EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Volume 11 Number 40 November 10, 2003 ISSN 1068-2341

The Relationship between Exposure to Class Size Reduction and Student Achievement in California

Brian M. Stecher
Daniel F. McCaffrey
Delia Bugliari

RAND
Santa Monica, California

Citation: Stecher, B. M., McCaffrey, D.F. & Bugliari, D. (2003, November 10). The relationship between exposure to class size reduction and student achievement in California. *Education Policy Analysis Archives, 11*(40). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n40/.

Abstract

The CSR Research Consortium has been evaluating the implementation of the Class Size Reduction initiative in California since 1998. Initial reports documented the implementation of the program and its impact on the teacher workforce, the teaching of mathematics and Language Arts, parental involvement and student achievement. This study examines the relationship between student achievement and the number of years students have been exposed to CSR in grades K-3. The analysis was conducted at the grade level within schools using student achievement data collected in 1998-2001. Archival data collected by the state were used to establish CSR participation by grade for each school in the state. Most students had one of two patterns of exposure to CSR, which differed by only one year during grade K-3. The analysis found no strong association between achievement and exposure to CSR for these groups, after controlling for pre-existing differences in the groups.

Introduction

In 2002, Florida voters passed a comprehensive, class size reduction amendment, making Florida the most recent state to adopt this popular, but expensive, educational improvement strategy. During the 1990s, class size reduction (CSR) policies were proposed or adopted by more than a score of states. California was the most dramatic example. In 1996, California enacted SB 1777, providing a substantial incentive for school districts to reduce their class sizes from an average of roughly 30 students per class to 20 or fewer. With the signing of this bill, districts in 1996-97 were provided with nearly $1 billion in education funds to reduce class size in grades K-3. The funding then increased to roughly $1.5 billion in the second year (1997-98), and it has continued at this level in subsequent years. In addition to the state initiatives, the federal government invested more than $1 billion annually in the reduction of class size during the Clinton administration.

Despite the continuing enthusiasm among educational policymakers, the value of large-scale CSR efforts remains unproven. The relationship of class size to student performance has been studied for over 30 years with mixed results. (See Bohrnstedt and Stecher, 1999, for a comprehensive review of the literature.) Earlier findings regarding the efficacy of class size reduction were mixed, but recent high-profile studies, especially those related to the Tennessee STAR (Student/Teacher Achievement Ratio) project, have tipped the policy scales firmly in favor of smaller classes (Mosteller, 1995; Finn, 1998; Finn and Achilles, 1999). In this controlled experiment, researchers have found both short-term and long-term achievement gains associated with smaller class sizes in grade K-3 (Nye, Hedges, and Konstantopoulos, 1999). In fact, a recent study by Krueger and Whitmore (1999) shows that students who were in smaller classes in K–3 as part of the Tennessee STAR project were more likely to take high school courses known to lead to college attendance and to take college entrance examinations. Importantly, in all the STAR-related studies the gains were larger for minority and lower socio-economic students than for others.

Can these effects be achieved on a large scale? The experience of California offers important insights into class size reduction as a statewide policy. The size and complexity of initiating a class size reduction program in the nation's most populous state and the diversity of California's classrooms represent an important, real-world, test of the effectiveness of CSR as a broad-based policy. This paper presents the results of the most recent analysis of the relationship between the level of exposure to CSR and student achievement in California.

Summary of Previous Findings from California

The CSR Research Consortium, a group of California research and policy organizations, (Note 1) evaluated California's CSR program beginning in 1998. In the first two Class Size Reduction (CSR) evaluation reports (see Bohrnstedt and Stecher 1999; Stecher and Bohrnstedt, 2000), researchers estimated the impact of CSR on student achievement by comparing the Stanford Achievement Test, 9th Edition (SAT-9) test scores of third-grade students taught in reduced size classes with those of third-grade students taught in non-reduced size classes. (Note 2) Pre-existing differences between the CSR and non-CSR students were adjusted for statistically using student and teacher background characteristics as well as scores from fourth- and fifth-grade students who had little or no exposure to CSR.

Stecher, McCaffrey and Burroughs (1999) and Stecher, McCaffrey, Burroughs, Wiley and Bohrnstedt (2000) found that students who were exposed to CSR in third grade performed better than those who were not. This was true in 1997-98, when both groups of third grade students had little or no prior exposure to CSR, and it was true again in 1998-99, when both groups had one to two years of prior exposure. The differences in scores were equivalent to effect sizes of about 0.04 to 0.1 standard deviation units. In 1998-99, the differences were larger for mathematics and language than for reading and spelling. The researchers found that the effects of such "one-year" differences in CSR exposure were similar regardless of a school's population demographics, i.e., regardless of a school's percentage of minority, (Note 3) low-income, (Note 4) or English learner (EL) students. (Note 5) In 1998-99 the effects were somewhat larger in schools with the highest percentages of minority, low-income, or EL students, but the differences in scores were not statistically significant.

There was evidence that CSR effects persisted after students had returned to non-reduced classes for one year. Restricting their attention to students enrolled in the same school for three or more years, Stecher, et al., (2000) found that third graders who were in reduced classes in 1997-98 scored higher than their counterparts in non-reduced classes. Then, in 1999, after both of these groups had been in non-reduced fourth-grade classes, the first group again outperformed the second, and the difference was 0.04 standard deviation units. These fourth-grade effects were observed for students exposed to CSR solely in third grade and for students exposed to CSR in both second and third grade. There were no such effects, however, for students whose exposure was in second grade only.

In those selected cases where the California results could be directly compared with the results of the Tennessee STAR project., the findings were similar. The important exception is that there was no interaction between class size effects and demographic factors in California, while in Tennessee it was found that class size reduction had roughly twice as great an effect for minority students as for non-minority students. Unfortunately, because the researchers did not have achievement data prior to the introduction of CSR and did not have student achievement data from kindergarten and first grade students, they were unable to estimate the cumulative effects of four years of exposure to CSR in California's schools. The size of this effect was one of the chief findings from the Tennessee

STAR study.

For a number of reasons, it was not possible to use the same approach to judging the impact of CSR on achievement in subsequent evaluation reports. By 2000-01, CSR had been implemented in over 95 percent of the third-grade classes in California, leaving too few untreated students to serve as a comparison group. Furthermore, some or all of the upper-grade (i.e., fourth- and fifth-grade) students in most schools had participated in reduced size classes in earlier years, so their test results could not be used to control for pre-existing differences. Thus, the analytic strategies used in the first two evaluations of the California CSR program were no longer applicable in subsequent years.

However, the large but uneven growth in participation in CSR over time provided an opportunity to look at the impact of CSR on achievement in a different manner. From 1996-97 to 2000-01, CSR went from partially implemented in two grade levels to almost fully implemented in four grade levels (kindergarten through third grade). In the third evaluation report, Stecher, Bugliari, and McCaffrey (2002) used statewide test results to compare achievement results among cohorts of students who had different patterns of exposure to CSR. Trends in achievement that corresponded to patterns of exposure provide evidence in support of the hypothesis that CSR improves achievement; trends that have no relationship to CSR participation offered no such support.

Focusing on statewide average achievement scores during the period 1997-98 to 2000-01, the researchers compared the average achievement of successive cohorts of students as they moved through the system with their average exposure to CSR. Successive cohorts of students had higher achievement during this period, which suggests that one or more of the state educational reforms (which include CSR, new curriculum standards, a statewide standardized testing program, the end of bilingual education, and high stakes accountability) had a positive effect. However, the trend in test scores over this period was unrelated to the trend in CSR exposure, so the researchers could not make a strong case that CSR was chiefly responsible for achievement gains.

