

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

Matthew M. Chingos

The Brown Center on Education Policy
The Brookings Institution
1775 Massachusetts Ave NW
Washington, DC 20036

mchingos@brookings.edu

Forthcoming in *Economics of Education Review*

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 student achievement.

JEL classifications: I20, I22, I28

Keywords: efficiency, input output analysis, productivity, resource allocation

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

Voters who provide political support to CSR initiatives must expect that state governments can improve student achievement by providing resources that must be spent on a specific policy, but there is little empirical evidence on this question. The most credible previous studies of CSR at the K–12 level 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

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). Studies of class size in postsecondary education include Becker and Powers (2001) and Kakkelenberg, Dillon, and Christy (2008).

classes (22–25 students) by about 0.22 standard 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. Cho, Glewwe, and Whitley (Forthcoming) apply one of Hoxby's methods to data from Minnesota and find evidence that reducing class size increases test scores, but the effects are very small in magnitude. Rivkin, Hanushek, and Kain's (2005) study of Texas schools, which also used variation over time in class sizes, found positive effects of smaller classes in fourth and fifth grades—although the fourth-grade effects are half the size of the STAR effects and the fifth-grade effects are even smaller.⁴ An important difference between the quasi-experimental studies and the STAR study is that schools that participated in the STAR experiment received additional resources to reduce class size, while the Connecticut, Minnesota, and Texas schools in the quasi-experimental studies did not (and thus likely had to divert resources from elsewhere when natural population variation led to smaller classes).⁵ 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.

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

⁴ One difference between Hoxby (2000) and Rivkin et al. (2005) is that the former uses variation in class size between cohorts, whereas the latter uses variation in the class sizes experienced by the same cohorts over time (e.g., due to changes in the number of students who move in or out of the school).

⁵ Connecticut and Minnesota schools that experienced increases in enrollment would have received additional funding on a per-pupil basis from the state and federal governments, but in most cases this would have been less than total per-pupil expenditures (leading to a decrease in overall per-pupil spending in the school). Additionally, Hoxby's regression discontinuity method exploits large changes in class size that result from very small changes in enrollment (and thus very small changes in resources).

There is also very little evidence on the effects of CSR in middle and high school. Rivkin et al. (2005) find no evidence of class-size effects on seventh-grade test scores. The only other high-quality study of class size in these grades is Dee and West's (2011) analysis of eighth-grade students in a nationally representative dataset. Dee and West compare the outcomes of the same students who had attended different size classes in different subjects. They find no impact of class size on test scores (except for a small effect in urban schools), but do find modest effects on non-cognitive skills such as student attentiveness and attitudes about learning. The current study extends this literature by estimating class-size effects on both cognitive and non-cognitive outcomes in the middle school grades (in addition to examining the later elementary grades).

Most class-size 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.

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 the hiring of inexperienced teachers (particularly in the first few years of the policy).⁶ However, Jepsen and Rivkin’s study is limited by data constraints—student achievement data are only available in the period after CSR implementation began. 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.⁷ 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 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. The district-level analysis, which focuses on

⁶ 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.” Other unintended negative consequences of the California policy included an increase in class size in grades four and five (Sims 2009) and the use of multi-grade classrooms (Sims 2008).

⁷ 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).

grades six through eight, yields no statistically significant effects but does not have sufficient statistical power to rule out small positive effects. I am able to examine test scores in grades three through eight using the school-level analysis, and find no evidence of positive effects and some evidence of negative effects. In general, the standard errors are small enough that even modest positive effects can be ruled out. I also do not find any significant evidence of heterogeneous effects. However, I do find some evidence of effects on non-cognitive outcomes, including student absenteeism in elementary school and incidents of crime and violence in middle schools.

2. Evaluating Florida's CSR Policy

In November 2002, Floridians narrowly voted (52 to 48 percent) 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 estimated at about \$20 billion over the first eight years, with continuing operating costs of about \$4 billion per year in subsequent years.⁸ 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.

⁸ "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 fell from 24.2 in 2003 to 18.6 in 2009.⁹ 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.¹⁰ 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.¹¹

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

⁹ 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.

¹⁰ 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.

¹¹ 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.¹²

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, 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

¹² Between 2001 and 2009, fourth-grade math and reading scores increased by 0.70 and 0.43 student-level standard deviations, respectively. Eighth-grade math and reading scores increased by 0.26 and 0.29 standard deviations, respectively.

in student achievement at 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 in May 2003, just four months before districts had to begin reducing class size. 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.¹³ 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 was measured at the classroom level.¹⁴

2.1. District-Level Analysis

I classify districts into two groups: 1) 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 2) 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

¹³ 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.

¹⁴ 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.

official average class sizes for 2003 (the year immediately preceding implementation of CSR) published by the Florida Department of Education (FLDOE) to classify districts, and only include the 67 regular school districts (which are coterminous with counties).¹⁵ 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.¹⁶

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 six 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 in grades four to eight 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.¹⁷

¹⁵ 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.

¹⁶ Below I show that my results are robust to including all charter schools or excluding all charter schools.

¹⁷ 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 approximately 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, and below I show that the greater CSR in treated districts was concentrated entirely in grades six to eight, so I focus on these grades in the district-level analysis. Additionally, 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 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

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).

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.

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 comparison districts documented in the regression results reported below.

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 Table A1, show that the estimated “effect” of CSR on district-level percent black, percent Hispanic, and percent eligible for free or reduced-price lunch is

¹⁹ When I use the additional years of data that are available for eighth-grade math and reading scores, 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. These results are available from the author upon request.

statistically insignificant and substantively small, as would be expected if the model assumptions hold.

Table A1 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 However, it is not clear whether this result is a fluke of demographics or an impact of CSR. It seems unlikely that families would leave districts because of CSR. In any case, 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.

The fourth column of Table A1 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 (although the effect on salaries is not statistically significant). In the first year of CSR, teacher salaries (adjusted for experience and education) 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.

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.

2.2. School-Level Analysis

I conduct a similar analysis using variation in CSR implementation at the school level. Beginning in 2007, individual schools were 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.

