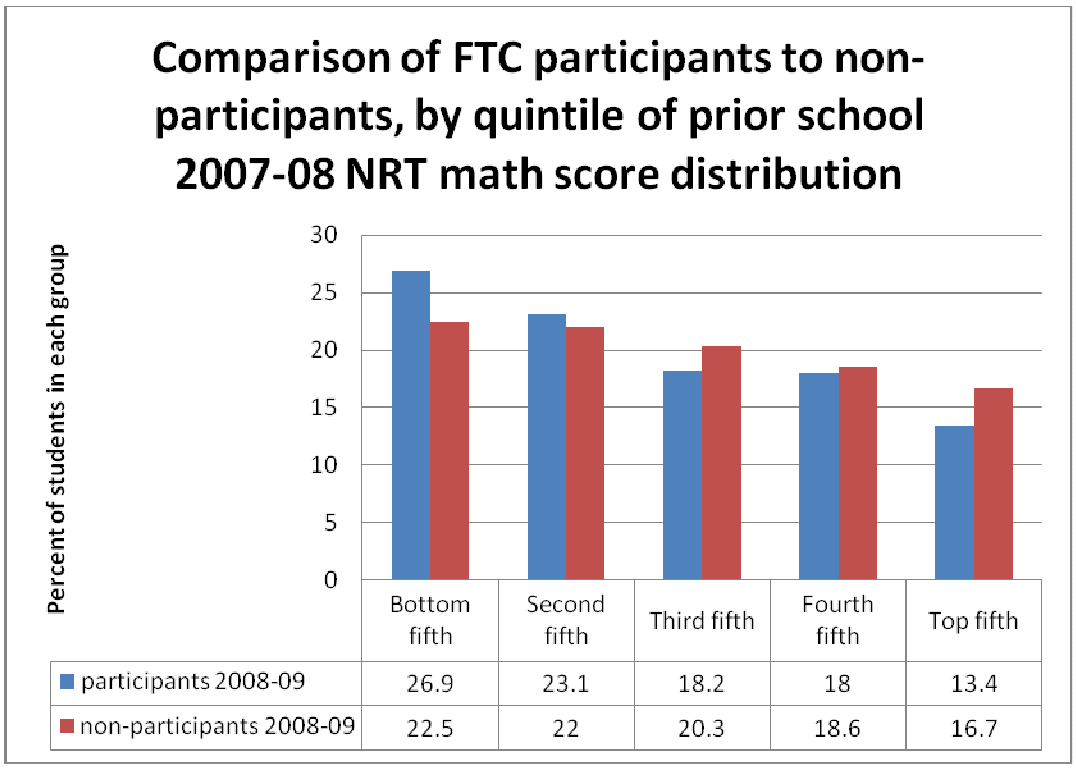
population of potential program participants who spent the prior year in a public school. We also limit the analysis to students who will be in grade 10 or below in 2008-09, so that this reflects the set of students for whom a test score is possible. By these standards, there were 1,629 new students in the FTC Scholarship program from this sample and 563,829 students who remained in the public schools and continued on subsidized school lunches in 2008-09.

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 2007-2008 school year, participants were more likely to qualify for free lunch as of the last survey taken, 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 Stanford mathematics and the Stanford reading tests, 2008-09 non-participants out-performed 2008-09 scholarship participants in the 2007-08 school year, when both groups were still attending public schools. 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.

The mean differences in 2007-08 performance between public school students who would ultimately participate in the FTC Scholarship Program in 2008-09 and those who are plausibly income-eligible but who remained in Florida public schools in 2008-09 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 selecting into the program. It is clear that these two possibilities have very different implications for the interpretation of differential selection into the program.

It is certainly the case that FTC Scholarship Program participants come disproportionately from lower-performing schools. For instance, amongst the elementary school students new to the program in 2008-09, 40.1 percent came from schools graded "A" by the Florida Department of Education in 2008, as compared with 51.1 percent of those public school students eligible for free or reduced-priced lunches who did not participate. At the other extreme, 7.2 percent came from schools graded "D" or "F" by the Florida Department of Education in 2008, as compared with 5.1 percent of those public school students eligible for free or reduced-priced lunches, and 36.8 percent came from schools graded "C" or below by the Florida Department of Education in 2008, as compared with 27.8 percent of those public school students eligible for free or reduced-priced lunches.



