CAPSTONE REPORT

CSR Research Consortium

a partnership researching California's class size reduction reform

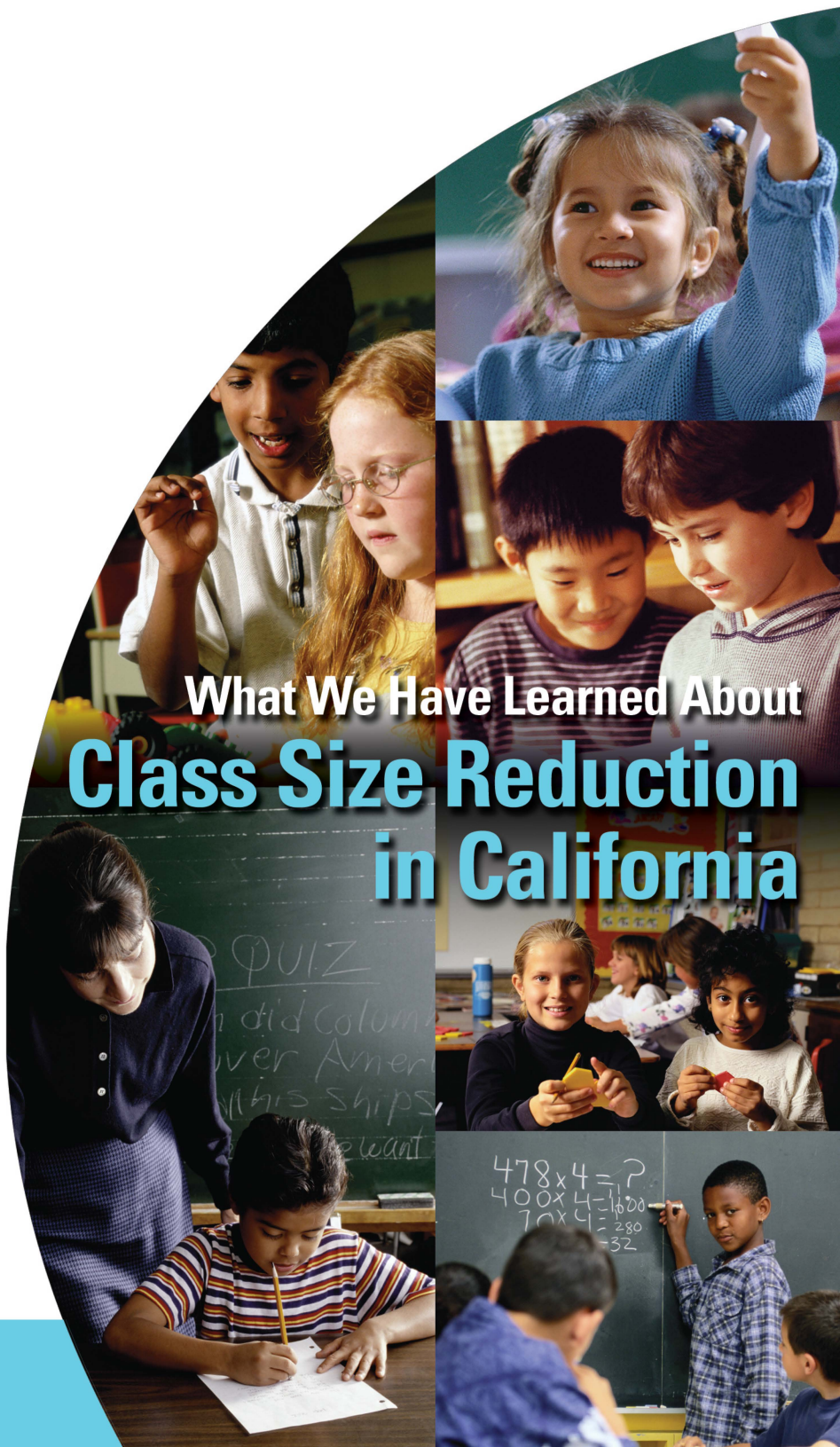
American Institutes for Research

RAND

PACE

EdSource

WestEd



# What We Have Learned About Class Size Reduction in California


Editors

George W. Bohrnstedt

Brian M. Stecher



SEPTEMBER 2002

CAPSTONE REPORT

CSR Research Consortium

a partnership researching California's class size reduction reform

# What We Have Learned About Class Size Reduction in California


## Authors and Consortium Members

**AIR**

George W. Bohrnstedt
Lawrence W. Gallagher
Hiroyuki Hikawa
Jamie Shkolnik
Ed Wiley

**RAND**

Catherine H. Augustine
Dionne Barnes
Delia Bugliari
J. R. Lockwood
Daniel F. McCaffrey
Brian M. Stecher
Marika J. Suttorp
Georges Vernez

**PACE**

Gerald C. Hayward, Sacramento
Michael W. Kirst, Stanford University

**EdSource**

Mary Perry
Miriam Rokeach
Jackie Teague
Trish Williams

**WestEd**

Joan McRobbie



## Consortium Funders

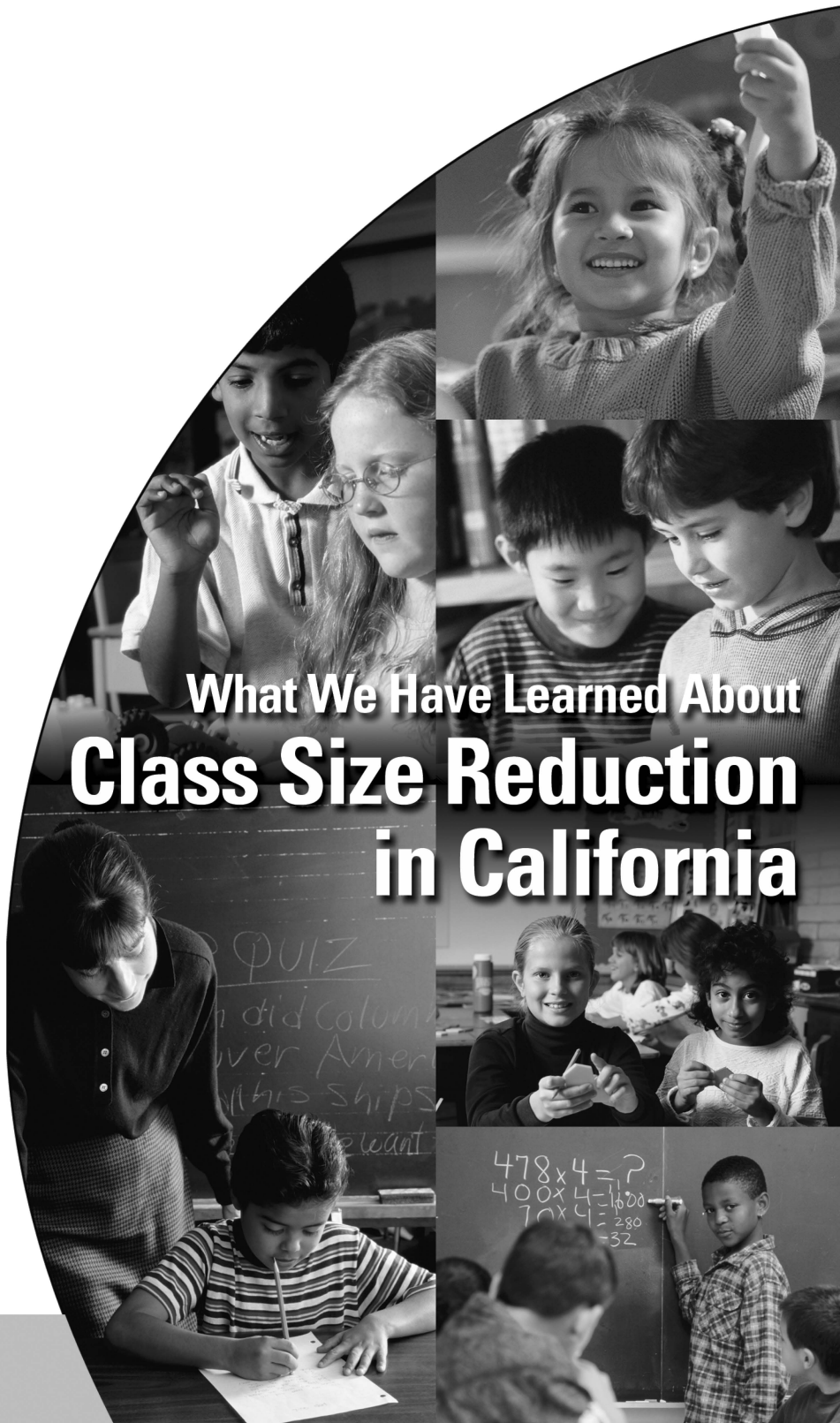
California Department of Education
The Walter and Elise Haas Fund
The William and Flora Hewlett Foundation
The Walter S. Johnson Foundation
The San Francisco Foundation
The Stuart Foundation


**www.classize.org**

CAPSTONE REPORT

CSR Research Consortium

a partnership researching California's class size reduction reform




# What We Have Learned About Class Size Reduction in California

SEPTEMBER 2002

CAPSTONE REPORT

CSR Research Consortium
a partnership researching California's class size reduction reform


**California Department of Education**

Delaine Eastin
State Superintendent of Public Instruction

**Office of Policy and Evaluation**

William Padia
Director

**School Facilities and Planning Division**

Duwayne Brooks
Director



**Suggested Citation**

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



**For More Information**

Contact:

George W. Bohrnstedt, AIR
650-493-3550
Brian M. Stecher, RAND
310-393-0411



For ordering information on this report, contact:

Ed Schrufer
American Institutes for Research
1791 Arastradero Road
Palo Alto, CA 94304
eschrufer@air.org
650-493-3550 ext. 8118

**This report is also available on the World Wide Web at www.classize.org**

# Acknowledgments

The Class Size Reduction (CSR) evaluation was conducted by a research consortium comprising the American Institutes for Research (AIR), RAND, WestEd, Policy Analysis for California Education (PACE), and EdSource. The effort owes its success to the cooperation and hard work of dedicated staff from all these organizations. We want to acknowledge the efforts of Joan McRobbie of WestEd, Michael Kirst and Gerald Hayward of PACE, and Trish Williams and Jackie Teague of EdSource.

While we and the other authors of this report take full responsibility for its content, we would like to credit the Consortium Advisory Group members for their invaluable advice on the study's design as well as their comments on an earlier draft of this report. The Advisory Group members are listed on the inside back cover.

The evaluation effort received generous support from public and private sources. Planning funds were provided by the Walter S. Johnson Foundation, the William and Flora Hewlett Foundation, and the San Francisco Foundation. Support for the design of the project came from the California State Board of Education. First-year project costs were underwritten by the California Department of Education and three California foundations: the Walter and Elise Haas Fund; the William and Flora Hewlett Foundation; and the Stuart Foundation. During the second year, the Walter and Elise Haas Foundation provided unrestricted funds for analysis of the CSR data. In addition, a grant from the U.S. Department of Education, Office of Educational Research and Improvement, allowed data on instructional practices in both reduced and non-reduced classes to be collected. We are grateful to all of these agencies and organizations for their financial support.

This year our analyses of state data were supplemented with information gathered from school districts. We are indebted to the many districts that shared their internal evaluation reports with us. A general summary of district findings, prepared by EdSource graduate research intern Miriam Rokeach, is provided separately in *What We*

*Have Learned About Class Size Reduction: Technical Appendix* (available on the Web at www.classize.org).

In addition, seven districts supplied us with specially compiled student and teacher data sets for an analysis of achievement gains and teacher characteristics. (This analysis is also reported in the *Technical Appendix*.) We want to express our gratitude to the districts for cooperating in the research and to their staff who assembled the data for us and helped interpret the results.

We also thank project staff at each of the Consortium organizations for their contributions to the research and to the generation of this report.

At the American Institutes for Research, we thank Ed Schrufer and Pam Wong for their administrative help. Special thanks go to Ed Wiley, who helped with many aspects of the project, including the design of several analyses that appear in this report; and to Jamie Shkolnik, who provided invaluable help with analysis design and writing in the third and fourth years of the project. Thanks are also due to Hiro Hikawa for carrying out various analyses, and to Freya Makris for assistance with a number of research tasks. Special thanks to Larry Gallagher who took the lead in help to conceptualize and analyze our study of CSR and teacher mobility. Finally, Michelle Bullwinkle deserves very special thanks for the design and execution of the report.

At RAND, Donna White helped with production of text and figures, Eileen LaRusso created the cover and title page layouts, and Jeri O'Donnell edited the final version of this report.

At EdSource, we thank deputy director Mary Perry for her skilled rewrite of Chapter 3 and for editing all three chapters.

We appreciate the administrative coordination and good humor provided by Diana Smith, Mary Gardner, and Terry Alter across the three offices of PACE.

We also thank WestEd staff Priya Armstrong for handling the full range of meeting logistics, Liza Cardinal for helping coordinate press relations, Maria Villegas for research assistance, and Christian Holden

for designing and producing the Consortium's Web site: www.classize.org.

We want to acknowledge the encouragement and support of William Padia, Director of the Office of Policy and Evaluation for the California Department of Education (CDE); Cathy George, Larry Boese, and John Bovine, CDE's monitors for the CSR contract; and other staff at CDE who provided support and advice throughout this fourth year of the project. We appreciate their review of an earlier draft of this report.

George W. Bohrnstedt
Co-principal Investigator
American Institutes for Research

Brian M. Stecher
Co-principal Investigator
RAND

# Executive Summary

The mid-1990s found California worried about the education its students were receiving. Standardized tests provided evidence that the state's students were losing ground compared to their counterparts across the country. The results of the 1994 National Assessment of Educational Progress (NAEP) released in 1995 only reinforced the concern—California's fourth graders had tied for last place in reading among the 39 states that participated in NAEP.

A task force assembled by the California Department of Education, called for among other reforms, smaller classes—a move strongly favored not only by the teachers' unions, but also by parents and teachers. California elementary schools had the largest class size in the country—averaging 29 students. Evidence from the Tennessee STAR experiment had shown rather clearly that elementary students in the primary grades did better academically when in small versus larger classes in K–3, and the difference was greatest for inner-city and minority students.

All that was missing to put class size reduction into place was political will and the money to do so. The dot-com boom of the 1990s solved the latter problem by providing a windfall of tax revenues, most of which were required by law to be spent on elementary and secondary education. Republican Governor Pete Wilson and the Democratically controlled legislature seized the moment and passed SB1777 (O'Connell) in July 1996. The law provided districts with $650 per student for each K–3 classroom with 20 or fewer students, providing they first reduced all first grade classes in a school, followed by all second grades and finally by either kindergarten or third grade classes. The cost to the state in the first year was roughly $1 billion dollars and in the current year, roughly $1.6 billion.

The California Department of Education and a group of California foundations awarded contracts to the American Institutes for Research, who along with RAND headed up a consortium to evaluate the effects of class-size reduction on achievement, on the quality of the state's teaching corps, on special needs students, and on other practices. The Consortium, which also included Policy Analysis

for California Education (PACE), WestEd, and EdSource, has produced three evaluation reports thus far. This is the fourth and final report. In it, we summarize previous findings and discuss new research done in the final year of the contract; we also include a set of policy recommendations and conclude with lessons learned.