In the district-level analysis, I included all 67 regular school districts in order to maximize the sample size. However, the number of schools is much larger so I am able to select treatment and comparison groups that are as similar as possible in terms of their pre-treatment observable characteristics. Most Florida students in grades K–8 attend an elementary school (typically K–5) and then a middle school (typically 6–8). Consequently, I identify treatment and comparison groups separately for grades 3–5 and grades 6–8.²¹

For grades 3–5, I identify schools that serve students in all three of these grades. I then classify schools using similar definitions to the district-level analysis, based on the official FLDOE calculations of school-level average class sizes for 2006. An important difference in this case is that treated schools must be treated in both relevant grade ranges (PK–3 and 4–8) and comparison schools must be untreated in both ranges. As a result, I remove from the analysis schools that are treated in one grade range but not in the other. These schools muddy the

²¹ The small number of schools that include grades in both the 3–5 and 6–8 ranges could be included in both parts of the school-level analysis.

comparison because they may face incentives to reduce class size in some grades by increasing class size in other grades. This method identifies 921 comparison schools and 383 treated schools serving grades 3–5. These schools differ somewhat in terms of observable characteristics, so I select 383 of the comparison schools using nearest-neighbor matching.²²

I repeat this procedure for grades 6–8, identifying schools that serve all three grades in this range and classifying them into treated and comparison groups based on their average class size in grades 4–8. For the small number of school that also serve grades PK–3, I eliminate those whose treatment status in the early grades does not match their status in the later grades. This procedure identifies 574 comparison schools and 182 treated schools. I select 182 of the comparison schools using the same matching methods as in the grades 3–5 analysis.

The matching method selects a set of comparison schools that are more similar to the treated schools than the full group of potential comparison schools. However, conducting the analysis on all schools, separately for grades PK–3 and 4–8, produces qualitatively similar results.²³

The school-level estimation strategy 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, average class size in the comparison schools in grades 3–5 remained the same, while the treated schools reduced average class size in these grades by 3.3

²² The matching is conducted using propensity scores estimated based on school-level percent black, percent Hispanic, percent eligible for free or reduced-price lunch, and average FCAT reading and math scores in 2006. Each treated school is matched to the comparison school with the most similar estimated propensity score (without replacement).

²³ These results are available from the author on request.

students. In grades 6–8, class size fell by 1.1 in the comparison schools and 2.8 in the treated schools. Pre-treatment (2006) summary statistics for treated and matched comparison schools are shown in Table 2. In both sets of grades, the treated and comparison groups are very similar in terms of demographic breakdowns and standardized test scores, although the treated schools in grades 3–5 are modestly larger and serve a larger share of Hispanic students.

The final column of Table 2 shows comparable descriptive statistics for all schools serving students in the same grades. For grades 3–5, the treated and comparison schools are largely similar to all schools in the state in terms of demographics and student achievement. In the case of grades 6–8, the comparison and treated schools differ modestly from state averages in that they serve students who are less likely to be eligible for free or reduced-price lunch (by about 5 percentage points) and who scored higher on the state tests (by 0.10–0.15 standard deviations). These differences are modest in size, but suggest more caution in interpreting results for these grades than for the earlier grades.

The results of falsification tests reported in the first three columns of each panel of Table A2 indicate that CSR had only negligible “effects” on the demographic composition of treated schools (some of the coefficients are statistically significant, but small in size). And unlike in the district-level analysis, CSR had no statistically significant impact on enrollment.

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 descriptive statistics reported on Table 2 do not support either hypothesis for grades 3–5, as the treated schools are similar to all schools statewide. As noted above, the treated schools for grades 6–8 appear to serve more advantaged students than schools statewide, although the differences are modest in size. As a further check on this hypothesis, below I show CSR effects for various subgroups of students. If less advantaged students benefit disproportionately from CSR, then the subgroup results will show that even if the overall effect is biased downward by the tendency of more affluent middle schools to adopt CSR earlier than other schools.

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/ethnicity, 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.²⁴ 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.

²⁴ Florida Department of Education, “Class Size Reduction in Florida’s Public Schools,” available at <http://www.fldoe.org/ClassSize/pdf/csfaqfinal.pdf>.

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 and non-cognitive outcomes from the FLDOE's Florida School Indicators Reports.

4. Results

4.1. 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. I first estimate the impact of a district being required to reduce class size on district-level average class sizes, both overall and by grade.²⁵ 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.²⁶

²⁵ I use my own calculations from the EDW data because the FLDOE only reports average class size by grade grouping, not by individual grade. Although the official FLDOE class size averages do not line up perfectly with those I am able to calculate from the EDW database, they are strongly correlated ($r=0.82$ in grades four to eight, weighting each district by its enrollment).

²⁶ 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.

Class sizes in grades four and five were reduced by similar amounts in the treated and comparison districts.²⁷ Consequently, I will only present and discuss results for the middle school grades.²⁸ The three-year average relative reduction in class size for these three grades was 2.3 students.

Table 4 presents my preferred district-level estimates of the effect of the CSR mandate on FCAT math and reading scores in grades six to eight. The 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.²⁹

It is instructive to examine all of the coefficient estimates reported for my preferred model. The coefficient on *YEAR* in the first column of Table 4 indicates that, prior to the introduction of CSR, math scores were increasing by about 0.03 standard deviations per year in the comparison districts. The coefficient on $T \times YEAR$ (0.009) indicates that this pre-treatment achievement trend was similar 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 (but declined subsequently), although of course this increase cannot be causally linked to CSR (and the additional funding provided to comparison districts) given the

²⁷ 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.

²⁸ The results for grades four and five, which are generally statistically insignificant from zero, are available from the author upon request.

²⁹ 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.

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 0.008 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 0.060 standard deviations and positive effects larger than 0.077 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. The estimates for reading scores indicate a negative but statistically insignificant ($p=0.12$) effect of 0.07 standard deviations, with a 95 percent confidence interval that ranges from -0.168 to 0.020. Results disaggregated by grade are less precisely estimated in general but are all negative, and one (seventh grade) is statistically significantly negative.

These main results are robust to a variety of alternative specifications.³⁰ Table A3 shows the results from six other alternative specifications. Using results from the Stanford Achievement Test, a low-stakes test that was administered along with the FCAT until 2008, yields similar results. Similar results are also 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

³⁰ A potential limitation of the district-level analysis is that only three years of data are used to estimate the pre-treatment trend. For eighth-grade math and reading scores, five years of pre-treatment data are available. 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. Although restricting the analysis to two years of pre-treatment data noticeably changes the results (the point estimates in both subjects are positive but statistically insignificant), 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. All of these results are available from the author upon request.

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 (although the point estimate for math scores is modestly larger).

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 Table A4. The three-year effects are statistically indistinguishable from my preferred estimates, although the one-year math effect is positive and statistically significant.

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.³¹ I estimate results disaggregated by gender, race/ethnicity, and eligibility for free or reduced-price lunch (FRL). However, the estimates (not shown) are too imprecisely estimated to be statistically significantly different from each other. The only statistically significant CSR effects are large negative effects on the math and reading scores of Hispanic students.

³¹ 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.

Finally, I examine the effect of CSR on several non-cognitive outcomes.³² The results (not shown) offer no evidence that CSR affected student absenteeism, incidents of crime and violence, or student suspension rates.³³ All of the estimated effects on these undesirable outcomes are statistically indistinguishable from zero, although the point estimates are positive.

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 2.3-student decrease in class size for one year to increase achievement by about 0.04 standard deviations.³⁴ I cannot reject (with 95 percent confidence) an effect of this magnitude middle school 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.