Yet, aggregate analyses do not tell the whole story. For example, the state level analysis could not control for external effects, such as student mobility. Neither did it permit the researchers to examine the influence of student or teacher background characteristics. The present study addresses these limitations by analyzing trends in exposure and achievement at the school level, where more data are available to refine the comparisons and control potentially confounding factors.

Methods

Achievement Data

Beginning in 1998, California students in grades 2-11 have been required to complete the SAT-9 annually in the spring. The test results are reported in the summer and fall, and they are made available for research purposes in the public release California Standardized Testing and Reporting (STAR) data files. All analyses reported below use the public release STAR data (http://www.cde.ca.gov/statetests/).

As part of STAR testing, students complete standardized multiple-choice tests in mathematics, reading, language and spelling. We focus here on mathematics, reading and language. We use SAT-9 scale scores (rather than raw scores, percentile ranks, or normal curve equivalents) as measures of achievement in these analyses because scale scores are designed so that score differences are comparable for the entire range of scores. In addition, the scales are equated across grade levels, facilitating cross grade comparisons.

School Sample

The initial school sample included 4,961 elementary schools in the STAR data files from school years 1997-98 through 2000-01. We excluded those schools for which the STAR file in any year contained scores for 10 or fewer students and those schools for which the STAR files were missing basic demographic data (gender, ethnicity, English language fluency status) on all students. These criteria excluded 2,069 school (42 percent), leaving 2,892 schools in our analysis file.

Despite the exclusions, the schools in our sample closely resemble the schools in the state as a whole in terms of student demographic characteristics. Table 1 shows the comparison between the sample schools and the whole state in terms of participation in CALWORKS, eligibility for free or reduced priced lunches, race/ethnicity, and language status for the 1999-2000 school year. The mean values for sample schools are within one to three percentage points of the mean values for the state as a whole on all variables, so the generalizability of the results from our analyses are not limited by the populations served by sampled schools.

Table 1. Demographic Characteristics of Sample Schools and All Elementary Schools

Demographic feature	All elementary schools[a]	Analysis sample schools
Percent CALWORKS participants	13.59 (12.88)	14.64 (12.97)
Percent free or reduced price lunch eligible	51.99 (30.27)	53.76 (30.17)
Percent white	38.39 (29.38)	34.86 (28.66)
Percent Hispanic	40.98 (29.31)	43.19 (29.60)
Percent African American	8.13 (12.60)	9.28 (14.00)
Percent Asian	7.76 (12.01)	7.88 (11.40)
Percent minority	61.61 (29.38)	65.15 (28.66)
Percent ELL	27.14 (24.10)	29.29 (24.30)
Total enrollment	609.94 (282.39)	660.40 (276.38)

[a]State sample includes 4,761 elementary schools open since 1996 with CDS codes.

Class Size Reduction Participation

Class size reduction began with the 1996-97 school year, one year prior to STAR testing. By the 1999-2000 school year over 90 percent of all students in kindergarten through third grade were participating in CSR. However, for earlier cohorts, CSR participation varied across schools. This variation provided an opportunity to compare achievement with CSR exposure. The first step in our analysis, therefore, was to determine CSR participation by grade and school year for each of the 2,892 schools in the analysis file. We focused on CSR participation for three cohorts of student--those who entered kindergarten in 1995-96 (K95), 1996-97 (K96) or 1997-98 (K97). These three cohorts of students reached the third grade in 1999, 2000, and 2001 and they are the only cohorts with exposure to CSR for whom we have SAT-9 scores in both second and third grade.

For each elementary school in California we developed an indicator of CSR participation by grade level by year. Unfortunately, the state did not collect comparable information about CSR participation every year, so we had to use multiple data sources to infer CSR status. The primary data for assessing CSR status were the individual student SAT-9 answer files, which included indicator variables for CSR participation for every student. We also used teacher reports of classroom enrollment from the CBEDS Professional Assignment Information Form (PAIF). A third source of information was the district level J-7 CSR report, which describes district participation in CSR for the 1996-97 and 1997-98 school years (http://cde.ca.gov/csr/). The J-7 information was only useful when participation was uniform across the district. Finally, the CBEDS School Information File (SIF) data contain school and grade level CSR indicators for the 1998-99, 1999-2000, and 2000-01 school years.

The CSR indicator development process began with the student-level STAR data file. If 10 percent or fewer students within a grade at a school were coded as participating in the CSR program (either option 1 or 2), we classified that grade as not reduced. If 90 percent or more students within a grade at a school were indicated as in the CSR program, we classified that grade as reduced. We classified a grade as undetermined by STAR if between 10 percent and 90 percent of students were indicated as CSR. Let $C_{gjt,STAR}$ denote the CSR status for grade g = kindergarten, 1, 2 or 3, in school j and school-year t = 1996-97, 1997-98, 1998-99, 1999-00 or 2000-01. $C_{gjt,STAR}$ equals "R" if we determine the school had reduced classes for grade g in year t; $C_{gjt,STAR}$ equals "N" if not reduced and "U" if undetermined.

Because the STAR data did not permit clear classification for every school, grade level, or school year, i.e., in some instances $C_{gjt,STAR}$ equals "U," we turned to other sources to make our final determination of CSR participation. The PAIF data provide the number of students in each teacher's classroom and the number of teaching assignments. The distribution of students across classrooms for teachers with multiple assignments cannot be determined from the PAIF. Therefore, for determining CSR participation we used only teachers with a single teaching assignment. Also, some teachers report over 50 students or fewer than 14 students in their classroom. We excluded these teachers from the classification process, arguing that they represented data errors or nontraditional education assignments.

A school was judged to have reduced size classes for a given grade in a given year if over 65 percent of included teachers in that grade reported 21 or fewer students. If fewer than 35 percent of included teachers in a grade reported 21 or fewer students, we classified that grade as not reduced. We classified a grade as undetermined by PAIF if between 35 percent and 65 percent of the classes were reported as having 21 or fewer students. We let $C_{gjt,PAIF}$ denote the CSR status as determined by the PAIF where the variable again takes on the values of "R," "N," and "U" for reduced, not reduced or undetermined.

We also created variables for the CSR participation as determined by the SIF ($C_{gjt,SIF}$) and the J-7 data ($C_{gjt,J7}$). $C_{gjt,SIF}$ equals "U" for the 1996-97 and 1997-98 school years for all grades and schools because grade-level CSR indicators were not added to SIF until 1998-99. Finally, $C_{gjt,J7}$ takes on values "R" and "N" only if the district had uniform CSR practices at a grade level across all schools.

For final CSR classification, we compared the CSR indicators based on STAR, PAIF, SIF and J-7. In the majority of cases, all determinable sources agreed, $C_{gjt,STAR} = C_{gjt,PAIF} = C_{gjt,SIF} = C_{gjt,J7}$ or some variables equaled "U" and the remaining variables agreed. In these cases we assigned the common value to the CSR indicator. In the cases of disagreement, we examined the longitudinal trend in CSR indicators before making a final determination. For example, if $C_{gjt,STAR}$= R and $C_{gjt,PAIF}$= N for year *t* we checked the data for the previous year (*t* - 1). If $C_{gjt-1,STAR} = C_{gjt-1,PAIF}$ = R, then we decided that the school probably had reduced class size in year *t* as well. Schools for which we were unable to resolve data conflicts confidently were excluded from the final analytic file. We excluded 543 schools because of indeterminate CSR status, leaving a sample of 2,349 schools. The excluded schools constituted 19 percent of the 2,892 schools that met the data and size conditions described above. The remaining schools constituted 47 percent of the original sample.

