NBER WORKING PAPER SERIES

WHAT IS THE TRADEOFF BETWEEN SMALLER CLASSES AND TEACHER QUALITY?

Christopher Jepsen
Steven Rivkin

Working Paper 9205
http://www.nber.org/papers/w9205

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2002

Any opinions expressed are those of the authors alone and do not necessarily reflect any position of the Public Policy Institute of California or the National Bureau of Economic Research. We would like to thank Mark Baldassare, Julian Betts, Gary Hart, Kim Rueben, Jon Sonstelie, and seminar participants at the Public Policy Institute of California, Stanford University, and the Fall 2001 APPAM meetings for their many helpful comments. We would also like to thank Anne Danenberg for invaluable data support and Pedro Cerdan and Jackie Chou for excellent research assistance.

© 2002 by Christopher Jepsen and Steven Rivkin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

What is the Tradeoff Between Smaller Classes and Teacher Quality?
Christopher Jepsen and Steven Rivkin
NBER Working Paper No. 9205
September 2002
JEL No. I2, H4, J4

ABSTRACT

This paper investigates the effects of California's class size reduction program on teacher quality and student achievement in an effort to gain a comprehensive understanding of the impact of a large-scale decrease in class size. It uses year-to-year differences in class size generated by natural variation in enrollment and the state's class size reduction program to identify both the direct effects of class size reduction and accompanying changes in teacher quality. The results show that, all else equal, smaller classes raise third-grade mathematics and reading achievement, particularly for lower-income students. However, the expansion of the teaching force required to staff the additional classrooms appears to have led to a deterioration in average teacher quality in schools serving a predominantly black student body. This deterioration partially or, in some cases, fully offset the benefits of smaller classes, demonstrating the importance of considering all implications of any policy change.

Christopher Jepsen
Public Policy Institute of California
500 Washington Street
Suite 800
San Francisco, CA 94111
jepsen@ppic.org

Steven Rivkin
Amherst College
Department of Economics
P.O. Box 5000
Amherst, MA 01002-5000
and NBER
sgrivkin@amherst.edu

Encouraged by results from the Tennessee class size experiment, California enacted legislation in 1996 that reduced K-3 class sizes by roughly ten students per class at an annual cost of over $1 billion. Although the Tennessee experiment and most recent statistical analyses find that smaller classes benefit students in the early grades (particularly those who are economically disadvantaged), little evidence exists on the effects of a large-scale reduction in class size that requires the hiring of a large number of new teachers. In addition to adding many rookie teachers, such a rapid expansion of the teaching force may lead to a decline in the average quality of new hires that offsets at least some of the benefits of smaller classes over the longer term, particularly in high-poverty urban schools that have the most difficulty attracting and retaining teachers. Ross (1999) describes the influx of inexperienced, non-certified teachers into elementary schools in South Central Los Angeles following class size reduction (CSR). This influx, he reports, was prompted in part by the departure of many experienced teachers to newly created positions in more affluent communities.

As Heckman (2001) emphasizes, it is important to consider both direct and indirect effects of a policy in order to gain a comprehensive understanding of its impact. With that in mind, this paper investigates the effects of the California CSR program on student achievement with special attention to the link between class size and teacher quality. A number of factors hinder an analysis of the overall effect of CSR. Because California did not administer statewide examinations until the 1997-1998 school year, no baseline measure of achievement prior to CSR is available. Also, the statewide implementation of CSR makes it more difficult to separate its effects from other changes including expanded school accountability, the spread of charter schools, and test score inflation.

Previous evaluations of the effects of CSR compare achievement in schools that did and did not implement CSR in the initial year. In drawing these comparisons, the reports used demographic information and student performance in non-CSR grades as controls for potentially

confounding differences in students and schools.[1] Although such comparisons may well account for most of the systematic differences between timely and late adopters of CSR, these difference-in-difference estimators are likely to be contaminated by the effects of CSR on non-CSR grades. As we show below, evidence strongly suggests that the implementation of CSR also affected the composition of teachers and student achievement in grades not directly affected by CSR.