4.2. 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

³² For previous evidence on the effect of CSR on non-cognitive outcomes, see Dee and West (2011).

³³ 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.

³⁴ 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 2.3 (the amount of CSR in sixth to 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.

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 CSR funding from the state was provided in equal amounts to districts—not to schools—it is likely that districts were led to provide greater resources to the treated schools than to the comparison schools. If that is the case, 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 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 regressions results it is instructive to examine the trends in average class sizes and test scores at the treated and comparison schools. Figures 1a and 1b show average class size over time in grades 3–5 and 6–8, respectively. In the younger grades, pre-treatment class sizes largely paralleled each other until 2006, when there was a modest spike in the treated schools. Over the three years of school-level CSR implementation documented in these data, average class size converged in the treated and comparison schools. In the older grades, average class sizes were very similar pre-treatment and partially (but not fully) converged post-treatment.

If the smaller classes brought about by CSR had any effect on student achievement, we would expect to see test score trends change beginning in 2007 when class sizes began to converge between the treated and comparison groups. However, Figures 2a and 2b suggest that this did not occur. In the earlier grades, test scores were increasing at a slightly faster rate at the treated schools than in the comparison schools during the early 2000s. However, the trends were parallel beginning in 2006 and did not diverge at all through 2009. In the middle school grades,

test-score trends in the treated and comparison schools were largely parallel beginning in 2003, with no divergence in 2007–2009.

These graphical results are confirmed by the formal regression analysis. Table 5 shows that, prior to school-level CSR implementation, class size in grades 3–5 was decreasing by 0.9 students per year in the comparison schools and 0.6 students per year in the treated schools. By the third year of school-level CSR implementation, class size fell by 3.6 students more in the treated schools than in the comparison schools. In the middle school grades, the three-year effect was somewhat smaller, at 2.1 students. Consequently, if CSR had an effect on achievement in any of the grades from three 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).

Another reason why we might expect to find class-size effects in the school-level analysis is that treated schools reduced class sizes in grades K–2 by a larger amount than comparison schools. For example, the three-year effect of being in the grades 3–5 treatment group on average class size in first and second grade was 2.2 and 2.6 students, respectively. Test scores are only available beginning in third grade, but after three years of implementation the third-grade test scores will also reflect any lasting gains from CSR in first and second grade.

The results for FCAT math and reading scores, shown in Table 6, indicate that even small positive effects can be ruled out in grades 3–5. The estimated three-year effect on math scores is negative, and positive effects larger than 0.018 can be ruled out with 95 percent confidence. The effect on reading scores is both negative and statistically significant. Small differences between the treated and comparison schools in their pre-treatment achievement trends suggests caution in interpreting the negative estimates as causal. In the middle school grades, where there were no

pre-treatment differences in achievement trends, the point estimates are close to zero and positive effects larger than 0.040 in math and 0.035 in reading can be ruled out with 95 percent confidence.

Results disaggregated by grade (not shown) do not follow any discernible pattern. Results for third-grade students are particularly interesting because the three-year effect will reflect the impact of student exposure to smaller classes in first, second, and third grade (three of the four grades included in the Tennessee STAR experiment). But these results mirror the combined estimate for grades 3–5, with a point estimate close to zero for math and a statistically significant negative effect for reading.

These results are robust to many but not all alternative specifications. Controlling for prior-year test scores requires dropping data from 2001 as well as students with missing prior-year data in later years. This sample restriction increases the coefficients, and the addition of prior-year control variables (including test scores) increases them further in the elementary grades but not in the middle school grades.³⁵ Results from five other alternative specifications are reported in Table A5. The statistically significant negative effects on reading scores in grades 3–5 are robust to including district-specific linear time trends (to control for district-level policy changes), to dropping the two earliest years of data, to excluding charter schools, and to weighting all schools equally.

However, the negative effects are not robust to a standard difference-in-differences specification. The difference-in-differences estimates are generally close to zero and precisely estimated. 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

³⁵ These results are available from the author upon request.

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) generally yields results that are close to zero, although this estimation indicates positive CSR effects on math scores in grades 3–5 (Table A6). However, for this subject-grade combination there were differences of a similar magnitude in multiple pre-treatment years.

The school-level CSR effects do not differ markedly by student demographics, although Table A7 suggests that the negative effects on reading scores in grades 3–5 are concentrated among black and Hispanic students and students eligible for free or reduced-price lunch. A similar but weaker pattern appears for the middle school grades (Table A8), as might be expected given that the overall estimates are closer to zero in these grades. In other words, I find no evidence that the overall null findings for CSR in Florida mask positive effects for specific subgroups of students.

I find evidence of CSR effects on some non-cognitive outcomes, as shown in Table A9. These results indicate that CSR reduced the absence rate in grades 3–5 by 0.4 percentage points (about 9 percent of its mean) and the incidents of crime and violence per 1000 pupils in middle schools by 1.9 (about 30 percent of its mean). However, I do not find any effects on middle school absence rates or student suspension rates.³⁷

The negative CSR estimates produced by the school-level analysis are not sufficiently robust to be taken as definitive, but are consistent enough to pose an empirical puzzle. The most

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

³⁷ I do not present results for incidents of crime and violence and suspension rates for elementary school students because these behaviors are quite rare in the elementary grades.

theoretically plausible explanation for how a reduction in class size might produce negative effects is that teacher quality fell as a large number of new teachers needed to be hired quickly, as occurred in California (Jepsen and Rivkin 2009).³⁸ A full exploration of the impact of Florida's CSR policy on teacher hiring (and sorting among schools) is beyond the scope of the present study, but I do estimate CSR impacts on four measures of teacher experience: share of teachers new to the teaching profession, share of teachers new to the school, mean total teaching experience, and mean experience in current job.³⁹

These results, which are shown in Table 7, do not follow a consistent pattern. For example, they indicate that school-level CSR decreased average total experience of teachers in grades 3–5, but increased average teacher experience in their current jobs. There are plausible theories that would fit this pattern of results (such as schools hiring more teachers but also retaining more teachers in their current jobs), but the key point is that they do not provide a compelling explanation for the negative test-score effects documented above.

Another potential explanation is that the relative reduction in teacher salaries at treated districts continued during the school-level implementation period and led to relative salary reductions in districts containing concentrations of treated schools. I test this hypothesis by

³⁸ Another explanation is that Florida's CSR policy created incentives for schools to use more mixed-grade classrooms at the cutoffs between grade ranges. For example, the class size maximum is 18 students in third grade and 22 students in fourth grade. A class that includes third- and fourth-grade students would be subject to the 22-student maximum as long as a majority of students are in fourth grade. Sims (2008) found negative impacts of such CSR-induced mixed-grade classrooms in California.