## What Did We Learn?

The major findings from this four-year contract can be summarized as follows:

1. *Implementation of CSR occurred rapidly, although it lagged in schools serving minority and low-income students.* Districts reacted quickly to the opportunity presented by the CSR program when it was enacted in July 1996. CSR implementation was virtually complete for first and second grades by the second year of the program, and for kindergarten and third-grade students by the fourth year of the program. Implementation was slower, however, for schools with higher percentages of minority students and of low-income students, partially because schools in urban districts had more difficulty acquiring the needed space to expand the number of classrooms.

2. *Our analyses of the relationship of CSR to student achievement was inconclusive.* Student achievement has been increasing since the first administration of the SAT-9 in 1997, but we could find only limited evidence linking these gains to CSR. We found a positive association in 1998 between third-grade class size and SAT-9 scores after controlling for differences in student and school characteristics. However, the size of this CSR effect was small. In the following year, 1998–99, these positive differences persisted when students who had been in reduced size third-grade classes moved to the fourth grade and regular size classes. The spring 1999 SAT-9 results showed that fourth-grade students who had been in reduced size third-grade classes scored higher than those who had not been in such classes. By 2001, CSR implementation was nearly complete, and as a result we could not examine differences in SAT-9 scores between students who were and were not in reduced size classes. Instead, we tracked achievement gains between cohorts of students with incrementally different patterns of CSR exposure to CSR from kindergarten through third grade.

Although both overall exposure to CSR and statewide average test scores increased across cohorts, the magnitude of the changes in test scores did not track with the incremental changes in CSR. Thus, attribution of gains in scores to CSR is not warranted. More refined school-level analyses also failed to find meaningful differences in second- or third-grade scores of students with an additional year of CSR exposure in first grade compared to students who participated only in grades 2 and 3. We could not determine whether our ability to link CSR to achievement was due to weakness of the effect of incremental differences in CSR or to design limitations (or a combination of both). We were also limited in our ability to determine how much of the recent gain in achievement was attributable to CSR and how much was linked to other initiatives,

3. *CSR was associated with declines in teacher qualifications and a more inequitable distribution of credentialed teachers.* Reducing class size required an enormous increase in the number of K–3 teachers in California. Between 1995–96, the year before CSR implementation, and 1998–99, the third year of the program, the total number of K–3 teachers increased 46 percent, from 62,226 to 91,112. To meet the increased demand for teachers, many districts hired teachers without full credentials. As a result, the proportion of K–3 teachers who were not fully credentialed (e.g., teachers with intern or emergency credentials) increased from 1.8 percent before the program started to 12.5 percent in the second year of the program. Most of the uncredentialed teachers were hired by schools serving the most disadvantaged students, in part because these schools were slower to implement CSR, and more certificated teachers had already been hired elsewhere. In 2000–01, more than one in five K–3 teachers were not fully credentialed in schools with high percentages of low-income, EL, minority, or Hispanic students (primarily large and urban).

4. *CSR had only a modest effect on teacher mobility.* One of the fears was that class-size reduction would result in two types of teacher mobility—teachers from urban schools moving into suburban schools and upper grade elementary teachers moving into K–3. While there was some initial increase, the effect was small and soon disappeared. Approximately 7 percent of first-grade teachers in 1995–96 (the year prior to CSR) had been teaching in a

different school the previous year. That percentage rose to 11 percent in 1996–97 and dropped back down to 5 percent by 1999–00. The same pattern was true in the other elementary grades. The school transfer rate was small especially when compared to the 46 percent increase in the number of K–3 teachers during this period.

5. *CSR implementation did not affect special education identification or placement.* There was some concern that reducing class size might affect the number of students referred for special education assessment, the number of students identified as needing special education services, or the number of special education students placed in special day classes (instead of in general education classrooms). Our analyses of statewide enrollment data showed no evidence that the rates for either were affected by CSR.

6. *Students in reduced size third-grade classes received more individual attention, but similar instruction and curriculum.* Compared to teachers with larger classes, teachers of reduced size classes were more likely to say they know what each student knows and can do, that they provide feedback on writing assignments within one day, that they give more individual attention to students, and are able to meet the instructional needs of all students. Teachers in reduced size classes also reported fewer behavior problems and reported that students were more likely to complete the lesson for the day and less likely to be "off task" for more than 5 minutes. But teachers in both reduced and non-reduced size third-grade classes reported spending similar amounts of time and covering similar amounts of curriculum in language arts and in mathematics.

7. *Parents liked reduced size classes.* Based on survey results, parents of third-grade students in reduced size classes rated selected features of their child's education higher than did parents of children in non-reduced size classes. The differences in rating of classroom size were particularly pronounced, with parents of children in reduced size classes reporting satisfaction levels far higher than parents of children in regular size classes. However, parents of children in both reduced and non-reduced size classes expressed equal satisfaction with the qualifications of their children's teachers.

8. *Classroom space and dollars were taken from other programs to support CSR.* Most districts in our statewide sample reported incurring operating costs for CSR that exceeded state payments for it, and these funding problems persisted, or even worsened, in recent years. Districts attempted to overcome budget shortfalls created by CSR by reducing funds for facility maintenance and administrative services. About one-third of such districts also reduced resources for professional development, computer programs, or libraries. To be able to implement the program, many schools reported having to reallocate full-sized classrooms that had been designated for special education back to K-3 classrooms, thereby forcing special education classes to use alternative spaces. CSR implementation also preempted space from such uses as music and arts, athletics, and childcare programs.

9. *In spite of budget shortfalls districts are not projecting CSR cutbacks for 2002–03.* In spite of the fact that the state of California is projecting a significant budget deficit for 2002–03, and that many districts are forecasting total revenues that will not meet projected expenses, none of the 38 districts we surveyed in 2002 indicated that they are contemplating elimination of CSR in the immediate future. Some did indicate, however, that cuts to the CSR program were a possibility and would continue to be discussed as their budgets were developed. However, it would be a "last resort" change given the popularity of CSR with parents and teachers.

## Recommendations

CSR is an enormously popular program in California among elementary parents and teachers. It is also clear that local educators and parents may value reduced class sizes for many reasons other than improved achievement as measured by statewide test scores. Therefore, maintaining small K–3 classes in California is likely to remain a priority. Nonetheless, we believe that some changes in the program should be considered based on our evaluation and on research done on CSR in other states. In addition, the state policy environment of 2002 is markedly different from that which existed in 1996 when CSR was introduced and implementation began. The

state has started moving toward a systemic standards-based system, with a strong emphasis on high expectations, accountability, and accompanying rewards and sanctions based on growth in student achievement.

1. *Improve the effectiveness of the current CSR program by integrating and aligning it with other reforms.*

   The Consortium is impressed with the need to link CSR to the state's overall strategic direction—i.e., to end its current status as a freestanding categorical program by integrating it into and aligning it with the state's standards-based policies. Such a shift, we believe, would allow CSR to better support the state's standards-based reform strategy and might prompt better results from the CSR investment. Schools may be able to use other elements of standards-based reform—e.g., additional funding for the turnaround of low-performing schools—in ways that allow them to take fuller advantage of the opportunities small classes have been shown to present in some states, especially for low-income and minority students. In short, integrating CSR into the state's evolving standards-based reform policy could significantly bolster California's ability to meet its objective of improving student achievement.

2. *Be explicit to the field about the assumptions underlying state reimbursement of CSR, while also taking steps to determine the real costs as well as the cost-effectiveness of CSR.*

   There exists a fundamental difference in the way CSR is viewed by state policymakers and school district personnel. State officials describe the program as an incentive program, not a state mandate. As such, they argue that districts have the option to participate or not. They further argue that the state provides adequate resources through the combination of CSR funding and general purpose funding. Many districts, on the other hand, feel that the state indicated an intent to fully fund the program when it adjusted the funding upward in 1997. State support has not kept up with costs since then, and districts believe that the state should once again provide adequate resources for full funding. In either case, whether costs are fully reimbursed or not, the rules regarding appropriate cost attribution should be explicit, and

districts ought to have reasonably predictable revenue streams so that they can make informed choices about implementing CSR. While determining costs attributable to specific programs is not a simple matter, a careful cost review could illuminate this issue and result in a single set of rules relating to cost attribution. Even more important, having solid cost data would assist state and local policymakers in determining the cost effectiveness of CSR compared with other possible reform expenditures.

3. *Provide more local flexibility within the current CSR program by allowing a school-wide average of 20 in grades K-3.*

   Along the lines suggested by the Legislative Analyst's Office—and consistent with the recommendations made by the CSR Consortium in previous reports—local districts should be given the flexibility to vary class sizes by up to two per class as long as the class size average within a school remains 20 or less. Making the class size cap of 20 applicable to a school rather than each class within a school would give schools a modicum of additional flexibility while only modestly affecting the way the limit is applied.

4. *Further test CSR's potential to improve the achievement of low-income/minority students by providing additional resources to create and evaluate pilots with even smaller class sizes in selected schools.*

   Based on the evidence from the Tennessee STAR experiment, it appears that class size reduction can be an especially effective policy strategy for raising the achievement of the most at-risk students if class sizes are reduced even further for that group and if those classes are staffed by skilled and qualified teachers. It is possible to conduct carefully controlled experiments to examine the difference moving to a class size of 15 or fewer would make, beginning with those schools that serve the largest number of low-income and minority students

5. *Further test and evaluate cost-neutral alternative CSR strategies by providing incentives to a small number of districts to experiment.*

   By allowing a relatively small number of school districts to use their CSR funding to create randomized trials of other small class size arrangements, the state could compare the effectiveness of the

current CSR program with alternative CSR designs. Participating districts would be required to randomly assign schools, or classes within schools, to the current CSR program structure versus an alternative CSR model. Researchers could then track the changes in student achievement for the alternatives. The state should also consider allowing districts to compare one or more non-CSR uses of the funds against the current CSR model, again with the requirement that this must be done using randomized trials. Both of these options have the virtue of providing information about the cost-effectiveness of the alternatives (since all would have the same cost), something that could not be done as part of the current evaluation. The major incentive to districts to participate would be the ability to design their CSR or other programs to meet local needs. A second incentive would be additional state funds for participating districts for the purposes of technical assistance in putting together the research design, and for evaluating the effects of the alternative uses of the CSR money.

6. *Further explore why and how CSR works by identifying best instructional practices in small classes.*

   We do not know enough yet about the conditions under which CSR is most effective in improving student achievement; as a result we can offer little guidance about how to make it work better. We need more research on understanding what classroom practices are most effective in small classrooms and whether these differ from best practices in larger classes. We know that CSR had some effects on instructional practice in California, but we do not know what type of changes in classroom teaching would be needed to maximize the benefits of the reform.

7. *Before undertaking any statewide effort to expand CSR to additional grades, policymakers should ensure the state has sufficient facilities and qualified teachers.*

   The state has taken substantive action to remedy the facility and teacher shortage problems originally created by K–3 CSR and address these issues more generally. There has been a thoughtful, concerted effort in recent years to establish new policies related to teacher preparation, credentialing, recruitment, and retention. As to California's school facilities crisis, significant progress has been

made since CSR was signed into law. Meanwhile, however, some school districts in California continue to be severely constrained by the capacity of their facilities. It is unclear whether the state's efforts will prove fully effective in remedying the current problems, much less in providing the kind of capacity that would be needed for CSR expansion.

8. *Improve the state's ability to determine the effectiveness of its education reforms by investing in an enhanced education data system.*

   California's education data system must be redesigned to allow researchers to link teachers and children with their achievement scores over time in order to better measure student gains from year to year. In this way, the state can more accurately measure the effect specific reforms or variables have on student achievement. The creation of such a data system requires an adequate investment of time, money, and political will on the part of the state and local school districts, but ultimately promises important returns related to school effectiveness.

Beyond the specific recommendations we make above related to California's current K–3 CSR program, we think that the state's experience provides some broader lessons for policymakers. Whether California embarks on additional CSR initiatives or undertakes other large-scale reform interventions, these lessons, as outlined in Chapter 3, can serve as important guides for developing effective education reform policy:

## Conclusion

The results of this evaluation, a changing policy context, and new research all provide justification for re-examining the current CSR policy in California. As suggested in our full report, the state can change some of the particulars of the CSR program without abandoning its commitment to smaller K–3 classes. Through carefully controlled pilot programs we can also learn more about what is working, what is not, and why.

## Report Overview

In this fourth and final report on the California Class Size Reduction (CSR) initiative, we summarize what we have learned since we began our CSR evaluation 1997. This report is divided into three chapters.

Chapter 1 examines the context within which our evaluation took place. In addition to recapping the history of California's CSR initiative, it includes a discussion of what state leaders' expectations were when CSR was passed. Further, in order to broaden the information base for our recommendations, Chapter 1 describes research on class size reduction that has been conducted elsewhere.

Chapter 2 summarizes findings from our previous CSR evaluations as well as several new analyses. These findings relate to program implementation issues, parent support, relationship of CSR to academic achievement, changes in teacher qualifications, teaching practices and resource issues, and the effect of CSR on the identification of special education students. The three new analyses in this report are[1]

- An additional analysis of the relationship between CSR and student achievement in California using school-level data.
- A teacher-flow analysis that compares the pre-CSR year-to-year movement of teachers among schools and among grade levels within schools with the year-to-year movement of teachers that occurred in each of the CSR years.
- A telephone survey of superintendents for estimates of the CSR program's effects on district budgets in 2001–02 and on budget projections for 2002–03.

[1] We also undertook a fourth analysis in which we examined the relationship between teacher qualifications and student performance in reduced size classes using a small, nonrandom sample of California districts whose student-level achievement data were linked across two academic years. The use of only one year of linked data resulted in our finding relatively little variation in achievement. Probably as a result of this restricted variation, our analyses of teacher characteristics yielded only a few statistically significant findings, the overall pattern of which was not easily interpretable. The detailed study results are reporting separately, in *What We Have Learned About Class Size Reduction in California: Technical Appendix* (available on the Web at www.classize.org).

Chapter 3 asks: Given the findings and limitations of the research design, what policy implications might be drawn for California? In addition, the chapter looks at how CSR fits into the larger context of standards-based education reform in California today, and at the lessons that California can draw from the CSR experience for future policy decisions.

# Table of Contents

# Chapter 1

## *Context for Evaluating Class Size Reduction in California*

### Background

In the spring of 1996, California's public education system faced numerous challenges. One challenge was overcrowding. During the preceding 5 years, the state's K–12 enrollment had increased by 12 percent. Besides needing to accommodate more students, schools had to serve students with more-complex learning needs, since many of the newcomers were not native English speakers. At the same time, the state was experiencing a growing shortage of certificated teachers, especially teachers credentialed to teach non-English speakers. Concern about how schools were faring increased as evidence from an array of standardized tests indicated that the state's elementary students were losing ground in academic achievement compared with their peers nationwide. Concern gave way to alarm when the results of the 1994 National Assessment of Education Progress (NAEP) were released in 1995: in reading, California's fourth graders had tied for last place among 39 participating states.

