

EXECUTIVE SUMMARY

Brian M. Stecher and George W. Bohrnstedt

Policymakers who supported the enactment of the Class Size Reduction (CSR) initiative had high expectations that the \$1.5 billion invested annually would yield noticeable improvements in student achievement and in education more generally. By 1997–98, CSR has brought reduced size classes to over 1.6 million students and showed small benefits in terms of student achievement. However, its rapid implementation may have increased underlying inequities in the state’s education system that may threaten the state’s ability to reap the full rewards of this huge investment. Because of the speed with which the California CSR program was implemented, it is still very much a “work in progress,” and a few more years will be needed before it is possible to judge its overall impact. This report, which is the first in a four-year effort to evaluate the CSR program, provides a description of CSR in its first and second years (1996–97 and 1997–98) that illustrates the following points:

- Almost all first- and second-grade students and almost two-thirds of all kindergarten and third-grade students are in reduced size classes.
- Districts serving low-income students,¹ minority students, or English language learners (ELL)² have been slower to implement CSR and have received disproportionately less CSR revenue as a result.
- To implement CSR this rapidly, many schools have taken space from other programs to use for classrooms. Facility shortages in schools serving low-income, minority, or ELL students mean that these schools have reduced class size more slowly and have faced greater demands on existing space.
- The costs of CSR have exceeded the revenues in many districts, and many have taken resources from other programs to make up the deficit. Districts serving low-income, minority, or ELL students have higher costs and are more likely to have taken resources from other activities to support the implementation of CSR.

¹ Students are referred to as low-income 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.

² Students for whom English is a second language and who are not fully English proficient are often referred to as either limited English proficient (LEP) or English language learners (ELL). The latter term is used throughout this report.

- The K–3 teacher workforce in California has increased by 38 percent, but the overall preparation level of K–3 teachers has declined. Schools serving low-income, minority, or ELL students have suffered a greater decline in the qualifications of teachers than other schools have.
- Teaching practices are very similar in reduced and non-reduced size classes, but teachers in reduced size classes spend somewhat less time on discipline and somewhat more time working one-on-one with problem readers and attending to individual students’ personal concerns.
- Parents of students in reduced size classes have more contact with teachers and are more satisfied with their children’s education.
- There is a small positive gain in achievement associated with being in a reduced size class, and this gain is realized by all groups of students.

The report also identifies issues for educators and policymakers to address as implementation continues.

History of the California CSR Program

The California CSR program was initiated in 1996, as California was making the transition from recession to booming economy and tax revenues were increasing.³ A decade-long decline in student achievement reached the point of alarm when the mean scores of California fourth-grade students on the 1994 National Assessment of Educational Progress (NAEP) reading assessment tied for last place out of 39 participating states. Moreover, concern was growing over the state’s persistent achievement gap—African American and Hispanic students and students living in inner cities were achieving at considerably lower levels than the rest of California’s students.

In a bold step, Governor Wilson and the legislature agreed on an infusion of funds to reduce class size from an average of about 30 students to 20 or fewer students in kindergarten through third grade. To our knowledge, this CSR initiative is the largest state educational reform in history, costing over \$1 billion per year and affecting over 1.6 million students. CSR is a voluntary program, and districts have the option of reducing class size if they so desire. However, the incentive to reduce class size is great, particularly after years of tightening school budgets. In 1996–97, districts were reimbursed a flat rate of \$650 for each child in a reduced size class (an amount that did not fully cover the cost of the reform in many districts). Some \$400 million was made available for new facilities as well. The next year the state raised the rate to \$800 per child and allowed schools to use any funds not applied directly to K–3 CSR for further facility improvements.

³ Proposition 98 requires that a fixed percentage of revenues be spent on education. As a result, when there is a surplus as there was in 1996, over one-half of the increased revenue had to be earmarked for education.

Expectations

Policymakers have high expectations for CSR, based in large part on the results of a class size reduction experiment conducted in Tennessee from 1985 to 1990. The Tennessee Student/Teacher Achievement Ratio (STAR) project produced relatively large achievement effects for all students based on scores on the Stanford Achievement Test (SAT-9).⁴ Moreover, low-income and minority student gains were almost twice as large as gains for the rest of the sample. The STAR experiment provided little information about factors that enhanced achievement in smaller classes, but it did create high expectations regarding the likely achievement benefits of CSR.

However, there are substantial differences between STAR and the California CSR program that may warrant some tempering of expectations. The Tennessee program was a small, highly controlled experiment involving fewer than 10,000 students. The California program has been implemented on a huge scale, with few guidelines, and involves more than 1.6 million students. In Tennessee, reduced size classes had 15 students on average; California's have 20. Most students in Tennessee were white, with only about 25 percent being African American. California's students are more diverse; fewer than one-half are white, and a substantial proportion are ELL students. The schools in Tennessee had classrooms available to accommodate the larger number of reduced size classes, could hire certified teachers to staff the new classes, had sufficient textbooks and materials, and shared a common curriculum with an aligned state test. By contrast, many California schools lacked both the space and teachers to create new classes. CSR exacerbated an existing teacher shortage, meaning that some districts could only implement CSR if they hired teachers on emergency permits. Many of the new classes lacked adequate books and materials, the state curriculum was under revision, and there was no state test. Given these differences, the California CSR program must be judged on its own terms, not as a replication of the Tennessee experiment.

It also is important to note that CSR is not being implemented in a vacuum. Numerous other major educational reforms are occurring in California, including changing curriculum standards, state assessments, bilingual education guidelines, teacher certification procedures, and student promotion policies, to name just a few. These programs interact in complex ways, making it difficult to attribute changes to any single effort.