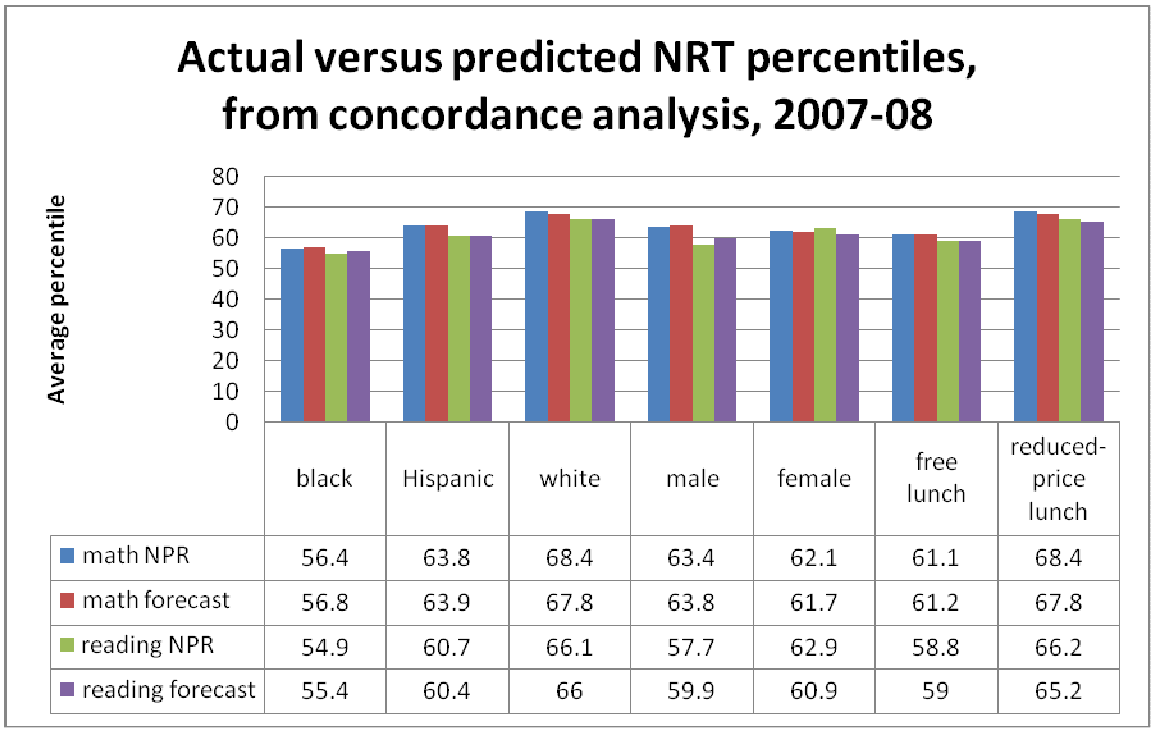
It is also the case that, regardless of the performance level of the public school that FTC Scholarship Program participants came from, these students tended to be lower-performing before they entered the program. As can be seen in the accompanying figure, 26.9 percent of students who would select into the program were in the bottom fifth of their prior public school's mathematics test score distribution, while only 22.5 percent of free- or reduced-price lunch students were in the bottom fifth of the distribution in the prior public school. (Similar differences are present in terms of reading scores.) At the top of the test score distribution, only 13.4 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 16.7 percent of free- or reduced-price lunch students in the top fifth of the distribution in the prior public school. Clearly, program participants are being drawn from lower-performing schools, and from relatively lower-performing students in their schools.

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

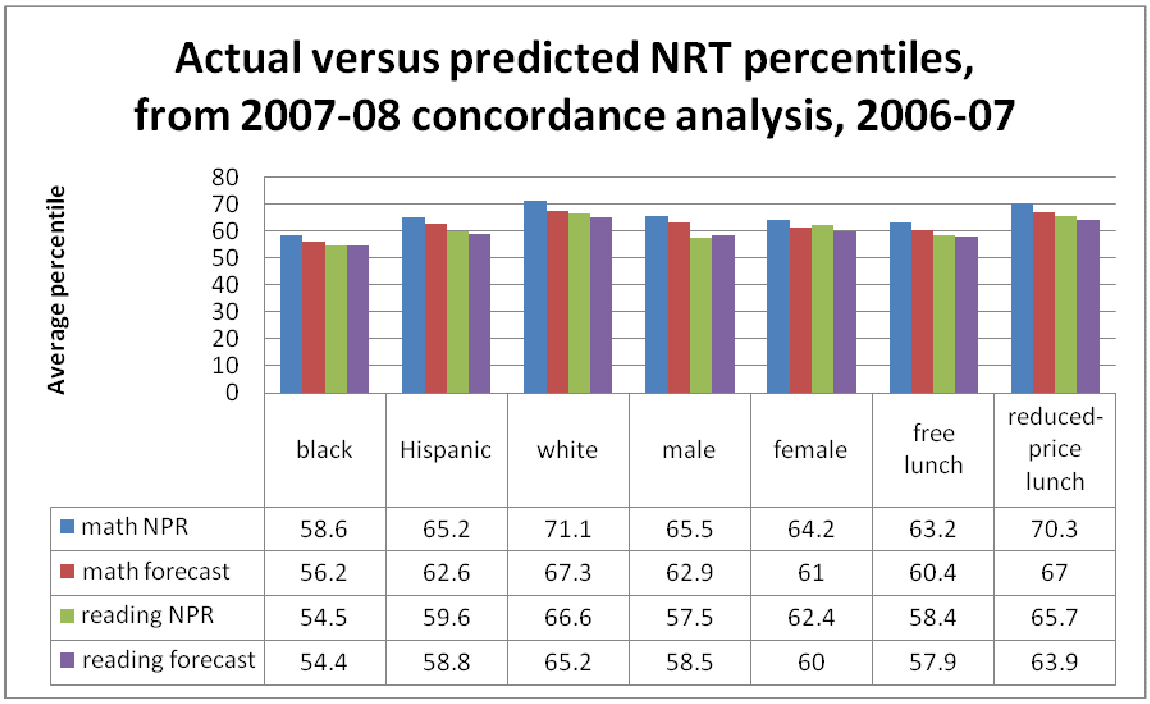
NRT score to accompany a given FCAT developmental scale score for a given grade level. Because students from different groups might have different concordances between the two tests, the predictions were made using the set of students who were eligible for subsidized school means in both 2007-08 and 2008-09. The results of this concordance analysis are highly robust to other population definitions.



	black	Hispanic	white	male	female	free lunch	reduced-price lunch
math NPR	56.4	63.8	68.4	63.4	62.1	61.1	68.4
math forecast	56.8	63.9	67.8	63.8	61.7	61.2	67.8
reading NPR	54.9	60.7	66.1	57.7	62.9	58.8	66.2
reading forecast	55.4	60.4	66	59.9	60.9	59	65.2

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.



