

Evaluation of Florida's Corporate Tax Credit Scholarship Program  
First Follow-Up Report – Participation, Compliance and Test Scores in 2007-08

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

June 16, 2009

## **Executive summary**

This is the second in a series of reports evaluating Florida's Corporate 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.

During the 2007-08 academic year, David Figlio, the Project Director, collected test score data from private schools participating in the CTC Scholarship Program in real time for the 2007-08 year and attempted to retroactively collect test scores from the 2006-07 year. Because the 2007-08 academic year is the first year in which data collection was fully controlled, measures of test score gains from 2006-07 to 2007-08 should only be properly interpreted as descriptive. True test score gains might have been larger or smaller. While it is inappropriate to draw conclusions about the efficacy of the program based on these incomplete data, it is still possible to compare test score gains for descriptive purposes based on the available data.

### Compliance with program testing requirements, 2007-08:

- Compliance with program testing requirements was very high. Private schools provided usable test scores for 92.7 percent of program participants in grades 3-10. Another 3.6 percent of participants were ineligible for testing or were not enrolled in the school at the time of testing. The 1.0 percent rate of illness/absence compares with the public school illness/absence rate.
- The vast majority (70.9 percent) of test-takers took the Stanford Achievement Test. Other popular tests were the Iowa Test of Basic Skills (19.9 percent) and the TerraNova (3.8 percent).
- Scholarship students whose test scores were received are modestly more advantaged than are those scholarship students whose scores were not received.

### Selection into the CTC 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, 2007-08:

- The typical student in the program scored at the 44.8th national percentile in reading and the 46.3rd 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 2008.
- The mean reading gain for program participants is -0.1 national percentile ranking points in reading and -0.9 national percentile ranking points in mathematics. In other words, the typical student participating in the program tended to maintain his or her relative position in comparison with others nationwide.
- Test score gains for program participants are similar in magnitude to comparable students in the public schools. Because the retroactive 2006-07 test score collection was imperfect, it will be necessary to wait until the collection of 2008-09 scores before one can determine whether program participants' gains are larger, smaller, or about the same as public school students' gains.

## **I. Background**

This is the second in a series of reports evaluating Florida's Corporate 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.

This report presents the results of the real-time test score collection in 2007-08. This report details key information about program participation and test scores, and evaluates the validity of the retrospectively-collected data as a baseline for analysis.

## **II. Test score collection in 2007-08**

### *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 2007-08 academic year was the first 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 2008 the Independent Research Organization contacted the 829 private schools that had participating students in grades three through ten during the 2007-08 school year. The Florida Department of Education provided the Project Director with a list of all participating students in 2007-08; of these, 10,734 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*

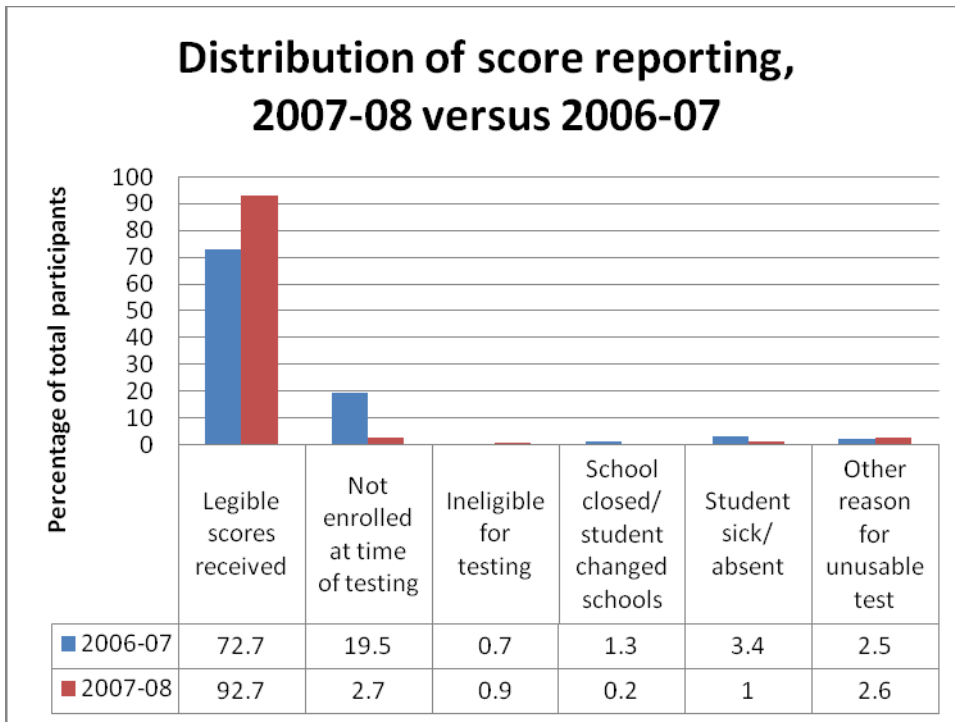
In over 99 percent of cases, schools submitted photocopies of official score sheets provided to them by the relevant testing company (e.g., Harcourt). In a small number of schools, the schools scored the tests themselves and forwarded to the Project Director detailed information regarding the nature 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 829 schools with students in the relevant grades in 2007-08, 3 schools either closed or left the program; in some of these cases the Project Director was still able to retrieve some of last year's test scores from the students' current schools. This left 826 schools that had students in the relevant grades last year and continued to participate in the program in 2008-09. At the time of writing, all but one of these schools provided some evidence of test administration, though in a small number of cases the schools administered unapproved examinations or did not provide sufficient data to compare students to national norms. In these cases, the Florida Department of Education has pursued disciplinary action against non-compliant schools.

Of the 10,734 students in relevant grades participating in the program in 2007-08, the Independent Research Organization received valid, legible test scores for 9,949 students, or 92.7 percent of all expected students; virtually all of these scores were from tests administered by the private schools themselves. This is a 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 underscores the importance of collecting test score data in real time, and test score collection for 2008-09 is already underway in real time. Therefore, the 2008-09 school year will be the first in which it will be possible to calculate student test score gains for a nearly complete set of students participating in the program in tested grades.

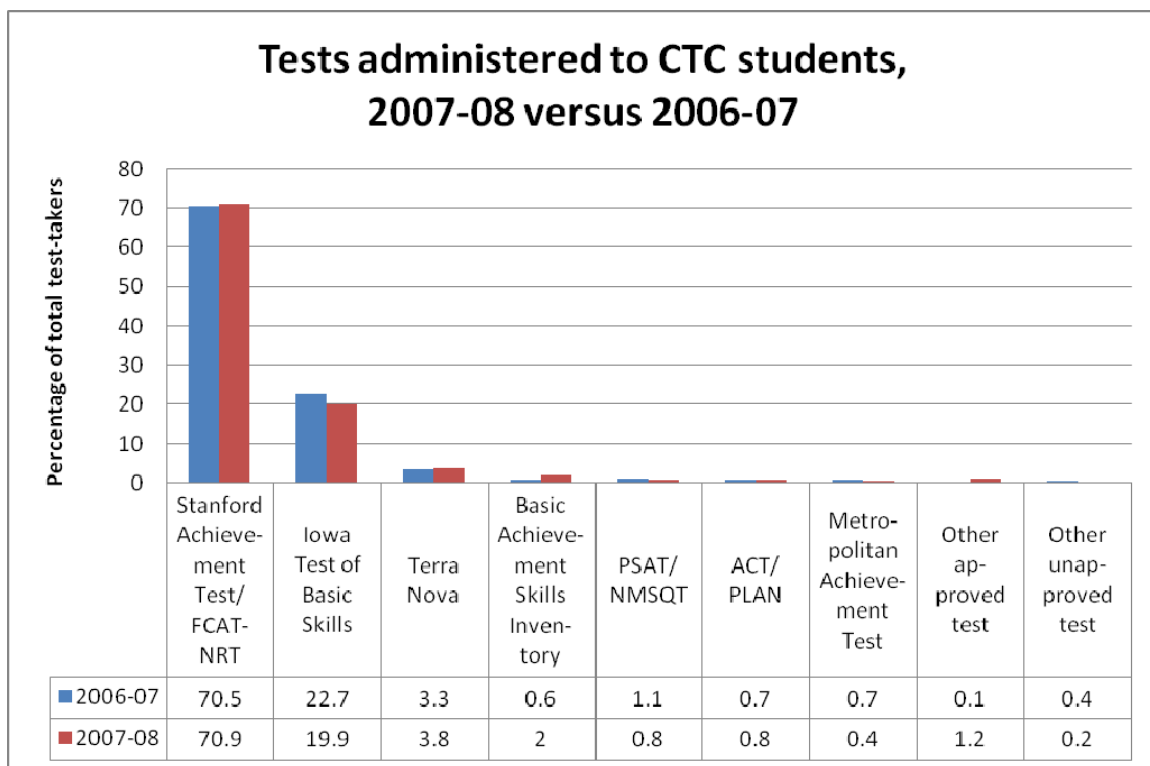
A large fraction of the students without valid, legible test scores in the 2007-08 data collection were not enrolled in the program at the time of testing; 2.7 percent of the 10,734 students potentially eligible for testing either left the program prior to test administration or arrived in the school following test administration. This is a substantial decline in this figure from the retroactive 2006-07 collection, in which 19.5 percent of students were not enrolled in the private school at the time of data collection. The primary reason for this dramatic improvement in the rate of test-taking is likely that private schools, prompted in real time about the necessity of submitting test score reports, were apparently considerably more likely to test those students arriving in the school after the beginning of the school year or those who left the school before the regular test administration period. Another 0.9 percent of students were found to be ineligible for testing pursuant to s. 330.287(8)(c)(2); this proportion is roughly similar to that which had been seen in 2006-07. This left 10,350 testing-eligible students enrolled in the program at the time of testing; these students are potentially eligible to take tests under the program.