Results

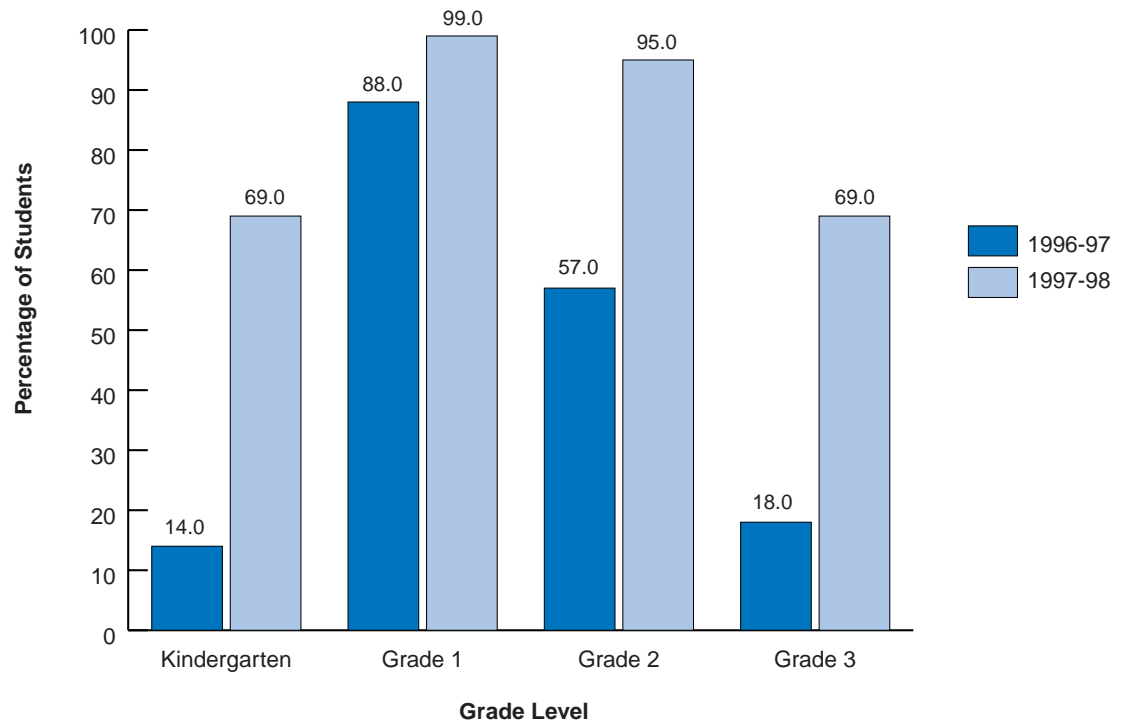
Rapid Statewide Implementation

Districts responded remarkably quickly to implement CSR. Figure 1 shows that almost 90 percent of first-grade students were in reduced size classes in the first year of the program (1996–97). By the second year almost all of the first- and second-grade students in California were in classes with 20 or fewer students. The speed of implementation is itself evidence of people's belief in the benefits of reduced class size,

⁴ The average effect size for STAR was 0.25 (one-quarter of a standard deviation), which is noteworthy for an education intervention. The effect size is a standardized way of expressing the difference between treatment and control groups that permits comparisons between programs that use different outcome measures.

the attractiveness of the funding, and the fact that the reform does not involve specialized preparation or materials. Districts had to make choices about the speed and grade-level order of implementation. Some districts consulted with parents, but most parents were not involved in implementation decisions. In fact, almost one-quarter of the parents with children in the third grade had not heard of the CSR program by the second year.