CSR Exposure by Cohort

For each of the three focal cohorts, K95, K96 and K97, Tables 2, 3 and 4 present the distribution of CSR exposure across the final sample of schools. Table 2 shows that nearly 90 percent of the schools in the sample had one of two patterns of CSR exposure for the K95 student cohorts: CSR in grades 2 and 3 only (22.3 percent) or CSR in grades 1, 2 and 3 (66.8 percent). For the K96 cohort there was even less variation in CSR exposure. Table 3 shows that these students participated in CSR for grades 1, 2 and 3 in almost every school (89.9 percent). By the K97 cohort, Table 4 shows that more schools introduced CSR in kindergarten, and the schools fell, almost exclusively, into one of two patterns of CSR exposure: kindergarten through grade 3 (38.8 percent) or grades 1, 2 and 3 (59.9 percent).

Table 2. Distribution of CSR Exposure for Cohort K95

Exposure pattern	Number of schools	Percent of sample
Indeterminate	20	0.9
None	25	1.1
Grade 3 only	7	0.3
Grade 2 only	66	2.8

Grades 2 and 3	525	22.3
Grade 1 only	10	0.4
Grades 1 and 3	5	0.2
Grades 1 and 2	105	4.5
Grades, 1, 2 and 3	1,569	66.8
Kindergarten and grade 3	1	0.0
Kindergarten, grades 2 and 3	1	0.0
Kindergarten, grades 1, 2 and 3	15	0.6

Table 3. Distribution of CSR Exposure for Cohort K96

Exposure pattern	Number of schools	Percent of sample
Indeterminate	12	0.5
None	1	0.0
Grade 3 only	1	0.0
Grades 2 and 3	12	0.5
Grade 1 only	4	0.2
Grades 1 and 3	1	0.0
Grades 1 and 2	50	2.1
Grades, 1, 2 and 3	2,112	89.9
Kindergarten, grades 2 and 3	1	0.0
Kindergarten, grades 1, 2 and 3	155	6.6

Table 4. Distribution of CSR Exposure for Cohort K97

Exposure pattern	Number of schools	Percent of sample
Indeterminate	7	0.3
Grades 2 and 3	1	0.0
Grades 1 and 2	20	0.9
Grades, 1, 2 and 3	1,406	59.9
Kindergarten, grades 2 and 3	1	0.0
Kindergarten, grades 1 and 2	3	0.1
Kindergarten, grades 1, 2 and 3	911	38.8

Grouping Schools by CSR Exposure

We focused our analyses on four groups of schools with distinctive patterns of CSR exposure. These 1,918 schools constitute 82 percent of the schools in the final analysis sample and 40 percent of the schools in the original sample. Table 5 shows these four patterns. Because few schools had any of the remaining exposure patterns, we restrict the study to schools in these four groups.

Table 5. Distribution of CSR Exposure for All Three Cohorts

Group	K95	K96	K97	Number of schools
A	1, 2, 3	1, 2, 3	1, 2, 3	877
B	2, 3	1, 2, 3	1, 2, 3	348
C	2, 3	1, 2, 3	K, 1, 2, 3	152
D	1, 2, 3	1, 2, 3	K, 1, 2, 3	541

Demographic differences across groups (described below) led us to focus our primary comparisons of outcomes on Group A and Group B. These two groups contain 1,225 schools. In Group A, students who entered kindergarten in 1995-96, 1996-97 or 1997-98 had reduced-size classes in grades 1, 2 and 3 (but not kindergarten). Group B schools serve a similar population of students, but the three cohorts had different exposure to CSR. Students entering kindergarten in 1995-96 had two years of exposure to CSR in second and third grade, those entering in subsequent years had an additional year of CSR in first grade.

Student Sample

As noted above, our analyses are restricted to students in the K95, K96 and K97 cohorts. From these cohorts we included only those students who: 1) attended the same school for kindergarten through second or third grade, depending on the grade of the outcome used in the analysis; 2) did not have a test identified as "Out of Level"; and 3) were not identified as receiving Special Education services. We also excluded students when their STAR data CSR flag was inconsistent with the data from the vast majority (over 90 percent) of their fellow students in the same grade and school. For example if the STAR student data file indicated that for a particular school over 90 percent of third graders in a cohort were in reduced size classes, then we excluded any third graders from that school and cohort for whom the STAR data indicated they were not in reduced size classes.

Table 6 contains summaries of the student demographic characteristics and teacher qualifications of the identified cohorts of students in schools in the four groups. Groups A and B are similar in terms of students and teacher characteristics, while Groups C and D are distinctly different. Schools in Groups A and B have greater percentages of minority students, EL students, and students from families receiving public assistance than schools in Groups C and D. Groups A and B also are similar in terms of teacher characteristics, and they have fewer teachers who are fully-credentialed than schools in Groups C and D. These differences make comparisons between Groups C and D and the other groups difficult because such comparisons would confound student demographics and teacher qualifications with CSR effects. Therefore we focus only on Groups A and B.

There is one instance in which schools in Groups A and B differ with respect to teacher credentials that only is apparent when the data are disaggregated by cohort. Group B schools have more uncredentialed first-grade teachers than Group A schools for cohorts K96 and K97. This difference appeared when Group B introduced CSR at first grade, and it probably is a result of these schools hiring new teachers in the tight teacher labor market that followed the introduction of CSR. (See Tables 7-12 for student and teacher characteristics disaggregated by cohort and grade level.)

Table 6. Average Student and Teacher Characteristics for Cohorts K95, K96 and K97, by Group

Group	Student characteristics[a]			Teacher characteristics[b]	
	Minority %	EL %	AFDC %	Experience	Credential
A	66.84	33.23	20.40	13.30	89.13
B	69.23	32.06	21.09	13.25	88.51
C	57.66	25.91	18.38	13.46	93.10
D	51.99	20.67	18.26	13.52	94.71

[a]Average for the three cohorts during their kindergarten, first, second, and third grades.

[b]Average years of experience for teachers of the identified cohorts of students; percentage of teachers of the identified cohorts of students with full credentials.