³⁹ These variables are calculated based only on teachers of the core subject areas covered by the CSR policy, and weight teachers based on the total number of minutes students spent with them each week. New teachers are identified as those with zero or missing total experience who were not employed by the Florida public schools in the previous year. Teachers new to the school are identified as those who in the previous year were employed in a different school or were not employed by the Florida public schools (teachers who were employed in more than one school are excluded from the calculation of this variable). The total experience and experience in current job variables are top-coded at 40 and are recoded as missing in the small number of cases when the indicated experience is greater than 60 years.

linking the district salary data discussed above to the treated and comparison schools.⁴⁰ The average salary data I compute are based on estimates of district salary schedules and thus should not be affected by the experience and education levels of teachers in a given district.

I find that from 2006 to 2009 inflation-adjusted salaries decreased by about 2 percent in the treated schools and increased by about 2 percent in the comparison schools, a difference of 4 percent. I did not find any such relative difference for the middle grades. If relative salary reductions lead to relative decreases in teacher quality or teacher effort, then the salary data provide suggestive evidence for why the estimated CSR impact is negative in the elementary grades but not in the middle school grades.

5. Conclusions

The results from both the district- and school-level analyses indicate that the effects of mandated CSR in Florida on student achievement 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, middle school student achievement in the treated districts was about the same in math and lower in reading (although not by a statistically significant amount) than it would have been had these districts received equivalent resources without a CSR mandate. These effects correspond to a relative reduction in class size of two or three students.

A significant limitation of the district-level analysis is that small positive effects of CSR can only be ruled out in the case of middle school reading, but this is not the case in the school-level analysis of grades 3–5. 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

⁴⁰ Specifically, I calculate the student-weighted average salary (measured at the district level) of the treated and comparison groups of schools.

than they would have been had these schools not been required to reduce class size. The corresponding analysis of middle schools produces estimates that are close to zero, which are precise enough to rule out three-year effects larger than about 0.04 standard deviations.

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.⁴¹ 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 likely 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.

A naïve extrapolation from the STAR results would lead one to expect a reduction of 3.6 students (the three-year effect on class size in grades 3–5 in the school-level analysis) for one year to improve test scores by 0.06 standard deviations.⁴² Such an effect is easily ruled out in both subjects in these grades. The smaller reduction in middle school of 2.1 students would be predicted to have a one-year impact of 0.04 standard deviations. Such an impact can only be

⁴¹ 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).

⁴² The second column of Table IX of Krueger (1999) indicates that a seven-student reduction in class size increased test scores by 4.2 percentile points after one year. This corresponds to 0.018 standard deviations per one-student reduction in class size

ruled out with 90 percent confidence in grades 6–8. However, these predicted effects are conservative (i.e., probably understated) in the sense that the estimated three-year effects capture any benefits of CSR for students in both the listed grade and earlier grades. For example, the three-year effect for fifth-grade students who remained in the same school for grades 3–5 captures the cumulative benefits of an average reduction in class size of 1.3 students in third grade, 2.5 students in fourth grade, and 3.6 students in fifth grade. Consequently, we should expect even larger impacts yet the data indicate effects that were very small at best.

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 groups in this study. Finally, very small positive effects cannot be ruled out in some grade-subject combinations. 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.

Vitae

Matthew M. Chingos is a Fellow in the Brookings Institution's Brown Center on Education Policy. He received a B.A. in Government and Economics and a Ph.D. in Government, both from Harvard University.

Acknowledgements

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, Eric Brunner, David Deming, Jeremy Finn, Josh Goodman, Larry Katz, Paul Peterson, Martin West, two anonymous referees, and seminar participants at the American Enterprise Institute, the Brookings Institution, and 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.

References

- Angrist, J.D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2), 533–575.
- Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics*, 111(1), 97–133.
- Becker, W.E., & Powers, J.R. (2001). Student performance, attrition, and class size given missing student data. *Economics of Education Review*, 20(4), 377–388.
- Bohrnstedt, G.W., & Stecher, B.M., eds. (2002). *What we have learned about class size reduction in California*. Palo Alto, CA, CSR Research Consortium.
- Cho, H., Glewwe, P., & Whitley, M. (Forthcoming). Do reductions in class size raise students' test scores? Evidence from population variation in Minnesota's elementary schools. *Economics of Education Review*.
- Dee, T., & Jacob, B. (2009). The impact of No Child Left Behind on student achievement. *National Bureau of Economic Research Working Paper No. 15531*. Cambridge, MA.
- Dee, T., & West, M. (2011). The non-cognitive returns to class size. *Education Evaluation and Policy Analysis*, 33(1): 23–46
- Education Commission of the States. (2005). State class-size reduction measures. Denver, Colorado, Education Commission of the States.
- Glass, G.V., & Smith, M.L. (1978). *Meta-analysis of research on class size and achievement*. San Francisco, Far West Laboratory.
- Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–77.
- Hoxby, C.M. (2000). The effects of class size on student achievement: new evidence from population variation. *Quarterly Journal of Economics*, 115(4), 1239–1285.
- Jepsen, C., & Rivkin, S. (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, T.J., & Stager, D.O. (2005). Using imperfect information to identify effective teachers. Unpublished Paper.
- Kokkelenberg, E.C., Dillon, M. & Christy, S.M. (2008). *Economics of Education Review*, 27(2), 221–233.

Krueger, A.B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 115(2), 497–532.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.

Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives*, 23(4), 211–30.

Sims, D. (2008). A strategic response to class size reduction: combination classes and student achievement in California. *Journal of Policy Analysis and Management*, 27(3), 457–478.

Sims, D. (2009). Crowding Peter to educate Paul: lessons from a class size reduction externality. *Economics of Education Review*, 28(4), 465–473.

Wößmann, L., & West, M. (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 1a. Average Class Size, Treated and Comparison Schools, Grades 3–5

