

Program on Education Policy and Governance Working Papers Series

**The Impact of a Universal Class-Size Reduction Policy:
Evidence from Florida's Statewide Mandate**

Matthew M. Chingos

Program on Education Policy and Governance
John F. Kennedy School of Government
Harvard University
chingos@fas.harvard.edu

Last revised: August 2010

PEPG 10-03

**The Impact of a Universal Class-Size Reduction Policy:
Evidence from Florida's Statewide Mandate***

Matthew M. Chingos
Program on Education Policy and Governance
John F. Kennedy School of Government
Harvard University
chingos@fas.harvard.edu

Last revised: August 2010

Abstract

Class-size reduction (CSR) mandates presuppose that resources provided to reduce class size will have a larger impact on student outcomes than resources that districts can spend as they see fit. I estimate the impact of Florida's statewide CSR policy by comparing the deviations from prior achievement trends in districts that were required to reduce class size to deviations from prior trends in districts that received equivalent resources but were not required to reduce class size. I use the same comparative interrupted time series design to compare schools that were differentially affected by the policy (in terms of whether they had to reduce class size) but that did not receive equal additional resources. The results from both the district- and school-level analyses indicate that mandated CSR in Florida had little, if any, effect on cognitive and non-cognitive outcomes.

* I am grateful to Tammy Duncan and Jeff Sellers at the Florida Department of Education for providing the data used in this paper. For helpful comments I thank Steve Ansolabehere, David Deming, Jeremy Finn, Josh Goodman, Larry Katz, Paul Peterson, Martin West, and seminar participants at Harvard University. Financial and administrative support was provided by the Program on Education Policy and Governance at Harvard. I also gratefully acknowledge financial support from the National Science Foundation's Graduate Research Fellowship program.

1. Introduction

In recent decades, at least 24 states have mandated or incentivized class-size reduction (CSR) in their public schools (Education Commission of the States 2005). These policies presuppose that resources provided to reduce class size will have a larger impact on student outcomes than resources that districts can spend as they see fit. The idea that local school districts know best how to allocate the limited resources available to them suggests that unrestricted resources will be spent more efficiently than constrained resources. However, there are also reasons to expect that this may not be the case. Collective bargaining may constrain schools from optimally allocating resources if additional unrestricted state funding is seen as an opportunity for employees to demand higher salaries. Alternatively, districts may pursue different goals than the state government. For example, the state may prioritize student achievement while districts may place greater emphasis on extracurricular activities such as athletics.¹

Although there are reasons to expect that state governments may well improve student achievement by providing resources that must be spent on a specific policy such as CSR, there is little empirical evidence on this question. The most credible previous studies of CSR in the United States have focused on either randomized experiments or natural (plausibly exogenous) variation in class size.² Krueger's (1999) analysis of the Tennessee STAR experiment finds that elementary school students randomly assigned to small classes (13–17 students) outperformed their classmates who were assigned to regular classes (22–25 students) by about 0.22 standard

¹ A related idea is that the median voter in the state may have different preferences than the median voter in the school district.

² For examples of the earlier (primarily observational) literature on class size reduction, see Glass and Smith (1978) and Hanushek (1986). Two examples of high-quality international evidence on class size are Angrist and Lavy (1999) and Wößmann and West (2006).

deviations after four years, although this effect was concentrated in the first year that students participated in the program.³ But Hoxby's (2000) examination of natural class size variation in Connecticut (resulting from population variation) finds no evidence of class size effects. An important difference between these studies is that schools that participated in the STAR experiment received additional resources to reduce class size, while the Connecticut schools in Hoxby's study did not (and thus likely had to divert resources from elsewhere when natural population variation led to smaller classes). Thus one potential interpretation of the divergent results is that the positive effects found in the STAR experiment were at least partially made possible by the additional resources. But whether unconstrained resources would have had an even larger impact is still an open question.

These studies are also necessarily confined to estimating the partial equilibrium effect of varying class size, which may not be the same as the total effects of large-scale CSR policies. The most widely cited example of a possible general equilibrium effect is that reducing class size on a large scale will require schools to hire a large number of new teachers, many of whom may not be as effective as the teachers hired previously (particularly if salaries are held constant or decreased). Additionally, large-scale CSR may affect the sorting of teachers across schools—for example, by creating new positions at affluent schools that may be attractive to experienced teachers currently serving disadvantaged populations. However, such effects are not certain outcomes of CSR. Ballou (1996) finds evidence that schools do not always hire the applicants for teaching positions with the strongest academic records, and Kane and Staiger (2005) report that average teacher quality did not decline in the Los Angeles public schools after the district tripled the hiring of elementary school teachers following California's CSR initiative.

³ For a discussion of earlier class size experiments (mostly on a smaller scale), see Rockoff (2009).

In the only existing evaluation of a large-scale CSR policy, Jepsen and Rivkin (2009) find evidence in California that third graders did benefit from CSR, but that those gains were dampened by a decrease in teacher quality at schools that serve minority populations.⁴ However, Jepsen and Rivkin’s study is limited by data constraints—the primary outcome examined is the school-level percentage of third-grade students that scored above the 50th percentile on math and reading tests.⁵ Additionally, the counterfactual in the California study does not reflect what outcomes in schools would have been had they received equivalent resources that were not tied to CSR. Thus, there is very little evidence on the overall effects of large-scale CSR policies and essentially no evidence on the effect of CSR as compared to equivalent additional resources.

This paper contributes to this literature by using a rich administrative dataset to examine Florida’s CSR mandate, a voter-passed initiative that amended the state constitution to require that class sizes be reduced until they fall below set maxima. The implementation of Florida’s policy lends itself to a comparative interrupted time series (CITS) research design at two levels of aggregation. I first examine the district-level implementation of the policy (2004–2006), which required greater CSR in some districts than others but provided similar additional resources to all districts.⁶ I also examine the first three years of school-level implementation of the policy (2007–2009), which required varying amounts of CSR across schools but likely led

⁴ In a related study, Bohrnstedt and Stecher (2002) are unable to find a link between class-size reduction and student achievement in California, and also report that “CSR was associated with declines in teacher qualifications and a more inequitable distribution of credentialed teachers.”

⁵ This also makes it difficult to compare the magnitude of the California estimates to other studies, which have primarily focused on the effect of class size on average test scores rather than the percent of students who scored above some threshold.

⁶ Throughout this paper I refer to school years using the calendar year of the spring (e.g., 2004 refers to the 2003–04 school year).

districts to allocate greater additional resources to schools that were required to reduce class size than schools that were not required to do so.

The results of both analyses suggest that mandated CSR in Florida had little, if any, effect on student achievement in math and reading in fourth grade through eighth grade. Most estimated effects are not statistically significant from zero, with standard errors such that even modest positive effects can be ruled out. I also do not find any significant evidence of heterogeneous effects or effects on non-cognitive outcomes such as student absenteeism, suspensions, and incidents of crime and violence.

2. Evaluating Florida's CSR Policy

In November 2002, Floridians narrowly voted to amend their state constitution to set universal caps on class size in elementary and secondary schools. The amendment specifically mandated that, by the beginning of the 2010–11 school year, class sizes were to be reduced to no more than 18 students in prekindergarten through third grade, 22 students in fourth through eighth grade, and 25 students in ninth through twelfth grade. The total cost to implement this policy, which is constitutionally mandated to be the responsibility of the state government, is currently estimated at about \$20 billion over eight years, with continuing operating costs of about \$4 billion per year in subsequent years.⁷ Florida's class-size reduction (CSR) policy, while popular with many teachers and parents, has remained controversial due to its substantial cost, an issue which has become even more salient as the current economic downturn has placed great strain on the state budget.

⁷ "2009–10 Florida Education Finance Program," DOE Information Database Workshop, Summer 2009, available at <http://www.fldoe.org/eias/databaseworkshop/ppt/fejp.ppt>.

Students in Florida experienced substantially smaller classes as a result of the CSR mandate. According to official statistics from the Florida Department of Education (FLDOE), average class size in core classes in grades four to eight (the focus of this paper) fell from 24.2 in 2003 to 18.6 in 2009.⁸ Calculations from my extract from the FLDOE's Education Data Warehouse (EDW) indicate that this decrease occurred fairly evenly across groups of students defined in terms of their race/ethnicity and socioeconomic status, although the decrease was modestly larger for regular education students than for special education students.⁹ These calculations also do not show any evidence that average class sizes in non-core subjects (i.e., subjects not covered by the CSR mandate) increased—in fact they decreased, although not by as much as in subjects covered by the CSR mandate.¹⁰

Student achievement in Florida was increasing during the years both prior to and following the introduction of CSR in 2004. The National Assessment of Educational Progress (NAEP) scores of students in fourth grade increased dramatically over the last decade, with Florida surpassing the national average in reading in 2003 and in math in 2005. Between 1996 and 2009, fourth-grade math scores increased by 0.84 standard deviations, while fourth-grade

⁸ Core classes, which include all subjects areas affected by the CSR mandate, include language arts/reading, math, science, social studies, foreign languages, self-contained, special education, and English for speakers of other languages.

⁹ Using the EDW student course files, I calculate the average size of the core classes attended by each student (weighting each class by the number of minutes per week the student spent in the class and dropping as outliers classes with fewer than five or more than 40 students). These data indicate that statewide average class size in grades four to eight fell by 5.3 students from 2003 to 2009 (the change in the corresponding official FLDOE statistics, which are calculated using a modestly different formula, for this period is 5.6). This decrease was smaller for special education students, who experienced an average decrease of 3.4 (from 20.6 to 17.2), as compared to 5.7 (from 26.0 to 20.3) for regular education students. Students eligible for the free or reduced-price lunch program experienced an average decrease of 5.2 (from 24.7 to 19.5), as compared to 5.6 (from 26.2 to 20.6) among ineligible students. The decreases for black, Hispanic, and white students were 5.1, 6.4, and 5.2, respectively.

¹⁰ Average class size in all non-core classes in grades six to eight (I exclude grades four and five because of the prevalence of self-contained classrooms) fell from 26.0 in 2003 to 24.0 in 2009, a decrease of 2.0. Average class size in art and music classes fell by 2.1. Average class size in core classes in these grades fell by 5.4.

reading scores increased by 0.39 standard deviations between 1998 and 2009. Over the same time periods, the NAEP scores of eighth-grade students in math and reading increased by 0.39 and 0.26 standard deviations, respectively. Scores on Florida's Comprehensive Assessment Test (FCAT) posted similarly large increases over this period.¹¹

A naïve approach to estimating the effect of CSR would be to examine whether the rate of increase in student achievement accelerated following the introduction of CSR, but this method would be misleading because CSR was not the only major new policy in Florida's school system during this time period. First, the A-Plus Accountability and School Choice Program began assigning letter grades (and related consequences) to schools in 1999, and the formula used to calculate school grades changed substantially in 2002 to take into account student test-score gains in addition to levels. Second, several choice programs were introduced: a growing number of charter schools, the Opportunity Scholarships Program (which ended in 2006), the McKay Scholarships for Students with Disabilities Program, and the Corporate Tax Credit Scholarship Program. Finally, beginning in 2002 the "Just Read, Florida!" program provided funding for reading coaches, diagnostic assessments for districts, and training for educators and parents.

In order to identify the effect of mandated CSR as compared to unrestricted additional financial resources, a credible comparison group must be identified. This paper compares students who were more affected by the policy because they attended districts or schools that had pre-policy class sizes further from the mandated maxima with students that were less affected because they attended districts or schools that were already in compliance with the class size policy. Specifically, I compare the deviations from prior trends in student achievement at

¹¹ Between 2001 and 2009, fourth-grade math and reading scores increased by 0.70 and 0.43 standard deviations, respectively. Eighth-grade math and reading scores increased by 0.26 and 0.29 standard deviations, respectively.

districts/schools that were required to reduce class size to deviations from prior achievement trends at districts/schools that were not required to reduce class size. In the case of the district-level analysis, these two groups of districts received the same amount of resources (per student) to implement the CSR policy.