Table 7. Percentage of Students in Cohort Whose Families Receive AFDC During Four Years, by Group

Group	Cohort	Kindergarten	First grade	Second grade	Third grade
A	K956	24.84	23.71	20.60	19.12
	K967	23.71	22.06	19.12	17.43
	K978	22.06	19.12	17.43	15.62
B	K956	25.36	24.34	22.01	20.06
	K967	24.34	22.99	20.06	17.74
	K978	22.99	20.06	17.74	15.39
C	K956	23.11	22.67	17.82	17.12
	K967	22.67	20.12	17.12	15.00
	K978	20.12	17.12	15.00	12.64
D	K956	21.39	20.97	21.70	17.44
	K967	20.97	19.16	17.44	15.13
	K978	19.16	17.44	15.127	13.16

Table 8. Percentage of Minority Students in Cohort During Four Years, by Group

Group	Cohort	Kindergarten	First grade	Second grade	Third grade
A	K956	64.79	65.69	63.78	67.55
	K967	65.69	66.57	67.55	68.42
	K978	66.57	67.55	68.42	69.57
B	K956	66.57	67.98	62.68	70.36
	K967	67.98	69.20	70.36	71.57
	K978	69.20	70.36	71.57	72.92
C	K956	55.10	56.45	52.88	58.64
	K967	56.45	57.08	58.64	59.86
	K978	57.08	58.64	59.86	61.21

D	K956	49.11	49.84	56.78	52.05
	K967	49.84	50.80	52.05	53.14
	K978	50.80	52.05	53.14	54.32

Table 9. Percentage of EL Students in Cohort During Four Years, by Group

Group	Cohort	Kindergarten	First grade	Second grade	Third grade
A	K956	32.51	32.92	33.26	33.37
	K967	32.92	33.33	33.37	33.48
	K978	33.33	33.37	33.48	33.40
B	K956	30.79	31.81	29.59	32.50
	K967	31.81	32.24	32.50	32.86
	K978	32.24	32.50	32.86	32.97
C	K956	24.50	25.45	23.58	26.26
	K967	25.45	26.49	26.26	26.52
	K978	26.49	26.26	26.52	27.15
D	K956	19.00	19.72	24.88	20.35
	K967	19.72	20.50	20.35	20.83
	K978	20.50	20.35	20.83	20.96

Table 10. Average Years of Teaching Experience for Teachers of Cohort During Four Years, by Group

Group	Cohort	Kindergarten	First grade	Second grade	Third grade
A	K956	16.08	13.24	12.85	12.51
	K967	16.87	11.15	12.61	12.94
	K978	14.58	10.72	12.59	13.45
B	K956	15.54	13.33	13.11	12.18
	K967	16.45	11.25	12.69	12.70
	K978	15.14	10.75	12.85	12.99
C	K956	16.30	13.04	12.38	13.86
	K967	15.58	11.17	12.62	14.44
	K978	13.24	11.53	12.96	14.42
D	K956	16.42	11.99	13.04	13.75
	K967	15.33	11.93	12.98	14.03
	K978	12.88	12.27	13.34	14.24

Table 11. Percentage of Teachers of Cohort with Full Credentials During Four Years, by Group

Group	Cohort	Kindergarten	First grade	Second grade	Third grade
A	K956	98.06	95.21	87.34	85.21
	K967	96.22	85.56	87.10	87.05
	K978	88.09	85.11	86.65	88.01
B	K956	98.78	95.74	86.56	84.18

	K967	96.26	83.11	86.09	86.28
	K978	88.96	82.29	85.75	88.11
C	K956	98.99	97.31	92.66	90.27
	K967	97.85	92.24	89.47	92.27
	K978	91.98	89.21	91.67	93.27
D	K956	98.45	96.58	93.17	94.05
	K967	97.01	93.63	94.41	94.28
	K978	92.39	94.02	94.19	94.34

Table 12. Parameter Estimates (Standard Errors) for Model 1

	Grade 2			Grade 3		
	Math	Reading	Language	Math	Reading	Language
Mean Group A, K95	569.5 (0.9)	571.5 (1)	583.3 (0.9)	603.3 (1)	608.8 (1.1)	607.6 (1)
Difference, Group A K96 less K95	7.3 (0.3)	5 (0.3)	4.4 (0.3)	6.9 (0.3)	4.6 (0.3)	5.8 (0.3)
Difference, Group A K97 less K95	13.2 (0.4)	10.9 (0.4)	9.1 (0.4)	12.3 (0.4)	9.4 (0.3)	10.4 (0.4)
Difference between Groups K95	-8.6 (1.7)	-4.5 (1.8)	-5.4 (1.7)	-5.8 (1.8)	-6.7 (2)	-7.3 (1.8)
Group B linear trend	1.9 (0.5)	-0.1 (0.5)	0.5 (0.5)	0.7 (0.5)	0.7 (0.5)	0.8 (0.5)
Effect of additional year CSR at grade 1	-0.9 (0.7)	1.7 (0.7)	0.9 (0.7)	0.7 (0.7)	-1.1 (0.7)	-0.8 (0.7)

Note: The difference parameter estimates of the Difference, Group A K96 less K95 and the Difference, Group A K97 less K95 contain the Group A linear trend and the common (across Groups) cohort deviations from the linear trend.

Group A schools had between 53,000 and 59,000 students per cohort when the cohorts reached grade 2, and between 46,000 and 48,000 students per cohort when the cohorts reached third grade. For Group B, the numbers of second graders per cohort ranged from 23,000 to 25,000 and the number of third graders per cohort ranged from 19,000 to 21,000. The samples are smaller in third grade than in second grade because they are restricted to students who attended the same school for one additional year.

Analysis

Our goal is to determine if cohort-to-cohort variation in CSR exposure predicts cohort-to-cohort variation in test scores. On the basis of the exposure patterns presented in Table 5, we note that a comparison of schools across years, groups and cohorts can only provide data on the effects of a one-year variation in exposure to CSR. Larger differences in exposure do not exist among comparable groups of schools. In addition, other reforms and changes were taking place during this period that might have affected test scores. As a result, a simple comparison of scores for students in the K95 cohort with scores for students in the K96 or K97 cohorts might confound CSR effects with these other changes. More complex comparisons, however, can isolate the effects of CSR with less confounding of alternative effects. For example, because the exposure to CSR was the same for all three cohorts in Group A, these schools provide a measure of the effects of factors unrelated to CSR on the trend in scores over these three years. Similarly, differences between K96 and K97 scores in Group B schools also are unrelated to CSR because exposure was the same for these two cohorts (but not for the K95 cohort). Thus, differences among these five cohorts in Groups A and B can be used to estimate the effects of other programs and the effects of cohort-to-cohort variation.

On the other hand, the students in the Group B-K95 cohort had one year less CSR exposure during first grade than the students in the two later Group B cohorts and than students in all three cohorts in Group A. By comparing scores for the Group B-K95 students to those of other students, we can observe differences between groups with varying exposure to CSR. However, we must make judicious use of the data from the other students to limit the confounding effects of other programs and cohort-to-cohort variation in scores. The following list of comparisons with Group B-K95 highlights the assumptions about groups and time trends that are required for the comparisons to provide unconfounded estimates of the CSR effect. It also points out the comparisons that we believe provide the best estimates of the CSR effect.

Comparison 1: Compare Group B-K95 scores to Group B-K96 or Group B-K97 scores. The comparison yields unconfounded estimates of the CSR effect if we assume that, in the absence of CSR, scores do not change systematically over time. However, research has consistently shown that score gains occur in the years following the introduction of a new, high-stakes testing program even in the absence of other initiatives. Thus, this assumption seems unwarranted, i.e., scores are likely to change over time even in the absence of CSR. In fact, this change is evident in Group A where CSR exposure is constant. As a result, we will not use these within-Group B comparisons as an estimate of the CSR effect.

Comparison 2: Compare Group B-K95 scores to Group A-K95 scores. This comparison yields unconfounded estimates of the CSR effect if we assume that, in the absence of CSR, the groups would have the same scores on average. At first this assumption seems reasonable because the schools in the two groups are very similar on student demographic and teacher characteristics. However, the schools in Group A implemented CSR more quickly than schools in Group B, and the factors that led to this alternative behavior might be related to average scores. Thus, we do not think this assumption is warranted. (Alternatively, comparison of Group B-K95 to Group A-K96 or K97 would be affected both by time trends and cross group differences. The required assumptions for unconfounded estimation are not tenable in these comparisons either.)