**Figure 1—
Percentage of Students in Reduced Size Classrooms, by Grade Level and Year**

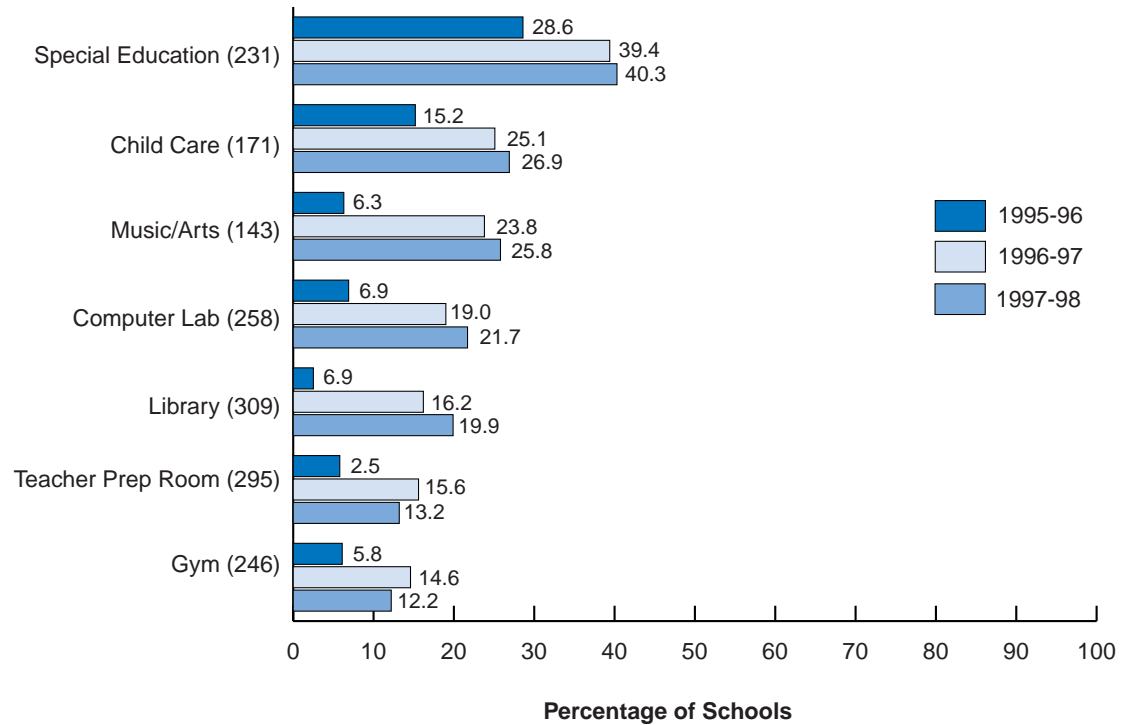


Source: California Department of Education. Retrieved February 24, 1999 from the World Wide Web: <http://www.cde.ca.gov/ftpbranch/sfpdiv/classize/facts.htm>.

Shortage of Facilities

Although implementation was swift, many schools could not implement CSR fully because of a shortage of facilities. Increasing enrollments had already caused some schools to convert facilities from other educational uses to classrooms in 1995–96, but this conversion was exacerbated with the advent of CSR in 1996–97. In many cases, schools reassigned space from other programs to create new classrooms. Figure 2 shows that by 1997–98 more than one-quarter of schools had taken space away from special education, child care, and music and arts programs, and more than one-fifth had converted library space and computer labs for classroom purposes. Perhaps not surprisingly, schools serving higher proportions of low-income, minority, or ELL students, which were already experiencing overcrowding, implemented CSR more slowly.

**Figure 2—
Percentage of Schools Reporting Educational Space Pre-empted for
Classrooms, by Type of Space**

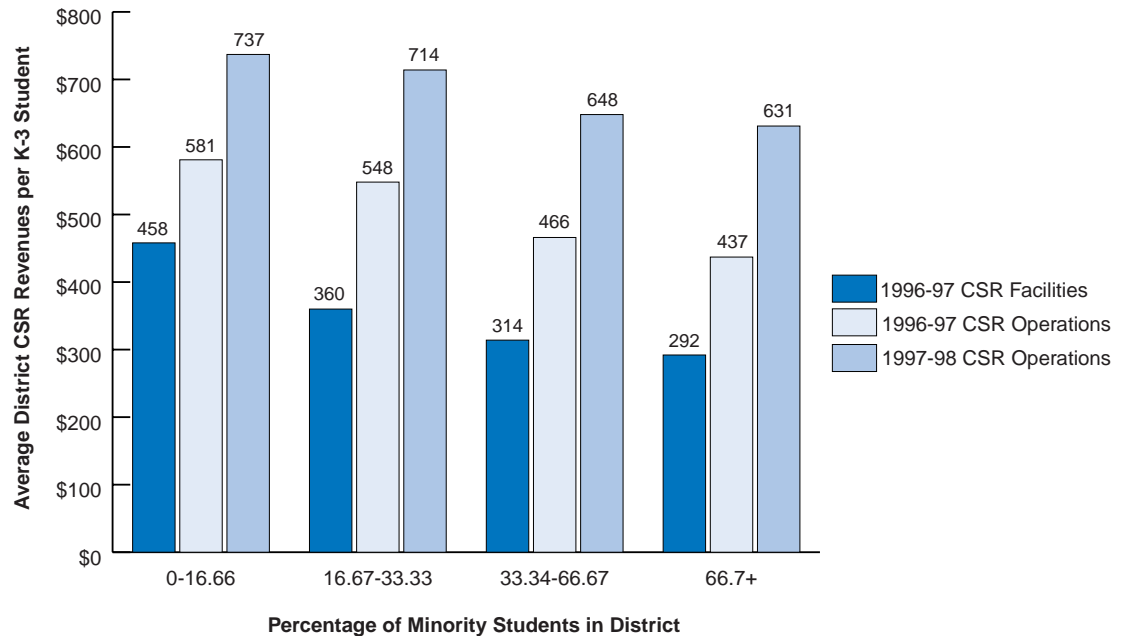


Note: Numbers in parentheses indicate the number of schools among the 336 responding to the survey which reported having such a facility in 1995.
Source: CSR Consortium 1998 Survey of Principals.

Unequal Revenues

Since funding was linked to implementation, districts that were slower to implement the program received less CSR revenue. As Figure 3 shows, on a per-pupil basis, proportionately more of the CSR resources for teachers and facilities went to districts serving few minority students than to districts serving many minority students. Similarly, districts serving larger proportions of low-income students or ELL students and urban districts received proportionately less of the CSR money during the first two years due to their slower rates of implementation.