Among the remaining students, 1.0 percent were sick or absent at the time that their school administered the test; in all of these cases, the school demonstrated that they did administer tests to other participating students at a designated time. (Note that this is less than one-third the rate of students missing tests for illness or absence as was observed in the retroactive 2006-07 data collection; schools apparently sought harder to administer makeup examinations in 2007-08.) The rate of illegible test reporting also plummeted from 2006-07 to 2007-08, from nearly 2 percent to just 0.1 percent. On the other hand, the fraction of students whose schools submitted an invalid or incomplete test increased from 0.6 percent to 1.4 percent. Several schools, accounting for 0.6 percent of students, also had their test scores destroyed in serious weather. More detail on test administration and score reporting can be found in Table II.1. This leaves 9,949 as the number of legible test scores received by the Independent Research Organization for the 2007-08 school year.

Table II.2 reports the distribution of tests taken by participating students. Of the students who have taken tests that were reported to the Independent Research Organization, 99.8 percent took a test approved by the Florida Department of Education.<sup>1</sup> The vast majority of the students (70.8 percent) took the Stanford Achievement Test, the nationally norm-referenced test administered to all public school students in Florida in 2007-08, while another 19.9 percent took the Iowa Test of Basic Skills and 3.8 percent took the Terra Nova test. The other students took a number of other tests, most notably the Basic Achievement Skills Inventory, taken by 2.0 percent of students.



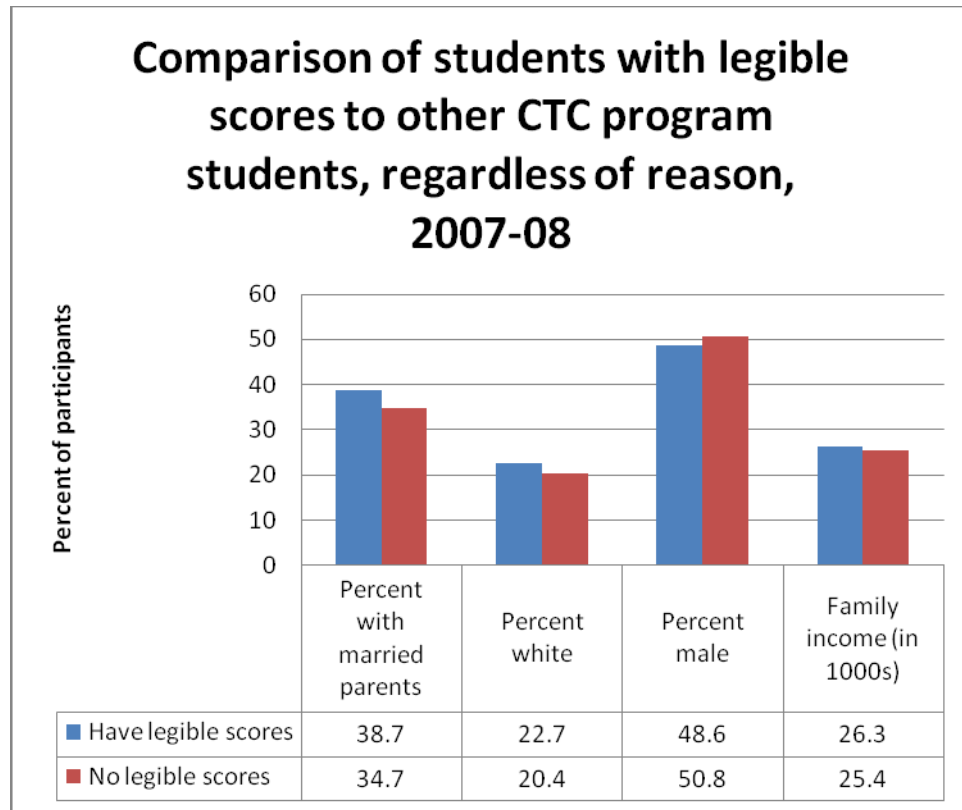
Schools have flexibility as to when they administer their exams, and 19 percent of participating students took their exam in the fall months. These scores are less likely to

<sup>1</sup> The Florida Department of Education has contacted the schools that administered an invalid test in 2007-08 and informed them of their testing requirements for 2008-09.

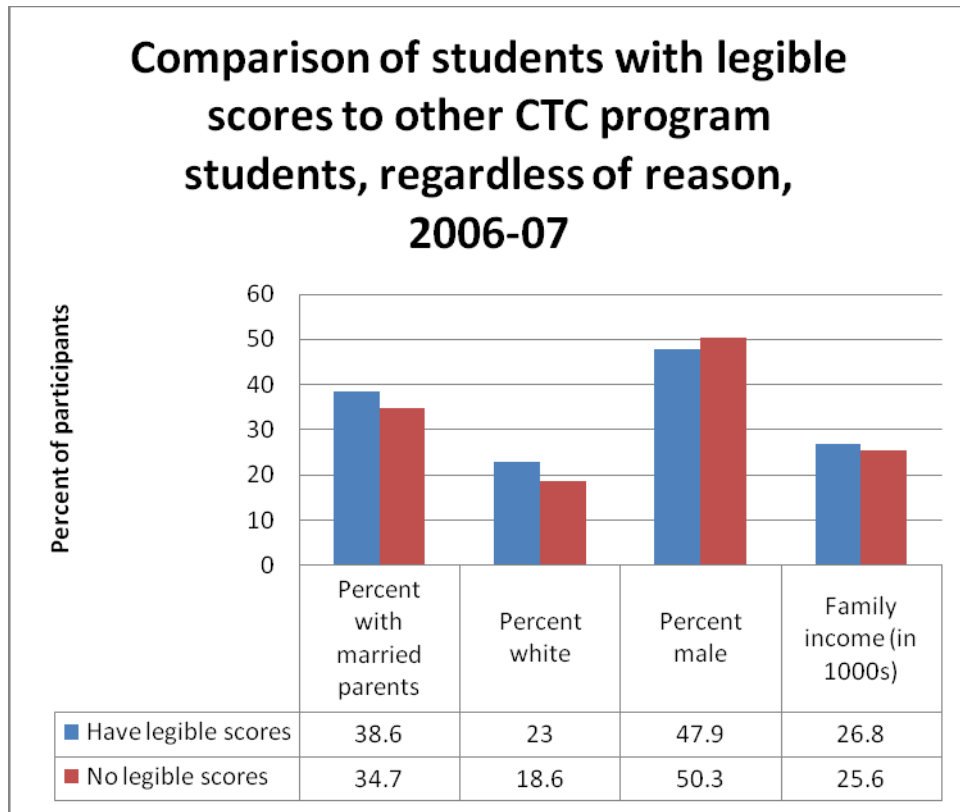
be directly comparable to public school students' tests than are those taken during the time immediately surrounding the public schools' test administration. The tests most typically taken in the fall months are the PSAT/NMSQT and the Iowa Test of Basic Skills. The latter case is driven strongly by Florida 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 2007-08, the first year of real-time test score collection in the program, the rate of successful test reporting was extremely high, and substantially higher than the retrospectively-collected 2006-07 test score data. However, because legible scores are still missing for over seven percent of the potentially-tested population (due in large part to students arriving at school after testing or leaving a school before testing, to students being sick or absent during the testing period, or to there being other test reporting issues) 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 2007-08.

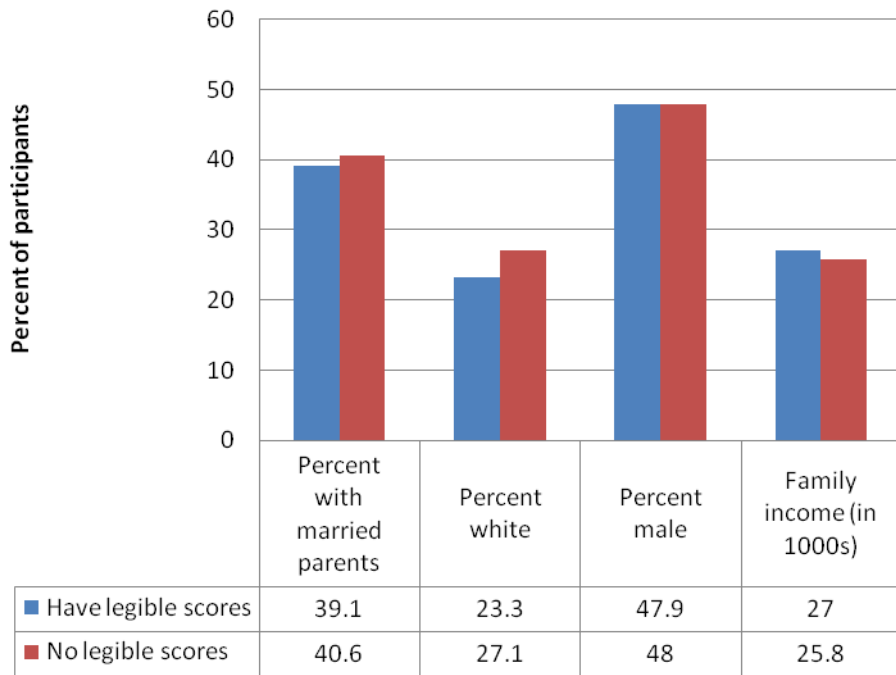


As can be seen from the accompanying figure, there is some evidence that students whose test scores were successfully reported are modestly more advantaged than other program participants whose scores were not successfully reported, based on data from the families' scholarship applications. Students 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 highly transient students are likely to be less advantaged, and are more likely to have not been tested 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.