This strategy takes advantage of the details of the implementation of the CSR mandate that were set by the state legislature. From 2004 through 2006, compliance was measured at the district level. Districts were required to reduce their average class sizes either to the maximum for each grade grouping or by at least two students per year until they reached the maximum. Districts that failed to comply faced financial penalties, so the vast majority complied.¹² Beginning in 2007, compliance was measured at the school level, with schools facing the same rules for their average class size that districts faced previously. Beginning in 2011, compliance will be measured at the classroom level.¹³

District-Level Analysis

I classify districts into two groups: comparison districts, which already had average class sizes beneath the mandated maxima for a given grade range in 2003, and thus were not required to reduce class size at all (although many did in anticipation of the school-level enforcement) and treated districts, which had average class sizes above the mandated maxima (and thus had to reduce class size to the maxima or by at least two students per year). I use the official average class sizes for 2003 (the year immediately preceding implementation of CSR) published by the

¹² For average class size in grades four to eight, 62 out of 67 districts were in compliance in 2004, 65 in 2005, and all 67 in 2006.

¹³ In the initial legislation, compliance was to have been measured at the classroom level beginning in 2009, but the legislation was twice amended by the state legislature to push back the deadline (and substantial rise in costs associated with implementing CSR at the classroom level) first to 2010 then to 2011.

Florida Department of Education (FLDOE) to classify districts, and only include the 67 regular school districts (which are coterminous with counties).¹⁴ Charter schools were not subject to the district-level implementation of CSR, so I exclude all charter schools that were in existence in 2003 from the district-level analysis.¹⁵

This strategy classifies as treated 59 out of 67 districts in prekindergarten to third grade, 28 out of 67 in grades four to eight, and 61 out of 67 in grades nine to 12. In the district-level analysis, I only examine students in grades four to eight, and thus only use the treatment groups defined by districts' average class sizes for those grades. These grades are the most amenable to my identification strategy because of the relatively even division of districts between treated and comparison groups and because all of the relevant grades are tested. On the other hand, almost all districts are treated in grades prekindergarten to three and very few districts are treated in grades nine to 12. Additionally, students are only tested in grades three to ten.

According to my calculations from the EDW, in the first year of district-level implementation (2004) average class size fell by 0.1 students in the comparison districts and 0.9 students in the treated districts. By the third and final year of district-level CSR implementation (2006), district-level average class size had fallen by 1.4 students in the comparison districts and 3.0 students in the treated districts. Thus, the treated districts reduced average class size by 1.6 students more than the comparison districts.¹⁶

¹⁴ The class size averages are available at <http://www.fldoe.org/ClassSize/csavg.asp>. The excluded districts are the four lab schools (Florida Agricultural & Mechanical University, Florida Atlantic University, Florida State University, and University of Florida) and the Florida Virtual School.

¹⁵ Below I show that my results are robust to including all charter schools or excluding all charter schools.

¹⁶ However, when instead I use the official FLDOE class size averages, I obtain modestly different results, which show a reduction of average class size by 2006 of 1.6 students in the comparison districts and 4.6 students in the treated districts, a difference of three students.

As discussed earlier, the amount of per-pupil funding allocated by the state for the purposes of CSR was roughly the same in all districts. Specifically, districts received roughly \$180 per student in 2004, \$365 per student in 2005, and \$565 per student in 2006. Thus even the comparison districts (which were not required to reduce class size at all) were given what essentially amounted to a block grant to do whatever they wished with. Some surely used it to reduce class size in anticipation of school-level enforcement, although the class size numbers suggest this behavior was modest and did not compromise the difference in changes in class sizes between the treatment groups.¹⁷ Some districts may have reduced the share of funding from local sources (property taxes) in response, although below I present evidence that this did not happen to a greater extent in the treated districts than in the comparison districts. Consequently, the district-level treatment effects should be interpreted as the effect of forcing a group of districts to reduce average class size, as compared to giving other districts similar resources but not requiring them to do anything in particular with those resources.

Table 1 presents summary statistics (weighted by district enrollment) for treated and comparison districts in the last pre-treatment year (2003). The only statistically significant differences between the two groups of school districts are in average class size and average teacher salaries.¹⁸ Classes were about 12 percent larger in treated districts than in comparison districts and teachers were paid about 13 percent more, so it is unsurprising that per-pupil spending was roughly the same in the two groups of districts. In terms of other observable

¹⁷ However, as I discuss below, my calculations from the EDW data suggest that in the district-level implementation period class size for grades four and five was reduced by similar amounts in the treated and comparison districts (although this was not the case for grades six to eight).

¹⁸ Average salary for each district-year is calculated as the average of median regular compensation (base salary) for teachers with a bachelor's degree (but no higher degree) in every experience cell from the first to twentieth year of experience, with missing values imputed using data from adjacent years or experience categories (i.e., the set of 20 medians is averaged for each district-year).

characteristics, differences are all statistically insignificant and most are substantively insignificant as well. The percent of students eligible for free or reduced-price lunch differed by only four percentage points. Student test scores were essentially identical, differing by only 0.01 and 0.03 standard deviations in math and reading, respectively. The only substantively meaningful difference is in enrollment. The average student in the comparison districts attended a district that enrolled 41,623 students in grades four to eight, as compared to an average of 63,202 students among treated districts. Figure 1, which shows the location of the treated and comparison districts, does not suggest any particular geographic pattern. For example, among the six largest cities, four are in treated districts and two are in comparison districts.

Any time-invariant characteristics of school districts that differ across treatment groups will be netted out by including district fixed effects in all specifications. Time-varying characteristics, including percent black, percent Hispanic, and percent eligible for free/reduced lunch, are controlled for in my preferred specification, which follows a comparative interrupted time series (CITS) setup very similar to that used by Dee and Jacob (2009):

$$A_{idt} = \beta_0 + \beta_1 YEAR_t + \beta_2 CSR_t + \beta_3 YR_SINCE_CSR_t + \beta_4 (T_d \times YEAR_t) + \beta_5 (T_d \times CSR_t) + \beta_6 (T_d \times YR_SINCE_YEAR_t) + \beta_7 Stud_{it} + \beta_8 Dist_{dt} + \delta_d + \epsilon_{idt} ,$$

where A_{idt} is the FCAT score of student i in district d in year t (standardized by subject and grade to have a mean of zero and standard deviation of one based on the distribution of scores in the pre-treatment years 2001 to 2003); $YEAR_t$ is the year (set so that 2001 is equal to 1); CSR_t is an indicator for whether the year is 2004 or later (indicating that CSR is in effect); $YR_SINCE_CSR_t$ indicates the number of years since CSR (pre-2004 is 0, 2004 is 1, 2005 is 2, and 2006 is 3); T_d is an indicator identifying districts in the treated group; $Stud_{it}$ is a vector of student-level characteristics (dummies for grade level, race/ethnicity, gender, free/reduced lunch

status, limited English proficiency status, and special education status); $Dist_{dt}$ is a vector of time-varying district-level characteristics (percent black, percent Hispanic, and percent eligible for free/reduced lunch); δ_d is a vector of district fixed effects; and ϵ_{idt} is a standard zero-mean error term. I estimate this equation separately by subject (reading and math) using data from 2001 to 2006.¹⁹ Standard errors are adjusted for clustering at the district level.

The coefficients of greatest interest are β_5 , which indicates the shift in the overall level of achievement (the change in the intercept) due to CSR and β_6 , which indicates the shift in the achievement trend (the change in the slope) due to CSR. I also present estimates of the total effect of the district-level implementation of CSR after three years, which is $\beta_5 + 3 \times \beta_6$.

Interpreting these coefficient estimates as the causal effects of mandated CSR (as compared to unrestricted additional resources) requires the assumption that, conditional on the control variables, the deviation from prior achievement trends at the comparison districts accurately approximates the deviation from prior trends that the treated districts would have experienced had they been provided with additional resources but not required to reduce class size. The fact that the two groups of districts are similar in terms of most of their observable characteristics supports this “parallel trends” assumption, as does the similarity of pre-treatment achievement trends in treated and control districts documented in the regression results reported below and depicted in Figures 2a and 2b. These figures show that treated and comparison districts had very similar achievement trends in eighth-grade math and reading scores during the period prior to CSR (2001–2003).

¹⁹ Below I show that, for selected grades and subjects for which additional years of data are available, the results are not sensitive to using four or five years of pre-treatment data instead of three. However, the results are sensitive to using only two years of pre-treatment data, as would be necessary if I were to control for prior-year test scores. Adding controls for prior-year test scores does not substantially change the results beyond those obtained using two years of pre-treatment data.

Another indirect test of the parallel trends assumption is to estimate the “effect” of CSR on variables that should not be affected by CSR. The results of these falsification tests, reported in the first three columns of Appendix Table 1, show that the estimated “effect” of CSR on district-level percent black, percent Hispanic, and percent eligible for free or reduced-price lunch is statistically insignificant and substantively small, as would be expected if the model assumptions hold.

Appendix Table 1 also shows the effect of CSR on enrollment in the district, per-pupil spending, and teacher salaries. The enrollment results indicate that CSR reduced enrollment in the treated districts (relative to what it would have been in the absence of treatment) by about four percent by 2006.²⁰ The fourth column shows that per-pupil spending did not change in the treated districts relative to the comparison districts, providing further evidence to support the interpretation of the district-level effects as the impact of mandated CSR as compared to equivalent additional resources. The final column suggests that CSR was accomplished partly by decreasing teacher salaries. In the first year of CSR, teacher salaries in the comparison districts increased by about two percent more (in real terms) than they had in previous years, whereas salaries in the treated districts fell by about seven percent.

School-Level Analysis

As a complement to the district-level analysis I also conduct a similar analysis using variation in CSR implementation at the school level. Beginning in 2007, individual schools were

²⁰ The coefficients on the other variables indicate that enrollment was growing by 1.9 percent per year in the comparison districts and 2.1 percent per year in the treated districts prior to CSR. After the introduction of CSR, enrollment grew by 2.4 percent per year in the comparison districts and 1.1 percent per year in the treated districts. In other words, these results do not indicate that CSR led to an absolute decrease in enrollment, but that it caused a smaller increase in enrollment than would have been experienced in the absence of CSR. This smaller increase in enrollment likely made it possible for treated districts to implement CSR at a lower cost than had enrollment increased at a faster rate.

required to reduce their average class sizes to the constitutionally mandated maxima or by two students per year until they were beneath the maxima. The state provided districts with approximately \$800 per student in 2007 and roughly \$1,000 per student in 2008 and 2009 to finance these reductions. As in the district-level implementation phase, districts were awarded roughly the same per-pupil amount regardless of how much they actually had to reduce class size to comply with the school-level requirements. However, districts were free to spend more of their CSR allocation on schools that needed to reduce class size as compared to schools that were already in compliance with the mandate.

Using the official FLDOE calculations of school-level average class sizes for grades four to eight in 2006, I classified schools into the same two groups using the same definitions as in the district-level analysis. This method identifies 2,106 comparison schools and 664 treated schools. The analysis is essentially identical to the district-level analysis, with school fixed effects in place of district fixed effects and school-level time-varying characteristics in place of the same variables measured at the district level. Standard errors are clustered at the school level.

Between 2006 and 2009, the comparison schools reduced their average class sizes by 0.5 students, while the treated schools reduced average class size by 3.1 students.²¹ Pre-treatment (2006) summary statistics for treated and comparison schools are shown in Table 2. The two groups are fairly similar in terms of per-pupil spending and demographic breakdowns, although the treated schools have modestly higher enrollment and test scores than comparison schools.²²

²¹ The official FLDOE statistics show a larger reduction of 3.4 students at the treated schools, as compared to 0.3 students at the comparison schools.

²² The treated schools were also substantially more likely to be treated in grades prekindergarten to three, which should increase the likelihood of finding CSR effects in fourth grade in 2008 and in fourth and fifth grade in 2009.