**Figure 3—
District CSR Funding in First Two Years of Implementation, by Percentage of
Minority Students**

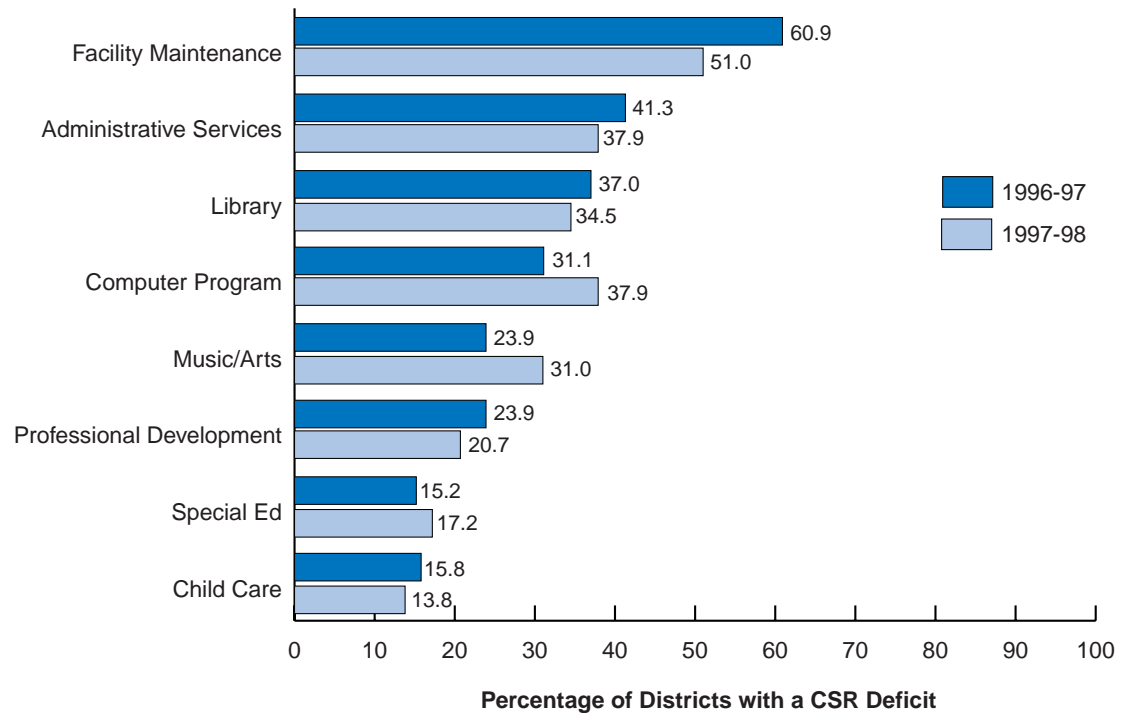


Source: California Department of Education. Retrieved February 24, 1999 from the World Wide Web: <http://www.cde.ca.gov/ftpbranch/sfpdiv/classize/facts.htm>.

Unequal Costs

Districts also encountered different costs to implement CSR. All districts earned the same amount of revenue per student in a reduced size class. However, districts with smaller classes and more available space prior to CSR needed fewer new teachers and less additional space, so their costs were lower. In addition, the cost of providing educational services differs from one district to the next, so the same increase in facilities and staff may cost more in some districts than in others. In fact, in 1996–97, a majority of superintendents reported that CSR funds were inadequate to cover the cost of implementation. Even after per-pupil funding was raised from \$650 to \$800 in 1997–98, more than 40 percent of superintendents still reported fiscal shortfalls. Districts that had deficits diverted resources from other functions, as shown in Figure 4. Districts serving low-income, minority, or ELL students were more likely to have costs that exceeded revenues, shortfalls that may persist even after districts implement CSR in all four grades.

**Figure 4—
Programs Reduced by Districts to Compensate for Insufficient CSR Funding**



Note: Superintendents who reported a CSR deficit were asked whether specified programs were somewhat reduced, significantly reduced, eliminated, or not affected. Figure 4 shows the proportion of districts that responded to the first three categories.

Source: CSR Consortium 1998 Survey of Superintendents.