The baseline report on 2006-07 test score collection reported a version of this analysis, but compared students with legible test scores to other program participants who were still enrolled in the questioned school during the testing period, as reported by the school. It is possible to repeat the same analysis for a fuller set of program participants. While the patterns appear similar between 2006-07 and 2007-08, upon closer inspection one observes that the differences between 2006-07 students with legible scores and other program students tend to be larger than the comparisons in 2007-08. The difference in family income, for instance, is over 30 percent higher in 2006-07 than in 2007-08, while the difference in percentage white is over 90 percent higher in 2006-07 than in 2007-08. That the differences are more pronounced in 2006-07 than in 2007-08 is unsurprising, given the dramatically larger number of untested eligible students, but this illustrates the importance of interpreting a 2006-07 baseline with considerable caution.

**Comparison of students with legible scores to other CTC program students, regardless of reason, 2007-08:  
Students with scores in 2006-07**



It is important to note that, conditional on having received legible test scores in 2006-07, students who have legible scores in 2007-08 are not necessarily more advantaged than are other program participants. While those with legible scores in 2007-08 come from modestly higher-income families, these students are also less likely to be white and are less likely to have married parents. It appears that students with legible scores in 2007-08 are likely comparable to those without legible scores; this is at least true for the set of students with legible scores in 2006-07 (the students for whom computing a gain score would be possible.) This implies that, while those students with legible scores in 2006-07 are not necessarily representative of the full set of program

participants in that year, those who continued in the program in 2007-08 with legible scores are apparently reasonably representative of returning students.

Table II.1: Test administration and score reporting, 2006-07 and 2007-08

Test score reporting status	Percent of total 2007-08	Percent of total 2006-07
<b>TOTAL NUMBER OF PARTICIPATING STUDENTS</b>	<b>10734 (100%)</b>	<b>9721 (100%)</b>
<b>NOT ENROLLED AT TIME OF TESTING</b>	<b>293 (2.73%)</b>	<b>1892 (19.46%)</b>
Student left program prior to school test administration	2.35%	15.34%
Student arrived in school after test administration	0.38%	3.19%
Student changed schools midyear between test windows	0.00%	0.94%
<b>INELIGIBLE FOR TESTING</b>	<b>91 (0.85%)</b>	<b>65 (0.67%)</b>
Student not in grades 3-10	0.54%	0.26%
Student certified to be disabled	0.31%	0.40%
<b>TOTAL NUMBER OF ENROLLED STUDENTS ELIGIBLE FOR TESTING</b>	<b>10350 (96.42%)</b>	<b>7764 (79.87%)</b>
<b>SCHOOL CLOSED / STUDENT CHANGED SCHOOL</b>	<b>16 (0.15%)</b>	<b>124 (1.28%)</b>
School closed or left program, no tests received	0.15%	0.88%
Student left program, but timing of test unknown	0.00%	0.08%
Student changed schools, no tests received	0.00%	0.32%
<b>USABLE TEST SCORE NOT RECEIVED</b>	<b>385 (3.59%)</b>	<b>573 (5.89%)</b>
Student certified to be sick/absent during testing period	0.98%	3.38%
Parent refused to test student	0.07%	Unknown
Test damaged in transport to scoring company	0.61%	0.00%
Incomplete test reporting (or invalid test used)	1.35%	0.55%
Test scores reported but school copy is illegible	0.14%	1.96%
Score missing, no explanation	0.45%	Unknown
<b>TOTAL NUMBER OF LEGIBLE SCORES RECEIVED</b>	<b>9949 (92.69%)</b>	<b>7067 (72.70%)</b>



Table II.2: Distribution of norm referenced tests administered to Corporate Tax Credit scholarship students

Test	Percentage of total tests 2007-08	Percentage of total tests 2006-07
Stanford Achievement Test/FCAT-NRT	70.83%	70.49%
Iowa Test of Basic Skills	19.89%	22.70%
Terra Nova	3.79%	3.25%
PSAT/NMSQT	0.82%	1.08%
ACT/PLAN	0.81%	0.74%
Metropolitan Achievement Test	0.41%	0.65%
Basic Achievement Skills Inventory	2.00%	0.62%
Other tests	1.45%	0.46%

### III. Test scores of 2007-08 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, 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.

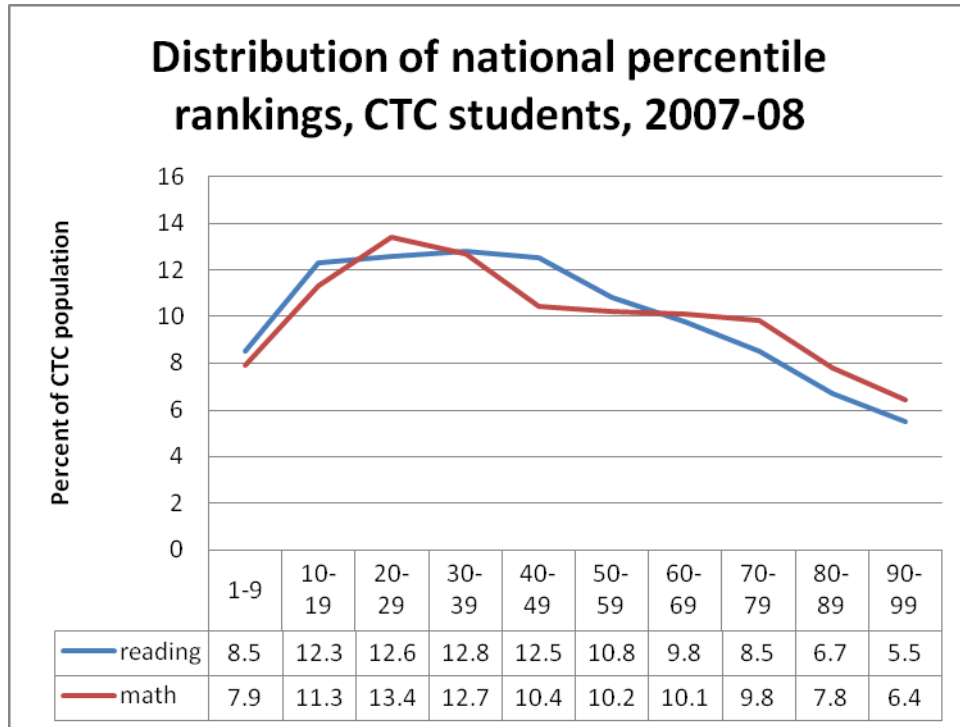


Table III.1 presents the basic distribution of national percentile rankings among CTC Scholarship students participating in the program in 2007-08, as well as those

students attending schools that administered the Stanford Achievement Test in spring 2008 -- the test administered to all public school students. It is apparent that reading and mathematics test scores are normally distributed in this population. The typical student in the program scored at the 44.8th percentile in reading and the 46.3rd 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.