In immediate response, the California Department of Education assembled the California Reading Task Force to identify the causes. State Superintendent Delaine Eastin, along with the teachers' unions, began calling for smaller classes—something parents and teachers strongly favored. In 1996, the state legislature took up the issue in earnest. The state's economy, spurred by the dot-com boom, had rebounded from the early 1990s recession. Under the terms of voter-approved Proposition 98 (as amended by Proposition 111), the resulting budget windfall had to be spent primarily on public education. Legislators began scrutinizing evidence from an educational experiment in Tennessee (described later in this chapter) that strongly supported class size reduction as a means of improving academic achievement, especially for inner-city and minority[2]

[2] Minority students are any students not classified at Caucasian. The largest groups of minority students are, in order of group size, Hispanics, Asian/Pacific Islanders, and African Americans.

students. At this point, California's elementary classes averaged 29 students, the largest average class size in the country.[3] The gap in achievement between low-income[4]—often African American and Hispanic—students and those from middle- and upper-income families was well documented.

Politics inevitably came into play. By targeting the money for class size reduction, Governor Pete Wilson and other state leaders could keep it out of districts' general funds and thus beyond the reach of collective bargaining agreements and teachers' unions (Parrish et al., 1999; Smith, 1998).

With all these forces converging, the legislature seized the moment and passed SB 1777 authorizing class size reduction in grades K–3 and appropriating $1 billion for its launch, making it one of the most expensive state education reforms in U.S. history. The governor signed the bill in July 1996, just 6 weeks before most schools re-opened for the fall.

## What Were California's Expectations for CSR?

The goal for California's Class Size Reduction (CSR) initiative, as stated on the California Department of Education's Web site, is to "increase student achievement, particularly in reading and mathematics, by decreasing the size of K–3 classes to 20 or fewer students per certificated teacher."[5] The literacy-bolstering intent was underscored by companion legislation that included AB 3482, the California Reading Initiative (CRI), which provided professional development designed to improve K–3 teachers' knowledge of language acquisition and literature as well as skill in reading instruction. CSR also was clearly seen as a way to increase parent and public support for California public school education.

On the face of it, reducing class size would be to everyone's advantage—teachers, parents, politicians, and especially the state's

[3] See http://nces.ed.gov/pubs2002/digest2001/tables/PDF/table069.pdf for 1993–94 average class sizes by state.

[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] See http://www.cde.ca.gov/classsize/facts.htm.

youngest students. But problems were apparent from the outset. In lieu of targeting funds to low-income or minority students, CSR was offered to every school and district. Instead of a formula addressing widely varying implementation costs across districts, funding was one-size-fits-all—$650 per student (almost $890 in 2001–02) for every K–3 student in a class of 20 or fewer. (See box below: *California's CSR Policy and Regulations.*) As a result, wealthier districts that already had smaller classes got an initial boon, while overcrowded districts dipped into their general funds to cover shortfalls. That is, the program was far from fully funded for many of California's districts. While the program was voluntary, the state's funding-starved schools were loath to turn away dollars (at the time, California was 41st among states in per-pupil spending). Perhaps more importantly, parent pressure to downsize quickly was immense, and in the weeks between the July passage of the bill and the start of the school year, the press covered school-by-school progress in adopting CSR like a horse race. By the end of the first year, 88 percent of the state's first graders and 57 percent of its second graders were in reduced size classes.

**California's CSR Policy and Regulations**

California's CSR initiative, by far the largest in the nation, is a voluntary incentive program in which the state provides per-pupil funding to districts for every student in kindergarten through third grade in a class of 20 or fewer. A class may have over 20 students on a particular day, but the average may not exceed 20.4 over the course of the school year.

In the first year, 1996–97, school districts received $650 for each participating student. The law now provides an annual cost of living adjustment, and by 2001–02, funding had increased to almost $890 per student.

At the program's launch, funding was available to reduce three of the four K-3 grades. State regulations specified the order for implementation: first grade had to be reduced first, followed by second grade, and then schools could choose to reduce class sizes in either kindergarten or third grade. In other words, for a school to get funding to reduce the size of kindergarten or third-grade classes, it first had to reduce all first- and second-grade class sizes. In 1997–98, state leaders committed to funding the program for all students in kindergarten through third grade. For the 1997–98 school year, teachers had to be hired by November 1, 1997, and CSR had to be implemented by February 1998. Starting in 1998–99, all class size reductions and teacher hiring had to be completed before the start of the school year. The regulations also required that teachers participate in specially designed professional development programs that focused on maximizing "the educational advantages of class size reduction."

CSR regulations did not prohibit districts from implementing in varying degrees in different schools. Some schools could implement fully while others in the same district might reduce classes in only one or two grade levels.

School districts could apply for both operations and facilities grants. The annual operations grant provided funds for full-day or half-day instruction in reduced size classes. In 1996–97, the grants were $650 and $325 per student for full-day and half-day participation, respectively. In 1997–98 these amounts increased to $800 and $400; and by 2000–01 they increased to $850 and $425. Full-day funding was called Option 1, and half-day funding was called Option 2. Option 2 was rare and is not included in this evaluation.

A one-time facilities grant of $25,000 per newly created classroom was available in 1996–97. Facilities grants increased to $40,000 per newly created classroom in 1997–98 and 1998–99. However, the $40,000 facilities grants for these 2 years were available to a district only if it had not reduced all of its eligible classes to 20 or fewer students and therefore did not use all the operational funds that it might have claimed.

The overnight need for approximately 18,000 new classrooms in a facility-challenged state led to expedient but compromised solutions—conversion of libraries, labs, and assembly stages into classrooms; switches to year-round calendars—some of which remain problematic. The already-existing problem of teacher shortages, quality, and distribution took on crisis proportions. The hiring of many new taxed schools' capacity to support and mentor teachers. Particularly troubling was the proliferation of emergency-permit teachers in high-poverty areas, which raised early concerns about

equity. Would low-income and minority students, who stood to gain the most from CSR, be least likely to benefit?

Despite such problems and concerns, CSR met with enormous local enthusiasm. For example, in a 1997 survey conducted among more than 5,000 parents in Santa Barbara County, parents voiced overwhelmingly their beliefs about smaller class sizes: that they improved one-on-one instruction between their children and teachers, that students learned more, that classrooms were under better control, and that fewer disruptions occurred (Cirone, 1997). Press reports quoted teachers around the state speaking of their new ability to give young students more individual attention and instruction.

Now, 6 years into the reform, the size of nearly all K–3 classes has been reduced. The program remains enormously popular with teachers and parents, but it has been expensive, and some have questioned whether the results have been worth the cost. The issue of cost-effectiveness looms especially large in light of the huge shortfall in tax revenues being projected for fiscal year 2002—a deficit estimated to be as large as $23.5 billion. Though the governor's budget message spares CSR from cuts, tough spending decisions will continue to be made, and parents and teachers are unquestionably nervous about the state's ability to continue supporting reduced size classes, as well as the ability of districts to pay for the remaining costs of CSR not covered by the state.

## What Does Research from Other States Say About Class Size Reduction?

Outside of California, much research has been conducted into various aspects of class size reduction, with a particular focus on how it affects student achievement. The research also provides some limited information on what types of state policies are most effective for helping schools realize the potential of reduced size classes.

## Effectiveness for Improving Academic Achievement

Hundreds of studies over several decades have examined the effects of reduced class size on student achievement. Some suggest a positive

impact; others find no evidence of any impact. Among the most influential research was Glass and Smith' s 1978 meta-analysis of 77 class size reduction studies, which concluded that "large [achievement] advantages [can be expected to occur] when class size is reduced below 20" (Glass and Smith, 1978, p. ii). In a 1982 follow-up report, Glass and associates reiterated the earlier findings and noted that of the more than 100 well-controlled comparisons, 81 percent favored smaller class sizes. They strongly suggested that class sizes needed to be reduced to fewer than 20 pupils for significant results to be observed (Glass et al., 1982).

But by far the most important and compelling evidence of class size reduction's positive effects on academic achievement comes from Tennessee's Student/Teacher Achievement Ration (STAR) project. It not only was a large-scale longitudinal study, but it also was and remains the only study of a large-scale class size reduction program that used an experimental research design. Beginning in 1985, 79 elementary schools agreed to participate. Students entering kindergarten were randomly assigned to one of three class types: a small class of 13–17 students, a regular class of 22–26 students, or a regular class of 22–26 students with a full-time teacher's aide. Importantly, teachers were also randomly assigned. Teachers and aides participating in the program did not receive special instruction or additional professional development training (Finn, 1998). Pupils stayed in the class type to which they were assigned through third grade, with a new teacher being randomly assigned each year. Achievement was measured at the end of each year.

Project STAR's major findings and those of other research to date include (Finn, 2002):

- Students in small classes performed better at all K–3 grade levels than did students in larger classes.
- Minority and inner-city children gained more from reduced classes than their White and non-urban school peers; indeed, the effects were two to three times as great.
- Teacher morale was higher in smaller than in larger classes.
- Teachers spent more time on direct instruction and less on classroom management in smaller versus larger classes.

- Students in smaller classes were more engaged in learning than were students in large classes.
- The earlier and longer the participation in small classes, the greater the effect on achievement.
- Students in small K–3 classes did better academically in grades 4, 6, and 8 than did students in larger K–3 classes.
- The more years students spent in small K–3 classes, the longer-lasting the benefits in later years of schooling.
- Students who had been in small K–3 classes were more likely to graduate from high school, to take college admissions examinations, and, in general, to take courses that prepared them for college than were those who had been in larger K–3 classes. Furthermore, these effects were stronger for minority students, thereby helping close the college preparation gap between African American and White students.

Hanushek (1998) argues that the effects in the Tennessee STAR project occurred primarily in kindergarten and first grade and that there is no evidence that additional years of class size reduction contribute incrementally to the effect of small classes in the early years. He acknowledges that the effects were greater for minority and disadvantaged students but then argues that "the effects appear small relative to costs of programs and alternative policy approaches" (Hanushek, 1998, p. 31). Krueger (2000) counters Hanushek's cost-ineffectiveness argument by showing that there may be significant long-term earning differentials for Tennessee students who were in small versus large classes given that they were more likely to take courses and entrance examinations that rendered them more college- and, therefore, more job-prepared.

Odden (1990); Hanushek (1998); Levin, Glass, and Meister (1984); and Levin (1988) all suggest that class size reduction may not be as cost-effective as other education interventions for improving academic outcomes. Levin's work, for example, suggests that lengthening the school year or using computer-assisted instruction may be more cost-effective than reducing class size (Levin, 1988; Levin, Glass, and Meister, 1984). However, each of these studies has important limitations. As a result, there is no "gold standard" study for drawing conclusions about the cost-effectiveness of reduced size classes.

Other key questions remain uninvestigated. Still unknown, for example, is *why* the class size effect occurred where it did. That is, changes in teacher practices or student behaviors that might account for the effect are not well understood. Furthermore, better research is needed to determine how small is small. Does one need classes of 20, 18, 15, or some other number of students?

## Effectiveness of State Policies

Research on the effective design of statewide class size reduction policy outside California also is sparse. Since the mid-1980s, at least 20 states have enacted CSR policies (see Appendix:), with many doing so in the mid- to late-1990s as the nationwide push for greater school accountability gained momentum. Drawing on the Tennessee STAR experiment, and flush with resources from a robust economy, many states embraced class size reduction as a means of addressing low academic performance and/or narrowing the achievement gap. No small factor was the political viability of class size reduction compared with many other interventions—smaller classes had the support of parents, teachers, Democrats, Republicans, and unions.

But exemplary as the STAR project is, it offers scant help in terms of how to design large-scale, statewide policy based on its findings. From STAR and other studies, policymakers generally decide to target the primary grades. But other issues are much less clear. How small is small enough—within state budgetary constraints? Are there ways to contain the costs while still getting the positive effects? Should reduced size classes be targeted to certain student populations? Are there ways around facility barriers? Should the reduced class strategy be coupled with other strategies? Should the program be optional or mandatory? Will the funding be flat or wealth-adjusted? Should there be a rigid cap, or will the number of students per class be flexible? Will smaller classes be self-contained or team-taught? These issues have been handled differently from state to state.

Despite the number of states now enacting reduced class size policies, very few have evaluated these policies' impact. Nevada's limited evaluations have been inconclusive. In 1989, Nevada began a phased-in reduction of primary-grade class size. Though researchers recently found evidence of a differential, positive effect on the achievement of

English language learners (Snow and LaMarca, 2001), achievement gains generally have been disappointing, and evaluation has not been comprehensive enough to indicate why. Utah has funded class size reduction since 1990, including some targeting of low-income students and flexibility in how districts and schools use the money. A 1997 study of five districts in Utah (Evans-Stout et al., 1997) found that the most successful schools combined reduced classes with teacher development, instructional improvement, and productive use of personnel and resources.

Similar findings were reported in Austin, Texas, where 2 of 15 low-performing schools showed dramatic gains while the other 13 saw achievement and attendance remain extremely low (Murnane and Levy, 1996). The two most successful schools combined reduced classes with other changes, such as new curricula and teaching methods, increased parent involvement, and health services. And in rural Burke County, North Carolina, where a phased-in program that began in 1991 combines classes reduced to 15 students (first through third grades) with comprehensive staff development, the result has been consistent and lasting achievement gains in reading and mathematics (Achilles, Harman, and Egelson, 1995; Egelson, Harman, and Achilles, 1996; Harman, Egelson and Achilles, 1997; Egelson and Harman, 2000).

Among the states other than California that have class size reduction programs, Wisconsin has conducted the most extensive and methodologically sound evaluation. Begun in 1996–97, Wisconsin's Student Achievement Guarantee in Education (SAGE) was a statewide pilot program (made permanent in 2001) targeted to schools and districts with high poverty rates. Phased in over 5 years, SAGE required that the student-teacher ratio be reduced to 15:1 in grades K–3. Participating schools were also required to implement a rigorous academic curriculum, provide before- and after-school activities, and implement professional development and accountability plans. As in the STAR project, the result was a significant, positive effect on academic achievement, with minority students benefiting the most. These state experiences, coupled with the STAR research, suggest several state policy lessons. STAR researchers, for example, note certain conditions without which the class size effect is unlikely to be realized. Chief among these are an adequate supply of qualified teachers and sufficient classroom space.

The STAR findings also provide support for differential targeting of resources to poor and minority students, and Wisconsin's SAGE experience has now corroborated these findings. Moreover, experiences in Utah, Texas, and Wisconsin all tend to support the coupling of reduced size classes with other reforms—notably curricular rigor and teacher professional development.

## How Does CSR Differ from Tennessee's STAR Project?

Tennessee's STAR project may have strongly influenced California's decision to initiate CSR, but comparisons of outcomes in the two states are problematic. There are important differences between California and Tennessee. It is instructive to examine just how different the "scope conditions" for implementation were between the two:

**Definition of a Small Class.** For the STAR project, the definition of a small class was 13–17 students, whereas a large class was 22–26. California's CSR program provides funding for classes of 20 or fewer, reduced from an average class size of 29. Given the costs, most California school districts keep their reduced size classes as close to 20 students as possible, a number closer to a large class in STAR.