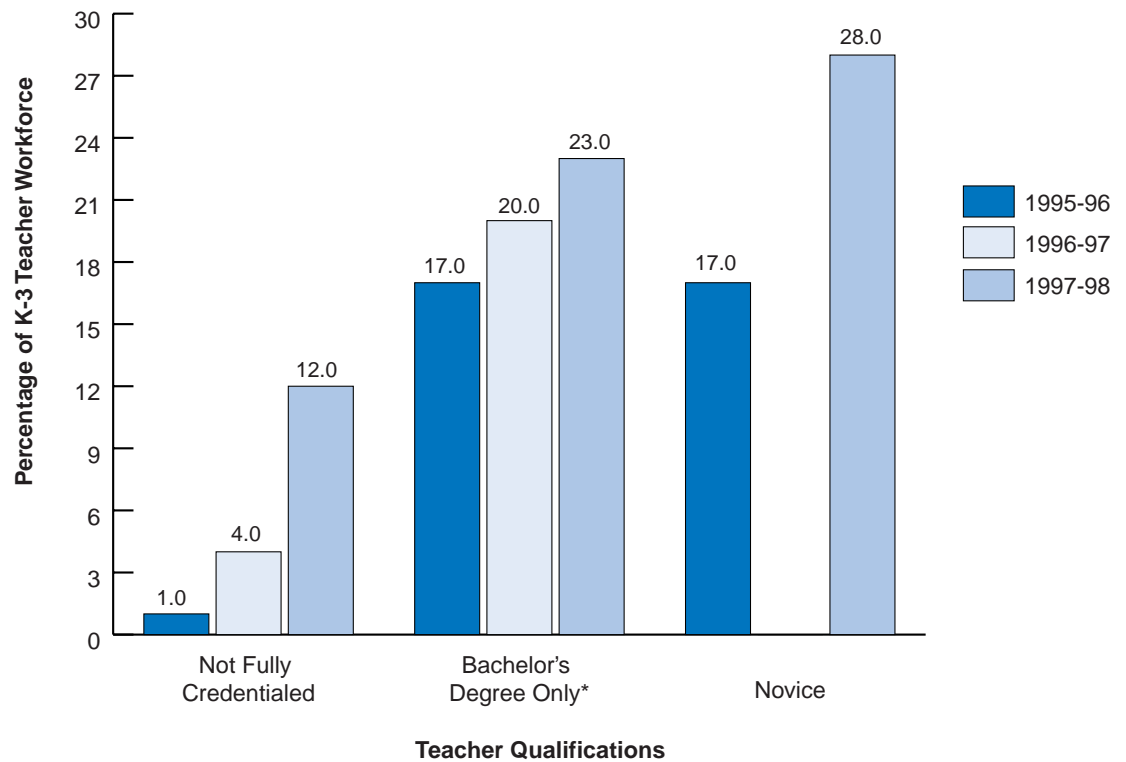
Mixed Interaction with Other Reforms

About two-thirds of districts and 85 percent of the elementary schools were implementing other curriculum and instructional reforms when CSR was enacted. Superintendents and principals reported that CSR interrupted these reform efforts by diverting administrators’ attention from the reforms, using facilities needed for the reforms, and creating new professional development needs that had to be met first. However, administrators also reported that CSR boosted teachers’ enthusiasm for other reforms and brought in teachers with new ideas that enhanced reform efforts.

Lower Level of Teacher Preparation

The rapid implementation of CSR resulted in a decline in the average education, experience, and credentials of K–3 teachers. In just two years, there was a 38 percent increase in the number of K–3 teachers in California and there were dramatic changes in the teacher labor market. Both were primarily due to CSR. The teacher workforce grew by 23,500 teachers during this period. Unfortunately, as Figure 5 shows, these new teachers reduced the qualifications of the teaching force as a whole.

**Figure 5—
Change in K–3 Teacher Qualifications**



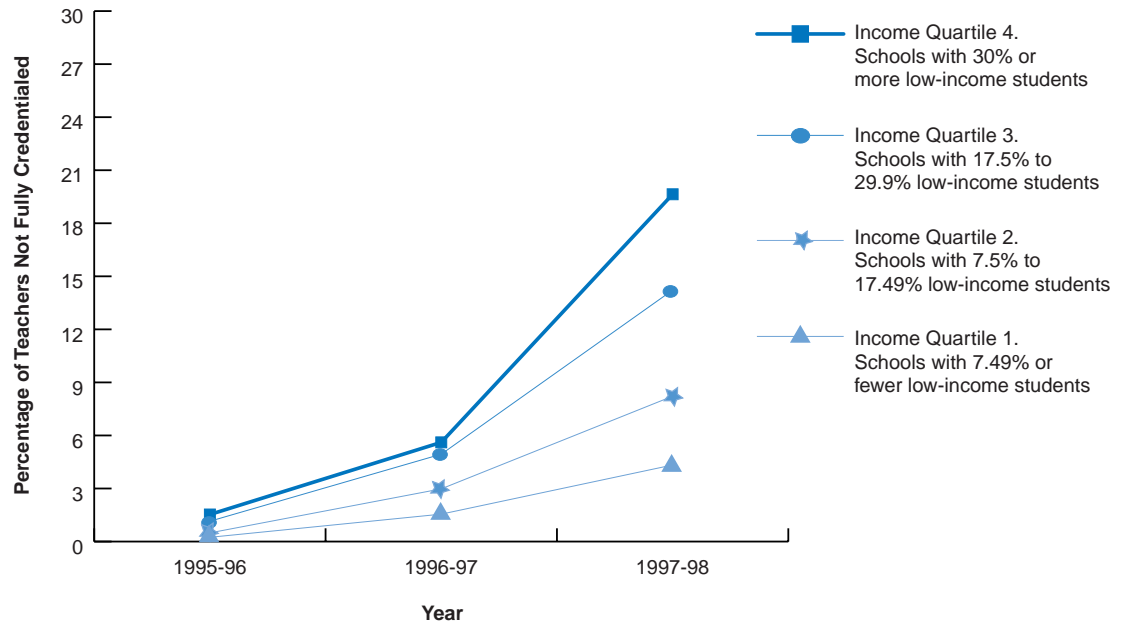
Note: 1996-97 percentage of novices omitted due to an unusually high proportion of missing experience data.

* The Bachelor's Degree Only category includes a small percentage of teachers, less than 0.6% of all K–3 teachers, who reported having less than a bachelor's degree.

Source: CSR Consortium analysis of California Department of Education, CBEDS-PAIF data.

The already weaker qualifications of teachers serving low-income and minority students became worse. Figure 6 shows that, prior to CSR, the top quarter of schools in terms of the percentage of low-income students had a slightly higher percentage of teachers lacking full credentials than did the remainder of schools. However, since the implementation of CSR this gap has increased almost tenfold, as the high-AFDC schools hired larger numbers of teachers without full credentials. The same is true for schools serving higher percentages of minority or ELL students. This decline in teacher qualifications was partly due to teachers moving to take jobs in other schools or districts, but it was primarily due to these schools being unable to compete for better-qualified new teachers. A similar change occurred in the distribution of specialist teachers with Bilingual Cross-Cultural Language and Academic Development (BCLAD) credentials required to teach in bilingual programs.

**Figure 6—
K–3 Teacher Credentialing in Schools with Different Proportions of Low-
Income Students**



Note: Differences between the top and bottom quartiles in the same years are statistically significant at the 0.01 level. Differences between years in the same quartile are also statistically significant at the 0.01 level. Source: CSR Consortium analysis of California Department of Education, CBEDS-PAIF data.