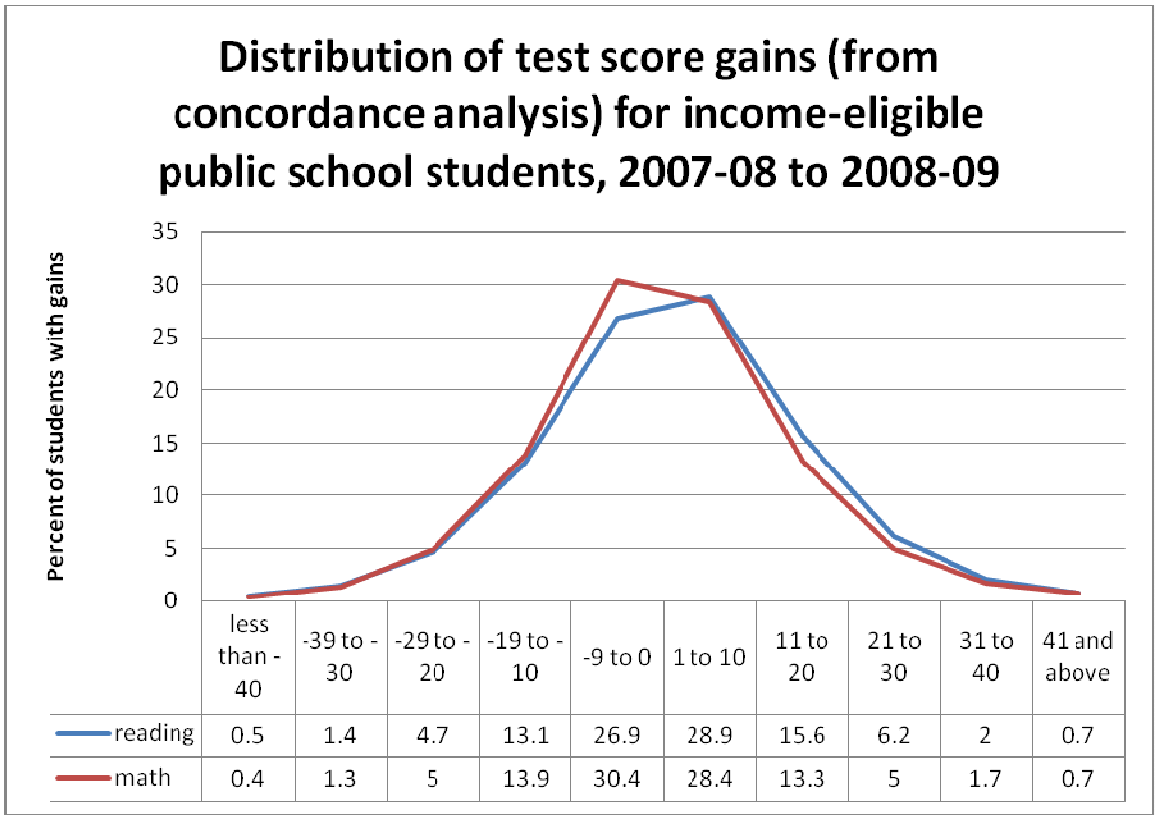
	black	Hispanic	white	male	female	free lunch	reduced-price lunch
math NPR	58.6	65.2	71.1	65.5	64.2	63.2	70.3
math forecast	56.2	62.6	67.3	62.9	61	60.4	67
reading NPR	54.5	59.6	66.6	57.5	62.4	58.4	65.7
reading forecast	54.4	58.8	65.2	58.5	60	57.9	63.9

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.

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 2007-08 and 2008-09. This report employs the concordance metrics described above to compute predicted NRT scores in 2007-08 and 2008-09 based on the student's actual FCAT scores in the two years. Comparing predicted scores is preferable to comparing actual NRT scores in 2007-08 to predicted NRT scores in 2008-09 because if one predicts scores in both years, the tests on which the scores are based remain the same. In practice, however, which comparison is made makes no difference: The mean public school test score gain based on forecasted NRT scores based on FCAT scores from both 2007-08 and 2008-09 is +1.1 percentile points in math and +2.1 percentile points in reading, while the mean public school test score gain based on actual NRT score taken in 2007-08 and forecast NRT score based on the FCAT score in 2008-09 is +1.1 percentile points in math and +2.0 percentile points in reading. Given that these means are basically identical, we

will focus on the more direct comparison of forecasts based on the FCAT in both 2007-08 and 2008-09 for the public school portion of this analysis.



	less than -40	-39 to -30	-29 to -20	-19 to -10	-9 to 0	1 to 10	11 to 20	21 to 30	31 to 40	41 and above
reading	0.5	1.4	4.7	13.1	26.9	28.9	15.6	6.2	2	0.7
math	0.4	1.3	5	13.9	30.4	28.4	13.3	5	1.7	0.7

As can be seen in the above graph, 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 2.2 percentile points higher than the mean gain amongst program participants in both reading and 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 schools, 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 some modest evidence that public school gains are

mildly larger than private school gains. We turn next to a more causal analysis to gauge the degree to which these differences in test score gains can be attributable to program participation.

V. Causal estimates of the effects of program participation on student test score gains using regression discontinuity models

As mentioned above, families choose to participate in the FTC Scholarship Program for a wide variety of reasons, and selection into the program is definitely not random. Indeed, there is strong evidence that those who participate in the program are substantially more disadvantaged and lower-achieving than are those who are likely income-eligible but do not participate in the program. This is not just a one-time phenomenon; the same selection factors were present in the analysis of the 2007-08 data for program participation.