**Demographics**. Tennessee's elementary school population is roughly 75 percent White. Of the non-White students, the vast majority is African American; only 2 to 3 percent are other minorities. California's enormously diverse population, by contrast, includes students from a broad array of racial, ethnic, and linguistic backgrounds. The state's 1.92 million K–3 public school enrollment for 2000–01 was 48 percent Hispanic, 32 percent White, 11 percent Asian/Pacific Islander, 8 percent African American, and 1 percent American Indian/Alaskan Native. One in every 4 students in California's public schools was classified as an English learner (EL);[6] the proportion of EL

[6] 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.

children in grades K–3 was more than 1 in 3.[7]

**Size of the Initiative**. The Tennessee project involved 6,000 K–3 students in 80 schools that had available space as an experiment to determine whether academic gains were associated with being in small versus large classes. As such it was an "effectiveness trial," to use the language of public health research. By contrast, California took CSR to scale as a fully implemented program. As a result, the first year of CSR in California involved over 150 times as many students as the Tennessee study—almost 1 million K–3 students—in 52,000 classrooms. At the end of the second year, roughly 1.59 million of the state's K–3 students were in reduced size classes in over 86,000 classrooms. By 2000–01, almost 1.86 million (97% of the state's 1.92 million K–3 students) were in almost 99,000 reduced size classes.

**Availability of Qualified Teachers**. Tennessee had no shortage of qualified teachers to staff the reduced size classes. By contrast, in 1996–97, the first year of CSR, California hired an additional 12,000 K–3 teachers as a result of increasing enrollments and CSR. Another 12,000 K–3 teachers were hired the following year, followed by an additional 5,000 K–3 teachers in 1998–99. In 1999–2000 the total went up by another 3,000; it then leveled off in 2000–01.

Because of this huge increase in the demand for teachers, schools had to be less selective in hiring. Before CSR, in 1995–96, almost all California teachers held full credentials. By 1997–98, however, there was a huge gap between the percentage of credentialed teachers in the quartile of schools with the most low-income students and the percentage of credentialed teachers in the quartile with the fewest low-income students. For example, a student at a school serving the neediest students would have a 1 in 5 chance of being taught by a teacher without full credentials, whereas a student at a school serving the fewest low-income students would have a 1 in 23 chance. McRobbie, Finn, and Harman (1998) underscore that an adequate number of qualified teachers is a key condition for realizing what STAR project showed to be "the class size effect."

[7] Taken from the CDE Web site, http://data1.cde.ca.gov/dataquest/StEnrAll.asp.

**Availability of Space**. Reducing class size creates a tremendous demand for additional classroom space. Tennessee had adequate space to house the reduced size classes, whereas California did not. As documented in Chapter 2, many schools gave up other space—e.g., libraries, computer labs, playground space, and child care facilities—to make room for new classes. Many also purchased or rented portable classrooms, partitioned classrooms, or acquired surrounding community facilities (e.g., community centers and churches) to be used as classrooms.

Because of these differences, or "scope conditions," between the California and Tennessee programs, the possibilities for comparison are clearly limited. We return to this point when examining the implications of our findings in Chapter 3.

## What Research Questions Are Examined in This Report?

Most evaluations of class size reduction conducted in other states have focused solely or primarily on the association between class size reduction and achievement. Our research consortium decided early on to take a much more systemic approach. We decided that it was important to look not only at academic achievement, but at other outcomes as well. This included the speed of implementation, barriers to implementation, the relationship of school characteristics and district resources on implementation rates, the role facilities played in implementation, how teacher supply affected CSR implementation, which teachers with which characteristics ended up in which districts, how CSR affected parental interest and participation in schools, and whether and how class size related to instructional practices. More details about the conceptual approach taken can be found in our three earlier evaluation reports: Bohrnstedt and Stecher, 1999; Stecher and Bohrnstedt, 2000; and Stecher and Bohrnstedt, 2002.

Using our conceptual model, our first three reports focused on research questions in seven areas:

- ***Rate of Implementation***—How quickly was CSR implemented? What percentages of students, by grade level, were in reduced size classes in each of the first 5 years of the program? How

did implementation rates vary as a function of a school's percentage of low-income, EL, or minority students?

- *Achievement*—What were the effects of CSR on student achievement? Did these effects differ by grade level, school characteristics, or student characteristics? Did these effects vary across years of CSR implementation?
- *Teacher Characteristics*—How did CSR affect the overall characteristics of California's K–3 teaching corps over the 5 years of implementation? Were changes in the workforce distributed proportionately across schools or were there differences associated with school characteristics such as the percentage of low-income, EL, minority, or Hispanic students?
- *Teaching Practices*—Did non-reduced and reduced size classes differ with respect to content covered in language arts and mathematics instruction? Did teachers in reduced size classes cover more topics or spend more time on individual topics? Did non-reduced and reduced size classes differ with respect to teaching practices? Were students grouped differently? Were there differences in students' learning activities? Were there differences in the amount of individualized instruction and feedback?
- *Parental Involvement and Satisfaction*—Did parents of children in non-reduced and reduced size classes differ in their education-related interactions with their children, and their children's teachers? Did they differ in terms of classroom volunteering? Did they differ in regard to satisfaction with their children's education?
- *Resource Allocation*—To what extent did state funding cover the districts' operating costs of CSR? Which districts benefited from surpluses? Which suffered shortfalls? To what extent did resource constraints (e.g., space, teachers, and supplies) affect implementation of CSR? How did districts and schools reallocate resources within these constraints?
- *Special Education*—Did CSR have an effect on the rate at which teachers referred students as possible candidates for special education services? Did CSR have an effect on the percentage of students identified as needing special education

services? Were students identified as needing special education services more likely or less likely to be placed in a separate special class after CSR had been implemented than they had been previously?

In this fourth and final report, we have added the following analyses:

- ***Additional Achievement Analyses***—What do more recent data reveal about the effects of CSR on achievement?
- ***Teacher Mobility***—Did teachers use the teaching shortage created by CSR as an opportunity to move from urban, high-minority, or high-poverty schools to suburban, low-minority, low-poverty schools? Did teachers in higher grade levels with large classes move to grades K–3?
- ***District Budgets and CSR***—Given California's looming state budget crisis, do superintendents plan to eliminate CSR in some grades?

# Chapter 2

## *Summary of Results*

### Research Methodology

A number of different data gathering and analysis methods were used to evaluate the effects of the CSR initiative. The core achievement data, updated each year of the evaluation, came from the California Department of Education. The initial files had socio-demographic data provided by the students when they took the Stanford Achievement Test (SAT-9), the only component of the California Standardized Testing and Reporting (STAR) program available at the time of our evaluation. We also were able to use California Basic Educational Data System (CBEDS) data to obtain characteristics such as credentials, years taught, and degrees for teachers in reduced and non-reduced size classes. We conducted two surveys of principals and superintendents to assess the impact of CSR on resources and space. Data for our analyses of instructional practices came from three surveys of teachers, the last two of which were linked longitudinally. We also observed third-grade classes, some of which were reduced in size and some of which were not. We selected a stratified random sample of districts (125), schools (625), and teachers (1,500) to make sure our results could be generalized to California. Response rates in 1997–98 ranged from 65 percent for teachers to 88 percent for superintendents. In 1998–99, the response rate for teachers was 80 percent. (We did not survey superintendents in 1998–99.) In 1999–2000, the response rates for teachers was 56 percent. Finally, in the first year of the evaluation only, we surveyed a sample of parents of third graders, some of whom were in smaller classes and others in larger classes.

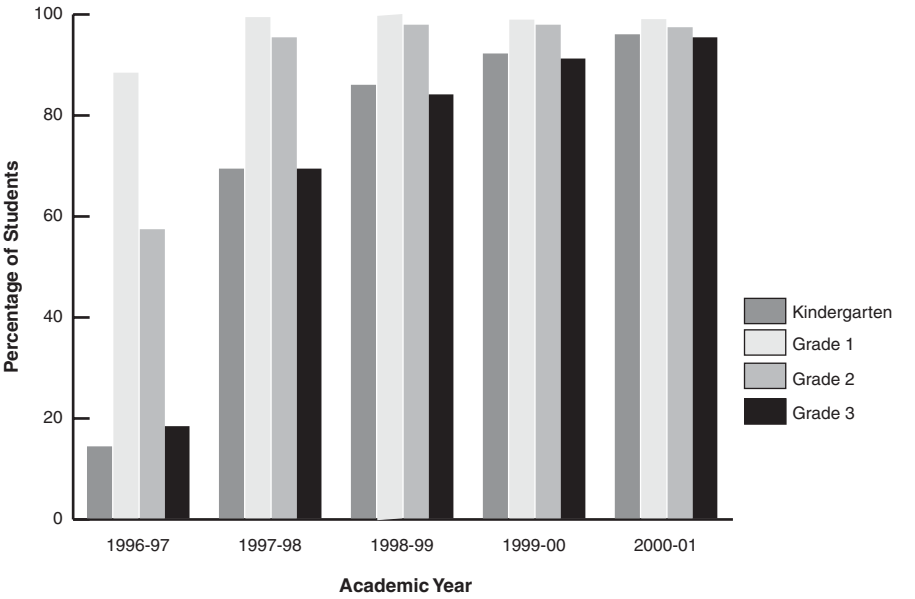
In the balance of this chapter we summarize results drawn from our three previous CSR evaluation reports (Bohrnstedt and Stecher, 1999; Stecher and Bohrnstedt, 2000; and Stecher and Bohrnstedt, 2002), as well as results from supplemental analyses we conducted during the fourth year of the evaluation. Detailed findings from the

fourth-year analyses are provided separately in *What We Have Learned About Class Size Reduction in California: Technical Appendix* available at www.classize.org).

## Implementation Occurred Rapidly, Although It Lagged in Schools Serving Minority and Low-Income Students

Districts reacted quickly to the opportunity presented by the CSR program when it was enacted in July 1996. CSR implementation was virtually complete for first and second grades by the second year of the program, and for kindergarten and third-grade students by the fourth year of the program (Figure 1).

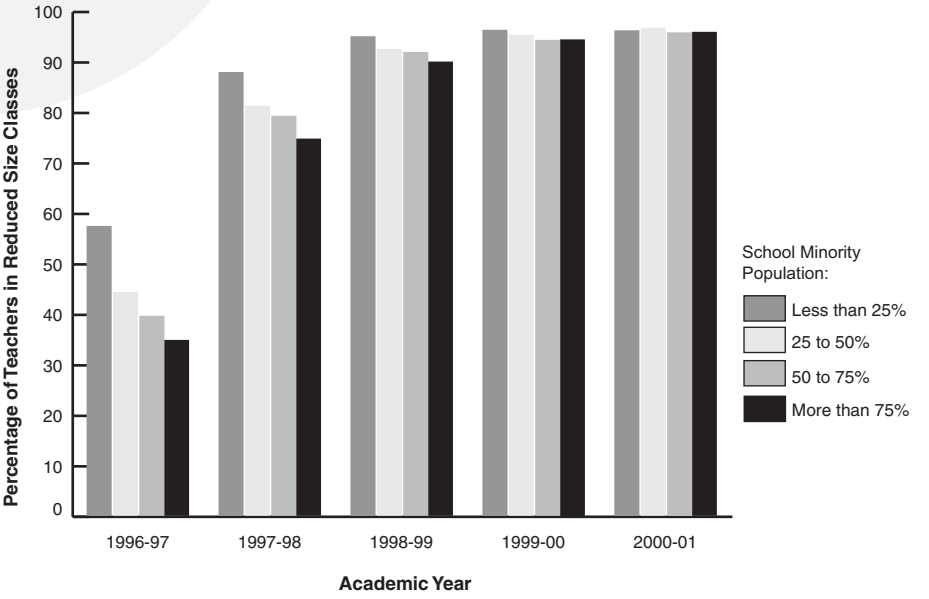
**Figure 1—**
**Percentage of Students in Reduced Size Classes, by Grade Level and Year**




Implementation was not equally rapid for all schools, however. In particular, it lagged in schools with higher proportions of students who might have been expected to benefit most from the program based on the results of the Tennessee STAR project—minority students and students from low-income families. For example, schools with higher percentages of minority students were slower in

reducing class sizes (Figure 2). Schools with higher percentages of students not fully proficient in English and schools with higher percentages of low-income students also were slower to implement CSR. The slowest implementation occurred in schools with a high proportion of Hispanic students. It was not until 1999–2000 that almost all of these schools had implemented the program in kindergarten through third grade.

**Figure 2—**
**Percentage of K–3 Teachers in Reduced Size Classes, by Percentage of Minority Students in Schools**




A number of factors appear to explain these initial differences in implementation. First, as noted in Chapter 1, CSR implementation occurred during a period of rapid growth in student enrollment, and hence a period of much strain on school facilities. Schools in general were crowded, and those serving minority, low-income, and EL students were often severely overcrowded. As a result, it was even harder for schools to find space to expand the number of classrooms. Second, the delay in implementation among schools serving Hispanic students may also have reflected shortages in teachers with cross-cultural and language certifications. Third, most districts reported that operating costs for the program exceeded state payments for it, and this appears to have affected how rapidly CSR was implemented.

Our review of the use of resources revealed that all three factors probably were at work.

## Results on CSR and Student Achievement Are Inconclusive

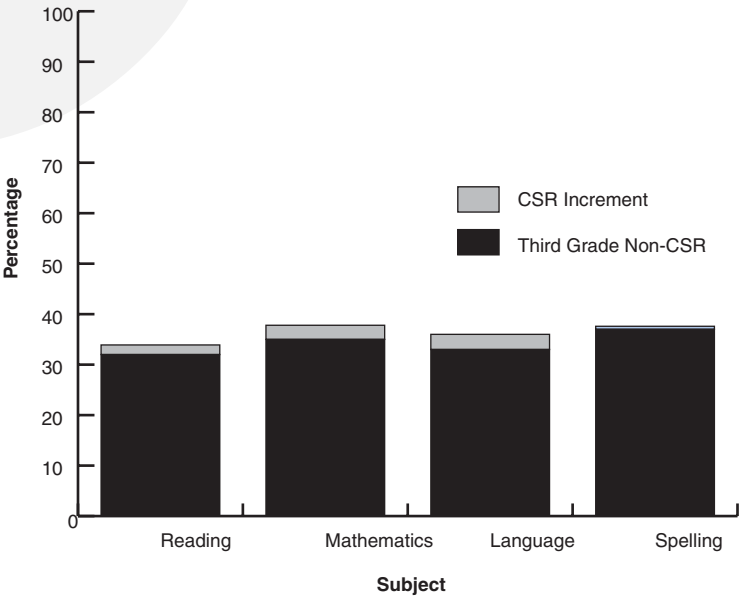
Student achievement has been increasing in California since CSR implementation, but we could find only limited evidence linking these gains to CSR. Whether our inability to link CSR to achievement was due to the weakness of the effect or to design limitations (or a combination of both) could not be determined. The pattern of CSR implementation also limited our ability to determine how much of the recent gain in achievement was attributable to CSR and how much was linked to other initiatives. The latter included new language arts and mathematics standards, the California Reading Initiative, the end of bilingual education, changes in social promotion and retention policy, and a new accountability system. In particular, because of the speed of implementation, we were not able to compare CSR with non-CSR students after 1998–99. Furthermore, California does not include kindergarten or first grade in its testing program, so we were unable to examine achievement in these two important grade levels.