Comparison 3: Compare the difference between Group B-K96 and Group B-K95 to the difference between Group B-K97 and Group B-K96. This comparison attempts to remove the time trend by using the difference between Group B-K97 and Group B-K96 scores as an estimate of the time trend between K95 and K96. The comparison yields unconfounded estimates if we assume that the time trend in scores is linear across the three cohorts. This is one of the estimates that will be presented in the Results section. (In Table 13, Comparison 3 is found in the row labeled Difference and the column labeled Group B.)

Comparison 4: Compare the difference between Group B-K95 and Group A-K95 to the difference between Group B-K96 and Group A-K96. (This is equivalent to comparing the difference between K96 and K95 for Group B to the difference between K96 and K95 in Group A.) This comparison uses differences across Groups in K96 to estimate differences across groups in K95. Alternatively, we can view this estimate as using Group A to estimate the time trend from K95 to K96. This estimate is unconfounded if we assume that, in the absence of CSR, group differences would be constant over time. (We could also include the K97 cohorts in these comparisons.) We also present this comparison in the Results section. (In Table 13, Comparison 4 is found in the row labeled K96 less K95 and the column labeled Difference.)

Comparison 5: Compare the difference in differences for Group B (i.e., compare the difference between K96 and K95 and the difference between K97 and K96) to the difference in differences for Group A. This model uses Group A to estimate the size of cohort-to-cohort deviations from a linear time trend in Group B. This model produces unconfounded estimates of the CSR effect if we assume that no interactions would exist in between groups and deviations from time trends in the absence of CSR. (In Table 13, Comparison 5 is found in the row labeled Difference and column labeled Difference.)

Because scores for students within the same school might be positively correlated and because schools vary in size, the simple average estimators described above might not be efficient. Therefore, we also fit a hierarchical linear model to estimate Comparison 5 while allowing for possible intra-school correlation. Model 1 for a score for the *k*th student in cohort *t* (*t* = 1 for K95, 2 for K96 and 3 for K97), school *j* of group *i*, y_{ijtk}, is given by

$$y_{ijtk} = \mu_{ij} + \beta_{ij} t + \varphi_1 I(t=1) + \varphi_2 I(t=2) + \varepsilon_{ijtk}$$
$$\mu_{ij} = \mu_i + \eta_{ij}$$
$$\beta_{ij} = \beta_i + \zeta_{ij}$$
$$\varepsilon_{ijtk}\ i.i.d.\ N(0,\sigma^2), \eta_{ij}\ i.i.d.\ N(0,\tau^2),\ \zeta_{ij}\ i.i.d.\ N(0,\nu^2),\ Corr(\eta_{ij}, \zeta_{ij}) = \rho$$

The functions I(*t* = 1) and I(*t* = 2) equal one if *t* = 1 or 2 respectively and zero otherwise. SAS Proc Mixed provided estimates of the coefficients of the random effects model. We also used fixed school effects models and the results were nearly identical. Sensitivity analyses were conducted to explore the effects of teacher credentials, and the results were essentially unchanged.

We fit Model 1 separately for grades 2 and 3. Individual student scores are not linkable over time in the STAR data, so growth modeling was not possible. Models of change for cohorts within school were feasible, but because we had no hypotheses on the effects of a year's delay in CSR for growth in the following two years, we looked only at the effects within each grade.

Results

CSR Effects on Math, Reading and Language Test Score

There is an upward trend in scores across cohorts K95, K96, and K97 in both Group A and Group B schools (see Figure 1). The top panel of the figure shows the box and whisker plots of the distribution of school mean math scores for the three cohorts of students from Group A schools. The dot corresponds to the median score, the upper and lower sides of the rectangle correspond to the 25th and 75th percentiles of the distribution, and the brackets at the ends of the whiskers correspond to the 5th and 95th percentiles of the distribution of scores. Dots beyond the whiskers are extreme outliers.

There is an obvious upward trend in scores across cohorts over time, as the distribution shifts to the right for each successive cohort. However, in Group A schools, all three cohorts experienced exactly the same pattern of CSR exposure (grades 1 through 3). Thus, the trend in scores is not related to changes in the level of CSR exposure.(Note 6) During the time period that our three study cohorts were in kindergarten through grade 3, California enacted several other statewide education initiatives, including the introduction of demanding new curriculum standards, a statewide standardized testing program with high-stakes accountability, and the end of bilingual education. All of these programs might contribute to rising test scores across cohorts, even if differences in CSR have no effect.

The lower panel in Figure 1 shows box and whisker plots for the cohorts in the Group B schools. The plots for Group B show a nearly identical trend to the plots for Group A, even though students in cohort K95 in Group B had one year less exposure to CSR than students in the other two cohorts in Group B. Figures for reading and language scores show similar patterns (see Figures 2 and 3). On the basis of this figure, it seems clear that the additional year of CSR in first grade did not have large effects on mathematics scores.

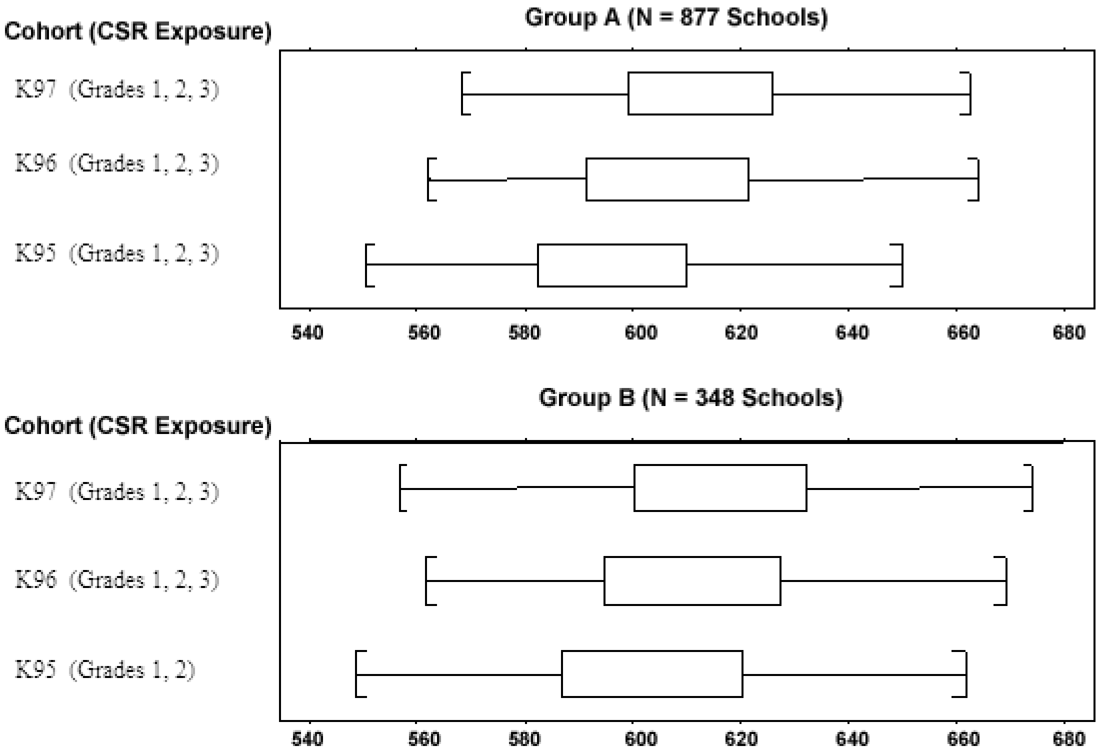

Figure 1. Third grade SAT-9 score distributions in mathematics for successive cohorts of students with constant vs. increasing CSR exposure.