The purest way to gain estimates of the causal effect of program participation on the scores of the participants is to conduct an experiment, in which people apply for scholarships and are randomly selected for participation in the program via lottery. Comparisons between program applicants that win the lottery and those that lose the lottery can then be interpreted as causal estimates of the effects of program participant on student outcomes. Such an experiment has high *internal validity* -- it can be clearly interpreted as causal -- but it may not have high *external validity* -- as the people who apply for a scholarship may not be representative of the overall candidate population. That said, at the least, this type of analysis would provide causal estimates of the effects of program participation for the set of people who wanted to participate in the program -- arguably still an important population.

Of course, participation in the FTC Scholarship Program is not governed by lotteries, and therefore an experimental evaluation of the consequences of participation is not possible. However, it is possible to evaluate the program using *quasi-experimental* statistical methods that emulate experimental conditions. This section of the report provides the best available attempt to use quasi-experimental methods to estimate the causal consequences of program participation. Specifically, we use a technique called *regression discontinuity design* to measure the effects of program participation.

Regression discontinuity methods are most useful when program participation is based on strict programmatic rules, where two very similar individuals who would be virtually identical *but for* where they stack up along the dimension where selection takes place end up receiving very different treatment. The FTC Scholarship Program is a perfect example of this type of situation: In order to participate, families must have incomes not greater than 185 percent of the poverty line. It is unlikely that an individual with family income of 186 percent of the poverty line is really any different than an individual with a family income of 185 percent of the poverty line, so if it is possible to directly compare these individuals we might be able to get stronger purchase on the causal question at hand.

One important potential problem with this type of analysis in the present setting is that we only observe family income for individuals who *apply* for the scholarship program. But in a world with perfect information, only income-eligible families would apply for scholarships. Therefore, this analytic approach will only work in the present situation if a sufficiently large number of people are confused about their family's potential eligibility for the program. This could happen if many people believe that

partial scholarships may be available, or that the scholarship income rules are not firm. One reason why this confusion could possibly take place is that there are different income cutoffs for participation depending on whether a student is a new or returning student; since some families can receive scholarships with incomes of 200 percent of the poverty line and others must have an income of 185 percent of the poverty line or below, some families may erroneously believe that they are eligible when they are not. Families may also be confused by the fact that the federal poverty line depends on household size rather than just family income. Any analysis would have to demonstrate that there are a sizeable number of people who applied for the program but could not participate because of ineligibility.

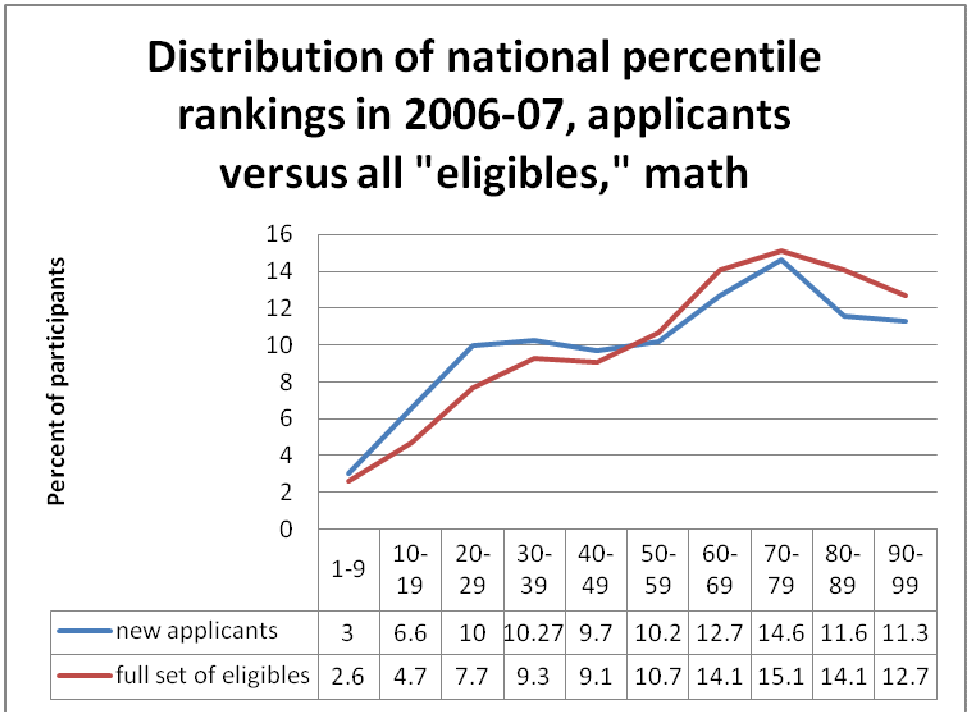
Another important potential problem with this type of analysis is that families may change their behaviors in order to qualify for the policy. In this case, a family with income that would be around 185 percent of the poverty line might choose to work less in order to qualify for the program, because the value of a scholarship to the family would generally be much higher than the lost wages associated with having an income of, say, 180 percent of the poverty line rather than 190 percent of the poverty line. If people are making these types of choices, one would observe the attributes of families just barely eligible to be quite different from families that are just barely ineligible. Therefore, any analysis would also have to gauge the degree to which this is the case.