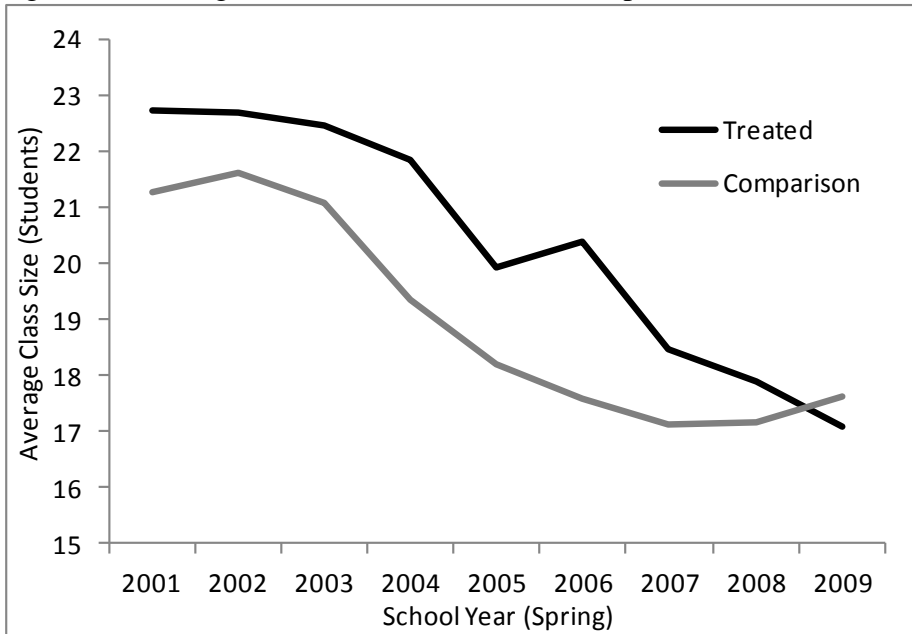


Figure 1b. Average Class Size, Treated and Comparison Schools, Grades 6–8

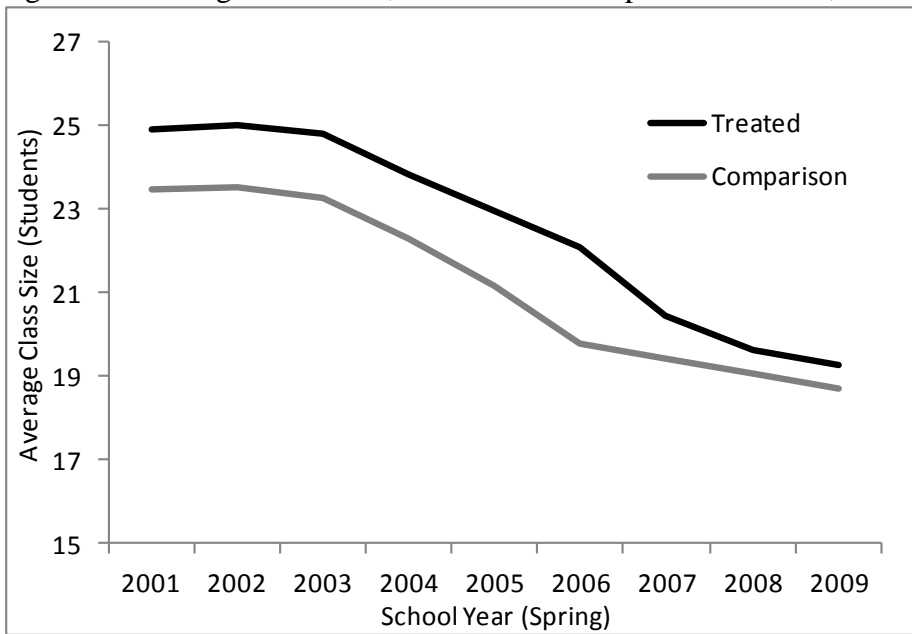


Figure 2a. Average FCAT Math/Reading Scores, Treated and Comparison Schools, Grades 3–5

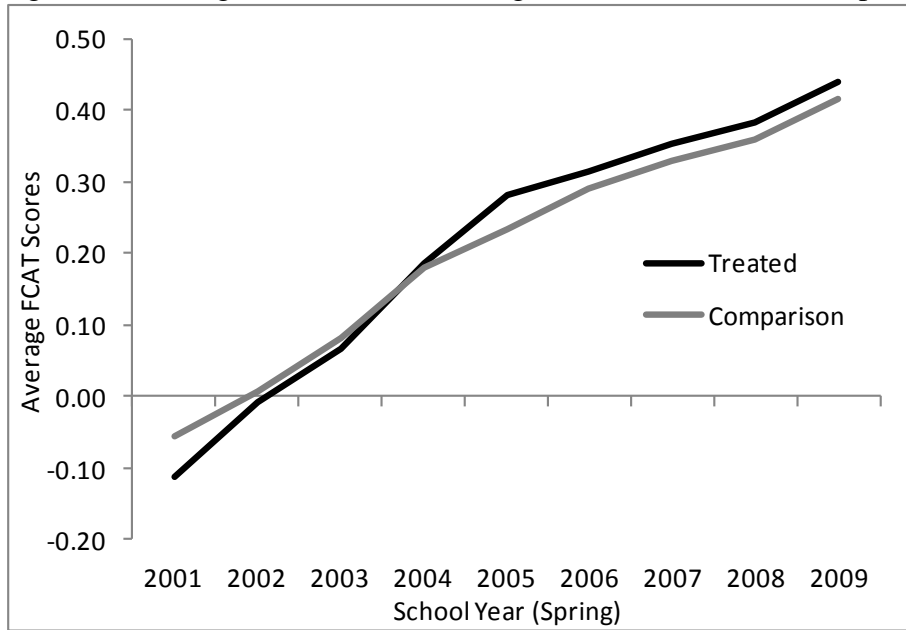


Figure 2b. Average FCAT Math/Reading Scores, Treated and Comparison Schools, Grades 6–8

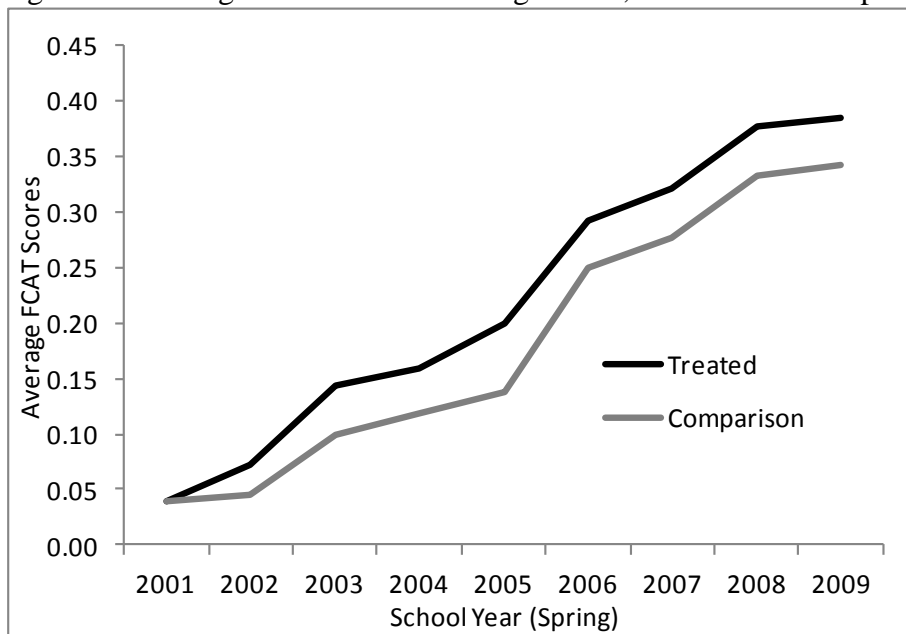


Table 1

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	0.058	0.027
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