The results of falsification tests reported in the first three columns of Appendix Table 2 indicate that CSR had only negligible “effects” on the demographic composition of treated schools (some of the coefficients are statistically significant, but trivial in size). And unlike in the district-level analysis, CSR had no impact on enrollment and had a positive impact on per-pupil spending in the first year of school-level implementation, but this increase faded out in the following two years. The latter finding indicates that the school-level results have a modestly different interpretation than the district-level results. Whereas the district-level results indicate the effect of CSR as compared to equivalent additional resources, the school-level results indicate the effect of CSR as compared to fewer additional resources in the first two years of school-level implementation (and similar resources in the third year).

The differing advantages and disadvantages of the district- and school-level approaches complement each other. The district-level approach has the substantial advantage of coming as a surprise to school districts, which probably could not have accurately anticipated whether the amendment would pass and how it would be implemented. The school-level approach clearly does not have this advantage, as schools (in cooperation with districts) likely anticipated the coming school-level implementation during the district-level implementation period. It is unclear in which direction this will bias my school-level results. If the “anticipatory CSR” occurred disproportionately in schools where students were most likely to benefit from it, then my school-level estimates will be biased downward (because the schools that remained to be treated in 2007 contained students that were less affected by CSR than their peers in the schools that reduced class size earlier and thus are included in my comparison group). However, the reverse could be true, such as if affluent schools with politically active parents pressured districts to reduce class size in their schools first (and if affluent students benefit less from smaller

classes, as some of the literature suggests), in which case my school-level estimates will be biased upward. The similar demographic breakdowns in treated and comparison schools reported in Table 2 do not support this hypothesis.

However, the school-level approach also has two key advantages. First, the larger number of schools provides greater statistical power for the detection of effects that may not be particularly large. Second, the fact that the school-level implementation came later (after the completion of district-level implementation) means that it is where one would expect to find larger general equilibrium effects (such as reduced teacher quality, if the pool of qualified applicants for teaching positions was depleted during the district-level implementation of CSR).

3. Data

The student-level data used in this study are from the K–20 Education Data Warehouse (EDW) assembled by the Florida Department of Education (FLDOE). The EDW data extract contains observations on every student in Florida who took the state assessment tests from 1999 to 2009.

The EDW data include test score results from Florida’s Comprehensive Assessment Test (FCAT), the state accountability system’s “high-stakes” test, and the Stanford Achievement Test (SAT), a nationally norm-referenced test that was administered to students at the same time as the FCAT until 2008 but was not used for accountability purposes. Beginning in 2001, students in third grade through tenth grade were tested every year in math and reading. The data also contain information on the demographic and educational characteristics of the students, including gender, race, free or reduced-price lunch eligibility, limited English proficiency status, special education status, days in attendance, and age.

In parts of the analysis I calculate class size from the EDW course files using the definitions published by the FLDOE.²³ According to these definitions, average class size is calculated “by adding the number of students assigned to each class in a specified group of classes and dividing this compiled number of students by the number of classes in the group.” Types of classes that are included in the calculation are language arts/reading, math, science, social studies, foreign languages, self-contained, special education, and English for speakers of other languages. I drop as outliers classes containing fewer than 5 or more than 40 students, although my results are not sensitive to this decision.

I obtain district- and school-level data on enrollment, student demographics (racial/ethnic breakdowns and percent eligible for free or reduced-price lunch), and per-pupil spending from the National Center for Education Statistics Common Core of Data and school-level data on accountability grades, per-pupil spending, and non-cognitive outcomes from the FLDOE’s Florida School Indicators Reports.

4. Results

District-Level Analysis

The legislation implementing CSR in Florida required districts to reduce their average class sizes in each of three grade groupings (including grades four to eight, which are the focus of this study) but left districts free to meet this goal in any way they wished. Although the official FLDOE class size averages do not line up perfectly with those I am able to calculate from the EDW database (as mentioned above), they are clearly correlated. It is instructive to estimate the impact of a district being required to reduce class size on district-level average class

²³ Florida Department of Education, “Class Size Reduction in Florida’s Public Schools,” available at <http://www.fldoe.org/ClassSize/pdf/csfaqfinal.pdf>.

sizes, both overall and by grade. The first column of Table 3 shows that average class size in both treated and comparison districts was essentially static before the introduction of CSR (see the coefficients on *Year* and $T \times Year$). As expected, average class sizes decreased after that, but to a greater degree in the treated districts than in the comparison districts. By 2006, average class size had fallen by 1.5 students more in the treated districts than in the control districts. This impact was concentrated in grades seven and eight, with a relative reduction of about three students, and to a lesser degree in sixth grade, which had a relative reduction of 1.6 students.²⁴ Class sizes in grades four and five were reduced by similar amounts in the treated and comparison districts.²⁵ Thus, in addition to presenting results that combine grades four to eight, I will also present results disaggregated by grade to see whether effects are concentrated in the middle school grades.

Figures 2a and 2b show the similar pre-treatment trends in eighth-grade FCAT scores at treated and comparison districts noted earlier as well as post-treatment achievement trends that do not diverge markedly. Beginning in 2005, the math trend for treated districts diverged from that of comparison districts, but only by about 0.03 standard deviations. There does not appear to be any divergence in reading achievement trends during the post-CSR period. This analysis is formalized using the regression model described above. Tables 4a and 4b present my preferred district-level estimates of the effect of the CSR mandate on FCAT math and reading scores. The

²⁴ I also estimated a version of this model that defined treatment (T_d) not as the dichotomous variable described above but as a continuous variable indicating by how many students each district was required to reduce class size. However, the estimates that correspond to those in Table 3 (not reported) were substantially weaker, suggesting that this measure of treatment intensity is not a good linear predictor of by how much districts reduced class size.

²⁵ Estimations using student-level average class size calculated from the EDW data (in place of district-level averages) yield qualitatively similar results. The class size results are slightly stronger when I examine average class size in general (e.g., self-contained), math, and reading classes rather than all core classes. The three-year effect of CSR is a reduction of 2.2 students in grades four to eight, 0.1 students in grades four and five, 2.2 students in grade six, 3.3 students in grade seven, and 3.5 students in grade eight.

test scores have been standardized by subject and grade using the pre-treatment (student-level) test-score distribution for ease of comparison with the rest of the class-size literature.²⁶

It is instructive to examine all of the coefficient estimates reported for my preferred estimates. The coefficient on *YEAR* in the first column of Table 4a indicates that, prior to the introduction of CSR, math scores were increasing by about 0.05 standard deviations per year in the comparison districts. The coefficient on $T \times YEAR$ (0.002) indicates that this pre-treatment achievement trend was nearly identical in the treated districts, which adds to the credibility of the parallel trends assumption made by my identification strategy. The coefficients on *CSR* and *YR_SINCE_CSR* indicate that math scores increased in the comparison districts after the introduction of CSR, although of course this increase cannot be causally linked to CSR (and the additional funding provided to comparison districts) given the myriad other reforms that were introduced in Florida around this time. However, this deviation from the pre-CSR achievement trend was fairly similar in the treated and comparison districts. By 2006, math achievement in the treated districts was only a statistically insignificant 0.035 standard deviations [$(T \times CSR) + 3 \times (T \times YR_SINCE_CSR)$] higher than it would have been had those districts received additional resources without a mandate to reduce class size. The standard error is such that negative effects larger than 0.026 standard deviations and positive effects larger than 0.096 standard deviations can be ruled out with 95 percent confidence.

The effect of CSR on math scores does not appear to vary by grade level. In particular, the effects are not larger for grades seven and eight (which saw the largest relative reductions in class size in the treated districts) than for the earlier grades. The estimates for reading scores

²⁶ Although the variation in my treatment variable is at the district level, it would be misleading to use the district-level standard deviation in test scores to interpret my estimates given that the district-average test-score distribution is highly compressed as a result of Florida's countywide school districts.

(Table 4b) follow a similar pattern, with a total effect point estimate of -0.001, with a 95 percent confidence interval that ranges from -0.085 to 0.083. Results disaggregated by grade are less precisely estimated (none are statistically significant) and although the point estimates vary somewhat it is clear that the effects are not larger for the middle school grades—in fact, the only negative point estimates are those for grades seven and eight. Combining grades seven and eight—the grades in which class size in the treated districts was reduced the most relative to the comparison districts—indicates that by 2006 CSR had *reduced* achievement by 0.087 standard deviations in reading, an effect that is statistically significant at the 5 percent level.

These main results are robust to a variety of alternative specifications. A potential limitation of the district-level analysis is that only three years of data are used to estimate the pre-treatment trend. For four subject-grade combinations, five years of pre-treatment data are available. Appendix Table 3 shows that for these subjects and grades, using four or five years of pre-treatment data produces similar results to using three years of pre-treatment data. However, the results are sensitive to using only two years of pre-treatment data. This is not surprising given the difficulty of estimating a trend from only two points, but it is relevant because any models that control for students' prior-year test scores (as is often done in the education literature) would necessarily be limited to two years of pre-treatment data. Appendix Table 4 shows that, for grades four to eight, restricting the analysis to two years of pre-treatment data noticeably changes the results. However, adding controls for prior-year tests scores (and other student characteristics that require prior-year data, including number of days absent the previous year, whether the student was repeating a grade, and whether the student moved between schools) causes only small additional changes to the results.

Appendix Table 5 shows the results from five other alternative specifications. Similar results are obtained when district-specific linear time trends are included, when all charter schools are excluded, when all charter schools are included, and when each district is weighted equally. The final column shows results from a traditional difference-in-differences specification, where the linear time trends are replaced with year fixed effects and a single $T \times CSR$ term is used to estimate an average effect of CSR over the three post-treatment years. This model controls for pre-treatment differences in achievement levels, but not for differences in pre-treatment trends. These results are qualitatively similar to my preferred estimates, as would be expected given the similarity of the pre-treatment achievement trends at treated and comparison districts.²⁷

I also ran a fully flexible specification that includes a full set of year dummies and treatment-year interactions. The coefficients on the treatment-year interactions, which indicate the treatment-comparison achievement difference in a given year compared to the last pre-treatment year (2003), are shown in Appendix Table 6. The three-year effects are similar to my preferred estimates, although the one-year math effect is positive and statistically significant.

Appendix Table 7 shows results that use scores from the Stanford Achievement Test (SAT), a low-stakes exam administered along with the FCAT, as the dependent variable. The results that combine grades four to eight are similar to the FCAT results, although the results by grade vary more. Two estimated effects—fourth-grade math and fifth-grade reading—are statistically significant, but the estimates for the grades where class size was actually reduced in

²⁷ When I use the standard difference-in-differences specification to estimate effects separately by grade (not shown), the positive fourth-grade math effect is now statistically significant and the seventh- and eighth-grade reading effect is no longer statistically significantly negative.

the treated districts relative to the comparison districts (seventh and eighth) are close to and statistically insignificant from zero.

Some previous literature finds that disadvantaged students (such as those that are members of underrepresented minority groups or are eligible for free/reduced lunch) benefit more from CSR than other students.²⁸ Appendix Table 8 shows results disaggregated by gender, race/ethnicity, and eligibility for free or reduced-price lunch (FRL). The point estimates for math scores are larger for blacks and Hispanics than for whites and for FRL students than for non-FRL students. However, these estimates are too imprecisely estimated to be statistically significantly different from each other. Point estimates for reading scores follow a similar pattern, except that the estimates for blacks and whites are similar. Examining subgroup results for grades six to eight only (not shown), the pattern for math scores is weaker, and all of the point estimates for reading are negative, with the largest negative effect (still statistically insignificant) occurring among black students. In all cases, estimates are similar for boys and girls.

Finally, I examine the effect of CSR on several non-cognitive outcomes.²⁹ Appendix Table 9 shows no evidence that CSR affected student absenteeism, incidents of crime and violence, or student suspension rates.³⁰ All of the estimated effects on these undesirable outcomes are statistically indistinguishable from zero, although the point estimates are positive.

²⁸ For example, Krueger (1999) finds that minority and free lunch students benefit more from attending a small class in the Tennessee class size experiment than other students. However, Hoxby (2000) finds no evidence of class size effects at schools with larger disadvantaged populations.