Little Change in Classroom Practices

There were few differences in instructional practices between non-reduced and reduced size third-grade classes.⁵ For example, teachers in reduced size classes did not spend significantly more time during regular lessons working individually with students, which is one way that smaller classes might promote achievement. Similarly, there were no differences in curriculum content and very few differences in the frequency of teachers' use of specific instructional strategies or student activities in either language arts or mathematics.

However, there were a few differences in classroom practices in third grade that are worth noting. When specifically asked about students who needed help with reading skills, teachers in reduced size classes reported that they spent more time giving sustained attention (five or more continuous minutes) to these students than did teachers in non-reduced size classes. Teachers in reduced size classes also reported spending more time addressing individual students' personal concerns and less time disciplining students than did teachers in non-reduced size classes.

⁵ The investigation of classroom practices focused on teachers in third grade.

Increased Parent Contact and Satisfaction

Parental contact with teachers and parental satisfaction with schools were higher among parents of third-grade students in reduced size classes. Among parents who responded to our survey,⁶ a somewhat larger percentage of parents with children in reduced size classes initiated contacts with teachers (74%, compared with 69% in non-reduced size classes) and were contacted by teachers (85%, compared with 81% in non-reduced size classes) during the 1997–98 school year. Parents of students in reduced size classes rated the overall quality of their children’s education as very good; parents of students in non-reduced size classes rated it as good to very good. Furthermore, the former group rated all aspects of educational quality higher than did the latter group.

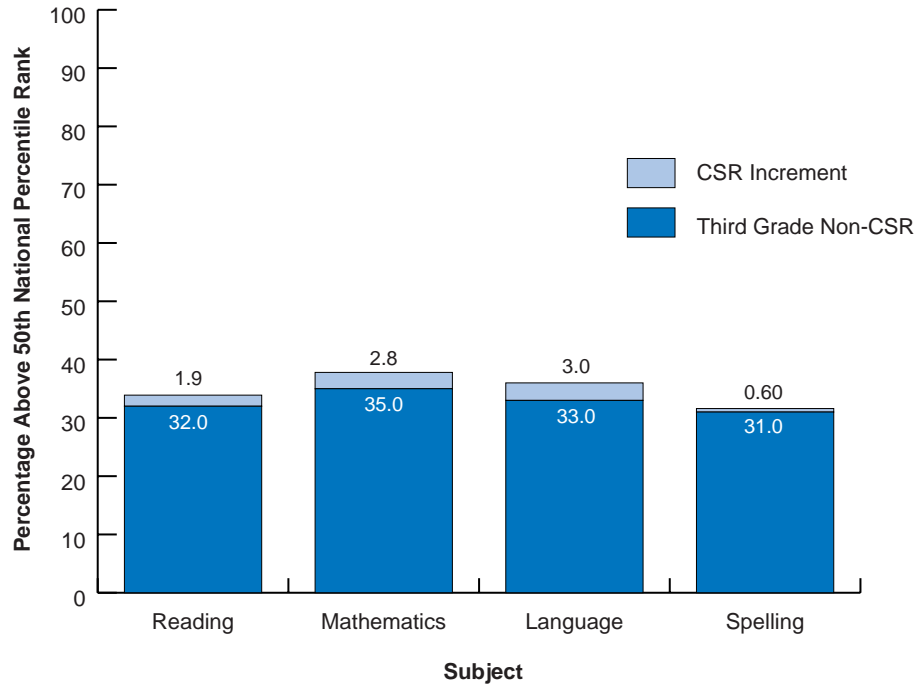
Small Improvement in Third-Grade Student Achievement

Because of rapid implementation of the reform, it only was possible to compare achievement in reduced and non-reduced size classes in third grade. Figure 7 shows that the percentage of students whose SAT-9 scores were above the 50th national percentile rank increased by 2–4 percentage points in reduced size classes in third grade. This small but statistically significant difference is equivalent to an effect size of 0.05–0.1 (one-twentieth to one-tenth of a standard deviation) in reading, mathematics, and language.⁷ There was no significant difference in spelling. Furthermore, differences in achievement were the same regardless of students’ race/ethnicity, family income level, or language status (i.e., minority, low-income, or ELL students gained as much as a result of CSR as did other students).

⁶ Completed surveys were received from 1,075 parents of third-grade students (slightly greater than half of those who were mailed surveys).

⁷ Another way to illustrate the size of this effect is to adapt an example from Mosteller’s review of the Tennessee STAR results (1995). Consider a pupil who, without special treatment such as attending small classes, would achieve about the average score, say, at the midpoint, or 50th percentile, of all students. What would a gain of 0.08 of a standard deviation (the average gain in mathematics) do for such a pupil? That pupil would move from the 50th percentile of all pupils up to the 53rd percentile, thus surpassing an additional 3 percent of the population beyond the 50 percent exceeded originally.