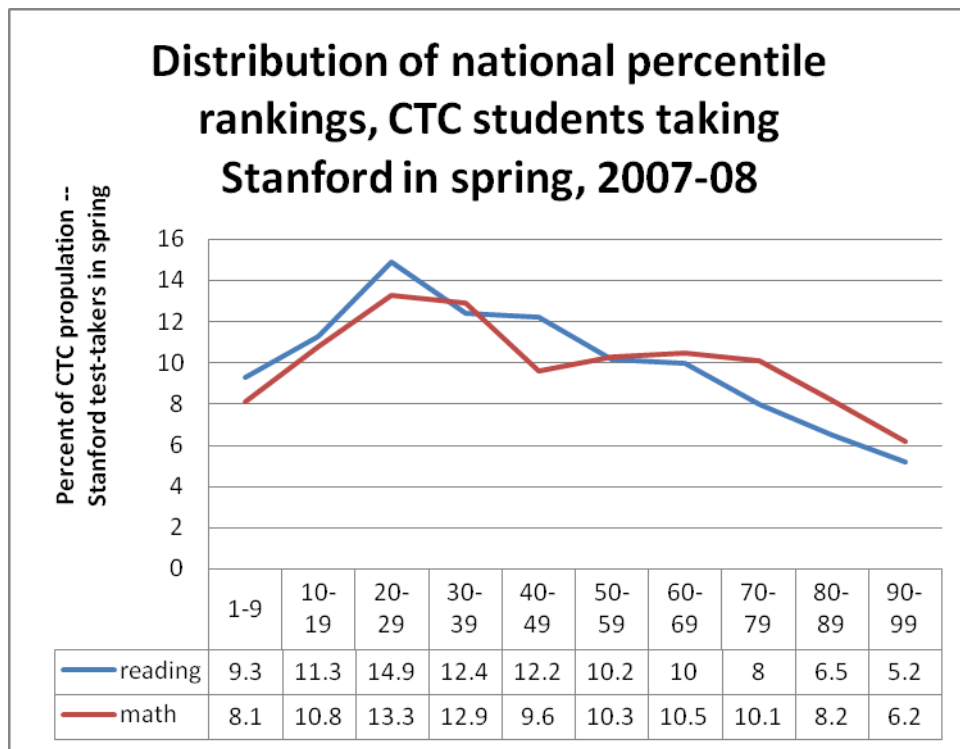
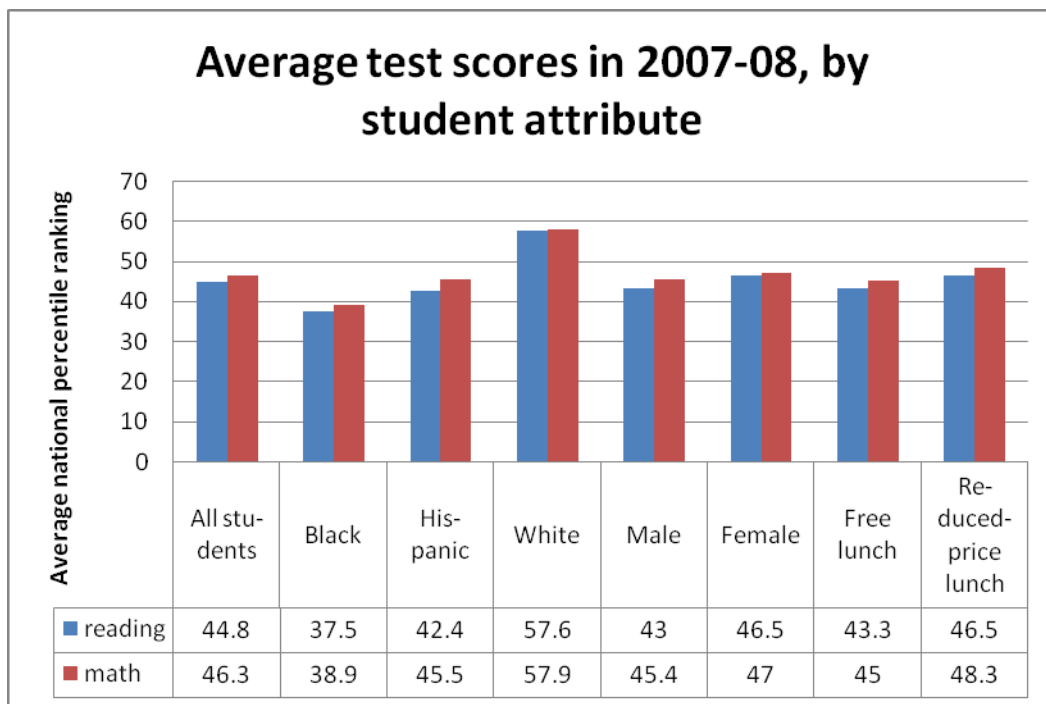
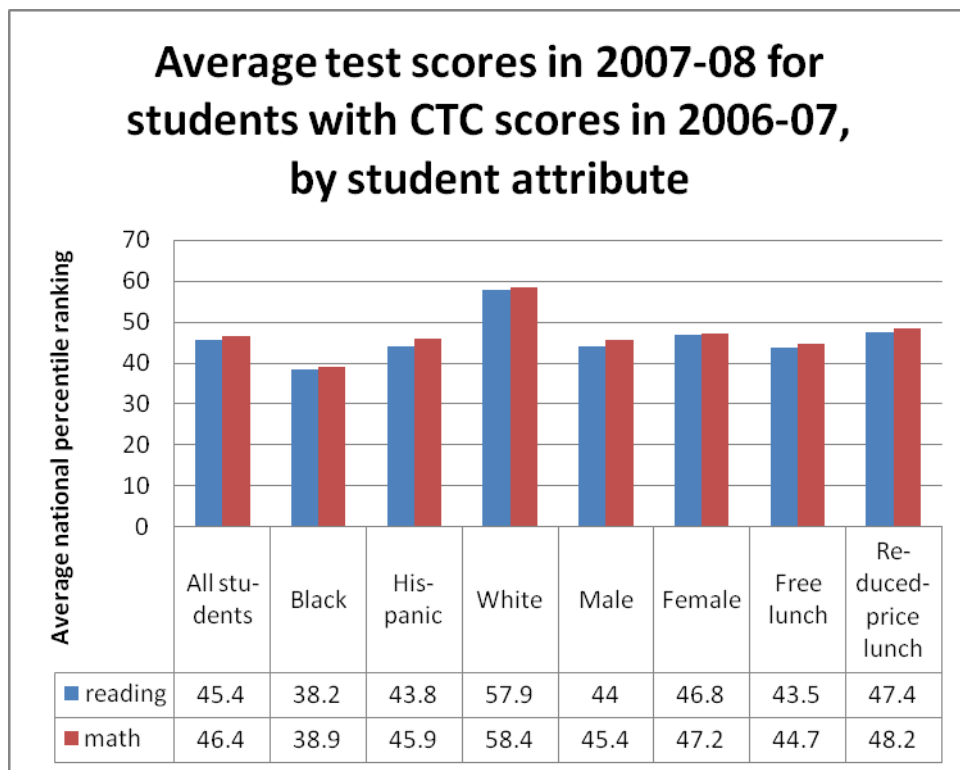


Table III.2 presents average norm referenced test scores, expressed in terms of national percentile rankings, for various subsets of the CTC Scholarship recipient population, stratified by race, sex, income, parental marital status and household size. Income is expressed in terms of fraction of the poverty 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, students from larger families tend to score better than do students from smaller families, and relatively high-income families tend to score better than do relatively low-income families. Students in schools that administer the Stanford test in the spring months mirror those in the CTC Scholarship population in general, although they tend to perform slightly worse (around one national percentile) in reading, though not in mathematics. In general, however, it appears that Stanford test-takers are highly similar in their scores to all other CTC Scholarship program participants.



One major purpose of this analysis is to measure student test score gains for students who remained in the program from one year to the next. The following graph presents a similar set of results for students who were tested under the CTC Scholarship program in 2006-07 and also were tested in 2007-08. One can observe that the average test scores for these students are reasonably similar to those reported in the preceding graph. The general patterns of results remain unchanged, regardless of whether students had previously taken a test in 2006-07 under the CTC Scholarship program.



*Test score gains for CTC 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 CTC 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 CTC Scholarship Program.

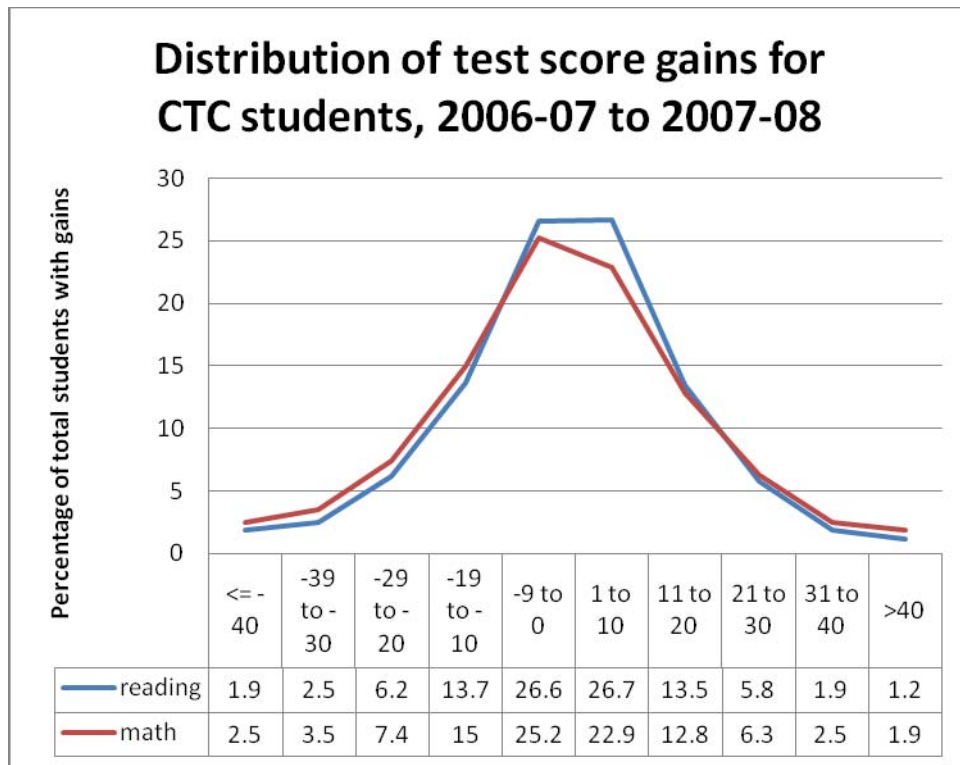
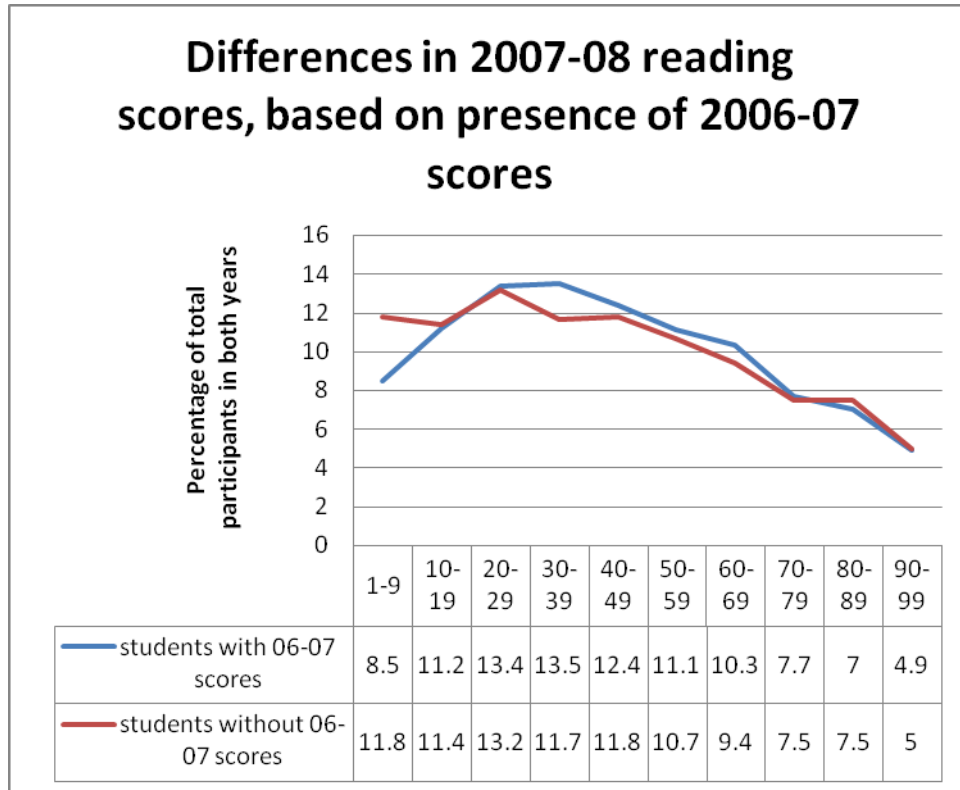