As a consequence, our initial analysis of the effects of CSR on achievement focused on third grade, enough students remained in non-reduced size classes for us to make comparisons. When we controlled for differences between students, we found that third-grade students in schools that had reduced classes scored slightly better on the SAT-9 than did third-grade students in schools where classes were not reduced in size.

We found that there was a positive association in 1998 between third-grade class size and SAT-9 scores after controlling for differences in student and school characteristics. However, the size of this CSR effect was small, particularly when compared to the size of achievement differences related to socio-economic status or race/ethnicity. In reading, for example, the proportion of third-grade students who scored above the national median (the 50$^{th}$ percentile) was 32 percent for those in a typical *non-reduced size* class and 34 percent for those in a typical *reduced size* class. The relationships between CSR and achievement were stronger in language and

mathematics than in reading and spelling (Figure 3). CSR implementation was associated with an additional 3 percent of students (approximately 15,000 students) testing above the national medians in these subjects. Unlike results in the Tennessee STAR project, the impact was felt equally for White and minority students and for high- and low-income students.

**Figure 3—**
**Percentage of Students Scoring Above National Median, by CSR Status and by Subject, on SAT-9 in 1998**




In the following year, 1998–99, these positive differences persisted when students who had been in reduced size third-grade classes moved to fourth-grade classes, which are not part of the CSR program and thus are regular size. The spring 1999 SAT-9 results showed that fourth-grade students who had been in reduced size third-grade classes scored higher than those who had not. The difference between the two groups of students was smaller than it had been in the third grade, but it was still statistically significant.

### New School-Level Analysis Finds No Relationship Between CSR Exposure and Student Achievement

By 2001, CSR implementation was nearly complete, so we could not examine differences in SAT-9 scores between students who were and were not in reduced size classes. Instead, we tracked achievement

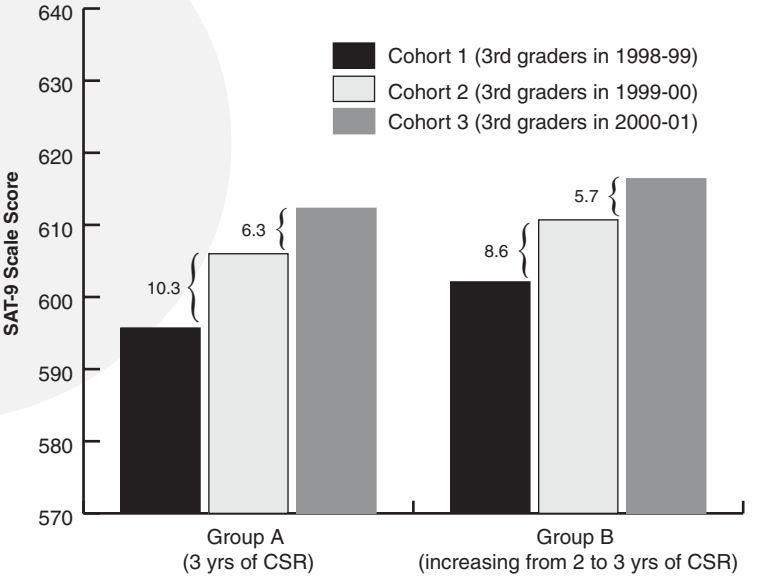
gains in cohorts of students with different amounts of exposure to CSR from kindergarten through third grade. Whether we used statewide average scores or conducted a more refined school-level analysis, we reached the same conclusion. We found no association between small differences in exposure to CSR (i.e., the total number of years a student had been in reduced size classes) and differences in academic achievement.[8] However, the observed differences in CSR exposure were small, most often equal to only 1 year.

In the more detailed, school-level analysis, we compared students in two groups of schools with similar student populations. In each group, we looked at second- and third-grade test scores from three successive cohorts of students: those who entered kindergarten in 1995–96, those who entered 1996–97, and those who entered 1997–98. In the first group of schools, group A, the three cohorts had the same level of CSR exposure; all students had reduced size classes in grades 1, 2, and 3. In the second group of schools, group B, the three cohorts had different exposure to CSR: Cohort 1 had two years of reduced size classes in third grade. Cohorts 2 and 3 had reduced classes in grades 1, 2, and 3.

Test scores increased over time in both groups of schools (Figure 4). Furthermore, the increase was almost exactly the same even though CSR exposure was constant in group A and increased in group B. In group A, Cohort 2 scored 10.3 points higher than Cohort 1, and Cohort 3 scored 6.3 points higher than Cohort 2. The comparable numbers for group B were 8.6 and 5.7, respectively. If CSR had an impact on scores, we would likely see greater increases in group B than in group A, but the added year of CSR did not change the pattern of achievement. In fact, the increase in scores in group A cannot be explained by CSR at all, because each successive cohort of students had exactly the same exposure. The results were similar when we examined reading and language scores and when we focused the analysis on schools with high percentages of minority students.

[8] Results derived from an analysis of state average scores are reported in Stecher and Bohrnstedt (2002). Results based on school-level analyses are reported in *What We Have Learned About Class Size Reduction in California: Technical Appendix* (available at www.classize.org).

**Figure 4—**
**Third Grade SAT-9 Scores in Mathematics in Schools with Similar Students but Different CSR Exposure**




The conclusions about the effect of CSR on academic achievement that can be drawn from 4 years of analysis are quite limited because of the rapidity with which CSR was implemented, gaps in the state testing program, and other research design considerations. For many people, the lack of a clear relationship between CSR and student achievement will be disappointing. The overall effects we did find were smaller than those found in the Tennessee STAR experiment, and we did not find greater effects among disadvantaged students. When comparing the two studies, it is important to remember that the treatment differences we measured in California were weaker than those measured in Tennessee. The Tennessee STAR research compared students in regular classes for 4 years (kindergarten through third grade) with students in reduced size classes for 4 years. We were comparing students whose exposure to CSR only differed by 1 year. It is also important to remember that there were substantial differences in the general educational context, student demographics, and specific class size reduction in the two state programs, as we noted in Chapter 1.

In 2001, we also collected and reviewed studies and formal evaluations that local school districts had conducted to assess the effects of CSR in their district.[9] We reviewed 14 formal reports and 20 informal letters, electronic mail messages, and phone responses that were received in response to a request sent to approximately 880 districts. The findings were generally positive, but the analyses were rarely rigorous. The districts had to address the same analytic challenges that confronted the statewide evaluation, and they had fewer analytic and statistical resources to bring to bear on the problems. These reports suggest that CSR may have been more effective in some districts than others, but because the studies varied in quality, they do not cause us to reconsider our conclusion about our overall statewide and school-level results.

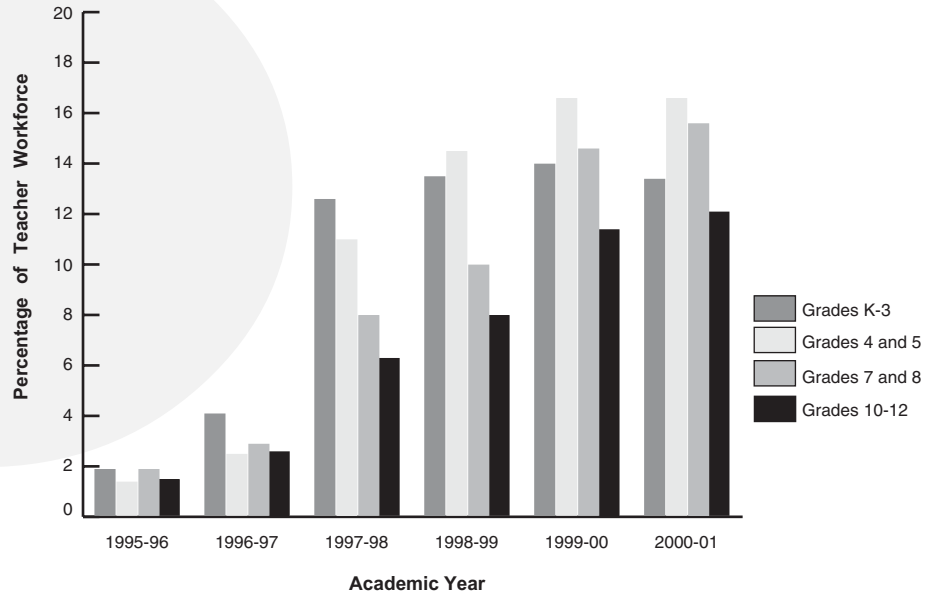
## CSR Was Associated with Declines in Teacher Qualifications and a More Inequitable Distribution of Fully Qualified Teachers

CSR implementation required an enormous increase in the number of K–3 teachers in California. Between 1995–96, the year before CSR implementation, and 1998–99, the third year of the program, the total number of K–3 teachers increased 46 percent, from 62,226 to 91,112. Since 1998–99, the number of K–3 teachers has remained relatively stable.

To meet the increased demand for teachers, many districts hired teachers who were not fully credentialed. As a result, the proportion of K–3 teachers who were not fully credentialed (e.g., teachers with intern or emergency credentials) increased from 1.8 percent before the program started to 12.5 percent in the second year of the program (Figure 5). The proportion of teachers without full credentials also increased rapidly in the upper elementary grades and at a somewhat slower rate in the secondary grades.

[9] The results are reported in *What We Have Learned About Class Size Reduction in California: Technical Appendix* available at www.classize.org.

**Figure 5—**
**Percentage of Teachers Not Fully Credentialed, by Grade and Year**

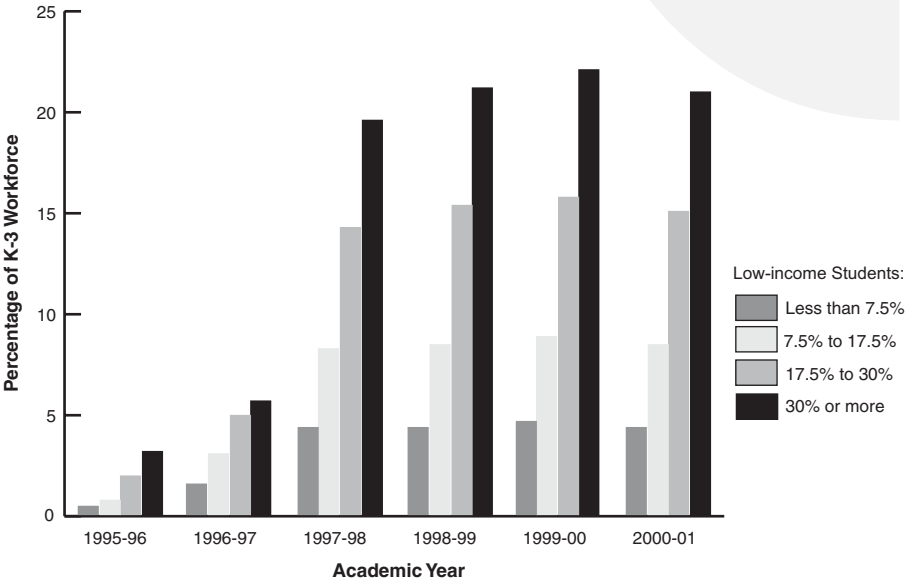



**Note**: We present no data on grades 6 and 9 because of the varying placement of these grades in California schools. Some districts place grade 6 in middle rather than elementary school, just as some place grade 9 in middle rather than high school. For more information on California credentials for different grade levels, see http://www.ctc.ca.gov/credentialinfo/credinfo.html (as of March 29, 2002).

Teachers without full credentials have become far more common in some types of schools than in others. In 1995–96, the year prior to CSR, there was very little difference among schools in the proportions of K–3 teachers who were not fully credentialed, because such teachers were rare. Even in the most disadvantaged schools, fewer than 4 percent of K–3 teachers lacked full credentials. By the third year of CSR implementation, however, K–3 teachers who were not fully credentialed had become much more common in the most disadvantaged schools, while remaining rather rare in other schools. For example, by 1998–99, K–3 teachers without full credentials increased to 21.2 percent in schools with at least 30 percent low-income students, but to only 4.3 percent in schools with less than 7.5 percent low-income students (Figure 6). California has enacted a number of programs to address the teacher shortage, and since 1999, the percentage of K–3 teachers lacking full credentials has decreased

slightly (Commission on Teacher Credentialing, 2002).[10] However, the number of such teachers remains high, and they continue to be distributed unevenly across schools. In 2000–01, more than 1 in 5 K–3 teachers were not fully credentialed in schools (primarily large and urban) with high percentages of low-income, EL, minority, or Hispanic students. Similar differences in teacher credentials were observed in the higher grades.[11]

**Figure 6—**
**Percentage of K–3 Teachers Not Fully Credentialed, by Low-income Students in School**




**Note**: These categories were selected to divide schools in the state into four roughly equal groups.

Credentials are just one indicator of teacher quality. In recent years, the gap between disadvantaged and other schools also increased on other indicators of teacher qualifications, though not as much as it did in teacher credentials. For example, in the most economically disadvantaged schools in 2000–01, about 1 in 4 K–3 teachers had

[10] California has also changed its credential system, and recent reports from the Commission on Teacher Credentialing indicate a decline in emergency permits in 2000-01.

[11] With cooperation from six districts, we conducted a supplemental study of teacher qualifications and student achievement gains. The results showed no consistent relationships between qualifications and student gains, but the study had a number of limitations. Findings are reported in *What We Have Learned About Class Size Reduction in California: Technical Appendix* (available at www.classize.org).

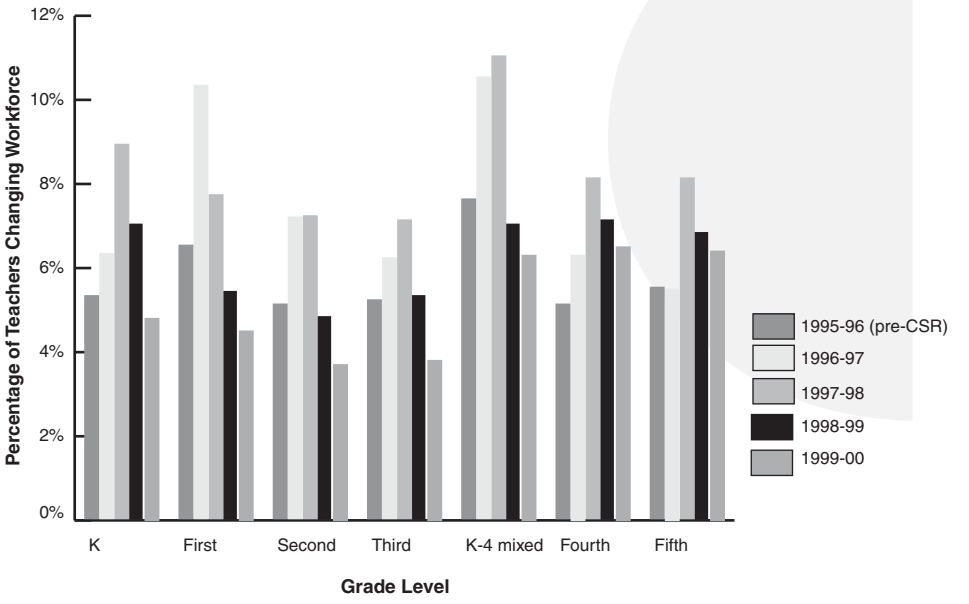
3 or fewer years of experience, compared with fewer than 1 in 5 teachers in the least economically disadvantaged schools. And about 1 in 3 teachers in the most economically disadvantage schools had fewer than 30 units of credit beyond a bachelor's degree (the category used for reporting), compared to about 1 in 6 in the least economically disadvantaged schools. The same patterns were found for teachers in the fourth and fifth grades.