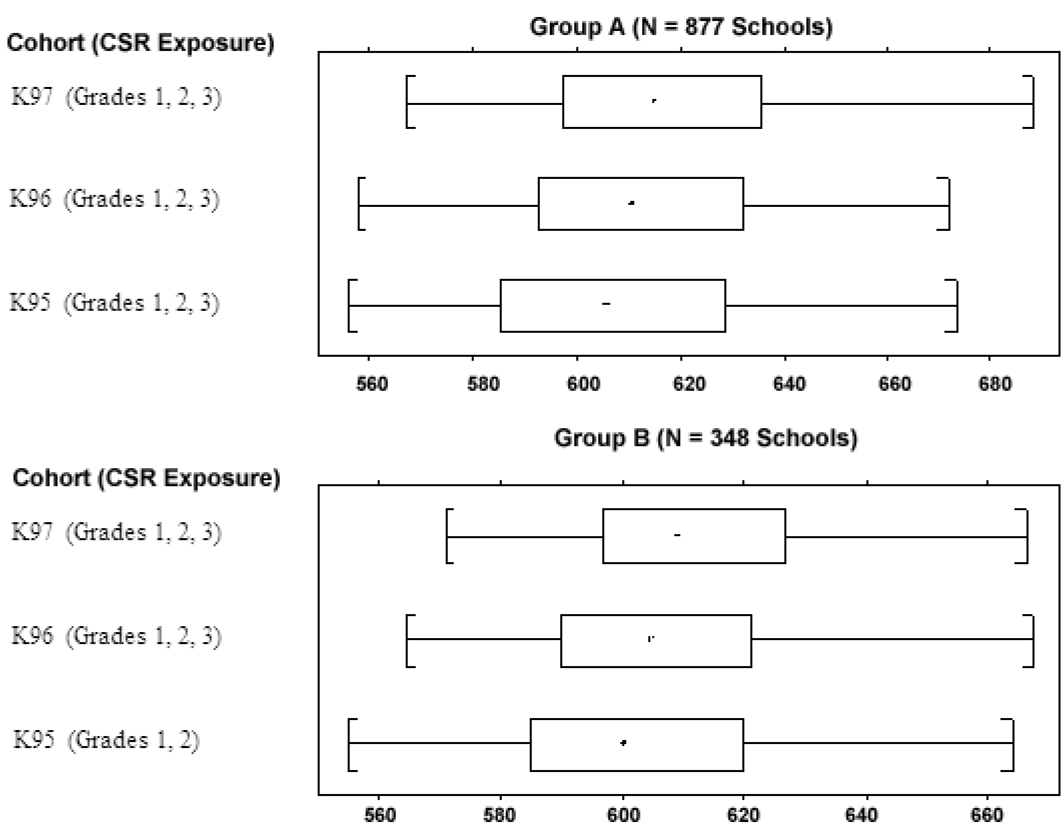

Figure 2. Third grade SAT-9 score distributions in reading for successive cohorts of students with constant vs. increasing CSR exposure.

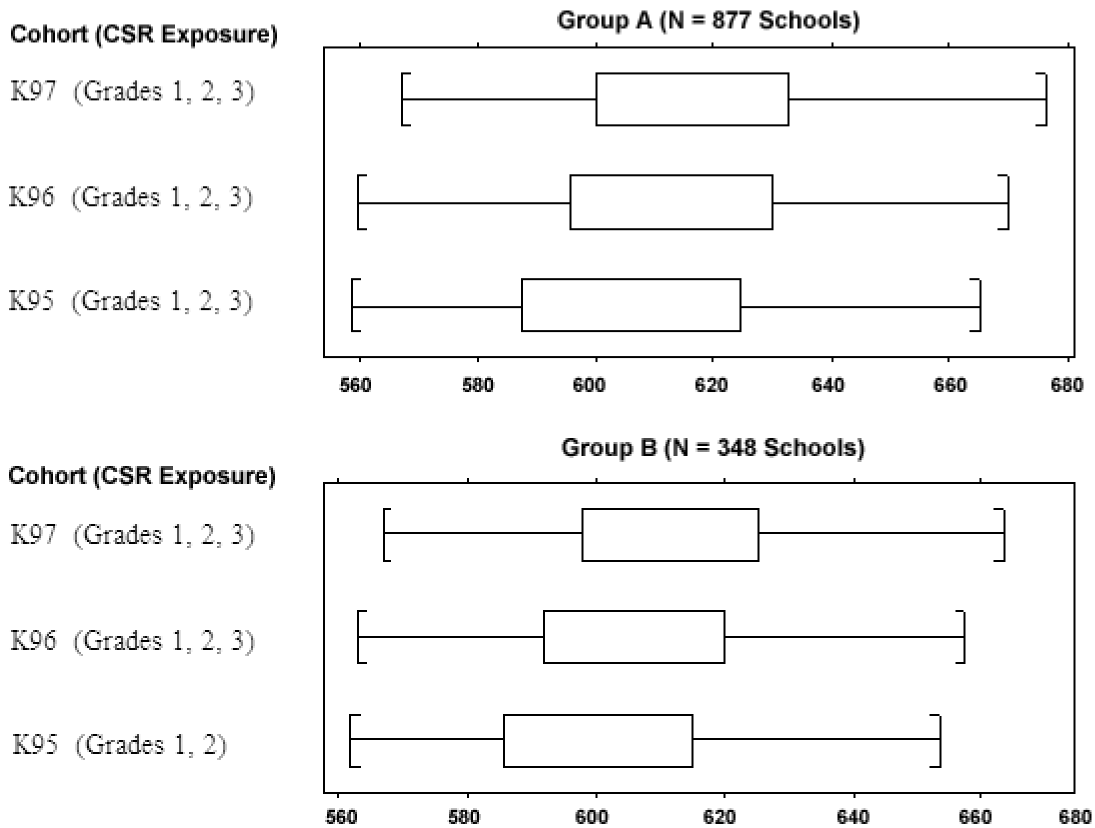

Figure 3. Third grade SAT-9 score distributions in language for successive cohorts of students with constant vs. increasing CSR exposure.

Table 13 provides further evidence that, for students in these cohorts and schools, the effects of an additional year of CSR were small. In Table 13a, the first row presents the differences between mean second-grade math scores for K96 and K95 for Groups A and B, and the difference between these differences. The second row contains the differences between mean second-grade math scores for K97 and K96 for the two groups and the difference between the differences. In the third row we have the difference of these two cohort-to-cohort differences in each Group and between the groups. Tables 13b-13f contain similar differences for grade 3 mathematics scores and for grades 2 and 3 reading and language scores.

The table contains the results of comparisons 3, 4 and 5 among cohort means by group, grade, and subject. For example, for Group B, the difference in mean scores for K96 and K95 is the difference between a cohort of students that participated in CSR in grades 1, 2 and 3 and a cohort that participated only in grade 2 and 3. Thus, the value of 6.49 from Table 13a represents in part an effect of one additional year of CSR when students were tested in second grade. It also includes other effects occurring during this time. Comparison 3 attempts to remove the time trend in this comparison by using the difference between K97 and K96 in Group B, which is found in Table 13a to be 8.05. Under the assumptions listed above, the difference between these two values produces an unconfounded estimate of the CSR effect as - 1.57 (the last row of Table 13a in the Group B column).