²⁹ For previous evidence on the effect of CSR on non-cognitive outcomes, see Dee and West (2008).

³⁰ The incidents of crime and violence and suspension variables are calculated by aggregating (to the district level) school-level data for schools that serve students in at least one of the grades four to eight but no students in grades nine to 12.

The district-level evidence suggests that mandated CSR did not have a positive effect on student achievement above and beyond the effect of equivalent additional resources. In some cases, modest positive effects cannot be ruled out. For example, extrapolating from the early grades in the Tennessee STAR experiment to middle school grades in Florida would lead one to expect a 2.6-student decrease in class size for one year to increase achievement by about 0.04 standard deviations.³¹ One cannot reject (with 95 percent confidence) an effect of this magnitude in seventh- and eighth-grade math, but can do so in reading. Additionally, the negative point estimates for middle school reading scores raise the possibility that comparison districts were able to spend the additional resources more productively than the treated districts, which were forced to spend it on CSR.

School-Level Results

Although the school-level analysis does not have the advantage of CSR coming as a surprise to schools (as it did to districts), it offers the advantages of much greater statistical power and the opportunity estimate the effect of CSR at a point when general equilibrium effects (such as reduced teacher quality) are likely to be greater. Because treated schools received more resources than comparison schools in two of the three post-treatment years that I examine, the results of this analysis should be interpreted as the effect of CSR that included additional resources greater than those received by the comparison schools. However, it will not be

³¹ This extrapolation is made by taking the one-year effect of being assigned to a small class in the STAR experiment from the first column of Table IX in Krueger (1999), converting it to standard deviation units, dividing it by seven (roughly the amount of CSR in Project STAR), and then multiplying it by 2.6 (the amount of CSR in seventh and eighth grade in Florida after three years of district-level implementation). However, such an extrapolation may be inappropriate because estimates from the STAR experiment include the effect of additional resources whereas my estimates from Florida do not.

possible to separate out general equilibrium effects from additional resource effects, and it should be noted that these two potential effects are expected to have opposite signs.

Before turning to the school-level results it is instructive to examine the impact of mandated CSR at the school level on average class size. Table 5 shows that, prior to school-level CSR implementation, class size was decreasing by 0.7 students per year in the comparison schools and 0.5 students per year in the treated schools. By the third year of school-level CSR implementation, class size fell by 2.7 students more in the treated schools than in the comparison schools. This effect was somewhat concentrated in fourth and fifth grade, where the effect was about three students (as compared to 1.3–1.9 students in grades six to eight).³² Consequently, if CSR had an effect on achievement in any of the grades from four to eight we would expect to find it in the school-level analysis (unlike in the district-level analysis, which showed that class size in grades four and five fell by similar amounts in the comparison and treated districts).³³

The results for FCAT math and reading scores, shown in Tables 6a and 6b, indicate that even small positive effects can be ruled out. Table 6a shows that math scores were increasing at similar rates in both treated and comparison schools prior to school-level CSR. But by the third year of school-level implementation, math scores fell by 0.024 standard deviations more in the treated schools than in the comparison schools, an effect that is statistically significant. Effects disaggregated by grade are almost all negative, but only the fourth-grade effect is statistically significant.

³² Estimations using student-level class-size data yield modestly larger three-year effects: 2.7 students in grades four to eight, 3.6 in grade four, 3.5 in grade five, 2.2 in grade six, 2.2 in grade seven, and 1.8 in grade eight.

³³ Another reason why we might expect to find class-size effects in the school-level analysis is that treated schools reduced class sizes in the earlier grades by a larger amount than comparison schools. For example, the three-year effect of being in the grade four to eight treatment group on average class size in third grade was 2.3 students (the effect on second-grade class size was 1.9 students).

The results for reading scores (Table 6b) indicate a similar negative effect of CSR (0.032 standard deviations), which is also statistically significant. Results by grade are all negative, with statistically significant estimates of negative effects of 0.059 and 0.041 standard deviations in fourth and fifth grade, respectively. The lack of positive CSR effects is robust to the inclusion of prior-year controls (including test scores), as shown in Appendix Table 10, although dropping the first year of pre-treatment data yields effects that are no longer statistically significantly negative.³⁴

Results from five alternative specifications are reported in Appendix Table 11. The statistically significant negative results are robust to including district-specific linear time trends (to control for district-level policy changes), to excluding charter schools, and to weighting all schools equally. However, they are not robust to dropping the first two years of pre-treatment data or to the standard difference-in-differences specification, although neither of these two alternative specifications yields a positive CSR effect.³⁵ The lack of negative findings in the different-in-differences specifications may reflect the slightly different pre-treatment achievement trends at treated and comparison schools in some subject-grade combinations, which are controlled for in my preferred estimates but not in the difference-in-differences estimates.³⁶ A fully flexible specification (with year dummies and treatment-year interactions)

³⁴ When results (not shown) are estimated separately for elementary and middle grades (four and five and six to eight, respectively) with varying numbers of years of pre-treatment data, the only robust statistically significant finding is a negative impact on elementary reading scores of 0.044–0.077 standard deviations.

³⁵ Difference-in-differences estimates disaggregated by grade (not shown) indicate a small positive effect of 0.023 standard deviations on fourth-grade math scores and 0.026 standard deviations on fifth-grade reading scores, although all other effects are trivially small and statistically insignificant.

³⁶ I do not report school-level results using Stanford Achievement Test (SAT) scores because Florida stopped administering the SAT in 2009.

generally yields results that are close to zero, although the three-year effect in elementary reading is negative and statistically significant (Appendix Table 12).

The school-level CSR effects do not differ markedly by student demographics, although Appendix Table 13 suggests that the negative effects in math and reading (in my preferred model) are concentrated among black and Hispanic students and students eligible for free or reduced-price lunch. I also do not find much evidence of effects on non-cognitive outcomes, with the only statistically significant effects in Appendix Table 14 indicating that CSR reduced the number of days elementary school students were absent by 0.3 percentage points (about 0.06 student-level standard deviations) after three years and the percent of students receiving an out-of-school suspension by 0.4 percentage points (about 0.06 school-level standard deviations) after one year.

An important limitation of the results reported above is that they do not directly examine the effect of CSR on students in the earlier elementary grades, which results primarily from the fact that Florida only begins testing students in third grade. In the district-level analysis it was not possible to examine third-grade students because almost all districts were in the treated group for this grade. However, in the school-level analysis it is possible to examine third-grade test scores (classifying schools into treated and comparison groups based on their 2006 average class sizes in grades prekindergarten to three).³⁷ The results, which are reported in Table 7, indicate that CSR decreased achievement in math and reading by 0.040 and 0.052 standard deviations respectively. Both effects are statistically significant, but the math effect is not robust to

³⁷ Regression results similar to those reported in Table 5 (not shown) indicate that average class size in third grade in the treated schools fell by 2.4 students more than in the comparison schools over the first three years of school-level implementation. The third-grade results will also reflect CSR in first and second grade, which experienced three-year relative reductions of 1.9 and 2.0 students, respectively.

excluding the first year of pre-treatment data and neither result is robust to a standard difference-in-differences specification (Appendix Table 15).³⁸

5. Conclusions

The results from both the district- and school-level analyses indicate that the effects of mandated CSR in Florida on cognitive and non-cognitive outcomes were small at best and most likely close to zero. The preferred estimates from the district-level analysis indicate that, after three years of implementation, student achievement in the treated districts was (a statistically insignificant) 0.035 standard deviations higher in math and no higher in reading than it would have been had these districts received equivalent resources without a CSR mandate. One might not expect a large effect given that over three years class size was only reduced by 1.5 students more in the treated districts than in the comparison districts, but I also find no evidence of positive CSR effects in grades seven and eight, where the relative reduction in class size was three students. In fact, the preferred reading estimate for these grades is negative and statistically significant.

One limitation of the district-level analysis is that small positive effects of CSR cannot generally be ruled out, but this is not the case in the school-level analysis. The latter results indicate that, after three years of implementation, math and reading scores at the treated schools were either no different from or modestly lower than they would have been had these schools received *less* funding per pupil and not been required to reduce class size. The school-level

³⁸ A fully flexible specification including year dummies and treatment-year interactions (not shown) yields estimates that are all close to zero and precisely estimated, with the exception of the three-year effect in reading, which is negative and statistically significant (with a point estimate of -0.025 standard deviations).

analysis can also be applied to third-grade math and reading scores, which yield similar estimates.

It is difficult to compare these results to others from the class size literature because most prior studies do not compare the effect of reducing class size to the effect of providing equivalent additional resources to schools. For example, in the STAR experiment Tennessee provided extra resources to schools to implement CSR, but these resources were concentrated on students assigned to small classes.³⁹ Thus, it is impossible to disentangle the effect of reducing class size from the effect of providing additional resources. In the present study, the students in the comparison districts all potentially benefited from the additional resources and thus the results indicate the marginal effect of reducing class size relative to the outcomes produced by equivalent resources. In the school-level analysis the comparison schools received less additional resources than the treated schools, but assuming that these resources have a positive effect implies even larger negative effects of CSR on student achievement than those reported above.

The findings reported in this paper do not apply to all aspects of Florida's CSR policy, particularly its coverage of prekindergarten to second grade and grades nine to 12. It may well be that the policy had a larger effect on these grades. And it remains a possibility that the resources provided to districts and schools as a result of the CSR mandate had positive effects on both the comparison and treated districts/schools in this study. But the results of this study do strongly suggest that large-scale, untargeted CSR mandates are not a particularly productive use of limited educational resources.

³⁹ An aide was provided to some regular size classes, although student achievement was not significantly higher in these classes than in regular size classes without an aide (Krueger 1999).

References

- Angrist, Joshua D. and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate The Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2): 533–575.
- Ballou, Dale. 1996. "Do Public Schools Hire the Best Applicants?" *Quarterly Journal of Economics*, 111(1): 97–133.
- Bohrnstedt, George W. and Brian M. Stecher, eds. 2002. *What We Have Learned About Class Size Reduction in California*. Palo Alto, CA: CSR Research Consortium.
- Dee, Thomas and Brian Jacob. 2009. "The Impact of No Child Left Behind on Student Achievement." Cambridge, MA: NBER Working Paper No. 15531.
- Dee, Thomas and Martin West. 2008. "The Non-Cognitive Returns to Class Size." Cambridge, MA: NBER Working Paper No. 13994.
- Education Commission of the States. 2005. "State Class-Size Reduction Measures." Denver, Colorado: Education Commission of the States.
- Glass, Gene V. and Mary L. Smith. 1978. *Meta-analysis of Research on Class Size and Achievement*. San Francisco, CA: Far West Laboratory.
- Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, 24(3): 1141–77.
- Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics*, 115(4): 1239–1285.
- Jepsen, Christopher and Steven Rivkin. 2009. "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size." *Journal of Human Resources*, 44(1): 223–250.
- Kane, Thomas J. and Douglas O. Staiger. 2005. "Using Imperfect Information to Identify Effective Teachers." Unpublished Paper.
- Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 115(2): 497–532.
- Rockoff, Jonah. 2009. "Field Experiments in Class Size from the Early Twentieth Century." *Journal of Economic Perspectives*, 23(4): 211–30.
- Wößmann, Ludger and Martin West. 2006. "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS." *European Economic Review* 50 (3) 695–736.