	Grades 3-5			
	Comparison	Treated	Difference	All Schools
Class Size (Official) in grades K-3	16.3	23.8	7.5**	18.9
Class Size (Official) in grades 4-8	18.4	24.8	6.4**	20.5
Class Size (Author's Calculation) in grades 3-5	17.6	20.4	2.8**	18.7
Enrollment	780	920	141**	800
Percent Black	0.24	0.21	-0.03	0.24
Percent Hispanic	0.29	0.41	0.12**	0.25
Percent Eligible for Free/Reduced Lunch	0.52	0.53	0.01	0.52
Accountability Grades	3.43	3.52	0.08	3.41
Percent New Teachers in grades 3-5	0.11	0.13	0.01	0.12
Average Teacher Experience in grades 3-5	10.3	9.8	-0.5	10.2
FCAT Math Scores (std.) in grades 3-5	0.32	0.34	0.02	0.30
FCAT Reading Scores (std.) in grades 3-5	0.26	0.29	0.03	0.25
Number of Schools (Unweighted)	383	383		1,864

	Grades 6-8			
	Comparison	Treated	Difference	All Schools
Class Size (Official) in grades 4-8	20.3	23.8	3.6**	20.7
Class Size (Author's Calculation) in grades 6-8	19.8	22.1	2.3**	19.9
Enrollment	1,200	1,300	41	1,200
Percent Black	0.22	0.20	-0.02	0.24
Percent Hispanic	0.25	0.31	0.06	0.24
Percent Eligible for Free/Reduced Lunch	0.44	0.42	-0.02	0.48
Accountability Grades	3.70	3.71	0.02	3.54
Percent New Teachers in grades 6-8	0.14	0.15	0.00	0.15
Average Teacher Experience in grades 6-8	7.4	8.0	0.7	7.8
FCAT Math Scores (std.) in grades 6-8	0.25	0.30	0.04	0.15
FCAT Reading Scores (std.) in grades 6-8	0.25	0.29	0.04	0.17
Number of Schools (Unweighted)	182	182		797

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 and accountability grades are from the Florida Department of Education (FLDOE); author's class size calculations, percent new teachers, average teacher experience, and FCAT scores (average of math and reading) 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.

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]**

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 are the author's calculations from the EDW and cover the period from 2001 to 2006.

Table 4

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

	Math Scores in Grade(s)				Reading Scores in Grade(s)			
	6-8	6	7	8	6-8	6	7	8
YEAR	0.034 [0.008]**	0.067 [0.009]**	0.023 [0.010]*	0.012 [0.008]	0.011 [0.007]	0.001 [0.008]	0.010 [0.008]	0.023 [0.008]**
CSR	0.049 [0.015]**	0.004 [0.008]	0.069 [0.015]**	0.077 [0.033]*	0.006 [0.018]	0.070 [0.011]**	-0.010 [0.013]	-0.043 [0.032]
YR_SINCE_CSR	-0.019 [0.014]	-0.052 [0.009]**	0.004 [0.014]	-0.008 [0.025]	0.029 [0.020]	0.036 [0.015]*	0.060 [0.020]**	-0.008 [0.026]
T x YEAR	0.009 [0.009]	0.014 [0.012]	0.011 [0.012]	0.001 [0.007]	0.018 [0.008]*	0.011 [0.009]	0.025 [0.008]**	0.019 [0.010]
T x CSR	0.008 [0.018]	-0.010 [0.018]	0.012 [0.018]	0.021 [0.038]	-0.022 [0.024]	-0.070 [0.032]*	-0.014 [0.018]	0.015 [0.034]
T x YR_SINCE_CSR	-0.000 [0.016]	0.008 [0.012]	-0.007 [0.016]	0.001 [0.026]	-0.017 [0.020]	0.015 [0.018]	-0.034 [0.020]	-0.030 [0.027]
Total effect by 2006	0.008 [0.034]	0.013 [0.035]	-0.010 [0.043]	0.023 [0.047]	-0.074 [0.047]	-0.025 [0.042]	-0.116 [0.051]*	-0.076 [0.056]
Observations (Student*Year)	3,296,513	1,095,097	1,111,338	1,090,078	3,301,663	1,096,173	1,112,510	1,092,980
R-squared	0.30	0.30	0.29	0.31	0.30	0.31	0.29	0.31

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 5

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

	Grades	
	3-5	6-8
YEAR	-0.9 [0.0]**	-0.8 [0.0]**
CSR	-0.9 [0.1]**	-0.5 [0.2]*
YR_SINCE_CSR	1.1 [0.1]**	0.4 [0.1]**
T x YEAR	0.3 [0.1]**	0.2 [0.1]*
T x CSR	-0.1 [0.2]	-0.9 [0.3]**
T x YR_SINCE_CSR	-1.2 [0.1]**	-0.4 [0.1]**
Total effect by 2009	-3.6 [0.3]**	-2.1 [0.4]**

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 are the author's calculations from the EDW and cover the period from 2001 to 2009.

Table 6

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

	Math		Reading	
	3-5	6-8	3-5	6-8
YEAR	0.085 [0.003]**	0.053 [0.003]**	0.077 [0.002]**	0.046 [0.003]**
CSR	-0.014 [0.007]	-0.011 [0.007]	-0.050 [0.006]**	-0.010 [0.007]
YR_SINCE_CSR	-0.030 [0.004]**	-0.017 [0.004]**	-0.043 [0.004]**	0.009 [0.004]
T x YEAR	0.007 [0.003]*	-0.001 [0.004]	0.009 [0.003]**	-0.003 [0.004]
T x CSR	-0.034 [0.010]**	-0.007 [0.009]	-0.011 [0.008]	-0.002 [0.008]
T x YR_SINCE_CSR	0.006 [0.006]	0.004 [0.006]	-0.012 [0.005]*	0.002 [0.006]
Total effect by 2009	-0.015 [0.017]	0.006 [0.018]	-0.046 [0.013]**	0.004 [0.016]
Observations (Student*Year)	2,199,287	2,659,264	2,201,767	2,662,444
R-squared	0.31	0.32	0.32	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 7

Effects of School-Level CSR on Average Teacher Experience

Grades 3-5				
	% New to Profession	% New to School	Total Experience	Exp in Current Job
T x CSR	0.028 [0.010]**	0.041 [0.016]**	-0.4 [0.2]	0.2 [0.1]
T x YR_SINCE_CSR	-0.012 [0.005]*	-0.018 [0.008]*	-0.1 [0.1]	0.9 [0.1]**
Total effect by 2009	-0.008 [0.012]	-0.012 [0.023]	-0.6 [0.3]*	3.0 [0.23]**
Grades 6-8				
	% New to Profession	% New to School	Total Experience	Exp in Current Job
T x CSR	0.002 [0.013]	-0.015 [0.020]	0.3 [0.2]	0.2 [0.2]
T x YR_SINCE_CSR	0.000 [0.006]	-0.007 [0.011]	-0.1 [0.1]	0.3 [0.1]*
Total effect by 2009	0.002 [0.016]	-0.035 [0.032]	-0.5 [0.4]	1.0 [0.4]*

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering at the school level appear in brackets. These data are calculated from the EDW (2002 to 2009 for percent new to profession or school and 2001 to 2009 for experience). All regressions include school fixed effects and controls for school-level percent black, percent hispanic, and percent eligible for free or reduced-price lunch.