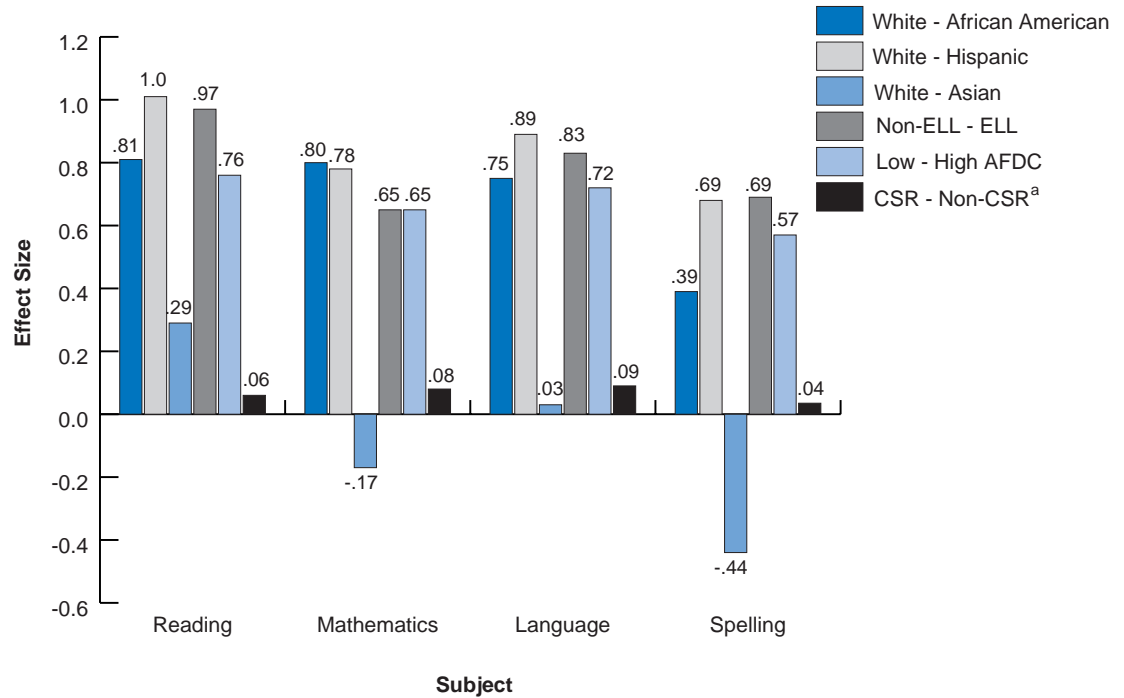
**Figure 7—
Percentage of Third-Grade Students Scoring Above the 50th National Percentile Rank (Median) on SAT-9**



Source: Consortium analysis of Standardized Testing and Reporting (STAR) public release data for 1997–98.

Figure 8 attempts to put these gains into perspective. The size of the achievement differences between third-grade students in non-reduced and reduced size classes is quite small compared with the size of the differences associated with student background factors. For example, as seen for reading achievement, the effect size when comparing whites and African Americans is 0.8, whereas the effect size associated with CSR on reading achievement, net background factors, is about 0.05.

**Figure 8—
Effect Sizes Associated with Difference in Third-Grade Students' Background
Characteristics**



^a Average from fourth- and fifth-grade reference groups.

Note: All differences shown are simple differences; we have not attempted to separate the effects of variables that are correlated, e.g., race/ethnicity and income.

Source: Consortium analysis of Standardized Testing and Reporting (STAR) public release data for 1997–98.

Important Caveats

Some will find the achievement results to be good news, particularly at this early stage in the implementation of CSR and especially given the difficulties some schools have encountered implementing CSR. Others may be disappointed that California's large investment of funds has produced relatively small effects and that the effects for low-income, minority, or ELL students were not larger than those for other students. We suggest caution in making too strong a judgment about the effect of CSR on achievement at this early point in time. No one has ever implemented a CSR reform on this scale before, and it is difficult to establish criteria for success at this juncture. Further experience is necessary to measure the cumulative effects of reduced size classes over time and to clarify what would be a reasonable standard for success.

We also urge caution about drawing cause-and-effect conclusions from these first-year findings. Although it is tempting to relate the findings concerning student achievement to those describing degree of implementation, teacher characteristics, or classroom practices, it would be incorrect to assume any cause-and-effect relationships on the basis of these results. Additional multivariate analyses will clarify some of these interrelationships, but limitations in the data will continue to frustrate causal interpretations.

Implications

Findings from the first year suggest a number of mid-course adjustments that might increase the benefits of CSR. The findings also have implications for future educational policymaking in California.

Reducing the Shortage of Classrooms

- The single greatest factor in the uneven implementation of CSR has been the shortage of facilities to accommodate new classes. Students most in need academically tend to attend the most overcrowded schools (i.e., schools that are least able to find the space to implement CSR). Despite the facilities funds that were available during the first two years of CSR and the recent state bond issue, the shortfall remains. Unless something is done to increase classroom space in overcrowded schools, these students will continue to be served in larger classes. Alternatively, these schools will take more space from other programs, which will have unknown consequences for students.

Addressing the Decline in Teacher Preparation

- The decline in the preparation level of K–3 teachers is substantial, and it is cause for concern. Fortunately, efforts already are under way to increase the level of professional development and support available to new and continuing teachers, and major changes have recently been made in the teacher licensing process. Nevertheless, CSR has changed the teacher workforce in substantial ways, and it is not clear that current professional development efforts are adequate to solve the problem.
- It may be necessary to invest additional resources to address the decline in teacher preparation that accompanied CSR. New funds might be used for a variety of efforts, such as bolstering school-based training programs, supporting professional development courses at alternative times, assisting universities to prepare more teachers to teach in smaller classes, providing incentives to attract qualified teachers to the schools most in need, or addressing other identified needs. The choice should be informed by further study of the needs of teachers and schools and the capacity of the existing system to meet them.

- Efforts should be made to learn more about the needs of the uncertified teachers hired in response to CSR and to investigate the effectiveness of current training and support efforts. Information about the problems encountered by new teachers and their response to existing training and support programs would be useful to help improve these efforts and design new ones.