This regression discontinuity analysis concentrates on students who spent 2006-07 in the public schools and applied for a scholarship for the 2007-08 academic year. All told, there were 4,544 students for whom this was true *and* the student took standardized tests in the public schools in 2006-07. Only a small majority of these program applicants

ultimately participated in the FTC Scholarship Program in 2007-08; 2,200 (48.4 percent) of applicants with public school test scores in 2006-07 did not participate in the program in 2007-08. Students might not participate in the program for any number of reasons, including an inability to find a good match with a school, financial constraints, and other reasons, but the proposed regression discontinuity model relies on there being a substantial number of applicants whose incomes rendered them ineligible to participate in the program. In the study population there were 330 (7.2 percent of the total) with family income above 185 percent of the poverty line. Of these, 10.6 percent had family incomes between 185 and 190 percent of the poverty line and 32.7 percent had family incomes between 185 and 200 percent of the poverty line. On the other hand. 26.4 percent had incomes over 250 percent of the poverty line, and 11.8 percent had incomes over 300 percent of the poverty line. The fact that there exists a reasonably large number of applicants above the family income cutoff implies that there may be sufficient sample size to conduct the proposed regression discontinuity analysis. Moreover, the threshold of 185 percent of the poverty line is consequential for applicants: Very few applicants with family incomes over 185 percent of the poverty line ultimately participate in the program for the first time in 2007-08. (We are excluding returning scholarship students from this analysis.)

On the other hand, in the matter of the external validity of the analysis, there is strong evidence to suggest that the applicants for the FTC Scholarship Program for the 2007-08 school year are not representative of the overall population of potential participants. The figure below presents the distributions of the public school mathematics NRT national percentile rankings in 2006-07 of new applicants to the

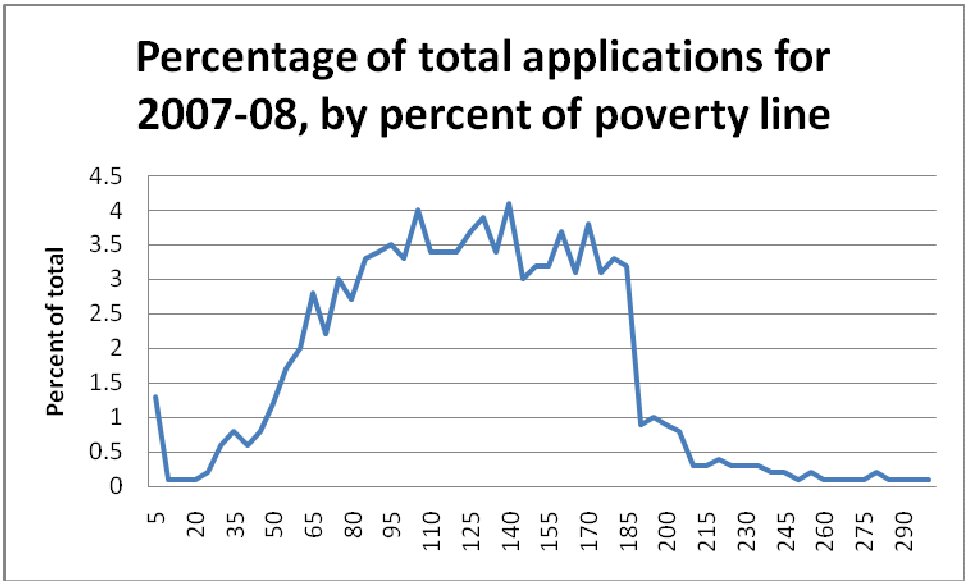
program for 2007-08 and the full set of potential program eligibles. As can be seen, the set of applicants tends to be considerably lower performing than the set of potentially eligible students. (The differences for reading are even more pronounced.) This result is unsurprising, given the previously-reported results regarding differential selection into the program, with program participants being considerably lower-performing in prior years than non-participants. That said, these results suggest that this analysis is probably best thought of as the estimated effects of program participation for the types of students who would apply to participate in the program. This may be exactly the right population to consider, but it is important to note that the results should not be seen as necessarily generalizable to the population of eligible students as a whole.



	1-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
new applicants	3	6.6	10	10.27	9.7	10.2	12.7	14.6	11.6	11.3
full set of eligibles	2.6	4.7	7.7	9.3	9.1	10.7	14.1	15.1	14.1	12.7

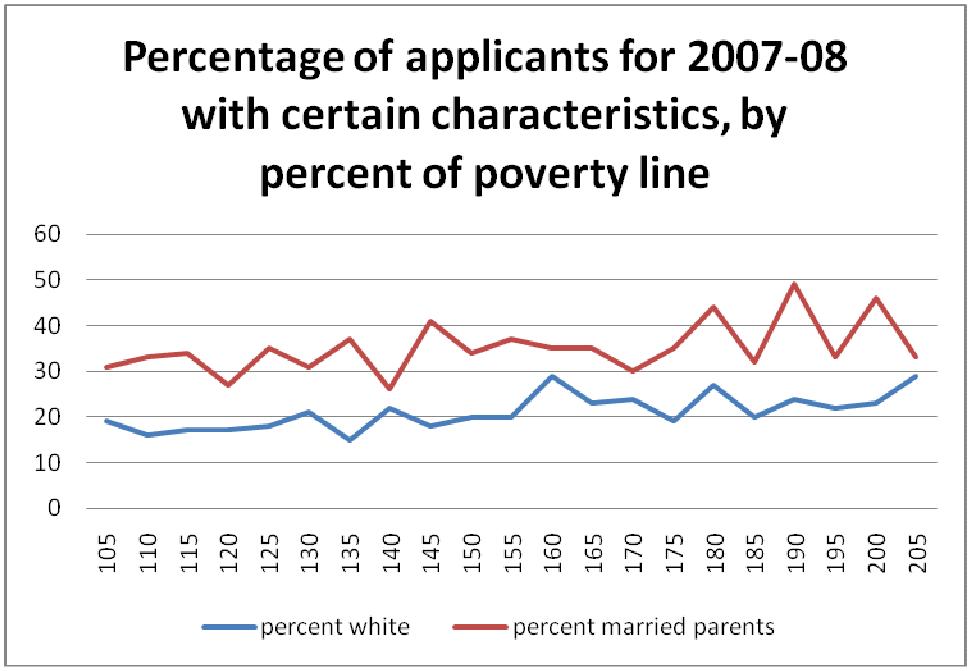
We next turn to the second potential threat to identification in the regression discontinuity model -- the potential for "bunching" of individuals just below the 185 percent of poverty threshold. If one were to observe a large number of applications with