Table III.3 presents information on the distribution of CTC Scholarship Program participants' test score gains in reading and mathematics for the set of 4,531 students with legible reading scores and 4,580 students with legible mathematics scores in both 2006-

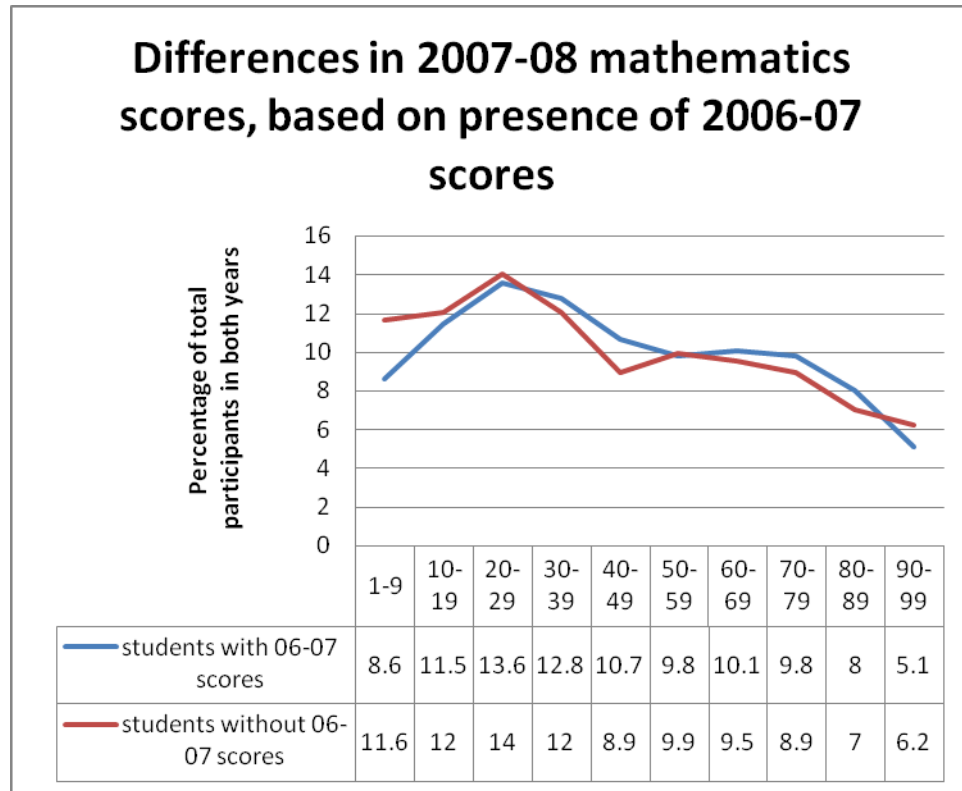
07 and 2007-08. The mean gain for program participants is -0.1 national percentile ranking points in reading and -0.9 national percentile ranking points in mathematics. In other words, the typical student participating in the program tended to maintain his or her relative position in comparison with 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 CTC 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, 10.6 percent of students participating in the program lost 20 or more percentile points in reading relative to the nation as a whole between 2006-07 and 2007-08, while 8.9 percent of program participants gained 20 or more percentile points in reading over this same time period. Furthermore, these comparisons are very similar when limited to students taking the Stanford Achievement Test during the spring: The distributions of Stanford test score gains are nearly identical, and the mean gains for Stanford-only test-takers are +0.2 national percentiles in reading and -0.4 national percentiles in mathematics.

These comparisons are also made more difficult because the participating students whose test scores were not collected in the retrospective test score collection of 2006-07 are not a random sample of the population of students participating in the program in both 2006-07 and 2007-08. As discussed above, students for whom legible test scores were submitted in 2006-07 were considerably more advantaged than were those for whom no legible test scores were submitted; the comparison in 2007-08, when test scores were collected in real time, is considerably less problematic. The potential problem with

missing baseline test scores is even more apparent when one compares average test scores in 2007-08 for students with valid, legible test scores received in 2006-07 versus those without these scores received. These comparisons are presented in Table III.4.







As is evident from the table, the students who participated in the program in both years, but who were missing test scores in 2006-07 tend to be lower-performing in 2007-08 than are those for whom valid legible test scores were received in both years. Indeed, the mean 2007-08 reading score for those with two years of valid tests is 1.2 percentile ranking points above the mean for those missing 2006-07 scores, and the mean 2007-08 mathematics score for those with two years of valid tests is 1.7 percentile ranking points above the mean for those missing 2006-07 scores. While these differences are not dramatic, they underscore the concerns inherent with the retrospective collection of 2006-07 test score data, and support the notion that test score gains computed using data collected from 2007-08 and 2008-09 will be more likely to be completely valid. That said, the students are sufficiently similar that it is still possible to analyze the gains from

2006-07 to 2007-08, with the caveat that these gains are unlikely to be fully representative of program participants.<sup>2</sup>

---

<sup>2</sup> Note that it is impossible to gauge whether the observed gains are understatements or overstatements of the true gains in test scores in the program. The missing students are less advantaged and lower-performing in 2007-08, and were likely lower-performing in 2006-07 as well, but this could lead to either higher or lower gain scores across the years.

Table III.1: Distribution of national percentile rankings for participants in the CTC Scholarship program, 2007-08

National percentile	Reading		Mathematics	
	Students with legible tests in 2007-08	Students with legible tests taking school-administered Stanford in spring	Students with legible tests in 2007-08	Students with legible tests taking school-administered Stanford in spring
1-10	8.5%	9.3%	7.9%	8.1%
11-20	12.3%	11.3%	11.3%	10.8%
21-30	12.6%	14.9%	13.4%	13.3%
31-40	12.8%	12.4%	12.7%	12.9%
41-50	12.5%	12.2%	10.4%	9.6%
51-60	10.8%	10.2%	10.2%	10.3%
61-70	9.8%	10.0%	10.1%	10.5%
71-80	8.5%	8.0%	9.8%	10.1%
81-90	6.7%	6.5%	7.8%	8.2%
91-99	5.5%	5.2%	6.4%	6.2%

Table III.2: Average national percentile rankings for participants in the CTC Scholarship program, 2007-08, for students of different background characteristics

Characteristic	Reading		Mathematics	
	Students with legible tests in 2007-08	Students with legible tests taking school-administered Stanford in spring	Students with legible tests in 2007-08	Students with legible tests taking school-administered Stanford in spring
All students	44.8	43.9	46.3	46.6
Black	37.5	37.1	38.9	40.2
Hispanic	42.4	41.6	45.5	45.4
White	57.6	56.5	57.9	57.5
Male	43.0	42.0	45.4	45.6
Female	46.5	45.6	47.0	47.4
Household size <=4	44.1	43.1	45.0	45.4
Household size >=5	45.9	45.0	48.4	48.5
Free lunch eligible	43.3	42.1	45.0	44.9
Reduced price lunch eligible	46.5	45.9	48.3	48.6
Parents married	49.1	47.9	50.6	50.6
Parents unmarried	42.1	41.4	43.5	44.1