Figure 1. Location of Comparison and Treated Districts

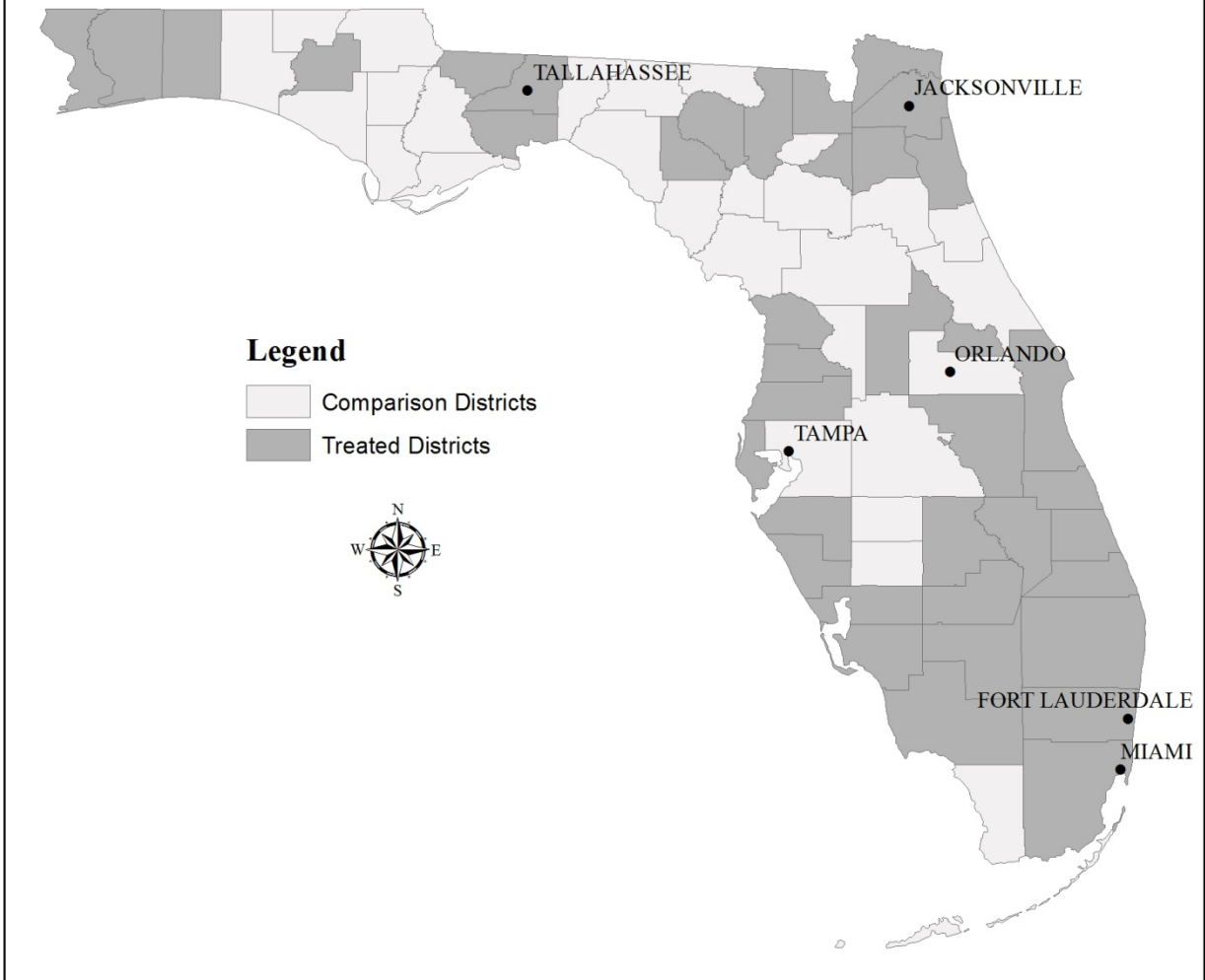


Figure 2a. 8th-Grade FCAT Math Scores, 2001-2006

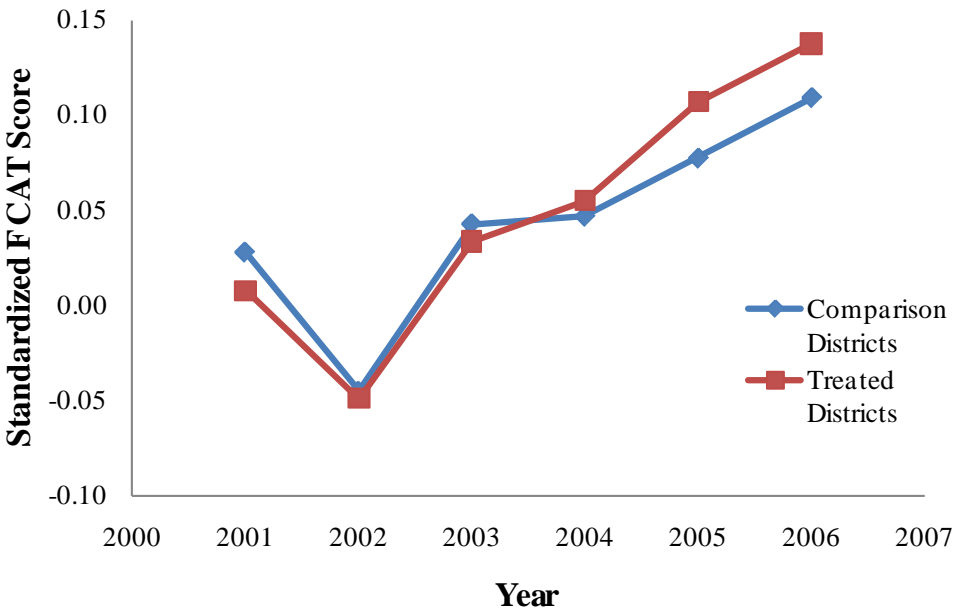


Figure 2b. 8th-Grade FCAT Reading Scores, 2001-2006

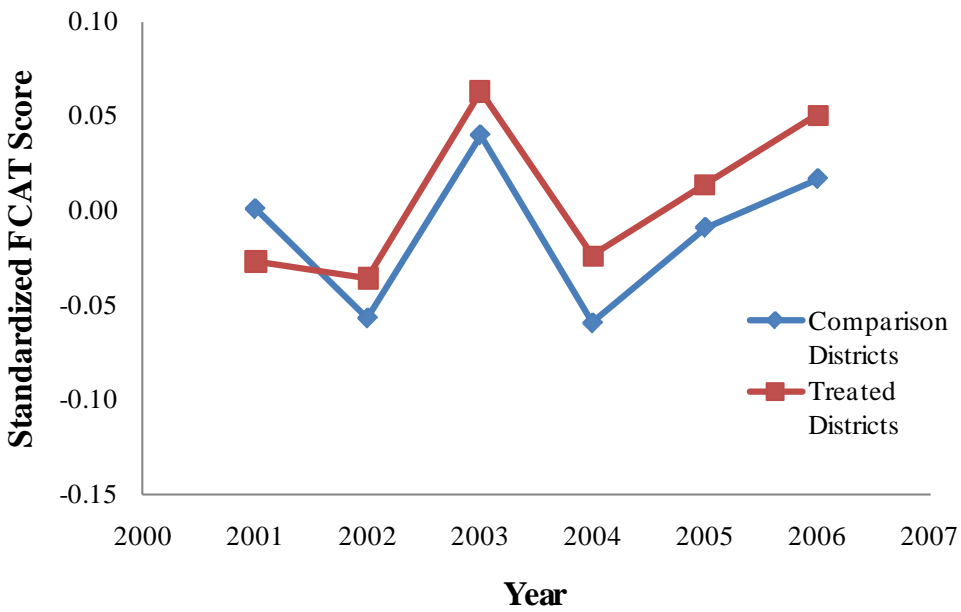


Table 1

Pre-Treatment (2003) Characteristics of Treated and Comparison Districts

	Comparison	Treated	Difference
Class Size (Official), Grades 4-8	21.3	25.4	4.1**
Class Size (Author's Calculation), Grades 4-8	20.4	22.9	2.6**
Per-Pupil Expenditure (2008 \$)	\$9,317	\$9,303	-\$14
Average Salary, Teacher with BA (2008 \$)	\$39,687	\$44,767	\$5,080**
Enrollment, Grades 4-8	41,623	63,202	21,579
Percent Black, Grades 4-8	0.24	0.25	0.01
Percent Hispanic, Grades 4-8	0.17	0.23	0.05
Percent Eligible for Free/Reduced Lunch	0.48	0.44	-0.04
Percent Districts Treated in Grades PK-3	0.97	1.00	0.02
Accountability Grades	3.15	3.00	-0.14
Percent New Teachers, Grades 4-8	0.06	0.05	-0.01
Average Teacher Experience, Grades 4-8	11.4	10.7	-0.7
FCAT Math Scores (Standardized), Grades 4-8	0.042	0.056	0.014
FCAT Reading Scores (Standardized), Grades 4-8	0.031	0.057	0.026
Number of Districts (Unweighted)	28	39	

Notes: ** $p < 0.01$, * $p < 0.05$; significance levels are based on standard errors that are adjusted for clustering at the district level. All statistics are weighted by district enrollment in grades four to eight. Official class size data and accountability grades are from the Florida Department of Education (FLDOE); author's class size calculations, average teacher salary, percent new teachers, average teacher experience, and FCAT scores are from the FLDOE's Education Data Warehouse (EDW); per-pupil expenditures, enrollment counts, and demographic breakdowns are from the Common Core of Data. Accountability grades are average of school-level grades (weighted by student enrollment, with A-F ratings placed on a 0-4 GPA-type scale). A district is identified as being "treated" in grades PK-3 if its average class size in those grades was more than 18 in 2003.

Table 2

Pre-Treatment (2006) Characteristics of Treated and Comparison Schools

	Comparison	Treated	Difference
Class Size (Official), Grades 4-8	18.9	24.4	5.5**
Class Size (Author's Calculation), Grades 4-8	18.6	21.6	3.0**
Per-Pupil Expenditure (2008 \$)	\$5,983	\$5,997	\$14
Enrollment, Grades 4-8	900	1,100	200**
Percent Black, Grades 4-8	0.25	0.21	-0.04
Percent Hispanic, Grades 4-8	0.21	0.34	0.13
Percent Eligible for Free/Reduced Lunch	0.51	0.48	-0.03
Percent Districts Treated in Grades PK-3	0.25	0.55	0.30**
Accountability Grades	3.40	3.56	0.16**
Percent New Teachers, Grades 4-8	0.12	0.13	0.01
Average Teacher Experience, Grades 4-8	10.4	10.1	-0.30
FCAT Math Scores (Standardized), Grades 4-8	0.201	0.322	0.121**
FCAT Reading Scores (Standardized), Grades 4-8	0.186	0.297	0.111**
Number of Schools (Unweighted)	2,106	664	

Notes: ** $p < 0.01$, * $p < 0.05$; significance levels are based on standard errors that are adjusted for clustering at the school level. All statistics are weighted by school enrollment. Official class size data, accountability grades, and per-pupil expenditures are from the Florida Department of Education (FLDOE); author's class size calculations, percent new teachers, average teacher experience, and FCAT scores are from the FLDOE's Education Data Warehouse (EDW); enrollment counts and demographic breakdowns are from the Common Core of Data. Accountability grades (A-F) are placed on a 0-4 GPA-type scale. A school is identified as being "treated" in grades PK-3 if its average class size in those grades was more than 18 in 2006.

Table 3

Effect of Required CSR at District Level on Average Class Size (Number of Students per Class)

	Grade(s):					
	4-8	4	5	6	7	8
YEAR	-0.1 [0.2]	0.1 [0.2]	0.1 [0.2]	-0.1 [0.2]	-0.2 [0.1]	-0.1 [0.2]
CSR	0.5 [0.3]	0.3 [0.8]	-0.4 [0.8]	0.1 [0.2]	0.5 [0.2]*	0.5 [0.3]
YR_SINCE_CSR	-0.6 [0.2]**	-1.2 [0.2]**	-0.7 [0.2]**	-0.3 [0.2]	-0.4 [0.2]*	-0.5 [0.2]*
T x YEAR	-0.0 [0.2]	-0.1 [0.2]	-0.2 [0.2]	-0.1 [0.2]	0.2 [0.2]	0.1 [0.2]
T x CSR	-0.5 [0.5]	-1.1 [0.8]	-0.3 [0.9]	-0.2 [0.4]	-0.6 [0.5]	-0.4 [0.5]
T x YR_SINCE_CSR	-0.4 [0.2]	0.6 [0.3]*	0.3 [0.3]	-0.5 [0.3]	-0.7 [0.3]*	-0.7 [0.3]*
Total effect by 2006	-1.5 [0.6]*	0.8 [1.4]	0.5 [1.4]	-1.6 [0.9]	-2.7 [0.6]**	-2.5 [0.6]**
Observations (District*Year)	402	402	402	402	402	402
R-squared	0.88	0.76	0.75	0.85	0.88	0.87

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. All regressions include district fixed effects and are weighted by district enrollment. Data cover period from 2001 to 2006.