Comparison 4 uses the difference between K96 and K95 in Group A to estimate the natural trend in scores, and adjusts the Group B differences accordingly. This produces an estimate of the CSR effect as 1.15 (the last column in the first row of Table 13a). As noted above, each estimate makes different assumptions about what has remained constant across time or groups. The estimate in the Group B column assumes that changes from cohort to cohort in Group B are constant except for CSR. The estimate in the K96 less K95 row assumes that changes from K95 to K96 are constant across Groups A and B except for CSR.

Comparison 5 assumes that, except for the effects of CSR, cohort-specific deviations from a linear trend are constant across groups. This difference of differences approach provides an estimate of the CSR effect equal to - 0.52. This value is computed as the difference of the values for Groups B and A in the last row of Table 13a. (The estimate is given in the Difference column of the Difference row of Table 13a.)

Table 13a. Second Grade Math

	Group A	Group B	Difference
K96 less K95	5.34	6.49	1.15
K97 less K96	6.39	8.05	1.67
Difference	-1.05	-1.57	-0.52

Table 13b. Third Grade Math

	Group A	Group B	Difference
K96 less K95	6.79	8.17	1.38
K97 less K96	6.53	7.29	0.71
Difference	0.26	0.93	0.67

Table 13c. Second Grade Reading

	Group A	Group B	Difference
K96 less K95	2.05	3.66	1.61
K97 less K96	6.26	6.05	-0.21
Difference	-4.21	-2.39	1.82

Table 13d. Third Grade Reading

	Group A	Group B	Difference
K96 less K95	4.63	4.04	-0.59
K97 less K96	6.23	6.77	0.54
Difference	-1.59	-2.72	-1.13

Table 13e. Second Grade Language

	Group A	Group B	Difference
K96 less K95	2.00	3.35	1.35
K97 less K96	5.25	5.54	0.29
Difference	-3.26	-2.20	1.06

Table 13f. Third Grade Language

	Group A	Group B	Difference
K96 less K95	6.03	5.83	-0.20
K97 less K96	5.78	6.55	0.76
Difference	0.24	-0.71	-0.96

The estimates in Table 13 ignore random school effects that are included in Model 1 to produce efficient estimates and test the null hypothesis that the effect is zero. The results of this model are reported in Table 14, and the full model estimates are included in Table 15. The estimated effects are uniformly small in absolute value ranging from - 1.1 to 1.7; these estimates are also small relative to the standard deviation in SAT-9 scores (about 40 scale score points). In addition, the effects across grades are offset--the negative estimate for math in grade 2 is followed by a positive estimate at grade 3, and the positive estimates for reading and language at grade 2 are followed negative estimates at grade 3. Overall, the estimates from Table 13 and Table 14 are very similar and suggest little CSR effect. We also explored school fixed effects models and the results were nearly identical to those in Table 15.

Table 14. Estimates of 95% Confidence Intervals of CSR Effects from Model 1

	Grade 2	Grade 3
Math	- 0.9 (- 2.3, 0.5)	0.7 (- 0.7, 2.2)
Reading	1.7 (0.3, 3.1)	- 1.1 (- 2.6, 0.3)
Language	0.9 (- 0.4, 2.2)	- 0.8 (- 2.3, 0.6)

We also conducted some sensitivity analyses to see whether these results were consistent for across student and teacher characteristics. We found similar results when we restricted the analysis to schools with more than 65 percent minority students, suggesting that the CSR effect was not larger for minority students. (This analysis included about one-half of the schools.) To address the possible bias introduced by the difference between Groups A and B in the change in the percentage of fully-credentialed first grade teachers, we restricted the analysis to schools with no change in the percentage of fully-credentialed teachers during this time period. The results of this analysis were similar, as well. Finally, we ran the analyses with both restrictions, and although the sample of schools was small, we saw no substantial differences in the results.

Caveats

These school-level analyses were less susceptible to confounding from external sources than the statewide analyses presented by Stecher, Bugliari and McCaffrey (2002). For example, we were able to control for student mobility by only including students who attended the same school from kindergarten through second or third grade. Yet, there are still limitations in these analyses. The greatest limitation comes from the lack of variation that existed in exposure to CSR. Our comparisons were limited to a one-year difference in exposure to reduced size classes among students whose total exposure was two or three years. The one-year difference occurred in first grade, and all students subsequently participated in reduced size classes in second and third grade--the points at which their achievement was measured. The Tennessee STAR experiment compared students who attended reduced size classes for four consecutive years with students who attended normal size classes for four consecutive years. They found that at least two years of exposure were needed to produce lasting differences. Those conditions for comparison did not exist in California.

There have also been modest changes in the demographic characteristics of students during this period that might have affected achievement trends. Table 15 shows selected demographic characteristics of California public school students during this time period. There has been a modest increase in the percentage of Hispanic students during this time period, but our differencing approach should have minimized the impact of this gradual change. Yet, our models were simple and did not adjust for demographic or other student background variables. Given the small size of effects and the general similarity of the comparison groups we used a simple analysis rather than complex models. However, small differences among the groups might have affected our results, and more complex models might have removed some of these differences.

Table 15. Demographic Characteristics of California Students, 1995-2000 (percentages)

			Race/Ethnicity				
School year	Total enrollment	Limited English Proficient (LEP)	Asian	Hispanic or Latino	African American	White (not Hispanic)	Other
1995–96	5,467,224	23.6	8.2	38.7	8.8	40.4	3.9
1996–97	5,612,965	24.2	8.2	39.7	8.7	39.5	3.9
1997–98	5,727,303	24.6	8.1	40.5	8.8	38.8	3.9
1998–99	5,844,111	24.6	8.1	41.3	8.7	37.8	4.2
1999–00	5,951,612	24.7	8.0	42.2	8.6	36.9	4.3
2000–01	6,050,895	24.9	8.0	43.2	8.4	35.9	4.5

Note. Starting in 1998–99, all figures include California Youth Authority (CYA) schools. “Other” includes American Indian or Alaskan Native, Pacific Islander, Filipino, and, beginning in 1998, Multiple or No Response.
Note. Source: California Department of Education, Education Demographics Unit.

There have been significant policy and program changes during this period that also affected student achievement. These changes include new state standards and curricula, revised grade-level promotion policies, a new test-based school-level accountability system with large rewards for increases in scores, and the elimination of traditional bilingual education programs. Because they occurred simultaneously, we used various forms of differencing to disentangle their separate effects and to isolate the unique contribution of CSR to score improvement during this period. However, the differencing requires many assumptions about the equivalence of groups and cohorts in the absence of CSR and the large of number of changes in other programs calls into question the validity of those assumptions.

In addition, there is some reason to doubt the validity of the score gains we used as the basis for these analyses. The California school accountability system has created a high-stakes atmosphere that may lead to changes in test scores that are independent of actual changes in achievement. The gains in SAT-9 scores observed in California are well within the range that might be associated with such score inflation. Again, differencing removes general trends due to score inflation but cannot account for differential inflation.