Table III.3: Distribution of gains in national percentile rankings for participants in the CTC Scholarship program, 2006-07 to 2007-08

Change in national percentile ranking	Reading		Mathematics	
	All students with gain scores	Students taking school-administered Stanford in spring	All students with gain scores	Students taking school-administered Stanford in spring
<= -40	1.9%	1.9%	2.5%	2.5%
-39 to -30	2.5%	2.3%	3.5%	3.6%
-29 to -20	6.2%	5.8%	7.4%	7.0%
-19 to -10	13.7%	13.4%	15.0%	14.5%
-9 to 0	26.6%	26.0%	25.2%	24.4%
1 to 10	26.7%	27.8%	22.9%	23.9%
11 to 20	13.5%	13.8%	12.8%	12.8%
21 to 30	5.8%	6.0%	6.3%	6.3%
31 to 40	1.9%	1.8%	2.5%	2.7%
>40	1.2%	1.2%	1.9%	2.3%

Table III.4: Distribution of national percentile rankings for participants in the CTC Scholarship program, 2007-08, by whether student's score was reported in 2006-07

National percentile	Reading		Mathematics	
	Students with legible tests in 2006-07	Students without legible tests in 2006-07	Students with legible tests in 2006-07	Students without legible tests in 2006-07
1-10	8.5%	11.8%	8.6%	11.6%
11-20	11.2%	11.4%	11.5%	12.0%
21-30	13.4%	13.2%	13.6%	14.0%
31-40	13.5%	11.7%	12.8%	12.0%
41-50	12.4%	11.8%	10.7%	8.9%
51-60	11.1%	10.7%	9.8%	9.9%
61-70	10.3%	9.4%	10.1%	9.5%
71-80	7.7%	7.5%	9.8%	8.9%
81-90	7.0%	7.5%	8.0%	7.0%
91-99	4.9%	5.0%	5.1%	6.2%

#### **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 CTC Scholarship Program students to those of comparable public school students. This report compares the distribution of test score gains between 2006-07 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 CTC 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.<sup>3</sup> In addition, selection into a public school comparison group is not random. All CTC 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 CTC Scholarship Program. Therefore, it seems to be clear that school meals recipients in the public schools are not a very effective comparison group for CTC Scholarship Program participants, because their family incomes are likely to be

---

<sup>3</sup> A technical description of selection into the CTC Scholarship Program is provided in David Figlio, Cassandra Hart, and Molly Metzger, "Who Uses a Means-Tested Scholarship, and What Do They Choose?" which is currently under preparation for publication in a peer reviewed scholarly journal. A brief summary of the key points of that paper is provided in this report.

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.

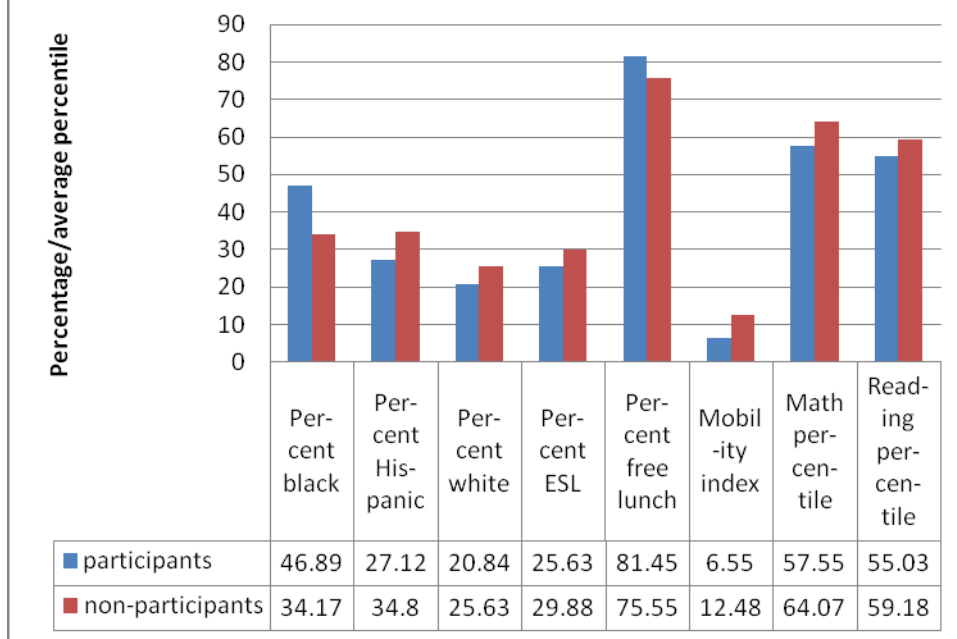
Taken together, these two factors indicate that direct comparisons of test score gains in the public sector versus CTC Scholarship Program participants, while informative, should not be interpreted as effects of the program on student test score gains. It will only be possible to determine causal effects of program participation on student test score gains following the collection of a second complete round of student test scores in 2008-09, coupled with the evaluation of more detailed information that could facilitate the construction of appropriate comparison groups. At present, any evaluation using 2006-07 and 2007-08 data to compare student test score gains can only be viewed as illustrative, rather than causal.

#### *Summary of key selection findings*

Before directly comparing student test score gains between CTC 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.



### Comparison of new CTC Program participants to "income-eligible" non-participants, 2007-08



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 2006-07 academic year in the Florida public schools, so one can compare the students who entered the CTC Scholarship Program in 2007-08 versus potentially comparable students who did not enter the program in that year. Table IV.1 presents some basic facts about CTC Scholarship Program participants relative to other potentially income-eligible students. One observes that CTC Scholarship Program participants differ from non-participants on many of the characteristics on which they are compared. 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.

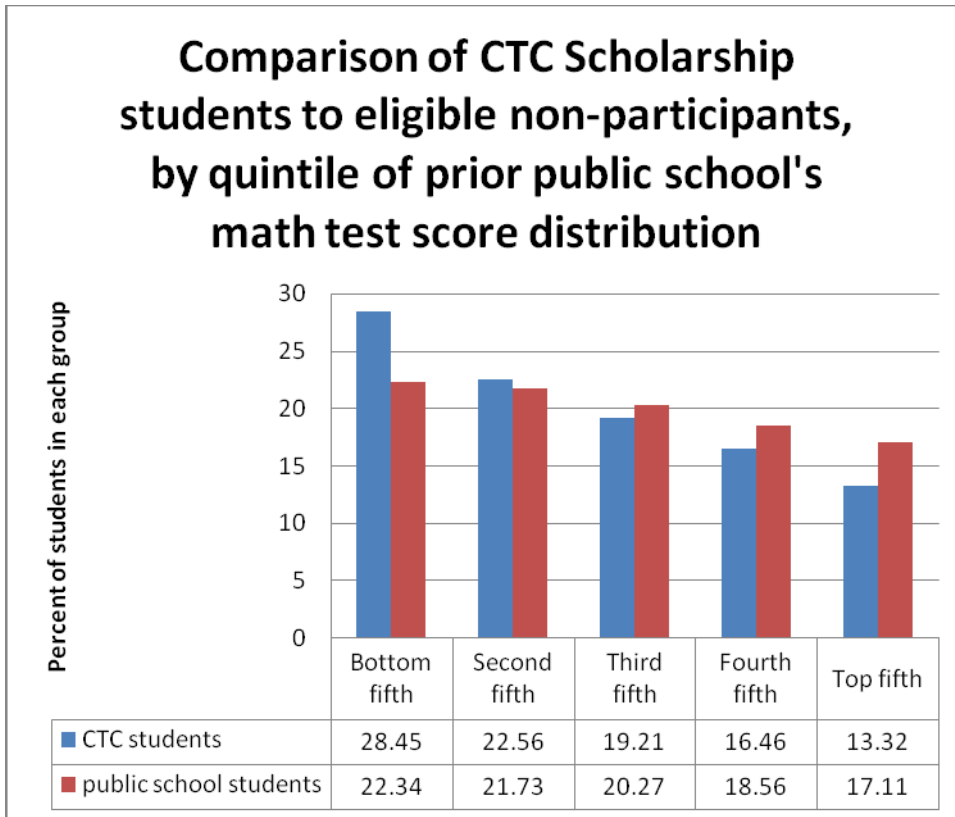
Scholarship participants are more economically disadvantaged than are non-participants on average. While all children in both the participant and non-participant groups were self-reported to be eligible for subsidized lunch at some point in the 2006-2007 school year, participants were more likely to 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. In a comparison of the percent of all observations about a student from 2003 to 2007 -- a measure reflecting the consistency of a child's poor or near-poor status -- scholarship participants qualified for free or reduced lunch a significantly higher percentage of the time than did non-participants, indicating more persistent low income for scholarship participants. That said, while the prior indicators suggest that scholarship participants are relatively more disadvantaged than are non-participants, they have significantly more stable prior schooling histories. Two different measures are used to compare students in terms of the percentage of times that they changed schools between surveys. In the first measure, all transitions between schools are employed; that is, a change in schools between fifth and sixth grades was counted as a transition, as well as school changes that occurred in non-transition grades. In the second measure, "expected" transitions are excluded. Using both measures, scholarship participants are less mobile than their non-participant peers.