Table 4a

Effects of District-Level CSR on FCAT Math Scores (Student-Level Standard Deviations)

	Grade(s)					
	4-8	4	5	6	7	8
YEAR	0.046 [0.009]**	0.080 [0.012]**	0.040 [0.017]*	0.070 [0.010]**	0.024 [0.010]*	0.012 [0.008]
CSR	0.036 [0.012]**	0.107 [0.016]**	0.018 [0.015]	-0.026 [0.012]*	0.042 [0.016]**	0.051 [0.033]
YR_SINCE_CSR	0.015 [0.010]	-0.006 [0.011]	0.025 [0.017]	-0.013 [0.012]	0.040 [0.013]**	0.030 [0.025]
T x YEAR	0.002 [0.009]	0.006 [0.012]	-0.005 [0.019]	0.008 [0.011]	0.007 [0.009]	-0.003 [0.008]
T x CSR	0.017 [0.014]	0.039 [0.027]	0.027 [0.017]	-0.009 [0.019]	0.013 [0.018]	0.023 [0.038]
T x YR_SINCE_CSR	0.006 [0.012]	0.003 [0.018]	0.002 [0.020]	0.016 [0.016]	-0.002 [0.013]	0.005 [0.026]
Total effect by 2006	0.035 [0.031]	0.049 [0.039]	0.034 [0.065]	0.038 [0.040]	0.008 [0.031]	0.039 [0.055]
Observations (Student*Year)	5,476,526	1,081,032	1,091,624	1,097,709	1,113,843	1,092,318
R-squared	0.27	0.26	0.25	0.28	0.27	0.29

Notes: ** $p < 0.01$, * $p < 0.05$; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2006.

Table 4b

Effects of District-Level CSR on FCAT Reading Scores (Student-Level Standard Deviations)

	Grade(s)					
	4-8	4	5	6	7	8
YEAR	0.032 [0.008]**	0.046 [0.015]**	0.075 [0.018]**	0.005 [0.008]	0.011 [0.008]	0.022 [0.008]**
CSR	0.028 [0.018]	0.247 [0.014]**	0.023 [0.028]	0.024 [0.014]	-0.058 [0.016]**	-0.084 [0.033]*
YR_SINCE_CSR	0.049 [0.012]**	-0.050 [0.018]**	0.026 [0.013]	0.096 [0.014]**	0.123 [0.019]**	0.051 [0.027]
T x YEAR	0.003 [0.010]	0.004 [0.017]	-0.023 [0.020]	0.003 [0.009]	0.018 [0.009]	0.013 [0.008]
T x CSR	-0.012 [0.018]	0.004 [0.022]	0.011 [0.029]	-0.069 [0.032]*	-0.012 [0.017]	0.019 [0.033]
T x YR_SINCE_CSR	0.004 [0.017]	0.001 [0.021]	0.038 [0.022]	0.025 [0.026]	-0.026 [0.023]	-0.024 [0.026]
Total effect by 2006	-0.001 [0.043]	0.008 [0.062]	0.124 [0.071]	0.007 [0.060]	-0.089 [0.059]	-0.052 [0.052]
Observations (Student*Year)	5,485,417	1,082,756	1,093,594	1,098,789	1,115,013	1,095,265
R-squared	0.26	0.26	0.27	0.27	0.25	0.27

Notes: ** $p < 0.01$, * $p < 0.05$; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2006.

Table 5

Effect of Required CSR at School Level on Average Class Size (Number of Students per Class)

	Grade(s):					
	4-8	4	5	6	7	8
YEAR	-0.7 [0.0]**	-0.7 [0.0]**	-0.7 [0.0]**	-0.6 [0.0]**	-0.6 [0.0]**	-0.6 [0.0]**
CSR	-0.6 [0.1]**	-0.8 [0.1]**	-1.0 [0.1]**	-0.6 [0.1]**	-0.8 [0.1]**	-0.4 [0.2]
YR_SINCE_CSR	0.6 [0.0]**	0.8 [0.0]**	0.8 [0.0]**	0.3 [0.1]**	0.3 [0.1]**	0.1 [0.1]
T x YEAR	0.2 [0.0]**	0.3 [0.0]**	0.2 [0.0]**	0.1 [0.1]	0.0 [0.1]	-0.0 [0.1]
T x CSR	-0.6 [0.2]**	-0.4 [0.2]	0.1 [0.2]	-0.9 [0.3]**	-0.6 [0.2]**	-0.8 [0.4]*
T x YR_SINCE_CSR	-0.7 [0.1]**	-0.9 [0.1]**	-1.0 [0.1]**	-0.3 [0.1]*	-0.4 [0.1]**	-0.2 [0.2]
Total effect by 2009	-2.7 [0.2]**	-3.2 [0.3]**	-3.0 [0.3]**	-1.9 [0.3]**	-1.9 [0.3]**	-1.3 [0.4]**
Observations (School*Year)	22,729	16,109	16,199	7,684	7,042	7,299
R-squared	0.73	0.55	0.57	0.75	0.79	0.77

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. All regressions include school fixed effects and are weighted by school enrollment. Data cover period from 2001 to 2009.

Table 6a

Effects of School-Level CSR on FCAT Math Scores (Student-Level Standard Deviations)

	Grade(s)					
	4-8	4	5	6	7	8
YEAR	0.070 [0.001]**	0.113 [0.002]**	0.059 [0.002]**	0.066 [0.002]**	0.067 [0.003]**	0.046 [0.002]**
CSR	-0.023 [0.003]**	-0.080 [0.006]**	0.016 [0.005]**	-0.092 [0.007]**	0.028 [0.006]**	0.025 [0.006]**
YR_SINCE_CSR	-0.020 [0.002]**	-0.029 [0.003]**	-0.032 [0.003]**	-0.005 [0.004]	-0.037 [0.003]**	-0.007 [0.003]*
T x YEAR	0.004 [0.002]	0.013 [0.003]**	-0.000 [0.003]	0.003 [0.004]	-0.000 [0.004]	0.001 [0.004]
T x CSR	-0.006 [0.006]	-0.008 [0.012]	-0.019 [0.010]	0.013 [0.012]	-0.007 [0.010]	-0.009 [0.010]
T x YR_SINCE_CSR	-0.006 [0.003]	-0.012 [0.006]*	0.007 [0.005]	-0.007 [0.007]	-0.002 [0.006]	-0.004 [0.005]
Total effect by 2009	-0.024 [0.010]*	-0.045 [0.016]**	0.002 [0.014]	-0.009 [0.017]	-0.013 [0.017]	-0.022 [0.016]
Observations (Student*Year)	8,293,046	1,646,586	1,651,779	1,668,180	1,673,878	1,652,623
R-squared	0.30	0.29	0.29	0.31	0.31	0.32

Notes: ** $p < 0.01$, * $p < 0.05$; robust standard errors adjusted for clustering at the school level appear in brackets. Dependent variables are FCAT developmental scale scores in math, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2009.

Table 6b

Effects of School-Level CSR on FCAT Reading Scores (Student-Level Standard Deviations)

	Grade(s)					
	4-8	4	5	6	7	8
YEAR	0.066 [0.001]**	0.080 [0.002]**	0.090 [0.001]**	0.066 [0.002]**	0.067 [0.002]**	0.031 [0.002]**
CSR	-0.025 [0.003]**	-0.115 [0.005]**	0.032 [0.005]**	-0.042 [0.006]**	0.015 [0.006]**	-0.005 [0.006]
YR_SINCE_CSR	-0.013 [0.002]**	-0.015 [0.003]**	-0.077 [0.002]**	-0.013 [0.003]**	-0.012 [0.003]**	0.045 [0.003]**
T x YEAR	0.006 [0.002]**	0.010 [0.002]**	0.014 [0.002]**	0.000 [0.003]	-0.000 [0.003]	0.001 [0.003]
T x CSR	0.002 [0.005]	0.015 [0.009]	-0.018 [0.009]*	0.015 [0.010]	0.016 [0.010]	-0.014 [0.010]
T x YR_SINCE_CSR	-0.011 [0.003]**	-0.025 [0.005]**	-0.008 [0.004]	-0.007 [0.006]	-0.010 [0.006]	-0.003 [0.005]
Total effect by 2009	-0.032 [0.010]**	-0.059 [0.013]**	-0.041 [0.011]**	-0.004 [0.015]	-0.013 [0.015]	-0.022 [0.014]
Observations (Student*Year)	8,304,672	1,648,737	1,654,664	1,669,663	1,675,505	1,656,103
R-squared	0.29	0.28	0.29	0.29	0.28	0.3

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. Dependent variables are FCAT developmental scale scores in reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2009.

Table 7

Effects of School-Level CSR in Grades PK-3 on 3rd-Grade
FCAT Scores (Student-Level Standard Deviations)

	Math	Reading
YEAR	0.109 [0.002]**	0.091 [0.001]**
CSR	-0.017 [0.007]*	-0.082 [0.005]**
YR_SINCE_CSR	-0.031 [0.004]**	-0.052 [0.003]**
T x YEAR	0.013 [0.003]**	0.007 [0.002]**
T x CSR	-0.007 [0.013]	0.010 [0.010]
T x YR_SINCE_CSR	-0.011 [0.007]	-0.021 [0.005]**
Total effect by 2009	-0.040 [0.017]*	-0.052 [0.012]**
Observations (Student*Year)	1,716,312	1,717,977
R-squared	0.29	0.27

Notes: ** $p < 0.01$, * $p < 0.05$; robust standard errors adjusted for clustering at the school level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2009.

Appendix Table 1

Effect of Required CSR at District Level on District Characteristics

	% Black	% Hisp	% FRL	Log(Enroll)	Log(PPS)	Log(Salary)
YEAR	0.000 [0.001]	0.010 [0.001]**	0.002 [0.007]	0.019 [0.004]**	-0.008 [0.010]	-0.003 [0.009]
CSR	-0.003 [0.001]**	-0.004 [0.001]**	0.004 [0.006]	0.004 [0.004]	-0.027 [0.031]	0.019 [0.007]*
YR_SINCE_CSR	-0.003 [0.001]	0.003 [0.001]**	0.003 [0.012]	0.005 [0.002]*	0.051 [0.027]	0.002 [0.010]
T x YEAR	-0.001 [0.002]	-0.000 [0.002]	0.005 [0.008]	0.002 [0.006]	0.017 [0.013]	0.003 [0.011]
T x CSR	-0.000 [0.001]	0.003 [0.001]*	0.013 [0.008]	0.005 [0.005]	0.007 [0.036]	-0.093 [0.064]
T x YR_SINCE_CSR	0.003 [0.002]	-0.003 [0.002]	-0.012 [0.013]	-0.016 [0.005]**	-0.004 [0.032]	0.016 [0.024]
Total effect by 2006	0.008 [0.004]	-0.007 [0.004]	-0.023 [0.033]	-0.042 [0.013]**	-0.006 [0.074]	-0.046 [0.029]
Observations (District*Year)	402	402	402	402	402	402
R-squared	1.00	1.00	0.96	1.00	0.80	0.68

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. All regressions include district fixed effects and are weighted by district enrollment. Data cover period from 2001 to 2006.