incomes just below this threshold, it could raise concerns that individuals are changing their income-earning behaviors in order to qualify for the scholarship. We would, of course, expect a considerable dropoff in applications immediately above the 185 percent of poverty threshold because those with incomes above that level are ineligible to participate in the program, but there would not ideally be a much larger number of applicants immediately below the income threshold versus farther away from the threshold. As can be seen in the graph below, there is no evidence of bunching of applications just below the 185 percent of poverty line. There is the expected sharp dropoff in applications at 185 percent of poverty, but one notes that the decline in applications among the income-ineligible is gradual. This provides some support for a regression discontinuity model.



A regression discontinuity design could also be challenged if applicants are fundamentally different above versus below the critical value of 185 percent of the poverty line. To gauge the degree to which this is true, we plot two applicant attributes -- race and parental marital status -- against income levels on a graph. Specifically, we

investigate whether the percentage of applicants who are white or the percentage of applicants with married parents appears to differ substantially around the critical threshold (also called the "discontinuity."). We limit the analysis to those above 100 percent of the poverty line and those below 210 percent of the poverty line so that we can hone in more clearly on the area around the discontinuity. As can be seen in the following graph, there is no apparent difference along either dimension around the discontinuity, implying that at least along these dimensions there is no fundamental difference between those with incomes just above the threshold and those with incomes just below the threshold. The applicant attributes become somewhat noisier once incomes are above the threshold -- as would be expected because of the smaller sample size -- but there is no evidence that the applicants are different in any substantial way.



Given that it has been established that applicant attributes appear to be similar around the discontinuity, and that there is no bunching of incomes around the discontinuity, it is now possible to measure whether student test score gains are different

on either side of the discontinuity. The most basic regression discontinuity model involves estimating a linear regression in which the dependent variable is the student's test score gain and there are two key explanatory variables -- the student's family income as a percentage of poverty and an indicator for whether the student's family is income-eligible for the FTC Scholarship Program (i.e., the income is 185 percent of the poverty line or below.) Other models described below also include student-level control variables and more complicated specifications of the relationship between family income and test score gains. The regression discontinuity model does not distinguish between eligible students who used the scholarship and eligible students who did not use the scholarship; rather, in order to identify causal effects, the eligibility criterion serves as an *instrumental variable* for the actual participation decision. But as mentioned above, participation conditional on eligibility (so long as an application was made) is nearly 60 percent, so this is a strong instrumental variable for participation.

In order to be in the regression discontinuity sample, one must observe public school test scores in 2006-07 as well as a test score gain in either the public sector or the private sector between 2007-08 and 2008-09. We exclude five students with family incomes over 500 percent of the poverty line, so the sample size for the analysis is 2,332 students in mathematics and 2,325 students in reading. Of these students, 177 students (176 in reading) are ineligible, based on their application data, to participate in the program. While this is a small sample, it is adequate to detect moderate differences in performance between program eligibles and program ineligibles.

Because the primary purpose of this report is not to provide a technical treatment of the causal estimated effects of program participation, we do not provide the technical

details of the model estimation, but rather present the results of the regression discontinuity analysis. The table below presents only the key coefficient estimates, standard errors and statistical significance levels of the estimated effects of program participation on reading and mathematics scores.

Model specification	Estimated effect on participation	Estimated effect on math gains	Estimated effect on reading gains
Linear model, no controls except for family income	0.580 (0.032) [p=0.000]	0.960 (1.549) [p=0.536]	-0.895 (1.459) [p=0.540]
Linear model, controlling for 2006-07 reading and math scores	0.580 (0.033) [p=0.000]	0.856 (1.546) [p=0.580]	-0.989 (1.459) [p=0.498]
Linear model, also controlling for student race, gender, household size, and family marital status	0.404 (0.029) [p=0.000]	2.174 (1.587) [p=0.171]	-0.126 (1.496) [p=0.933]
Quadratic model, also controlling for student race, gender, household size, and family marital status	0.350 (0.031) [p=0.000]	2.611 (1.695) [p=0.124]	0.142 (1.593) [p=0.929]
Cubic model, also controlling for student race, gender, household size, and family marital status	0.290 (0.034) [p=0.000]	2.394 (1.888) [p=0.205]	0.099 (1.782) [p=0.956]

In the table above, each cell represents the estimated effect of program participation on student test score gains between 2007-08 and 2008-09 in a regression discontinuity framework. Standard errors are in parentheses beneath coefficient estimates, and statistical significance levels are in square brackets. The first row presents estimated causal effects in a model with no control variables except for family income as a percentage of the poverty line, the variable used to determine program eligibility. As can be seen, being eligible to participate, according to our calculations, conditional on

application for 2007-08 is strongly related to participation in 2007-08 -- eligible participants are 58 percentage points more likely to participate in the program than are those who appear to be ineligible. This provides strong first-stage evidence that the regression discontinuity model provides a valid instrument for program participation. The other two columns suggest small and statistically insignificant positive effects of participation on mathematics gains and small and statistically insignificant negative effects of participation on reading gains. The estimated effects are smaller than one national percentile point in absolute value, suggesting little substantive difference between the outcomes of program participants and non-participants.