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, non-participants out-performed scholarship participants in the year prior to the comparison. 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 2006-07 performance between public school students who would ultimately participate in the CTC Scholarship Program in 2007-08 and those who are plausibly income-eligible but who remained in Florida public schools in 2007-08 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 CTC Scholarship Program participants come disproportionately from lower-performing schools. For instance, amongst the elementary school students new to the program in 2007-08, 46 percent came from schools graded "A" by the Florida Department of Education in 2007, as compared with 54 percent of those public school students eligible for free or reduced-priced lunches. At the other extreme, 9 percent came from schools graded "D" or "F" by the Florida Department of Education in 2007, as compared with 6 percent of those public school students eligible

for free or reduced-priced lunches, and 38 percent came from schools graded "C" or below by the Florida Department of Education in 2007, as compared with 28 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 CTC 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, 28 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 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 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 17 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.

### *Comparisons of Test Score Gains with Public School Students*

As the preceding discussion makes clear, direct comparisons of the test score gains of students participating in the program with like students not participating in the program are complicated by three factors: (1) the incomplete test score participation and/or reporting during the 2006-07 school year, which was retrospectively collected; (2) the fact that scholarship participants who qualify for the program are more disadvantaged than observed non-participants because a sizeable fraction of non-participants observed to be eligible for the program would likely be found to be ineligible; and (3) participating students in the program are considerably more likely to have been low-performing and disadvantaged, in comparison to observed eligible students, potentially reflecting the differences in qualifications from (2). Complication (1) cannot be addressed this year -- in retrospect, two full years of real-time-collected test score data are necessary for fully valid comparisons of test score gains between the public sector and scholarship participants, meaning that the first year in which this would be possible is after the 2008-09 academic year's data are collected. However, complications (2) and (3) can be at least partially considered by comparing scholarship participants to relatively disadvantaged students participating in the school meals program; and by comparing students across

programs with similar test score profiles in prior years. These are, of course, only partial comparisons, because there is no way to know whether public school students are indeed eligible for the program, but they can provide a way of gauging the degree to which the issue of un-audited public school program participation poses a problem for analysis.

This subsection will make these comparisons.

While these comparisons are intended to be as close to apples-to-apples as possible, it is important to recognize that they are not causal estimates of the effect of program participation on student outcomes. Causal comparisons require more complete modeling of the selection decisions into the scholarship program and fuller data from a baseline than is afforded using the 2006-07 school year test score collection. More compelling causal estimates of program participation will be possible following the collection of the 2008-09 school year's test score data. The comparisons in this subsection should be interpreted as purely descriptive in nature.

Table IV.2 presents information on mean test score gains on the nationally norm-referenced test (school administered for CTC Scholarship Program students; Stanford Achievement Test for public school students) in reading and mathematics between 2006-07 and 2007-08. For the purposes of comparison, all test score gains are presented in terms of changes from year to year in the national percentile ranking for the individual student. Students included in the comparison are students who spent both years in the CTC Scholarship Program or who spent both years in Florida public schools and were receiving subsidized school meals as of the last survey taken in each of the two school years. As is observed in the table, CTC Scholarship Program participants experience slightly smaller test score gains than do those in the comparison group in the public

schools, though the differences in gain scores are small for reading and inconsequential for mathematics. In mathematics, CTC Scholarship Program participants experience test score gains that are 0.42 points lower than comparison public school students, on average, a difference that is statistically indistinguishable from zero. Reading test score gain differences are statistically different from zero but still very modest in magnitude, with program participants averaging 1.62 points lower gains. That said, these differences do not represent causal effects of program participation, and may reflect differences in student demographics or other attributes.

As mentioned above, students participating in the CTC Scholarship Program are almost surely less advantaged from a socio-economic perspective than are participants in subsidized school meals programs, in part because their family incomes have been audited. One possible way to judge whether these differences in student advantage might influence differences in test score gains across the two sectors would be to compare test score gains for students of different income levels, and see whether students from relatively low-income families gain more or less than students from comparatively high-income families. Ideally, this would take place in both sectors, but audited family income data only exist for CTC Scholarship Program participants.<sup>4</sup> For public school students, free and reduced-price-lunch recipients' family incomes are self-reported, and the audits of these families found that free-lunch participants and reduced-price-lunch participants were both reasonably likely to be found to be income-ineligible for the

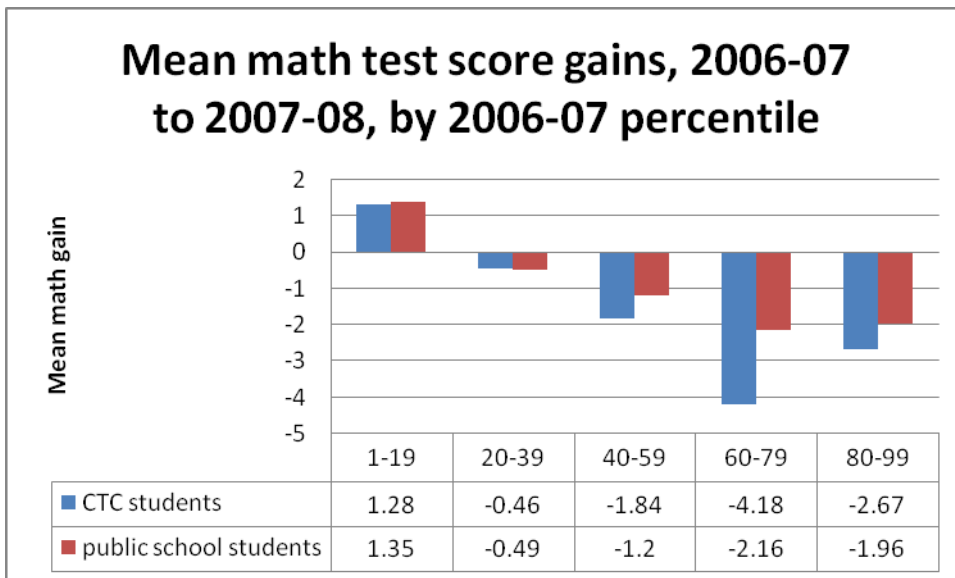
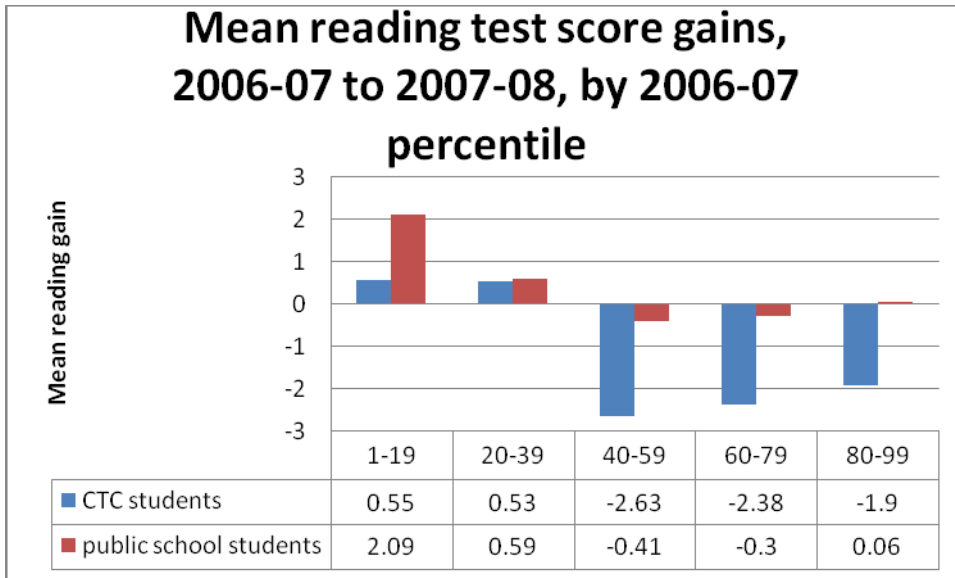
---

<sup>4</sup> That said, it is possible to compare test score gains for *program participants* of different family income levels. When making this comparison (reported in the table), there is no apparent relationship between family income and test score gains in reading and mathematics amongst the low-income population of scholarship participants. That is, among students certified to be eligible for and participating in the program, those from very low-income families perform at around the same level (in gain score terms) as do students from low-income families. However, it is impossible to extrapolate this finding to the public school sector.

program. Therefore, in order to gauge the degree to which these test score gain differences are due to differences in family advantage across the two sectors, it would be very useful to measure the test score gains for public school students who were part of the Department's three percent audit and whose family incomes were certified to be such that they could participate in the program. These data were not available, however, at the time of the report's writing. This would be a very useful comparison to make when measuring test score gains from 2007-08 to 2008-09.