Appendix Table 2

Effect of Required CSR at School Level on School Characteristics

	% Black	% Hisp	% FRL	Log(Enroll)	Log(PPS)
YEAR	0.002 [0.000]**	0.009 [0.000]**	0.011 [0.000]**	-0.005 [0.001]**	0.025 [0.001]**
CSR	-0.007 [0.001]**	0.001 [0.001]*	-0.041 [0.001]**	-0.002 [0.004]	0.105 [0.003]**
YR_SINCE_CSR	0.001 [0.000]*	-0.005 [0.000]**	0.014 [0.001]**	-0.014 [0.002]**	-0.024 [0.002]**
T x YEAR	-0.002 [0.001]**	-0.000 [0.001]	-0.006 [0.001]**	0.002 [0.002]	0.004 [0.001]*
T x CSR	0.001 [0.001]	-0.003 [0.001]*	-0.006 [0.003]*	0.002 [0.008]	0.070 [0.007]**
T x YR_SINCE_CSR	0.003 [0.001]**	-0.000 [0.001]	0.003 [0.001]*	-0.002 [0.005]	-0.021 [0.004]**
Total effect by 2009	0.009 [0.002]**	-0.004 [0.002]	0.003 [0.004]	-0.003 [0.012]	0.008 [0.008]
Observations (School*Year)	28,137	28,137	28,099	28,137	23,629
R-squared	0.99	0.99	0.95	0.96	0.80

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. All regressions include school fixed effects and are weighted by school enrollment. Data cover period from 2001 to 2009.

Appendix Table 3

District-Level Models with Additional Years of Pre-Treatment Data (Effects in Student-Level Standard Deviations)

	FCAT Math, Grade 5				FCAT Reading, Grade 4			
	Number of Years of Pre-Treatment Data				Number of Years of Pre-Treatment Data			
	5	4	3	2	5	4	3	2
T x CSR	0.017	0.014	0.027	0.037	0.008	-0.001	0.004	0.014
	[0.017]	[0.016]	[0.017]	[0.018]*	[0.029]	[0.024]	[0.022]	[0.023]
T x YR_SINCE_CSR	-0.010	-0.014	0.002	0.039	0.001	-0.006	0.001	0.042
	[0.017]	[0.015]	[0.020]	[0.030]	[0.013]	[0.015]	[0.021]	[0.026]
Total effect by 2006	-0.012	-0.027	0.034	0.153	0.012	-0.020	0.008	0.140
	[0.052]	[0.045]	[0.065]	[0.101]	[0.047]	[0.042]	[0.062]	[0.079]
Observations (Student*Year)	1,439,422	1,270,750	1,091,624	913,087	1,432,940	1,262,783	1,082,756	903,612
R-squared	0.28	0.26	0.25	0.24	0.29	0.27	0.26	0.25

	FCAT Math, Grade 8				FCAT Reading, Grade 8			
	Number of Years of Pre-Treatment Data				Number of Years of Pre-Treatment Data			
	5	4	3	2	5	4	3	2
T x CSR	0.022	0.017	0.023	0.025	0.030	0.025	0.019	0.025
	[0.033]	[0.036]	[0.038]	[0.040]	[0.027]	[0.029]	[0.033]	[0.036]
T x YR_SINCE_CSR	-0.001	-0.003	0.005	0.018	-0.019	-0.021	-0.024	-0.002
	[0.031]	[0.028]	[0.026]	[0.021]	[0.033]	[0.030]	[0.026]	[0.023]
Total effect by 2006	0.020	0.008	0.039	0.080	-0.026	-0.037	-0.052	0.018
	[0.072]	[0.060]	[0.055]	[0.045]	[0.078]	[0.070]	[0.052]	[0.054]
Observations (Student*Year)	1,414,552	1,258,940	1,092,318	925,331	1,417,616	1,261,749	1,095,265	928,155
R-squared	0.30	0.30	0.29	0.28	0.28	0.28	0.27	0.26

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 1999, 2000, 2001, or 2002 to 2006.

Appendix Table 4

District-Level Estimates that Condition on Prior-Year Controls (Effects in Student-Level Standard Deviations)

	FCAT Math, Grades 4-8				FCAT Reading, Grades 4-8			
T x CSR	0.017 [0.014]	0.021 [0.015]	0.019 [0.016]	0.029 [0.016]	-0.012 [0.018]	-0.005 [0.019]	-0.009 [0.021]	0.002 [0.011]
T x YR_SINCE_CSR	0.006 [0.012]	0.024 [0.014]	0.023 [0.015]	0.014 [0.019]	0.004 [0.017]	0.033 [0.016]	0.037 [0.020]	0.039 [0.022]
Total effect by 2006	0.035 [0.031]	0.093 [0.044]*	0.089 [0.045]	0.070 [0.066]	-0.001 [0.043]	0.094 [0.042]*	0.102 [0.048]*	0.118 [0.065]
Data from 2000-01 excluded?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Students missing prior-year data excluded?	No	No	Yes	Yes	No	No	Yes	Yes
Students prior-year controls included?	No	No	No	Yes	No	No	No	Yes
Observations (Student*Year)	5,476,526	4,599,367	4,049,020	4,049,020	5,485,417	4,607,296	4,054,914	4,054,914
R-squared	0.27	0.26	0.26	0.70	0.26	0.25	0.25	0.68

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Student prior-year controls include test scores in both subjects (and their cubed and squared terms), whether the student made a nonstructural or structural move from the previous year, the number of days the student was absent the previous year, and whether the student was repeating a grade. Data cover period from 2001 or 2002 to 2006.

Appendix Table 5

District-Level Analysis, Alternative Specifications (Effects in Student-Level Standard Deviations)

	FCAT Math, Grades 4-8				
	District Trends	No Charters	All Charters	Un-weighted	Diff-in-Diffs
T x CSR	0.019 [0.013]	0.016 [0.014]	0.016 [0.014]	0.010 [0.017]	0.036 [0.022]
T x YR_SINCE_CSR	0.001 [0.011]	0.007 [0.012]	0.006 [0.012]	-0.006 [0.014]	
Total effect by 2006	0.022 [0.029]	0.036 [0.030]	0.033 [0.030]	-0.008 [0.042]	
Observations (Student*Year)	5,476,526	5,448,411	5,589,472	5,476,526	5,476,526
R-squared	0.27	0.27	0.27	0.28	0.27

	FCAT Reading, Grades 4-8				
	District Trends	No Charters	All Charters	Un-weighted	Diff-in-Diffs
T x CSR	-0.005 [0.017]	-0.012 [0.018]	-0.014 [0.018]	0.006 [0.017]	0.005 [0.022]
T x YR_SINCE_CSR	-0.005 [0.014]	0.004 [0.017]	0.004 [0.017]	-0.015 [0.013]	
Total effect by 2006	-0.020 [0.036]	0.000 [0.042]	-0.002 [0.043]	-0.038 [0.039]	
Observations (Student*Year)	5,485,417	5,457,314	5,598,708	5,485,417	5,485,417
R-squared	0.26	0.26	0.26	0.27	0.26

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. "District Trends" also include district-specific linear time trends. "No Charters" indicates that all charter schools are excluded. "All Charters" indicates that all charter schools (including those in operation in 2003) are included. "Unweighted" indicates that each district is weighted equally. "Diff-in-Diffs" is a standard difference-in-differences specifications that also includes grade-by-year fixed effects. Data cover period from 2001 to 2006.

Appendix Table 6

District-Level Analysis, Fully Flexible Specification (Effects in Student-Level Standard Deviations)

	FCAT Math			FCAT Reading		
	Grades 4-8	Grades 4-5	Grades 6-8	Grades 4-8	Grades 4-5	Grades 6-8
T x YEAR2001	-0.006 [0.017]	-0.000 [0.027]	-0.010 [0.014]	-0.008 [0.017]	0.015 [0.032]	-0.024 [0.016]
T x YEAR2002	0.015 [0.017]	0.021 [0.024]	0.011 [0.015]	0.024 [0.013]	0.043 [0.022]	0.010 [0.012]
T x YEAR2004 (Effect by 2004)	0.030 [0.010]**	0.039 [0.013]**	0.025 [0.010]*	0.005 [0.009]	0.029 [0.015]	-0.008 [0.010]
T x YEAR2005 (Effect by 2005)	0.042 [0.023]	0.055 [0.026]*	0.036 [0.025]	0.007 [0.025]	0.039 [0.027]	-0.012 [0.027]
T x YEAR2006 (Effect by 2006)	0.045 [0.025]	0.045 [0.024]	0.044 [0.029]	0.016 [0.032]	0.046 [0.026]	-0.004 [0.040]
Observations (Student*Year)	5,476,526	2,172,656	3,303,870	5,485,417	2,176,350	3,309,067
R-squared	0.27	0.26	0.28	0.26	0.26	0.26

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include year dummies, district fixed effects, and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2006.

Appendix Table 7

Effects of District-Level CSR on Stanford Achievement Test (SAT) Scores (Student-Level Standard Deviations)

	SAT Math Scores in Grade(s)					
	4-8	4	5	6	7	8
T x CSR	0.021 [0.014]	0.030 [0.028]	0.039 [0.016]*	-0.014 [0.021]	0.016 [0.016]	0.041 [0.035]
T x YR_SINCE_CSR	0.005 [0.009]	0.013 [0.011]	0.014 [0.018]	0.011 [0.018]	-0.008 [0.012]	-0.010 [0.024]
Total effect by 2006	0.037 [0.030]	0.068 [0.022]**	0.080 [0.058]	0.019 [0.051]	-0.007 [0.043]	0.010 [0.051]
Observations (Student*Year)	5,429,421	1,075,344	1,085,155	1,087,317	1,101,688	1,079,917
R-squared	0.26	0.24	0.24	0.27	0.27	0.28

	SAT Reading Scores in Grade(s)					
	4-8	4	5	6	7	8
T x CSR	0.012 [0.013]	0.034 [0.029]	0.008 [0.018]	-0.018 [0.023]	0.018 [0.015]	0.019 [0.023]
T x YR_SINCE_CSR	0.008 [0.014]	0.005 [0.013]	0.046 [0.022]*	0.009 [0.015]	-0.020 [0.017]	-0.001 [0.024]
Total effect by 2006	0.036 [0.043]	0.050 [0.040]	0.145 [0.065]*	0.010 [0.047]	-0.041 [0.054]	0.015 [0.064]
Observations (Student*Year)	5,434,833	1,074,711	1,088,009	1,088,047	1,102,602	1,081,464
R-squared	0.27	0.25	0.28	0.27	0.26	0.27

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are SAT scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2006.

Appendix Table 8

Achievement Effects of District-Level CSR by Subgroup (Student-Level Standard Deviations)

	FCAT Math, Grades 4-8						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	0.021	0.012	-0.002	0.020	0.019	-0.010	0.030
	[0.014]	[0.014]	[0.019]	[0.011]	[0.017]	[0.017]	[0.024]
T x YR_SINCE_CSR	0.005	0.008	0.019	0.021	0.003	0.025	-0.008
	[0.010]	[0.014]	[0.014]	[0.025]	[0.011]	[0.019]	[0.011]
Total effect by 2006	0.035	0.035	0.054	0.085	0.028	0.066	0.006
	[0.025]	[0.037]	[0.046]	[0.076]	[0.025]	[0.061]	[0.025]
Observations (Student*Year)	2,688,064	2,788,462	1,278,901	1,175,282	2,782,828	2,691,303	2,772,051
R-squared	0.24	0.29	0.21	0.19	0.19	0.22	0.15
	FCAT Reading, Grades 4-8						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	-0.009	-0.016	-0.031	-0.008	-0.003	-0.042	-0.003
	[0.016]	[0.021]	[0.013]*	[0.014]	[0.020]	[0.019]*	[0.027]
T x YR_SINCE_CSR	0.003	0.005	0.010	0.025	-0.001	0.028	-0.008
	[0.015]	[0.019]	[0.014]	[0.039]	[0.012]	[0.025]	[0.012]
Total effect by 2006	0.000	-0.001	-0.002	0.068	-0.005	0.043	-0.027
	[0.038]	[0.048]	[0.049]	[0.113]	[0.028]	[0.071]	[0.030]
Observations (Student*Year)	2,691,893	2,793,524	1,281,329	1,176,822	2,787,415	2,696,245	2,775,878
R-squared	0.24	0.27	0.22	0.20	0.17	0.22	0.14

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include district fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2006.