The second and third rows of the table include additional control variables -- the second row includes 2006-07 national percentile rankings of reading and math NRTs taken in public schools, and the third row also includes controls for student race/ethnicity, gender, household size and parental marital status, all reported on the scholarship application. One observes that only controlling for 2006-07 test scores does nothing to the estimated effects of participation on test score gains. Further controlling for a richer set of covariates changes the estimated effect on reading gains from small, negative and insignificant to trivially positive and insignificant. The estimated effect on math gains doubles in magnitude, and while still modest, is close to statistical significance at traditional levels with its p-value of 0.124.

The fourth and fifth rows of the table present the same model specification as the third row, but with the relationship between family income as a share of poverty and test score gains modeled either as a quadratic function or a cubic function. The results are

substantively unchanged, though the statistical significance level of the estimates weakens somewhat.

The regression discontinuity model, while a substantial step forward from simple comparisons of test score gains between participants and non-participants, still could yield biased estimates. First, as mentioned above, while the concordance analysis used to provide comparable gains for public school students appears to have strong validity, the constructed NRT equivalents to FCAT scores are still just estimates, and the matches between actual and predicted NRT scores were somewhat better for reading than for mathematics. It is uncertain whether errors in the concordance analysis would bias these comparisons upward or downward. Second, if the FTC Scholarship Program is providing competitive pressure for public schools, the public school performance might be different than it would have been absent the FTC Scholarship Program. It is therefore important to interpret these results as the estimated effects of program participation for the types of students who apply to the program, and should not be seen as a more general effect of program participation.

In summary, the regression discontinuity model suggests that there is no discernible difference between FTC Scholarship Program participants and non-participants in terms of reading test score gains, and there may be modest positive effects of participation for mathematics gains. These differences, however, are still small in magnitude, are within the range of potential errors in the concordance analysis, and are not statistically significant at conventional levels, so they should be not be interpreted as strongly favorable, only potentially suggestive.

Recent research[4] indicates that the FTC Scholarship Program has led to modest but consistently statistically significant positive effects on the test score performance of students in Florida public schools. Therefore, the best interpretation of the findings of no substantial difference between FTC Scholarship Program participants and non-participants in the public schools is that students who have transferred to the private sector using a FTC Scholarship appear to be keeping pace with the gains observed in the public sector.

VI. Parental satisfaction

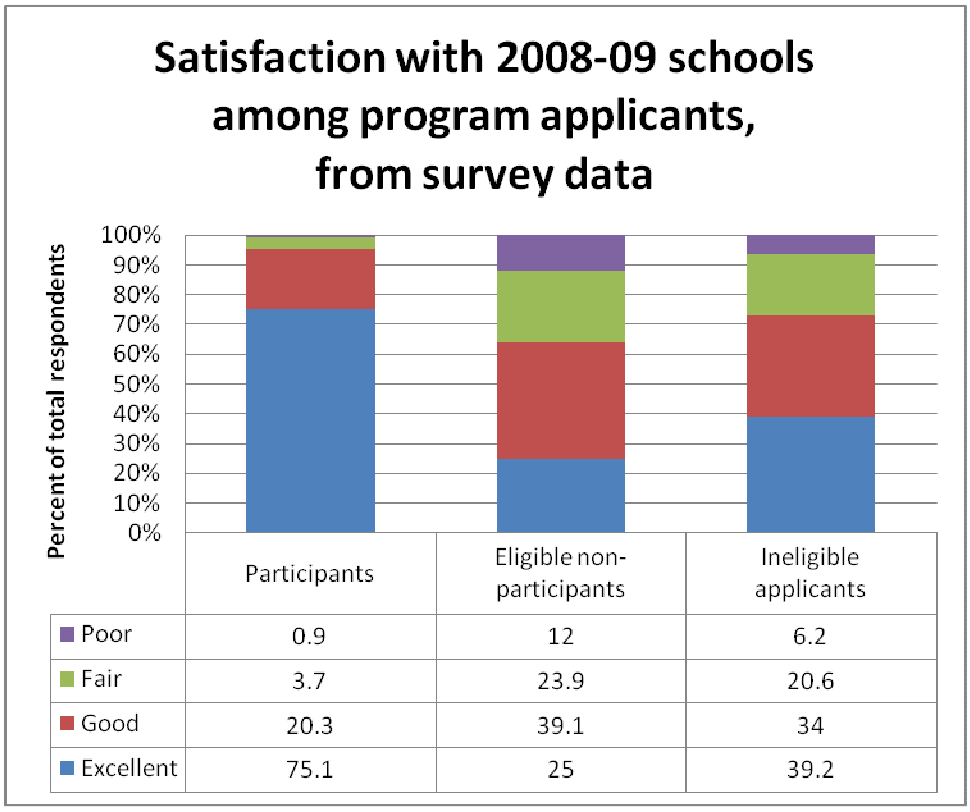
In March 2009, we conducted a survey of families who had applied to participate in the FTC Scholarship Program for the first time in the 2008-09 academic year. We randomly selected 1,994 households to receive a satisfaction survey from the full set of applications. Of these, 1,425 were for students eligible to receive a scholarship and 569 were for applicants ineligible to receive a scholarship. We asked all respondents to rate the quality of their child's school as "excellent," "good," "fair," or "poor." In addition, we asked respondents information about their race and ethnicity, degree of educational attainment, and other demographic and economic details.

We contacted the families using the home addresses on file at the time of the application, and followed up via mail and email two times each. We were limited in our ability to follow up with families because of budget constraints, but ultimately received 545 responses to our survey, a respectable 27.3 percent response rate. Our response rate is likely somewhat depressed due to the time lag between the initial scholarship

[4] David Figlio and Cassandra Hart, "Competitive Effects of Means-Tested School Vouchers," National Bureau of Economic Research working paper #16056, June 2010 (downloadable at http://www.nber.org/papers/w16056).