### New Analyses Show That CSR Had Only a Modest Effect on Teacher Mobility

We found little evidence of a sudden, large exodus of qualified teachers from one type of school or district to another in the wake of CSR.[12] For example, approximately 7 percent of first-grade teachers in 1995–96 (the year prior to CSR) had been teaching in a different school the previous year. That percentage rose to 11 percent in 1996–97 and had dropped down to 5 percent by 1999–2000 (Figure 7). The same pattern was true in the other elementary grades. That is, the inter-school transfer rate was small, especially when compared to the 46 percent increase in the number of K–3 teachers during this period.

[12] These results are reported in *What We Have Learned About Class Size Reduction in California: Technical Appendix* available at www.classize.org.

**Figure 7—**
**Percentage of Elementary Teachers Who Changed Schools from Prior Year**




The picture was similar for intra-school transfers. For example, 17 percent of first-grade teachers in 1995–96 had taught in the same school but at a different grade level the previous year. This percentage rose to 21 percent in the first year of CSR but then dropped to its earlier level.

Overall, these analyses suggest that the gap in teacher qualifications among schools increased because schools serving low-income, minority, or EL students were less able to hire qualified teachers to fill their new positions, rather than because of a rapid increase in transfers among qualified teachers already in the schools.

## CSR Implementation Did Not Affect Special Education Identification or Placement

To our knowledge, previous research and evaluations have not examined the effects of CSR programs on students receiving special education services. There are about 150,000 K–3 students in special education in California. There was some concern that reducing class size might affect the number of students referred for special education

assessment, the number of students identified as needing special education services, or the number of special education students placed in special day classes (instead of in general education classrooms). On the one hand, smaller classes afford teachers more opportunity to observe student behavior, which might lead to increased referrals to special education. On the other hand, smaller classes provide more opportunity for teachers to address individual differences, which might lead to decreased referrals to special education. Interviews with special education directors indicated the referral rate for special education assessments increased. Our analysis of statewide enrollment data showed no evidence that identification rates or special day class placement rates were affected by CSR.

## Students in Reduced Size Third-Grade Classes Received More Individual Attention, But Similar Instruction and Curriculum

According to survey responses, teachers in reduced size classes provided more individualized instruction for their students than did teachers in non-reduced size classes, but they did not report differences in what they taught or how they presented it.[13] Case studies of 16 teachers yielded similar results, but further study of the classroom practices data would be beneficial. Previous research on reduced size classes has offered little insight into why smaller classes might yield better achievement, so we examined a range of variables related to teaching, including allocation of class time, curriculum content, classroom organization, specific teacher and student activities, and student behavior.

In all three prior years of our evaluation, teachers of third-grade reduced size classes reported more sessions in which they provided individual instruction of 5 minutes or more to students having difficulties in reading. There were other indications that smaller classes were more personal, although not all of these were evident in each study year. Teachers of reduced size classes were more likely to say they

[13] In conducting these analyses, we controlled for teacher education, experience, and credential status, as well as for student demographic characteristics.

- Know what each student knows and can do
- Provide feedback on writing assignments within 1 day
- Give more individual attention to students
- Meet the instructional needs of all students

Conversely, teachers in non-reduced size classes were more likely to say that they would like to give more individual attention to their students but do not have time to do so, and that they find it hard to meet the instructional needs of all their students.

Teachers in reduced size classes reported fewer behavior problems among students and that students were more likely to complete the lesson for the day. They also reported that students were less likely to compete with one another for the teacher's attention, to engage in exclusionary behavior, to disrupt the work of other students, and to be "off task" for more than 5 minutes.

However, teachers in both reduced and non-reduced third-grade classes reported spending similar amounts of time and covering similar amounts of curriculum in language arts and in mathematics. For example, both groups devoted about 1 hour each day to mathematics and about 1.5 hours to language arts. Both groups assigned similar amounts of homework as well, and both spent similar amounts of time covering each of the mathematics and language arts topics we listed. Teachers of reduced size classes were somewhat more likely to say they had enough time to explore curriculum topics fully, but overwhelming majorities of both groups of teachers said they had "to hurry all year long just to cover the basic things" students need to know.

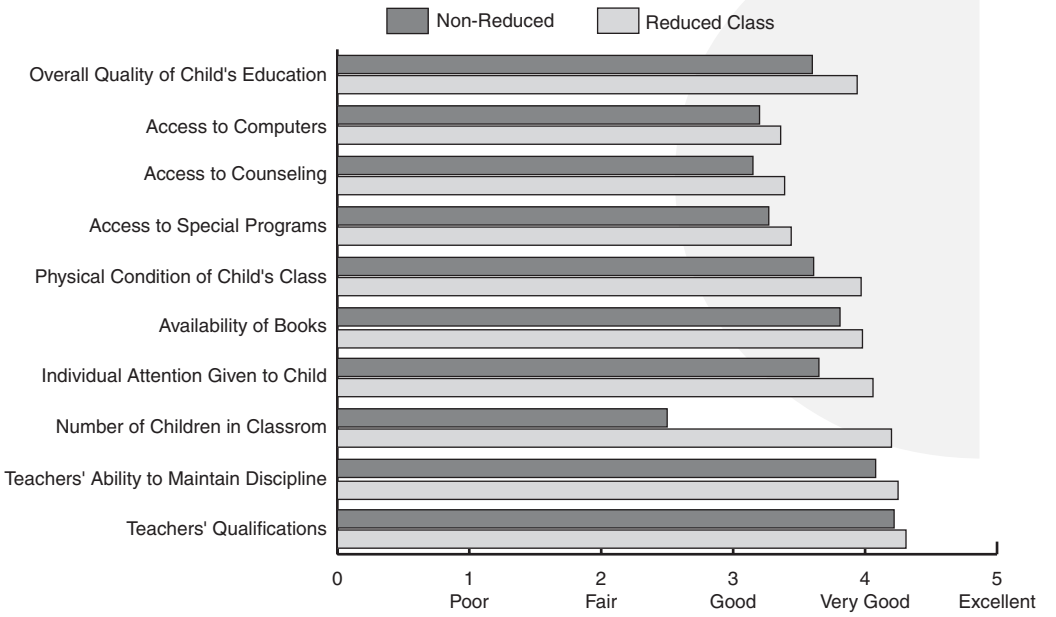
For the most part, teachers in reduced and non-reduced classes reported using selected instructional activities with equal frequency. In mathematics, these activities included using a calculator, using mathematics in the context of other subjects, doing mathematics worksheets, using patterns to discover mathematical relationships, and practicing computational skills. In language arts, the activities included having guided discussion about reading, discussing new or difficult vocabulary, working in a reading book, and listening to the teacher read stories. There were a few exceptions for teachers in reduced classes:

- In mathematics, they were more likely to have students work with measuring instruments, play with math-related games, and work with manipulative aids.
- In language arts, they were more likely to have students read aloud to a partner, work on phonics, and write narrative or descriptive material for which they were encouraged to use invented spellings, if needed.

## Parents Rated Reduced Size Classes Higher on Satisfaction Factors

The reform was widely praised by parents and teachers. Parent opinions about CSR were surveyed during the first year of the evaluation. With only one exception (teacher qualifications), parents of third-grade students in reduced size classes rated selected features of their child's education higher than did parents of children in non-reduced size classes (Figure 8). The differences in rating of class size were particularly pronounced, with parents of children in reduced classes reporting satisfaction levels far higher than those of parents of children in non-reduced classes. However, parents of children in both reduced and non-reduced classes expressed equal satisfaction with the qualifications of their children's teachers.

**Figure 8—**
**Third-Grade Parent Ratings of School Characteristics, by Child's Class Size**



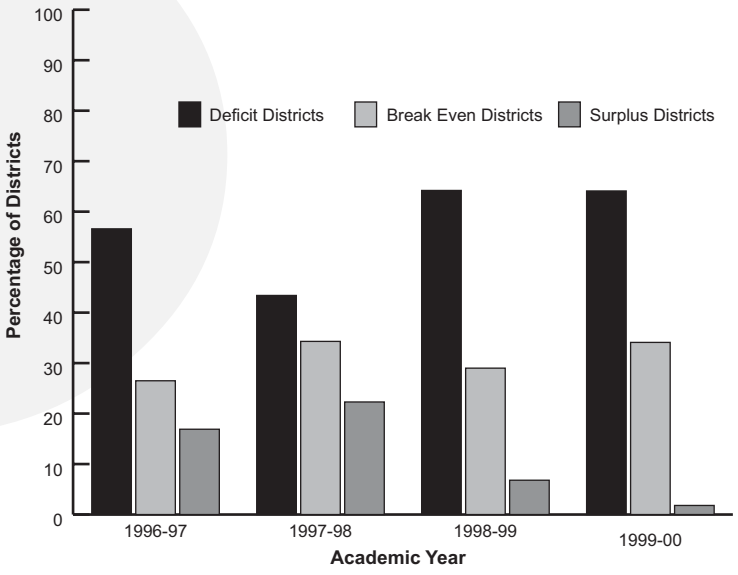

**Note**: All differences between ratings are statistically significant except in the case of teacher qualifications.

Parental reports of contacts with their child's school also favored reduced size classes to some extent. Seventy-four percent of parents of children in reduced size classes reported initiating contacts with their child's teacher, compared to 69 percent of parents with children in regular classes (a difference that was statistically significant). Other measures of involvement, such as volunteering in class, showed no difference.

## Classroom Space and Dollars Were Taken From Other Programs to Support CSR

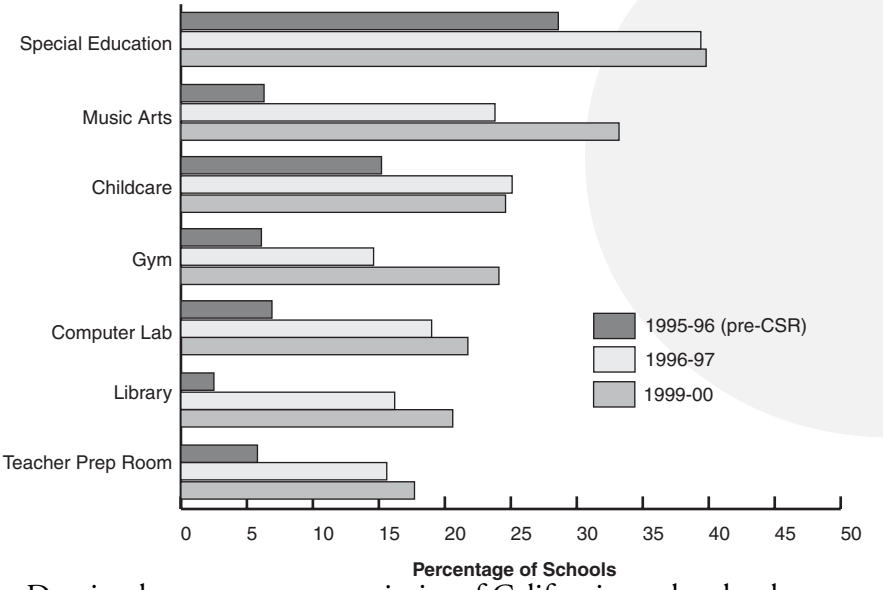
Most districts in our statewide sample reported incurring operating costs for CSR that exceeded state payments for it (Figure 9), and these funding problems persisted or even worsened in subsequent years. Most districts reporting a CSR funding shortfall attempted to overcome it by reducing funds for facility maintenance and administrative services. About one-third of such districts also reduced resources for professional development, computer programs, or libraries.

**Figure 9—**
**District Reports of CSR Funding Adequacy**




Although few districts reduced *funding* for after-school care, child care, or special education programs to compensate for CSR funding shortfalls, many districts reduced *classroom space* for such programs to make space available for CSR. Overcrowding placed a premium on space in many schools even before CSR. For example, in 1995–96, the year before CSR implementation, growing student enrollments had forced nearly 30 percent of California schools to reallocate full-sized classrooms that had been designated for special education, forcing special education classes to use alternative spaces (Figure 10). CSR implementation also accelerated the preempting of space from a variety of other uses, including music and arts, athletics, and child care programs. Given these space and funding constraints, it is not surprising that slower-implementing schools most often cited space and funding as reasons for the delay.

**Figure 10—**
**Percentage of Schools Preempting, for Classroom Use, Space Designated for Other Purposes, by Academic Year**




Despite these pressures, a majority of California grade school principals and superintendents reported that they would not favor reallocating CSR funds to other educational programs, in almost all cases. One notable exception: Most principals and many superintendents said they would divert or reallocate CSR funding to upgrade teacher training if they were given this option.

## In Spite of Budget Shortfalls, Few Districts Project CSR Cutbacks for 2002–03

The state of California is projecting a significant budget deficit for 2002–03, and many districts are forecasting total revenues that will not meet projected expenses. News reports indicate that a few districts are contemplating elimination of CSR in some grades to save money on salaries. To investigate this issue, we conducted a telephone survey of 38 districts around the state in April 2002.[14] Most districts indicated that they were facing budget shortfalls for the upcoming

[14] The telephone survey was conducted prior to the announcement of the governor's proposed budget for 2002–03. The results are reported in *What We Have Learned About Class Size Reduction in California: Technical Appendix* available at www.classize.org.

year but did not intend to reduce CSR participation. Some indicated, however, that such a reduction was possibility and would continue to be discussed as their budgets were developed.

## Study Limitations Are Important for Interpreting Findings

As we noted in our previous CSR evaluation reports, limitations in the data, the research design, and the analyses reduce the certainty one can attach to our findings. It is important to reiterate a few of these limitations. More complete discussions can be found in the previous CSR evaluation reports and in Technical Appendix supplement to this report.

First, the reader is reminded that our results apply to class size reduction as implemented in California. As we pointed out in Chapter 1, the California CSR program is different from class size reduction efforts in other states in a number of important respects, so our results do not generalize to class size reduction in general. In particular, reduced size classes in California contained 20 students; in Tennessee, they contained 13–17 students. Second, the scale of California's program led to changes in the underlying conditions of education that had an influence on the results. In particular, the demand for space exceeded the supply, leading to uneven implementation, and the demand for teachers exceeded the supply, leading to the use of not-fully-credentialed teachers. These two sets of conditions interacted and led to unequal distributions of fully-credentialed teachers across schools. Achievement results must be understood in light of these unanticipated consequences. Third, the reform was implemented rapidly, leaving few opportunities to compare practices or outcomes in reduced versus non-reduced classes. Finally, the current state testing program does not include students in kindergarten or first grade and did not begin until the second year of CSR, further limiting our ability to assess the program's effects on student achievement.