Table A1

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]

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 A2

Effect of Required CSR at School Level on School Characteristics

	Grades 3-5				Grades 6-8			
	% Black	% Hisp	% FRL	Log(Enroll)	% Black	% Hisp	% FRL	Log(Enroll)
YEAR	-0.001	0.012	0.006	-0.012	0.003	0.009	0.010	-0.010
	[0.001]	[0.001]**	[0.001]**	[0.002]**	[0.001]**	[0.001]**	[0.002]**	[0.004]**
CSR	-0.014	-0.003	-0.038	-0.005	-0.015	0.001	-0.044	-0.042
	[0.002]**	[0.002]	[0.003]**	[0.008]	[0.002]**	[0.002]	[0.005]**	[0.015]**
YR_SINCE_CSR	0.003	-0.009	0.013	-0.009	0.001	-0.006	0.015	-0.017
	[0.001]**	[0.001]**	[0.002]**	[0.004]*	[0.001]	[0.001]**	[0.002]**	[0.007]*
T x YEAR	0.000	-0.003	-0.002	0.007	-0.003	-0.001	-0.003	0.008
	[0.001]	[0.001]**	[0.001]	[0.003]*	[0.002]	[0.001]	[0.002]	[0.005]
T x CSR	0.007	-0.001	-0.012	0.024	0.006	-0.000	-0.003	0.025
	[0.002]**	[0.002]	[0.004]**	[0.011]*	[0.003]*	[0.003]	[0.007]	[0.018]
T x YR_SINCE_CSR	0.000	0.004	0.002	-0.014	0.005	-0.000	0.002	0.005
	[0.001]	[0.001]*	[0.002]	[0.007]*	[0.002]**	[0.002]	[0.003]	[0.010]
Total effect by 2009	0.007	0.010	-0.005	-0.019	0.022	-0.002	0.003	0.041
	[0.004]	[0.004]*	[0.007]	[0.020]	[0.006]**	[0.005]	[0.010]	[0.029]

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 A3

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

	FCAT Math, Grades 6-8					
	SAT Scores	District Trends	No Charters	All Charters	Un-weighted	Diff-in-Diffs
T x CSR	0.012 [0.017]	0.018 [0.014]	0.017 [0.014]	0.016 [0.014]	0.001 [0.021]	0.041 [0.026]
T x YR_SINCE_CSR	-0.008 [0.016]	-0.001 [0.013]	0.004 [0.013]	0.002 [0.013]	0.003 [0.015]	
Total effect by 2006	-0.012 [0.041]	0.016 [0.037]	0.028 [0.035]	0.023 [0.034]	0.010 [0.049]	
Observations (Student*Year)	3,261,546	3,262,802	3,242,653	3,329,222	3,262,802	3,262,802
R-squared	0.29	0.28	0.28	0.28	0.28	0.29

	FCAT Reading, Grades 6-8					
	SAT Scores	District Trends	No Charters	All Charters	Un-weighted	Diff-in-Diffs
T x CSR	0.005 [0.013]	-0.014 [0.018]	-0.020 [0.020]	-0.022 [0.020]	0.003 [0.019]	0.010 [0.026]
T x YR_SINCE_CSR	-0.012 [0.015]	0.006 [0.010]	0.015 [0.012]	0.015 [0.011]	0.002 [0.014]	
Total effect by 2006	-0.032 [0.043]	0.004 [0.031]	0.025 [0.031]	0.022 [0.031]	0.009 [0.046]	
Observations (Student*Year)	3,264,686	3,267,501	3,247,327	3,334,191	3,267,501	3,267,501
R-squared	0.30	0.30	0.30	0.30	0.29	0.3

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. "SAT Scores" using Stanford Achievement Test Scores as the dependent variable. "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.

Table A4

District-Level Analysis, Fully Flexible Specification (Effects in Student-Level Standard Deviations), Grades 6-8

	Math	Reading
T x YEAR2001	-0.018 [0.017]	-0.036 [0.017]*
T x YEAR2002	0.012 [0.017]	0.013 [0.015]
T x YEAR2004 (Effect by 2004)	0.023 [0.011]*	-0.011 [0.011]
T x YEAR2005 (Effect by 2005)	0.032 [0.024]	-0.016 [0.028]
T x YEAR2006 (Effect by 2006)	0.041 [0.029]	-0.006 [0.038]
Observations (Student*Year)	3,296,513	3,301,663
R-squared	0.30	0.30

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.

Table A5

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

	Grades 3-5									
	FCAT Math					FCAT Reading				
	District Trends	Drop 2001 and 2002	No Charters	Un-weighted	Diff-in-Diffs	District Trends	Drop 2001 and 2002	No Charters	Un-weighted	Diff-in-Diffs
T x CSR	-0.034 [0.010]**	-0.026 [0.010]**	-0.036 [0.011]**	-0.028 [0.014]	0.011 [0.010]	-0.012 [0.008]	-0.010 [0.008]	-0.012 [0.008]	-0.012 [0.012]	0.005 [0.008]
T x YR_SINCE_CSR	0.007 [0.006]	0.013 [0.006]*	0.006 [0.006]	0.010 [0.008]		-0.011 [0.005]*	-0.013 [0.005]*	-0.012 [0.005]*	-0.006 [0.006]	
Total effect by 2006	-0.014 [0.017]	0.013 [0.018]	-0.017 [0.017]	0.003 [0.020]		-0.043 [0.013]**	-0.048 [0.014]**	-0.048 [0.013]**	-0.031 [0.015]*	
Observations (Student*Year)	2,199,287	1,711,270	2,135,254	2,199,287	2,199,287	2,201,767	1,713,290	2,137,616	2,201,767	2,201,767
R-squared	0.31	0.29	0.31	0.31	0.31	0.32	0.29	0.32	0.32	0.32