Another limitation is the restricted sample of the schools and students used in our study. Many schools did not have complete student demographic data, and they were eliminated from our sample. Others had too few valid test scores and were eliminated for this reason. Still other schools were dropped because of indeterminacy in CSR exposure. In addition our analyses focus on students who did not change schools during the K-3 years. The effects of CSR might be different for the schools and students we excluded from our analysis, but we do not have the data to determine the effects of these restrictions on our results. We do not have any good hypotheses about the likely direction of differences between the CSR effects in our sample and those for the entire state.

Finally, the available data do not allow us to judge the impact of the entire CSR program and its effects on students for the last five years. Rather, we look for evidence that reduced size classes can make a difference by testing whether additional exposure yields greater achievement. A positive result would be encouraging evidence that small classes are beneficial and that offering them to students in California could have positive effects. Our null finding, however, cannot be interpreted as evidence that the CSR program is not effective. Our results are consistent with at least two possible inferences: a.) reduced size classes have no effect, or b.) two, three or four years of exposure to reduced size classes do have a positive effect compared to no exposure but the difference between two years of exposure and three years of exposure is negligible. One should not make the most pessimistic interpretation of our results (e.g., that reduced size classes have no effect and therefore the entire CSR program is a failure). Rather we should make the most cautious interpretation that, in the context of a K-3 program of reduced size classes, a one-year incremental difference in exposure has no effect. K-3 CSR might have large positive effects on students but differential gains among students with small differences in exposure cannot be used as evidence of those larger effects.

Conclusions

The goal of this investigation was to determine the extent to which changes in achievement correspond to the implementation of the CSR program. The analyses show that scores at the elementary level have been rising at the same time that increasing percentages of students have been taught in reduced size classes. However, many other educational reforms were enacted during this period that might have contributed to the achievement gains, and it is impossible for us to determine how much the various factors may have influenced trends in overall student achievement. Our analyses that used differences in group means to control for the other factors showed that a one-year difference in exposure occurring in first grade is not associated with greater gains in achievement. Due to the rapidity of CSR implementation, we could not test the cumulative effects of two or three years of exposure. Thus, while the analyses presented in this chapter find no association between one year's difference in exposure and differences in achievement, we cannot draw any conclusions about the effects of CSR in larger doses.

Notes

This research was conducted under the auspices of the CSR Research Consortium, including RAND, the American Institutes for Research, Policy Analysis for California Education (PACE), WestEd, and EdSource. Findings were reported previously as a Technical Appendix to the Consortium’s final report *What Have We Learned About Class Size Reduction in California* (Bohrnstedt and

Stecher, 2002). The research was funded by the California Department of Education, the Walter and Elise Haas Fund, the William and Flora Hewlett Foundation, the Walter S. Johnson Foundation, the San Francisco Foundation, and the Stuart Foundation. The opinions expressed here are the authors'.

Endnotes

1. The CSR Research Consortium includes the American Institutes for Research (AIR), RAND Corporation, Policy Analysis for California Education (PACE), WestEd, and EdSource.
2. The Consortium's analyses were limited by the fact that there were no achievement data for kindergarten students or first grade students in any year, and there were no achievement data for any students prior to 1998.
3. Minority students are any students not classified as Caucasian. The largest groups of minority students are, in order of group size, Hispanics, Asian/Pacific Islanders, and African Americans.
4. Students are referred to as low-income or as being from low-income families in this report if state records classify them as receiving public assistance in the form of Aid to Families with Dependent Children (AFDC) or its successor in California, CalWORKS.
5. Students for whom English is a second language and who are not fully proficient in English are often referred to as limited English proficient (LEP), English language learners (ELL), and English learners (EL). We use EL throughout this report to reflect the usage in the California law that implemented proposition 227, a proposition passed by California's voters in 1998 that banned the implementation of bilingual education except under special parental waiver conditions.
6. Although the trend in scores is not related to level of CSR exposure, the size of gains might be sensitive to class size reduction overall. For example, the achievement gains for primary grades were larger than for upper elementary, where classes remained large. Small classes might allow teachers to better implement reforms or to respond more quickly to the incentives of the accountability system. However, we do not have adequate data to test for effects between grades; we can only compare differential amounts of CSR among students in the same grades.

References

Bohrnstedt, G. W. & Stecher, B. M. (Eds.) (1999). *Class size reduction in California: Early evaluation findings, 1996–1998.* Palo Alto, CA: CSR Research Consortium.

Bohrnstedt, G. W. & Stecher, B. M. (Eds.) (2002). *What have we learned about class size reduction in California*. Sacramento: California Department of Education.

Finn, J. (1998). *Class size and students at risk: What is known? What is next?* (No. AR 98-7104). Washington DC: Office of Educational Research and Improvement. U.S. Department of Education.

Finn J. and Achilles, C. (1999). Tennessee' class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21, 97-109.

Krueger, A. B., and Whitmore, D. M. (1999). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*

Mosteller, F. (1995, Summer/Fall). The Tennessee study of class size in the early school grades. *The*

Future of Children, 5, 113-127.

Nye, B., Hedges, L. V., and Konstantopoulos, S, (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee Class Size Experiment. *Educational Evaluation and Policy Analysis*, 212, 127-142.

Stecher, B. M. & Bohrnstedt, G. W. (Eds.) (2000). *Class size reduction in California: The 1998-99 evaluation findings*. Sacramento, CA: California Department of Education.

Stecher, B. M. & Bohrnstedt, G. W. (Eds.) (2002). *Class size reduction in California: Findings from 1999-00 and 2000-01.* Sacramento, CA: California Department of Education.

Stecher, B. M., Bugliari, D., and McCaffrey, D. F. (2002, February). *Achievement.* In B. M. Stecher & G. W. Bohrnstedt (Eds.) *Class size reduction in California: Findings from 1999-00 and 2000-01*. Sacramento, CA: California Department of Education.

Stecher, B. M., McCaffrey, D. M., and Burroughs, D. (1999). *Achievement*. In G. W. Bohrnstedt & B. M. Stecher (Eds.). *Class size reduction in California: Early evaluation findings, 1996–1998.* Palo Alto, CA: CSR Research Consortium.

Stecher, B. M., McCaffrey, D. F., Burroughs, D. Wiley, E., and Bohrnstedt, G. W. (2000). *Achievement.* In B. M. Stecher & G. W. Bohrnstedt (Eds.) *Class size reduction in California: The 1998-99 evaluation findings*. Sacramento, CA: California Department of Education.

About the Authors

Brian Stecher
Senior Social Scientist
RAND
1700 Main Street
PO Box 2138
Santa Monica, CA 90407-2138

Brian Stecher's research emphasis is applied educational measurement, including the implementation, quality, and impact of state assessment and accountability systems; the cost, quality, and feasibility of performance-based assessments, and the development and validation of licensing and certification examinations.

Daniel McCaffrey
Senior Statistician
RAND
201 North Craig Street, Suite 202
Pittsburgh, PA 15213-1516
Email: daniel_mccaffrey@rand.org

Dan McCaffrey's research includes studies related to education and health policy. His current projects involve value-added modeling for estimating teacher effects and propensity score methods for comparing nonequivalent groups in quasi-experiments. He is also interested in nonparametric methods for estimating the standard errors for models fit to clustered data.

Delia Bugliari
Senior Programmer/Analyst
RAND
1700 Main St
PO Box 2138

Santa Monica CA 90407-2138

Delia Bugliari specializes in analysis of education data on student achievement, school demographics, and teacher surveys. Her research interest includes missing data imputation in education and economic data.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

EPAA is published by the Education Policy Studies Laboratory, Arizona State University