It is also important to keep in mind that this evaluation did not examine numerous student and teacher outcomes that many would see as valuable and that might be important consequences of CSR. These include student motivation and engagement with school, student development of interpersonal skills and a sense of self, and a

teacher's ability to better meet the needs of individual students who are struggling.

In spite of these caveats, the most straightforward conclusion is that California's CSR program had at best a small positive effect on student achievement. Perhaps the results would have been more positive if the key conditions in the Tennessee STAR experiment (described in Chapter 1) had existed in California—for example, an available supply of fully certified teachers, especially in schools serving the most disadvantaged students.

# CHAPTER 3

## *Policy Lessons and Recommendations*

### Context Within Which CSR's Future Must Be Considered

California's CSR program represents the largest single reform effort in the state's history and is by far the largest class size reduction program that any state has undertaken. Following its introduction in 1996, it was implemented statewide with extraordinary rapidity, and it currently costs the state upwards of $1.5 billion a year.

As discussed in Chapter 2, we found that the possibilities for using student achievement data to measure the value of the state's investment are quite limited. While student test scores have continued to improve in the elementary grades in California since CSR implementation, our evaluation, for a variety of reasons, is not able to determine what portion of those gains may be attributable to CSR. Regardless, CSR is widely popular with parents and school staff, many of whom report marked improvements in classroom learning environments and working conditions for teachers. As also discussed in Chapter 2, other important student outcomes that may be affected by CSR were not measured.

Before examining the policy implications to be drawn from our findings, it is important to consider how things have changed in the 6 years since 1996 when CSR began.

#### California's Policy Context Has Changed Since 1996

CSR can be viewed as a categorical aid program, albeit a very expensive one, of the kind that began dominating the education policy landscape in the 1960s.[15] California's proclivity for categorical programs has resulted in the allocation of almost one-third of Proposition 98 funds for specified purposes.[16]

---

[15] Categorical programs generally provide earmarked funding that is allocated to school districts for specified or restricted uses and often comes with extra reporting and regulatory requirements.

[16] California's Proposition 98 is a constitutional guarantee that sets an annual minimum funding level for K–12 education.

Since CSR's 1996 adoption, however, there has been a major shift in policy focus for education in California. The state has started moving toward a systemic standards-based system, with a strong emphasis on high expectations, accountability, and accompanying rewards and sanctions based on growth in student achievement.

The philosophy behind standards-based reform calls into question the use of proscriptive categorical programs. The theory argues that the state should centralize policy on academic standards and student performance outcomes but decentralize policy on implementation. In other words, the state should decide "what" students need to know and be able to do, and local educators should decide "how" they will provide instruction and services to help students meet those state expectations.

The second major difference between the education policy environments of 1996 and 2002 is the condition of the California economy. In 1996, the problem was how to expend a surplus of state dollars on education as mandated by Proposition 98. CSR was a popular choice. In 2002, and for the next few years, the problem is reversed: How does the state absorb a $23.5 billion deficit (on a $147 billion base) and still maintain its commitment to education? The sudden drop in resources since 2001 has dramatically increased the pressure to show evidence that costly education programs are indeed leading to improved student outcomes.

## Experience in Other States Provides Additional Perspective

Another change in context arises from the fact that since 1996, research in other states (described in Chapter 1) has provided more information on the effects of class size reduction. Outside California, at least 20 states and a number of local school districts now have class size initiatives. Moreover, the federal government launched its own class size reduction program in 1999, targeting high-poverty schools. Though the evaluations of state and local efforts are few and often limited, the findings that do exist include have important implications for California's policy.

The CSR policies and programs that have shown the most relevant results are in Tennessee; Wisconsin; Utah; Austin, Texas; and Burke

County, North Carolina. Each of these jurisdictions has seen reduction program with one or more of the following key features:

- Restricts class size to approximately 15 students
- Targets the poorest students
- Is augmented by and integrated with other supportive reform strategies

Tennessee reduced its classes to 13–17 students, substantially smaller than is the case in California. In Wisconsin, targeting the neediest students with well thought-out and fully funded programs has led to good results. In Wisconsin, as well as in Utah, Austin, and Burke County, the evidence indicates that class size reduction has the greatest positive effect on student achievement when it is integrated with a more comprehensive set of initiatives aimed at academic improvement. (See Chapter 1 for details.)

Today, California faces some crucial questions. Should CSR remain a stand-alone categorical program or should it be integrated into the state's larger, overarching K–12 systemic standards-based reform? How does CSR fit into the current state policy context, compared with that of 1996? If state policy is projected 5 to 10 years into the future, what would be the desired outcomes from CSR? What changes in the program would lead to these outcomes? How can the quality of the data collected and the research conducted be improved to answer these and other important education policy questions?

## Consortium Recommendations to California Policymakers

CSR is an enormously popular program in California, especially among parents and teachers, as recent public outcries about possible cutbacks have made clear. In addition, it is important to note that even now, with full implementation of CSR, California's elementary student-teacher ratio remains among the highest of any state.[17] It is also clear that local educators and parents may value reduced class sizes for many reasons other than improvements in achievement that can be measured by statewide test scores. Therefore, maintaining small K–3 classes in California is likely to remain a priority.

[17] In addition, California's middle and high school student-teacher ratios are the highest in the nation.

Nonetheless, based on our evaluation and on research done in other states, we believe that some changes to the program should be considered.

**Recommendation One:**
*Improve the effectiveness of the current CSR program by integrating and aligning it with other reforms.*

We are particularly impressed with the need to link CSR to the state's overall strategic direction—i.e., to end its current status as a stand-alone categorical program by integrating and aligning it with the state's standards-based policies. Such a shift, we believe, would allow CSR to better support the standards-based reform strategy and might prompt better results from the CSR investment. Schools may be able to use other elements of standards-based reform—e.g., additional funding to turn around low-performing schools—in ways that allow them to take full advantage of the opportunities small classes have been shown to present in some states, especially for low-income and minority students. In short, integrating CSR with the state's evolving standards-based reform policies could significantly bolster California's ability to meet its objective of improving student achievement.

Other state and federal education reform initiatives are moving in a similar standards-based policy direction. The legislature's Joint Committee for a K–16 Master Plan is considering a pre-K–16 policy consistent with this thrust. The plan would integrate preschool/child care, elementary, secondary, and postsecondary education. It calls for a better integration of programs and initiatives within each segment, all with the common goal of improving student achievement, broadly conceived. Furthermore, the state's legislative analyst has proposed a dramatic consolidation of California's categorical programs (California Legislative Analyst's Office, 2002) into more generic groupings, such as "academic improvement" and "teacher support and staff development," that would mesh better with standards-based reform. At the federal level, Congress passed a major overhaul of its elementary and secondary education programs that expanded California's Title I allocation and loosened some of the categorical strings, while also linking funding to continuous improvement in student achievement.

**Recommendation Two:**

*Be explicit about the assumtions underlying state reimbursement of CSR and take steps to determine the real costs and cost-effectiveness of CSR.*

School administrators often mention unreimbursed local costs as a barrier to maintaining the existing CSR program. Surveys of district superintendents and principals found a large and growing number of schools reporting that CSR revenues did not adequately fund the program. However, our interviews with Sacramento legislative and executive branch staff found a degree of skepticism regarding these local school reports of inadequacy.[18] Moreover, state officials said that not only does CSR receive a separate cost-of-living adjustment (COLA), but that the state provides general-purpose funds above and beyond growth and COLA that could be used to offset the increased costs of CSR.

Much of the disagreement stems from a fundamental difference in the way CSR is viewed by state policymakers and school district personnel. State officials describe the program as an incentive program, not a state mandate. As such, they argue that districts have the option to participate or not. They further argue that the state provides adequate resources through the combination of CSR funding and general purpose funding. Many districts, however, feel that the state indicated an intent to fully fund the program when it adjusted the funding in 1997. State support has not kept up with costs since then, and districts leaders believe that the state should once again provide adequate resources for full funding. Regardless of whether or not costs are fully reimbursed, the rules regarding appropriate cost attribution should be explicit, and districts ought to have reasonably predictable revenue streams so that they can make informed choices about implementing CSR.

Determining costs attributable to specific programs is not a simple matter. Does one ascribe increases in salaries that districts give all teachers as a legitimate CSR cost? If districts choose to assign more-experienced (and hence more expensive) teachers to grades K–3, should the cost be reimbursed through CSR? A careful cost review

[18] As a part of this study, PACE conducted interviews with selected legislative and executive branch staff regarding the passage and implementation.

could illuminate this issue and lead to a single set of rules relating to cost attribution. Better expenditure data derived from a single set of assumptions will shed more light on this contentious issue.

Even more importantly, having solid cost data would assist state and local policymakers in determining the cost effectiveness of CSR compared with other possible reforms. The state knows precisely the annual cost it incurs from CSR, because it is a line item in the governor's budget. What is not known is the extent to which CSR encroaches, if it does encroach, on a district's general-fund budget. If the district cannot avoid taking funds from other programs to pay for CSR, those funds should be documented by the district and acknowledged by the state as part of the actual cost of CSR. Along the same lines, however, if CSR funding is adequate to fully fund the program, that too should be known.

Even if the state does not fully fund CSR costs, policymakers need to be aware of the real costs and benefits before moving to expand or reduce CSR. The development of more accurate cost data would provide the state and districts with important information about next steps in the ongoing effort to improve student achievement in the most cost-effective way.

**Recommendation Three:**
*Provide more local flexibility within the current CSR program by allowing a schoolwide average of 20 students in grades K-3.*

Along the lines suggested by the Legislative Analyst's Office — and consistent with the recommendations we made in our previous CSR evaluation reports— local districts should be given the flexibility to vary class size by up to 2 students per class as long as the class size average within a school remains at 20 students or fewer.

Having the class size cap of 20 apply to a school rather than to each class within a school would give schools a modicum of additional flexibility while only modestly affecting the way the limit is applied. The current lack of flexibility has had a particularly onerous impact on smaller schools, which have fewer options for assigning students. For example, if a school has only 42 third graders, we would argue that making two classes of 21 each would probably be sounder educationally than placing some students in a combined second- and

third-grade class (assuming, of course, that there is not a similar problem in the second grade). Schools currently face the dilemma of losing their CSR money by not adhering to the stringent requirements. Providing modest flexibility to meet unique local needs would not substantially alter the current CSR program's purpose or its cost to the state.

Again, if CSR is placed in the overall context of standards-based reform aimed at improving student performance, districts need reasonable flexibility in allocating general and categorical funds from various resources, both state and federal, to achieve that goal. CSR should be a means to an end, not the end in itself.

**Recommendation Four:**

*Further test CSR's potential to improve the achievmeent of low-income/minority students by providing additional resources to create and evaluate alternatives with even smaller class sizes in selected schools.*

The best scientific evidence available for the positive effects of class size reduction on the academic achievement of low-income/minority students comes from the Tennessee STAR study (see Chapter 1). In that study, in which class size was reduced to an average of 13–17, minority and low-income students showed gains in student achievement two to three times as great as those of White and non-urban students. Wisconsin's statewide SAGE program—which targeted schools and districts with high poverty rates—combined class sizes reduced to 15 in grades K–3 with a rigorous academic curriculum, before- and after-school activities, and strong professional development for teachers. The significant positive results corroborated those found in the Tennessee STAR studies, with minority students benefiting most. Research indicates that the added advantage of reduced size classes for minority and low-income students continues into high school as well (Krueger and Whitmore, 1999).

Based on this evidence, it appears that class size reduction can be an especially effective strategy for raising the achievement of the most at-risk students *if* the classes for this group are small enough and are staffed by skilled and qualified teachers. Aligning these efforts with other standards-based reforms and other support programs would likely also increase these positive effects.

K–3 class sizes in California are now almost uniformly 19–20 students. As a result, it is possible to conduct carefully controlled experiments to examine the difference made by moving to a class size of 15 or fewer, beginning with schools that serve the largest number of low-income and minority students. The state should undertake controlled field experiments to evaluate the costs and benefits of smaller classes for students most in need.

**Recommendation Five:**

*Provide incentives to a smaller number of districts to experiment so that cost-neutral alternative class size reduction strategies can be tested and evaluated.*

One of the major difficulties in evaluating California's CSR program was the fact that almost all school districts in the state reduced class sizes in a similar way over a similar time period. Such a statewide implementation does not provide researchers with a good comparison group for evaluating the CSR program. If the state were to allow a relatively small number of school districts to use their CSR funding to create randomized trials of other small class size arrangements, it could compare the effectiveness of the current CSR program with alternative class size reduction designs. Participating districts would be required to randomly assign schools, or classes within schools, to the current CSR program structure or an alternative model. Researchers could then track the changes in student achievement for the alternatives. The state should also consider allowing districts to compare one or more non-CSR uses of the funds against the current CSR model, again with the requirement that this be done using randomized trials.

Both of these options—the current CSR model against other models, and the current CSR model against alternative uses of CSR money—have the virtue of providing information about the cost-effectiveness of the alternatives (since all would have the same cost), something that could not be done as part of our evaluation. The major incentive to districts to participate would be the ability to design their class size reduction or other programs to meet local needs. A second incentive would be additional state funds for participating districts for technical assistance in putting together the research design and for evaluating the effects of alternative uses of CSR money. This strategy would also benefit the state in that it would support the development of a corpus of knowledge about which uses of CSR money are the most effective and under what conditions.

The following list gives examples of possible alternative class size reduction interventions the state might consider. Each is consistent with the available research and could be implemented by districts at a per-pupil cost similar to that of the current CSR program.

- **Target Specific Teacher Populations.** Some research suggests that student achievement is lower in classes whose teachers have fewer than 3 years of teaching experience. Further, teaching larger classes is presumed to be more difficult than teaching smaller classes, especially for newly credentialed teachers. The state could allow a school or district to randomly assign one half of its credentialed teachers with fewer than 3 years of experience to K–3 classes of fewer than 20 students (say, an average of 18) and the other half to classes of more than 20 (say, an average of 22). The school or district would still be required to maintain a schoolwide student-teacher average of 20:1 with no class be over, say, 24 students.
- **Kindergarten and First Grade at 15 Students per Teacher**. Some research from other states suggests that class size reduction has the greatest impact in kindergarten and first grade. For the same aggregate level of funding, schools could provide class sizes of 15 or fewer in kindergarten and first grade, and 25 or fewer in second and third grades. In this case, half the schools or classes (within grade level) would be randomly assigned to the current 20-student limit per class, and the other half would be assigned to the appropriate alternative.
- **Schoolwide Class Size Reduction.** We estimate that with the same amount of funding, a K–5 elementary school could reduce class size to 23 for all grades, and a K–6 elementary could reduce class size to 24. In this case, half the elementary schools in a district would be randomly assigned to the current CSR program, and the other half would implement schoolwide class size reduction.
- **Target Specific Student Populations**. As noted above, some research suggests that socio-economically disadvantaged students benefit more from class size reduction than do other students. The state could allow a district to reduce K–3 class size to below 20 for socio-economically disadvantaged students by allowing half of the classes to average 18 students

and the other half to average 22 (but no more than 24). This could be done by randomly assigning either schools or classes to one or the other condition. The schools would be required to maintain a district or school student-teacher average of 20:1 to ensure cost neutrality.