	Grades 6-8									
	FCAT Math					FCAT Reading				
	District Trends	Drop 2001 and 2002	No Charters	Un-weighted	Diff-in-Diffs	District Trends	Drop 2001 and 2002	No Charters	Un-weighted	Diff-in-Diffs
T x CSR	-0.006 [0.009]	-0.005 [0.008]	-0.009 [0.010]	0.019 [0.015]	-0.004 [0.010]	-0.002 [0.008]	0.004 [0.007]	-0.004 [0.008]	0.014 [0.014]	-0.011 [0.010]
T x YR_SINCE_CSR	0.006 [0.005]	0.005 [0.006]	0.004 [0.006]	-0.009 [0.008]		0.004 [0.006]	0.006 [0.006]	0.003 [0.006]	-0.005 [0.008]	
Total effect by 2006	0.012 [0.017]	0.009 [0.017]	0.002 [0.018]	-0.009 [0.024]		0.009 [0.016]	0.020 [0.018]	0.004 [0.016]	-0.002 [0.022]	
Observations (Student*Year)	2,659,264	2,090,717	2,556,346	2,659,264	2,659,264	2,662,444	2,093,314	2,559,533	2,662,444	2,662,444
R-squared	0.32	0.30	0.32	0.34	0.32	0.32	0.30	0.32	0.33	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 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.

Table A6

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

	Grades 3-5		Grades 6-8	
	Math	Reading	Math	Reading
T x YEAR2001	-0.031 [0.015]*	-0.051 [0.013]**	-0.000 [0.019]	0.002 [0.019]
T x YEAR2002	0.006 [0.014]	-0.006 [0.011]	0.013 [0.015]	0.025 [0.016]
T x YEAR2003	0.001 [0.012]	-0.028 [0.010]**	-0.002 [0.012]	0.015 [0.014]
T x YEAR2004	0.024 [0.011]*	-0.001 [0.010]	0.022 [0.012]	0.022 [0.015]
T x YEAR2005	0.035 [0.008]**	0.009 [0.006]	-0.003 [0.007]	-0.010 [0.009]
T x YEAR2007 (Effect by 2007)	0.001 [0.008]	-0.006 [0.006]	-0.001 [0.006]	-0.001 [0.005]
T x YEAR2008 (Effect by 2008)	0.023 [0.010]*	-0.004 [0.007]	0.001 [0.008]	-0.005 [0.006]
T x YEAR2009 (Effect by 2009)	0.029 [0.011]*	-0.011 [0.008]	0.005 [0.009]	-0.003 [0.008]
Observations (Student*Year)	2,199,287	2,201,767	2,659,264	2,662,444
R-squared	0.31	0.32	0.32	0.32

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.

Table A7

Achievement Effects of School-Level CSR by Subgroup (Student-Level Standard Deviations), Grades 3-5

	FCAT Math						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	-0.030	-0.038	-0.038	-0.046	-0.030	-0.026	-0.041
	[0.011]**	[0.012]**	[0.019]*	[0.017]**	[0.012]*	[0.014]	[0.011]**
T x YR_SINCE_CSR	0.008	0.005	0.019	0.003	0.011	0.006	0.010
	[0.006]	[0.007]	[0.012]	[0.009]	[0.006]	[0.008]	[0.006]
Total effect by 2009	-0.007	-0.022	0.018	-0.037	0.003	-0.008	-0.012
	[0.017]	[0.018]	[0.030]	[0.027]	[0.018]	[0.022]	[0.017]
Observations (Student*Year)	1,079,473	1,119,814	488,458	718,079	870,102	1,182,386	1,016,631
R-squared	0.29	0.32	0.23	0.28	0.22	0.25	0.21
	FCAT Reading						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	-0.006	-0.015	-0.030	-0.006	-0.009	-0.006	-0.015
	[0.009]	[0.010]	[0.016]	[0.014]	[0.010]	[0.011]	[0.009]
T x YR_SINCE_CSR	-0.012	-0.011	-0.005	-0.014	0.001	-0.014	-0.005
	[0.005]*	[0.006]*	[0.010]	[0.007]	[0.005]	[0.006]*	[0.005]
Total effect by 2009	-0.042	-0.049	-0.044	-0.049	-0.006	-0.047	-0.030
	[0.014]**	[0.015]**	[0.024]	[0.020]*	[0.014]	[0.017]**	[0.014]*
Observations (Student*Year)	1,080,824	1,120,943	488,872	718,738	871,306	1,183,750	1,017,739
R-squared	0.30	0.32	0.24	0.32	0.21	0.27	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.

Table A8

Achievement Effects of School-Level CSR by Subgroup (Student-Level Standard Deviations), Grades 6-8

	FCAT Math						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	-0.006	-0.008	-0.014	-0.024	0.011	-0.012	-0.002
	[0.010]	[0.010]	[0.017]	[0.014]	[0.011]	[0.012]	[0.010]
T x YR_SINCE_CSR	0.007	0.002	-0.003	0.006	0.009	-0.000	0.007
	[0.006]	[0.006]	[0.010]	[0.008]	[0.006]	[0.007]	[0.006]
Total effect by 2009	0.014	-0.002	-0.023	-0.007	0.037	-0.013	0.018
	[0.018]	[0.019]	[0.032]	[0.023]	[0.019]*	[0.022]	[0.017]
Observations (Student*Year)	1,316,016	1,343,248	547,120	720,336	1,258,947	1,170,695	1,488,264
R-squared	0.30	0.34	0.26	0.29	0.23	0.28	0.22

	FCAT Reading						
	Female	Male	Black	Hispanic	White	FRL	Non-FRL
T x CSR	0.002	-0.006	0.002	-0.030	0.016	-0.007	0.004
	[0.008]	[0.009]	[0.013]	[0.012]*	[0.010]	[0.010]	[0.009]
T x YR_SINCE_CSR	0.004	0.001	0.002	0.010	0.000	0.002	0.001
	[0.006]	[0.006]	[0.009]	[0.008]	[0.006]	[0.007]	[0.005]
Total effect by 2009	0.013	-0.002	0.006	0.001	0.017	-0.001	0.006
	[0.015]	[0.018]	[0.023]	[0.021]	[0.016]	[0.020]	[0.015]
Observations (Student*Year)	1,317,236	1,345,208	547,930	720,911	1,260,625	1,172,312	1,489,827
R-squared	0.30	0.32	0.26	0.34	0.20	0.29	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.

Table A9

Effects of School-Level CSR on Non-Cognitive Outcomes

	Grades 3-5		Grades 6-8		
	% Days Absent	% Days Absent	ICV per 100 pupils	% Students ISS	% Students OSS
T x CSR	0.000 [0.001]	0.001 [0.001]	-0.2 [0.6]	-0.008 [0.009]	-0.009 [0.006]
T x YR_SINCE_CSR	-0.001 [0.000]**	-0.000 [0.001]	-0.6 [0.3]*		
Total effect by 2009	-0.004 [0.001]**	0.000 [0.002]	-1.9 [0.8]**		
Level of Aggregation	Student	Student	School	School	School

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 six 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.