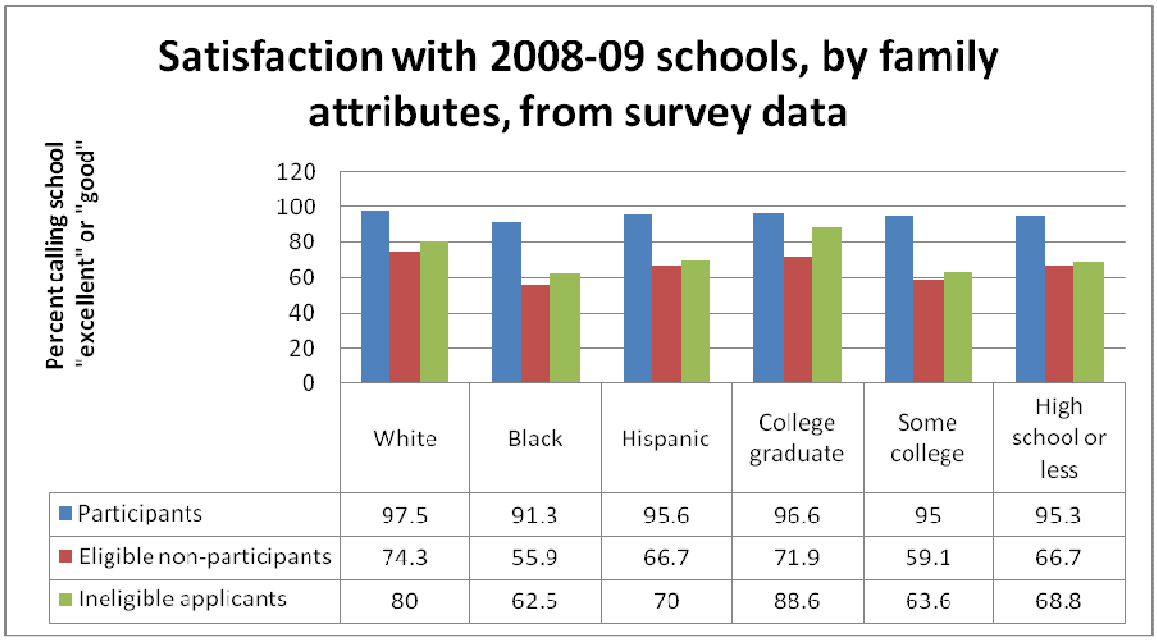
applications and the survey date, a significant factor especially in a mobile population like Florida's low-income community. The response rate for eligible students was 31.4 percent and the response rate for ineligible students was 17.2 percent. Due to the differences in response rates between eligible and ineligible families, the survey response differences should be interpreted cautiously. That said, they are still potentially informative.

We divide the set of respondents into three groups: those who were ineligible to receive a scholarship; those who were eligible to receive a scholarship and ultimately used the scholarship at a private school; and those who were eligible to receive a scholarship but did not use the scholarship. Of the 447 eligible students whose parents responded to the survey, 93 (20.8 percent) did not participate in the FTC Scholarship Program. The chart below describes the degree of parental satisfaction across these three groups. As can be seen, parents of students participating in the program are highly satisfied, relative to other families. This is not causal evidence as there are many reasons why a family would choose to participate in the program or to select a private school, but it does provide some suggestive evidence about the perceptions of parents regarding their children's schools. These results are consistent with other survey data in Florida and elsewhere suggesting that parents tend to be very happy with the choices they made in voucher and scholarship programs.



	Participants	Eligible non-participants	Ineligible applicants
Poor	0.9	12	6.2
Fair	3.7	23.9	20.6
Good	20.3	39.1	34
Excellent	75.1	25	39.2

We also investigate whether families with different backgrounds had different reactions to their children's schools in March of the 2008-09 school year. Specifically, we stratify families based on race and ethnicity (white, black and Hispanic) and based on the respondent's level of education (college graduate, some college or postsecondary education, or a high school degree or less.) In the figure below, we present the percentage of respondents of each type who rated their child's school as "excellent" or "good." As can be seen, participants across the considered dimensions rate their children's schools very highly. Minorities and less-educated households appear to be relatively unhappy with their children's schools when the children are not participating in the program. In summary, while readers should interpret these survey results cautiously, they do provide a fuller picture of the potential effects of participating in the FTC Scholarship Program for families that applied to participate.



	White	Black	Hispanic	College graduate	Some college	High school or less
Participants	97.5	91.3	95.6	96.6	95	95.3
Eligible non-participants	74.3	55.9	66.7	71.9	59.1	66.7
Ineligible applicants	80	62.5	70	88.6	63.6	68.8

VII. 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 2008-09, 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, though the raw gains are modestly smaller amongst scholarship participants than for non-participants. 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 low-income

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.

With this in mind, this report makes use of regression discontinuity models to estimate the causal impact of program participation. These models rely on data for those who apply for the program, so they may not be representative of the population of potentially eligible students (and there is evidence to suggest that applicants are indeed different from the overall population of free or reduced-price lunch recipients) and are best thought of as representative of the set of students who applied for the program. Nonetheless, the general pattern of small estimated effects of program participation on test score gains persists. In the regression discontinuity models, the estimated effects of program participation are modestly positive though statistically insignificant, and the results must be interpreted with considerable caution. That said, the first causal evidence regarding differential test score gains across the public and FTC Scholarship Program sectors indicates roughly comparable test score gains that are reasonably consistent and unlikely due to family income differences between participants and non-participants.