**Recommendation Six:**
*Further explore why and how class size reduction works by identifying best instructional practices in small classes.*

Not enough is known yet about why class size reduction works, which means we can offer little guidance about how to make it work better. More research is needed to understand which classroom practices are the most effective in small classes and whether these differ from best practices in larger classes. Our research describes what happened in California, but not what should have happened. For example, we found little difference in content coverage and instructional practices between reduced and non-reduced classes. However, we did find that teachers in reduced size classes were more likely to provide extra attention to students having problems in reading. Teachers in small classes also reported fewer disciplinary problems than did teachers in large classes. In short, we know that reduced size classes had some effects on instructional practice in California, but we do not know what type of changes in classroom teaching would be needed to maximize the benefits of the reform.

A better research base is needed. Once there is a better understanding of what works, it can be incorporated into training programs for new teachers and into professional development courses for experienced teachers.

**Recommendation Seven:**
*Before undertaking any statewide effort to expand CSR to additional grades, ensure that there are sufficient facilities and qualified teachers.*

Many of the findings in our previous reports focused on the CSR program's unintended consequences in relation to teacher quality and facilities. The state has taken substantive action to remedy the problems created by K–3 CSR and to address these issues more generally. There has been a thoughtful, concerted effort in recent years to establish new policies related to teacher preparation, credentialing, recruitment, and retention. The effect that these policies will have on the proportion and distribution of

uncredentialed teachers remains to be seen, however. As for California's school facilities crisis, significant progress in remedying the situation has been made since CSR was signed into law. In 1997, voters approved Proposition 1A, which set aside funds specifically for K–3 class size reduction. Proposition 39, passed in 2000, lowered the threshold for the passage of local bond measures. In addition, two huge statewide bond measures will appear on the November 2002 and March 2004 ballots. Meanwhile, however, some school districts in California continue to be severely constrained by the capacity of their facilities. It is unclear whether the state's efforts will fully eliminate the current problems, much less provide the kind of capacity that would be needed for CSR expansion.

If policymakers have a serious interest in reducing class sizes beyond third grade, they must first ensure that the state has created adequate capacity to provide the certified teachers and facilities such an effort requires.

**Recommendation Eight:**
*Invest in an advanced education data system so that the effectiveness of the state's education reforms can be better determined.*

This has been a recurring theme in each of our prior reports. California's education data system must be redesigned to allow researchers to link teachers and children with their achievement scores over time in order to better measure student gains from year to year. In this way, the state can more accurately measure the effects of specific reforms or variables on student achievement. Because the current data system lacks this capacity, we were unable to do some of the most interesting and important analyses, ones that would have allowed us to compare the relative effects of teacher qualifications and class size on student achievement statewide. The ability to ascertain the effect of teacher credentialing and experience on student achievement would be especially beneficial. The creation of such a data system requires an adequate investment of time, money, and political will by the state and by local school districts, but ultimately the investment promises important returns related to the effectiveness of education reform.

## Lessons Learned That Apply to Future Education Reform Efforts

Beyond the specific recommendations we make above related to California's current CSR program, we think that some broader lessons for policymakers are provided by the state's CSR experience. Whether California embarks on additional class size reduction initiatives or undertakes other large-scale reform interventions, we believe the following serve as important guides for developing effective education reform policy.

*Be clear about where a given initiative fits within the state's overall education policy plan.*

If the overarching goal is to improve student performance, the state legislature must carefully consider how a proposed initiative fits into the overall education plan. It is important that state leaders be clear about what education problem an initiative is trying to solve, and what system of supports might have to be in place for it to work.

We are encouraged that a joint legislative committee is developing a master plan for California education. Assuming that this plan is put into law, the next step is to assure that subsequent education initiatives fit within it. By way of contrast, the 2002–03 governor's budget contains over 70 categorical programs. Annual expenditures on these programs exceed $12 billion, and there is little coherence or coordination among them. As the state's legislative analyst has pointed out, most of these programs show no conclusive evidence of success, restrict local flexibility, fragment local programs, create negative incentives, and blur accountability for meeting student needs. If the state is to radically improve education outcomes for its students, a more systemic approach to education reform must be followed.

*Start small before creating a new, expensive statewide program.*

After the state decides what initiatives are needed to realize the goals embodied in its master plan, we encourage it to start with small, experimental versions of the programs before taking those programs to scale. If the state wants to assure that its education reform dollars are well spent, it should consider the use of controlled randomized field trials that can be evaluated prior to implementing dramatic

reforms statewide (see Bohrnstedt, Stecher, and Kirst, 2000). Once it is clear what works and in what situations, and what supports are required, the reform can be taken statewide with greater chances of effectiveness. It is also important to immediately incorporate what is learned into training programs for new teachers and into professional development courses for experienced teachers and administrators. Without this professional development, the program is unlikely to realize its full potential statewide and may even fail in some locations.

*Carefully examine the context before beginning implementation.*

Do not assume that any single policy intervention can be implemented in new ways or places without first considering the context in which it will be applied. Changes in one component can, and often do, have important effects on other components in the system. Indeed, perhaps the most important finding of our CSR evaluation is the unintended consequences we observed in areas other than student achievement. The rapid, massive introduction of the CSR initiative had deleterious effects on the types of teachers hired, classroom space, and financial resources in a number of districts and schools. Some of these effects could have been foreseen—and perhaps mitigated—if state leaders had taken a more studied approach. At a minimum, a statewide survey of districts could have been conducted to see what challenges and problems they would face in trying to implement CSR.

*Carefully assess the capacity to implement change.*

The number of California teachers who were less than fully qualified went up more than tenfold in the first 2 years of CSR. Further, the districts with the highest percentage of at-risk students were also the most likely to get the newest and less-than-fully-credentialed teachers. When initiating a new program, careful consideration is necessary to ensure that it does not exacerbate existing inequities. After all, it should be helping alleviate them. Overcrowded schools implemented the CSR reform more slowly and had to take facilities from other programs (such as art, music, and special education) to do so. Schools serving the highest percentages of at-risk students tended to be the most overcrowded. Because the state failed to adequately assess its capacity to implement reform, a program that had the opportunity to reduce inequity may actually have increased it (Stecher and Bohrnstedt, 2000).

*Match financial resources with student needs.*

CSR initially created windfalls for some districts, was revenue neutral for others, and placed a substantial financial burden on others. The districts serving the highest percentages of urban, low-income, minority, or EL students were the most likely to redirect funds that had been targeted for other educational needs in order to implement CSR. In this sense, CSR caused some of the better-off districts to gain, while the less well-off districts became further burdened.

One way to avoid this outcome is to use a funding formula based on need. The federal class size reduction funding formula (that mirrored the Title I formula in the distribution of funds) would have been untenable for a state program because it would have denied funding to California's more affluent districts. However, other approaches that acknowledge differences in need might have been possible as long as all districts received at least some funding. Another possibility would have been to stagger implementation so that the most-needy districts implemented the program first, thereby getting first access to the pool of available teachers. Implementation could then proceed in less-needy districts. Finally, it would have been possible to supplement the CSR program with funds distributed to districts on the basis of need, but with the proviso that they be used solely to provide an incentive for teachers to come to difficult-to-serve areas.[19]

We found that the achievement gains in California were the same regardless of student demographics. We know that less-than-fully-credentialed teachers were most likely to end up in districts with the largest percentages of at-risk students. It is interesting to speculate about what the effects of CSR on student achievement might have been in those districts had they been able to recruit teachers with the same qualifications as those in the wealthier districts. In hindsight, California might have benefited by providing full funding to schools positioned to gain the most from smaller classes based on the Tennessee results—i.e., those with the highest percentage of at-risk students. The greater availability of smaller classes would have provided an incentive to draw the best teachers into the most-challenging schools. Instead, we saw just the opposite happen.

[19] The state did eventually create such programs, including a variety of state and district incentives—such as signing bonuses and forgiving student loans—in return for new teachers agreeing to teach in low-performing schools.

*Allow some flexibility to accommodate local differences.*

Any educational intervention is unlikely to succeed to its full capacity if it does not allow for adaptations to local needs and context. For example, California's funding formula for CSR was built on the assumption that all schools would be able to assemble exactly 20 students into a classroom for an entire school year. Schools that exceeded the limit for 3 or 4 months would lose all their funding. Conversely, those that staffed for a smaller class size faced a higher cost for the program. California educators reported that many administrative challenges and fiscal inefficiencies resulted from keeping classes as close to 20 as possible without going over. After the first year of CSR, the Legislative Analyst's Office did a study of districts and found that the average size per reduced class was about 19, not 20 (California Legislative Analyst's Office, 1997). We found the same result in our year 3 analysis. Small schools were especially burdened by the rule. The smaller the school, the less likely the arithmetic of class size would work out easily. Programs that permit no local flexibility are likely to create unanticipated problems in some settings and be unresponsive to local needs.

*Allow sufficient time to implement large-scale initiatives.*

The CSR program was announced in July 1996 for implementation statewide in September 1996. This set off a mad scramble to convert space, obtain equipment, and hire teachers. The rapid start gave a dramatic advantage to districts that had excess capacity, had a ready pool of qualified applicants, were not struggling with other administrative issues, and could redirect resources to this effort quickly. Although there were certainly exceptions, these conditions generally put districts serving the most at-risk students at a disadvantage.

## Conclusion

The results of our evaluations, a changing state policy context, and new class size reduction research in other states—all of these provide justification for reexamining California's current class size reduction policy. As suggested above, the state can change some aspects of the CSR program—including aligning it with standards-based reform goals and policies now in place—without abandoning its commitment to smaller K–3 classes. Carefully controlled pilot programs can help determine what is working and why. Moreover, they can perhaps provide an answer to a pressing question that remains on the table: Is class size reduction a cost-effective strategy for improving student achievement, and, if so, under what conditions?

The CSR experience has taught California that a more systematic approach to large-scale educational improvement is needed. The more time and thought the state puts into the front end to ensure that it plans and funds adequately, allows sufficient time to implement, and conducts meaningful research, the more likely that future state programs will attain their objectives and justify the time and money invested in them.

# REFERENCES


Achilles, C.M., Harman, P., and Egelson, P. (1995). Using research results on class size to improve pupil achievement outcomes. *Research in the Schools, 2*(2), 23–30.

Bohrnstedt, G.W., Stecher, B.M., and Kirst, M. (2000). *Evaluating the effects of statewide class size reduction initiatives: The need for a systemic approach*. Paper presented at the National Invitational Conference on Taking Small Classes One Step Further.

Bohrnstedt, G.W. and Stecher, B.M. (1999). *Class size reduction in California: Early evaluation findings, 1996–1998*. Palo Alto, CA: American Institutes for Research.

Commission on Teacher Credentialing. (2002). *Meeting the Standard: Annual Report 2001*. Sacramento: Author).

California Legislative Analyst's Office, Analysis of the 2002–03 Budget Bill, February 2002. Sacramento.

California, Legislative Analyst's Office, Analysis of the 1997-98 Budget Bill, February 1997. Sacramento.

Cirone, B. (1997, September). Class size reduction: A one-year status check. *Thrust for Educational Leadership, 27*(1), 6–11.

Egelson, P. and Harman, P. (2000). Ten years of small classes in Burke County, North Carolina. In M. Wang and J. Finn (Eds.). *How small classes help teachers do their best* (pp. 279–297.) Philadelphia: Laboratory for Student Success at Temple University Center for Research Human Development and Education and the U.S. Department of Education.

Egelson, P., Harman, P., and Achilles, C.M. (1996). *Does class size make a difference?* Greensboro, NC: Southeastern Regional Vision for Education.

Evans-Stout, K., Fleming, N., Johnson Jr., B.L., Ehrlinger, E., Gray, J.M., Merrill, R., Parker, M., Roberts, R., and Stewart, T. (1997). *The*

*use of class size reduction funds in five Wasatch county districts.* Salt Lake City: Utah Education Consortium.

Finn, J.D. (2002). Class-size reduction in grades K-3. In A. Molnar (Ed.). *School reform proposals: The research evidence* (pp. 15-24.) Tempe, AZ: Education Policy Research Unit, Arizona State University.

Glass, G.V., and Smith, M.L. (1978). *Meta-analysis of research on the relationship of class size and achievement.* San Francisco: Far West Laboratory of Educational Research and Development.

Glass, G.V., Cahen, L.S., Smith, M.L., and Filby, N.N. (1982). *School class size: Research and policy.* Beverly Hills, CA: Sage.

Hanushek, E.A. (1998). *The evidence on class size* [Occasional Paper 98–1]. New York: W. Allen Wallis Institute of Political Economy, University of Rochester.

Harman, P., Egelson, P., and Achilles, C.M. (1997). *Results of a longitudinal application of class-size research findings.* Paper presented at the annual meeting of the American Educational Research Association.

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. Upcoming in the *Economic Journal*.

Krueger, A.B. (2002). *Economic considerations and class size.* Working Paper
# 447, Princeton University, Industrial Relations Center.

Levin, H.M. (1988). Cost-effectiveness and educational policy. *Educational Evaluation and Policy Analysis, 10*(1), 51–69.

Levin, H.M., Glass, G.V., and Meister, G.R. (1984). *Cost effectiveness of four educational interventions.* Stanford, CA: Stanford University. (ERIC Document Reproduction Service No. ED 246 533)

McRobbie, J., Finn, J.D., & Harman, P. (1998). *Class size reduction: Lessons learned from experience* (Policy Brief No. 23). San Francisco: WestEd.

Murnane, R.J., and Levy, F. (1996). Evidence from fifteen schools in Austin, Texas. In G. Burtless. (Ed.). *Does money matter? The effect of school resources on student achievement and adult success* (pp. 93–96). Washington, DC: Brookings Institution Press.

Odden, A. (1990). Class size and student achievement: Research-based policy alternatives. *Educational Evaluation and Policy Analysis, 12*(2), 213–227.

Parrish, T.B., Brewer, D., Kirst, M.W., Driver, C.E., Espinosa, H., Gill, B., Goodwin, J., Hawkins, E., Kaganoff, T., Krop, C., Kuo, V., Lam, I., Range, S.K., & Reichardt, R. (1999). *Class size reduction policies and rising enrollments: Cost implications over the next decade.* Palo Alto, CA: The American Institutes for Research.

Smith, D. (1998, February 22). New school proposals drawing mixed reactions. *Sacramento Bee,* p. A1.

Snow, M., and LaMarca, P. (2001). *Effects of class configuration and teacher characteristics on achievement in reduced size classes*. Carson City, NV: Nevada Department of Education.

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

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

## Advisory Group Members

CSRResearchConsortium

a partnership researching California's class size reduction reform

**AIR**
**RAND**
**PACE**
**EdSource**
**WestEd**

**www.classize.org**