Improving Teaching in Smaller Classes

- It would be valuable to learn more about instructional approaches that maximize the benefits of smaller classes. This question is largely unexplored, and the designers of professional development programs are largely without guidance. Further study of instructional practices might provide insights to improve professional development and support programs.

Increasing Program Flexibility

- Although the CSR program treats every school and district in the same manner, this “one size” intervention does not fit all districts equally well. Each district’s response has been defined by its local context, in which each confronts different local constraints and problems. It may be appropriate to acknowledge these different conditions by changing the program guidelines so as to, for example, give schools greater flexibility in how they use CSR funds to deal with shortages of facilities and teachers. For example, the Legislative Analyst’s Office (1997) suggested that the regulations be modified to permit individual classes to have as many as 22 students so long as the *district average* class size did not exceed 20.
- Similarly, the cost of implementing CSR has been different for each district. State resources might be stretched further or used more equitably if the allocation formula were changed to align CSR incentives more closely with actual costs.

Examining Cost-Effectiveness

- Although small achievement gains were observed in the first year, the larger question that must be addressed is whether the overall benefits are worth the cost. This entails measuring student achievement and other outcomes, assessing all the costs of implementing the reform (not just expenditures at the state or district level), and comparing these with other alternative interventions, such as whole school reform models that are beginning to demonstrate their effectiveness.
- It is too early to estimate the full cost of the reform. While the annual cost to the state treasury to fund CSR fully in grades K–3 is approximately \$1.5 billion, the total cost of the program is much greater. The total cost includes district and school resources taken from other uses to support CSR, as well as the

effects of CSR on the teacher workforce and on other programs. Moreover, these opportunity costs are not equally distributed in that they are borne to a greater extent by schools serving low-income, minority, or ELL students. It will take time to make a precise determination of these costs and to determine which communities have to bear them, but initial differences are quite large and deserve more attention.

Planning Future CSR Policy

- It would be helpful if the state developed a plan for future CSR efforts based on what has been learned so far in grades K–3. Such a plan would identify the key questions that should be answered and would delineate decision rules that follow from those questions. Among the issues to be incorporated into such a plan would be the pace of implementation, teacher supply and demand, the targeting of resources to serve students most in need (e.g., by earmarking more funds or allowing schools serving these students to implement earlier), and coordination with other reform efforts. Such planning would maximize the value of future CSR efforts in California and would be extraordinarily valuable for other states considering class size reduction programs.
- California is in the midst of trying to bring more coherence to its educational policymaking. Efforts to align curriculum and assessment to standards are an important part of this process. Recently, the Legislative Analyst’s Office called for the development of a master plan that would review state educational policies and suggest roles for different levels of government. In a similar vein, policymakers should carefully consider CSR’s role in any new comprehensive strategy for education and should think about how new policies will interact with CSR. An important issue is whether policies, singly or in combination, benefit all students and schools equally. Asking who gains and who loses is politically unpopular but essential for responsible policymaking. Many districts have put so many resources (financial, human, and physical) into CSR that they may be ill prepared to implement other educational reforms that have been passed recently. Urban districts, in particular, have been put under considerable stress by CSR. Given the multiplicity of program initiatives, it is important to monitor districts’ responses to new programs and watch for potential negative interactions.
- It is too early to tell how large the achievement benefits of CSR will be when the reform becomes stable and students spend more time in small classes. The one-year gains we observed fall below the (admittedly) arbitrary criteria of 0.2 for a “small” effect suggested by Cohen (1988), but they probably would be considered meaningful by most researchers and educators.⁸ For example, as a result of CSR about 6,000 more third-grade students are above the national median score in reading and about 9,000 more are above the national median in mathematics. The Consortium will be reporting more information about the cumulative impact of CSR on achievement during the next three years.

⁸ Mosteller (1995, p. 8) notes that “although effect sizes of the magnitude of 0.1, 0.2, or 0.3 may not seem to be impressive gains for a single individual, for a population they can be quite substantial.”

- In addition, efforts should be made to measure other potential benefits of CSR. These might include improvements in student socialization, teacher satisfaction, teacher retention, parent engagement, and public support.

Enhancing the “Evaluability” of CSR and Other Educational Reforms

- The CSR evaluation has been limited by the quality of information collected by the state. Improvements in the state educational data system would provide better information to judge the effectiveness of state policies. To be effective the system needs to be able to track students over time and to link teachers to students while still protecting individual privacy. It also is important to maintain consistency in labeling and classification of people and programs, and it would be invaluable to collect information on the implementation of and compliance with the panoply of state reform efforts.
- The CSR evaluation has also been hampered because it was not funded until the second year, when it was impossible to collect baseline data. State policymakers and taxpayers would benefit if evaluations were built into reforms when they are conceived rather than being added later. Every policy does not need to be evaluated, but those designed to have major systemwide impact are worthy of study.

Conclusion

As noted at the outset, CSR is a work in progress, and the Consortium will continue its evaluation over the next three years as the implementation proceeds. To date, CSR has had some clear successes, including the almost universal reduction in class size in first and second grade and the widespread reduction in kindergarten and third grade. There have also been small improvements in student achievement, which have occurred despite problems finding space and qualified teachers. These shortages have been felt the most in schools that serve higher proportions of low-income, minority, or ELL students. Although this study will continue for three more years, there is no reason for policymakers to wait to attend to these problems. They should also think carefully about the interactions of CSR with other reform initiatives and the need for better data for future analysis and policymaking. In time, there will be a better understanding of the effects of CSR on all students and of whether the benefits of the program are worth the cost.