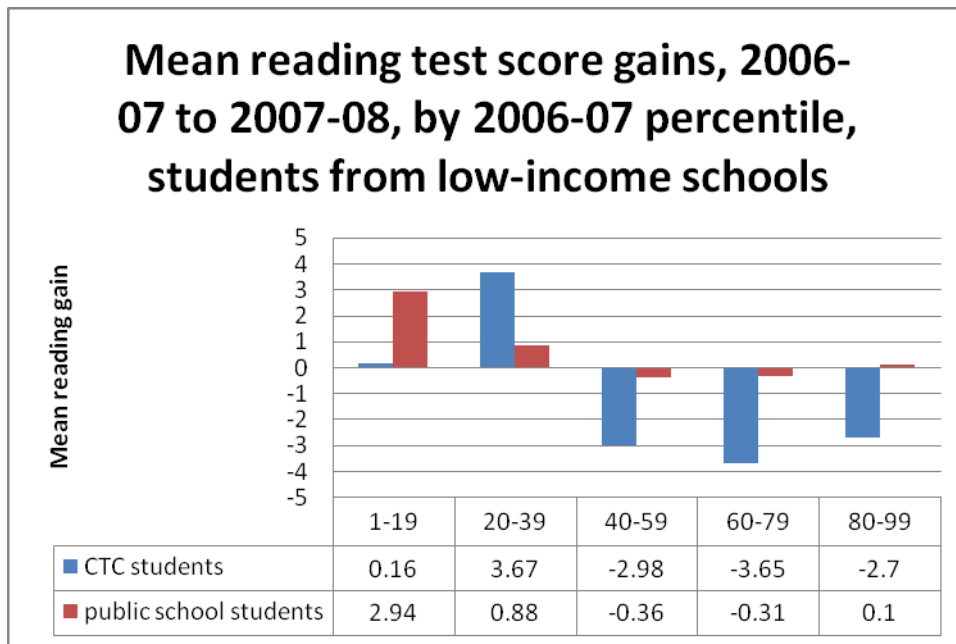
One can also compare the mean test score gains of students participating in the program to public school students in the comparison group, broken down by students' performance on the test during the 2006-07 school year. Table IV.3 presents mean test score gains for students in the CTC Scholarship Program and others in the comparison group of public school students for those in five quintiles of national percentile rankings in 2006-07. As can be observed in the table, CTC Scholarship Program participants and low-income public school students in the bottom 40 percent of the national percentile rank distribution experience similar gains, while public school students who had performed at a higher level in 2006-07 tend to have larger gains than do similar-performing students in the CTC Scholarship Program, particularly with regard to reading.



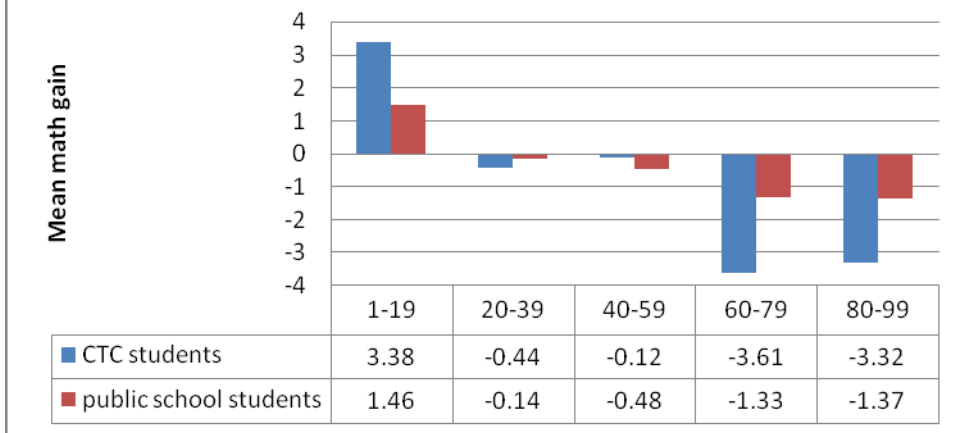


Public school students come from schools with widely varying socio-economic status, making it difficult to directly compare public school students with program participants. Therefore, Table IV.4 repeats the same comparison, but limits the analysis to students who were last observed in the quartile of schools serving the poorest student bodies in the state of Florida (with over 77 percent of students in the school receiving free or reduced-price lunches.) This analysis also tends to show that CTC Scholarship Program participants in the top quintiles of the 2006-07 test score distribution perform

somewhat worse than do similar-scoring public school students, but it also is more likely to show that CTC Scholarship Program students in some groups perform at higher levels on average than do similar-scoring public school students. So while this comparison continues the pattern of results of small test score gain advantages for public school students, it also highlights the fact that those relative gain measurements are fragile and small in magnitude relative to overall score variability.



### Mean math test score gains, 2006-07 to 2007-08, by 2006-07 percentile, students from low-income schools



To summarize the test score comparisons, CTC Scholarship Program students experienced comparably-sized, but modestly lower, average test score gains in reading and to a lesser degree in mathematics between 2006-07 and 2007-08. These differences are not uniform -- they are somewhat greater in magnitude for some groups than for others, and importantly, these group-specific differences change to some degree depending on the comparisons being made. Therefore, it is reasonable to conclude that CTC Scholarship Program participants' test score gains are in the same general ballpark, if on average modestly lower, as are those of comparison students in the public sector. As mentioned above, it is inappropriate to draw causal conclusions from these comparisons, however, and judgments regarding the relative efficacy of the two educational settings should be held until it is possible to analyze data based on two years of real-time-collected student test scores along with more formal causal modeling.

Table IV.1: Descriptive comparison of new CTC Scholarship Program participants in 2007-08 relative to non-participating school meals recipients

	New program participants ( <i>n</i> =3,278)	Non-participants ( <i>n</i> =995,933)
Male (%)	48.75	47.66
<b>Race</b>		
Asian/Pacific Islander (%)	1.77	1.78
Black (%)	46.89	34.17
Hispanic (%)	27.12	34.80
White (%)	20.84	25.63
<b>English Language status</b>		
English as a Second Language (%)	25.63	29.88
Limited English Proficiency (%)	13.21	13.99
<b>Socioeconomic status</b>		
Free lunch, last observation (%)	81.45	75.55
Reduced lunch, last observation (%)	16.41	20.43
Free/reduced lunch, 2003-07 (mean %)	91.04	88.87
<b>School mobility</b>		
School changes (%)	6.55	12.48
School changes, non-expected transitions (%)	6.01	10.32
<b>History of disability classification</b>		
Ever disabled (%)	5.74	6.26
<b>2006-07 Standardized test performance (national percentile)</b>		
Stanford Math, 2006-07	57.55	64.07
Stanford Reading, 2006-07	55.03	59.18

Table IV.2: Mean test score gains from 2006-07 to 2007-08, CTC Scholarship Program participants versus public school students receiving subsidized school meals

	Mean reading gain	Mean mathematics gain
CTC Scholarship Program participants	-0.13	-0.90
Public school students receiving meals	1.49	-0.48
<b>Among CTC Scholarship Program participants:</b>		
Income <100% of poverty	-0.06	-1.39
Income 100-125% of poverty	0.22	0.30
Income 125-150% of poverty	0.16	-0.92
Income 150-175% of poverty	-0.62	-1.00
Income >175% of poverty	-0.51	-1.25

Table IV.3: Mean test score gains from 2006-07 to 2007-08, CTC Scholarship Program participants versus public school students receiving subsidized school meals, by 2006-07 national quintile

National percentile in 2006-07	Reading		Mathematics	
	CTC Scholarship Program participants	Public school students receiving subsidized meals	CTC Scholarship Program participants	Public school students receiving subsidized meals
1-19	0.55	2.09	1.28	1.35
20-39	0.53	0.59	-0.46	-0.49
40-59	-2.63	-0.41	-1.84	-1.20
60-79	-2.38	-0.30	-4.18	-2.16
80-99	-1.90	0.06	-2.67	-1.96

Table IV.4: Mean test score gains from 2006-07 to 2007-08, CTC Scholarship Program participants versus public school students receiving subsidized school meals, by 2006-07 national quintile, students whose last observed public school was in lowest quartile of socio-economic status (over 77 percent free/reduced-price lunch)

National percentile in 2006-07	Reading		Mathematics	
	CTC Scholarship Program participants	Public school students receiving subsidized meals	CTC Scholarship Program participants	Public school students receiving subsidized meals
1-19	0.16	2.94	3.38	1.46
20-39	3.67	0.88	-0.44	-0.14
40-59	-2.98	-0.36	-0.12	-0.48
60-79	-3.65	-0.31	-3.61	-1.33
80-99	-2.70	0.10	-3.32	-1.37

## **V. Conclusion**

This report presents empirical evidence on the compliance and performance of private schools that participate in Florida's Corporate Tax Credit Scholarship Program. Analyzing data from 2007-08, the first year in which real-time test score collection was possible, there is strong evidence of high degrees of compliance with testing requirements for program participants. Retrospective analysis of the 2006-07 test score data collection, which was necessarily performed after that school year ended, indicates that 2006-07 may not be a reasonable baseline year for analysis of performance of program schools, due to large numbers of missing test scores for students who plausibly should have taken tests and the fact that these scores are apparently not missing at random.

With those provisos in mind, there is evidence that test score gains in the CTC Scholarship Program between 2006-07 and 2007-08, conditional on tests having been administered, are comparable in magnitude, if perhaps modestly smaller, than are gains in the public school comparison groups that have been constructed. 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 CTC Scholarship Program, and causal analysis of these differences would need to take this differential selection into account. This will be considerable more feasible following the collection of the 2008-09 school year test score data. That said, the first evidence regarding differential test score gains across the



public and CTC Scholarship Program sectors indicates roughly comparable test score gains that are reasonably consistent across different performance groups and unlikely due to family income differences between participants and non-participants.