Appendix Table 9

Effects of District-Level CSR on Non-Cognitive Outcomes

	% Days Absent, 4-8	% Days Absent, 4-5	% Days Absent, 6-8	ICV per 100 pupils	% Students ISS	% Students OSS
T x CSR	0.014 [0.010]	0.010 [0.009]	0.016 [0.011]	-1.6 [1.9]	0.001 [0.007]	-0.005 [0.006]
T x YR_SINCE_CSR	0.006 [0.007]	0.008 [0.006]	0.005 [0.007]	1.0 [0.6]	0.002 [0.005]	0.004 [0.005]
Total effect by 2006	0.032 [0.030]	0.034 [0.027]	0.031 [0.033]	1.4 [1.2]	0.007 [0.010]	0.008 [0.009]
Level of Aggregation	Student	Student	Student	District	District	District
Observations	5,402,992	2,140,351	3,262,641	536	536	536
R-squared	0.06	0.05	0.06	0.79	0.90	0.91

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. "% Days Absent" indicates the number of days the student was absent divided by the total number of days enrolled in the school (days absent plus days present) and is from the EDW data (2001 to 2006). "ICV per 100 pupils" indicate the number of incidents of crime and violence per 100 pupils. "% Students ISS (OSS)" indicate the percent of students that received at least one in-school (out-of-school) suspension. The ICV and suspension variables are calculated by aggregating school-level data for schools that serve students in at least one of the grades four to eight but no students in grades nine to 12. These data are from the FLDOE (1999 to 2006). All regressions include district fixed effects and controls for district-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Student-level (percent days absent) regressions also include controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status.

Appendix Table 10

School-Level Estimates that Condition on Prior-Year Controls (Effects in Student-Level Standard Deviations)

	FCAT Math, Grades 4-8				FCAT Reading, Grades 4-8			
T x CSR	-0.006	-0.002	-0.003	-0.006	0.002	0.007	0.002	-0.008
	[0.006]	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]	[0.006]
T x YR_SINCE_CSR	-0.006	-0.003	0.001	0.004	-0.011	-0.008	-0.002	0.002
	[0.003]	[0.003]	[0.004]	[0.003]	[0.003]**	[0.004]*	[0.004]	[0.002]
Total effect by 2009	-0.024	-0.012	-0.001	0.005	-0.032	-0.017	-0.005	-0.002
	[0.010]*	[0.010]	[0.011]	[0.008]	[0.010]**	[0.011]	[0.012]	[0.006]
Data from 2001 excluded?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Exclude students missing prior-year data?	No	No	Yes	Yes	No	No	Yes	Yes
Include student prior-year controls?	No	No	No	Yes	No	No	No	Yes
Observations (Student*Year)	8,293,046	7,414,547	5,283,476	5,283,476	8,304,672	7,425,220	5,290,581	5,290,581
R-squared	0.3	0.29	0.3	0.72	0.29	0.28	0.29	0.69

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Student prior-year controls include test scores in both subjects (and their cubed and squared terms), whether the student made a nonstructural or structural move from the previous year, the number of days the student was absent the previous year, and whether the student was repeating a grade. Data cover period from 2001 to 2009.

Appendix Table 11

School-Level Analysis, Alternative Specifications (Effects in Student-Level Standard Deviations)

	FCAT Math, Grades 4-8				
	District Trends	Drop 2001 and 2002	No Charters	Un-weighted	Diff-in-Diffs
T x CSR	-0.005 [0.006]	-0.000 [0.005]	-0.007 [0.006]	-0.002 [0.009]	-0.000 [0.006]
T x YR_SINCE_CSR	-0.005 [0.003]	-0.002 [0.004]	-0.007 [0.003]	-0.012 [0.005]**	
Total effect by 2006	-0.020 [0.010]*	-0.006 [0.010]	-0.027 [0.011]*	-0.037 [0.013]**	
Observations (Student*Year)	8,293,046	6,482,118	8,051,399	8,293,046	8,293,046
R-squared	0.30	0.29	0.30	0.35	0.30

	FCAT Reading, Grades 4-8				
	District Trends	Drop 2001 and 2002	No Charters	Un-weighted	Diff-in-Diffs
T x CSR	0.002 [0.005]	0.007 [0.005]	0.001 [0.005]	0.009 [0.007]	0.004 [0.006]
T x YR_SINCE_CSR	-0.010 [0.003]**	-0.009 [0.004]*	-0.012 [0.004]**	-0.022 [0.004]**	
Total effect by 2006	-0.029 [0.010]**	-0.019 [0.012]	-0.034 [0.011]**	-0.056 [0.011]**	
Observations (Student*Year)	8,304,672	6,491,854	8,062,619	8,304,672	8,304,672
R-squared	0.29	0.27	0.29	0.32	0.29

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. "District Trends" also include district-specific linear time trends. "Drop 2001 and 2002" indicates that the earliest two years of pre-treatment data are excluded. "No Charters" indicates that all charter schools are excluded. "Unweighted" indicates that each school is weighted equally. "Diff-in-Diffs" is a standard difference-in-differences specifications that also includes grade-by-year fixed effects. Data cover period from 2001 to 2009.

Appendix Table 12

School-Level Analysis, Fully Flexible Specification (Effects in Student-Level Standard Deviations)

	FCAT Math			FCAT Reading		
	Grades 4-8	Grades 4-5	Grades 6-8	Grades 4-8	Grades 4-5	Grades 6-8
T x YEAR2001	-0.019 [0.012]	-0.028 [0.012]*	-0.003 [0.017]	-0.036 [0.011]**	-0.062 [0.011]**	-0.010 [0.016]
T x YEAR2002	-0.002 [0.009]	-0.006 [0.011]	0.006 [0.014]	-0.010 [0.010]	-0.034 [0.009]**	0.012 [0.015]
T x YEAR2003	-0.008 [0.008]	-0.000 [0.010]	-0.008 [0.011]	-0.015 [0.009]	-0.051 [0.009]**	0.013 [0.013]
T x YEAR2004	0.032 [0.008]**	0.022 [0.010]*	0.039 [0.012]**	0.015 [0.010]	-0.017 [0.009]*	0.035 [0.014]*
T x YEAR2005	0.004 [0.004]	0.016 [0.006]*	-0.002 [0.006]	-0.010 [0.006]	-0.007 [0.005]	-0.011 [0.007]
T x YEAR2007 (Effect by 2007)	0.002 [0.004]	0.004 [0.006]	0.002 [0.005]	0.001 [0.004]	-0.007 [0.005]	0.006 [0.005]
T x YEAR2008 (Effect by 2008)	0.004 [0.005]	0.013 [0.008]	-0.000 [0.007]	-0.001 [0.004]	-0.010 [0.006]	0.003 [0.006]
T x YEAR2009 (Effect by 2009)	-0.002 [0.006]	0.010 [0.009]	-0.005 [0.008]	-0.011 [0.005]*	-0.016 [0.007]*	-0.006 [0.007]
Observations (Student*Year)	8,293,046	3,298,365	4,994,681	8,304,672	3,303,401	5,001,271
R-squared	0.30	0.29	0.31	0.29	0.28	0.29

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the district level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include year dummies, school fixed effects, and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2009.

Appendix Table 13

Achievement Effects of School-Level CSR by Subgroup (Student-Level Standard Deviations)

	FCAT Math, Grades 4-8						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	-0.008	-0.004	-0.010	-0.010	0.002	-0.004	-0.001
	[0.006]	[0.006]	[0.010]	[0.009]	[0.007]	[0.008]	[0.006]
T x YR_SINCE_CSR	-0.002	-0.010	-0.007	-0.012	0.001	-0.009	-0.002
	[0.003]	[0.004]*	[0.006]	[0.005]*	[0.004]	[0.004]*	[0.003]
Total effect by 2009	-0.015	-0.033	-0.032	-0.046	0.004	-0.030	-0.008
	[0.011]	[0.012]**	[0.018]	[0.015]**	[0.011]	[0.013]*	[0.010]
Observations (Student*Year)	4,072,038	4,221,008	1,902,963	1,874,377	4,098,156	4,086,849	4,139,068
R-squared	0.28	0.32	0.25	0.23	0.24	0.25	0.22
	FCAT Reading, Grades 4-8						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	0.004	0.001	-0.009	-0.002	0.007	0.008	0.008
	[0.006]	[0.006]	[0.009]	[0.009]	[0.006]	[0.007]	[0.006]
T x YR_SINCE_CSR	-0.010	-0.012	-0.009	-0.023	-0.001	-0.016	-0.006
	[0.003]**	[0.004]**	[0.005]	[0.006]**	[0.004]	[0.004]**	[0.003]
Total effect by 2009	-0.027	-0.037	-0.037	-0.070	0.003	-0.041	-0.011
	[0.010]	[0.012]**	[0.016]*	[0.016]**	[0.010]	[0.012]**	[0.010]
Observations (Student*Year)	4,077,048	4,227,624	1,906,032	1,876,339	4,104,273	4,093,333	4,144,037
R-squared	0.27	0.3	0.25	0.22	0.21	0.24	0.19

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets.

Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Data cover period from 2001 to 2009.

Appendix Table 14

Effects of School-Level CSR on Non-Cognitive Outcomes

	% Days Absent, 4-8	% Days Absent, 4-5	% Days Absent, 6-8	ICV per 100 pupils	% Students ISS	% Students OSS
T x CSR	0.001 [0.001]	-0.000 [0.000]	0.001 [0.001]	-0.9 [0.9]	-0.004 [0.002]	-0.004 [0.002]*
T x YR_SINCE_CSR	0.000 [0.000]	-0.001 [0.000]**	0.001 [0.001]	-0.3 [0.3]		
Total effect by 2009	0.001 [0.001]	-0.003 [0.001]**	0.003 [0.002]	-1.9 [1.7]		
Level of Aggregation	Student	Student	Student	School	School	School
Observations	8,157,155	3,240,715	4,916,440	20,326	15,485	15,485
R-squared	0.11	0.07	0.12	0.25	0.86	0.88

Notes: ** $p < 0.01$, * $p < 0.05$; robust standard errors adjusted for clustering at the school level appear in brackets. "% Days Absent" indicates the number of days the student was absent divided by the total number of days enrolled in the school (days absent plus days present) and is from the EDW data (2001 to 2009). "ICV per 100 pupils" indicate the number of incidents of crime and violence per 100 pupils. "% Students ISS (OSS)" indicate the percent of students that received at least one in-school (out-of-school) suspension. The ICV and suspension variables are calculated for schools that serve students in at least one of the grades four to eight but no students in grades nine to 12. These data are from the FLDOE (1999 to 2009 for ICV and 1999 to 2007 for suspension). All regressions include school fixed effects and controls for school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Student-level (percent days absent) regressions also include controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status.

Appendix Table 15

Effects of Required School-Level CSR in Grades PK-3 on 3rd-Grade FCAT Scores,
Alternative Specifications (Student-Level Standard Deviations)

	Drop 2001 Data		Standard Diff-in-Diffs	
	Math	Reading	Math	Reading
T x CSR	-0.002 [0.013]	0.016 [0.010]	0.026 [0.010]**	-0.000 [0.006]
T x YR_SINCE_CSR	-0.007 [0.007]	-0.017 [0.005]**		
Total effect by 2009	-0.024 [0.018]	-0.034 [0.013]**		
Observations (Student*Year)	1,538,357	1,540,176	1,716,312	1,717,977
R-squared	0.28	0.27	0.29	0.27

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. Dependent variables are FCAT developmental scale scores in math and reading, which are standardized by subject and grade based on the distribution of scores in 2001 to 2003. All regressions include school fixed effects and controls for student grade level, gender, race/ethnicity, free- and reduced-price lunch eligibility, limited English proficiency status, and special education status, as well as school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch. Difference-in-differences specifications also include grade-by-year fixed effects. Data cover period from 2001 to 2009.