Our approach also uses differences in the timing of CSR implementation but does not compare achievement across schools on the basis of the year a school reduces class size. Instead, we remove school fixed effects and use class size and teacher characteristic changes between 1997 and 1999 to estimate the effects of smaller classes and teacher characteristics on third-grade achievement. The class size coefficients can be multiplied by the average reduction in class size under CSR to obtain estimates of the gains from reducing class size for schools with particular demographic compositions.[2]

Although included variables provide some information on teacher quality, these and other observable characteristics explain only a small portion of the total variation in quality as measured by teacher value added to achievement (Rivkin, Hanushek, and Kain, 2000). Therefore we attempt to capture unmeasured changes in teacher quality in addition to the changes in observable characteristics by comparing the relationship between class size and achievement in schools that implemented CSR in 1998 or 1999 with the relationship for schools that had already implemented CSR as of 1997. This approach takes advantage of the fact that sources of within-school variation in class size for the years 1997 and 1999 differ systematically with the timing of CSR implementation. On the one hand, class size variation between 1997 and 1999 for early adopters (1996 or 1997) stems almost entirely from natural shifts over time in enrollment that are unlikely to be related to teacher quality changes. Estimates of class size effects for these schools should

[1] See Bohrnstedt and Stecher (1999) and Stecher and Bohrnstedt (2000).

capture the direct effect of class size holding teacher quality constant. On the other hand, the differences in class size between 1997 and 1999 for schools implementing CSR in 1998 or 1999 comes mostly from the hiring of additional teachers. Estimates of class size effects for this set of schools should combine the direct effect of class size with any related change in teacher quality. Any fall in average teacher quality following the implementation of CSR should reduce the magnitude of the class size estimates obtained from the sample of late adopters. However, this approach does not capture any decline in teacher quality following either the departure of teachers to other schools with newly created positions or exits out of fourth or fifth grade to take advantage of the newly smaller classes in the early grades. Therefore it provides a lower bound estimate of the total change in teacher quality that accompanied the implementation of CSR.

The results show that other things equal, smaller classes raise mathematics and reading achievement in third grade, particularly for lower-income students. However, the pattern of estimates also suggests that program implementation led to a decline in average teacher quality for low-income students, partially offsetting the gains from smaller classes. The decline is most noticeable in high percentage black schools that almost certainly face some of the most difficult teacher labor markets in the state. Because changes in teacher experience account for only a portion of this decline, the alarming possibility exists that class size reduction as implemented in California may not benefit or may even harm some of the most disadvantaged students in the state over the longer term by leading schools to hire teachers from lower down the quality distribution.

The paper is structured as follows. Section 2 describes the data set and test score instrument used in California. Section 3 develops a teacher demand and supply framework focusing on the role of teacher certification. It then documents changes over time in a number of teacher characteristics following the implementation of CSR. Section 4 develops the empirical

[2] Unpublished regression results for Texas find approximately linear class size effects on achievement in fourth and fifth grade for classes with between 10 and 35 students.

model used to analyze the relationship between achievement, class size, and teacher quality. Section 5 reports the results, focusing on differences by student demographic characteristics, district and the timing of CSR implementation. Section 6 summarizes the analysis and discusses policies related to both class size and teacher quality.

2. Data Sources

The data come from the California Department of Education (CDE) and include all public elementary schools (grades kindergarten through six) in California with the exception of charter and alternative schools (such as special education schools). Four years of data are used: 1990-1991, 1995-1996, 1997-1998, and 1999-2000 (for ease of discussion we subsequently refer to an academic year by the calendar year in which it begins). Micro-data are not released by the CDE, but aggregate information on students, teachers and schools is available for each grade.

The data set combines student demographic data and test performance with information on average class size and teacher experience, certification, and education. Beginning in 1997, each California public school student in second through eleventh grade took the Stanford Achievement Test Series, Ninth Edition, Form T (Stanford 9) multiple-choice test, administered by Harcourt, Brace & Co. Our measure of achievement for mathematics and reading test performance is percentage of non-limited English proficient (LEP) students who exceed the 50th percentile score based on national norms for each test.[3] Implementation of class size reduction in second grade was nearly universal as of 1997, meaning that there was little variation in the timing of second-grade CSR adoption. Therefore, the analysis of CSR focuses on third grade.

[3] Other test score results including individual or school mean scores are not available in 1997. Students classified as LEP are excluded because the testing requirements for LEP students changed between 1997 and 1999 and tests given in English are not good indicators of mathematics knowledge for non-English speakers.

3. California Public Elementary School Teachers: 1990-1999

The implementation of CSR led to a dramatic change in the composition of third-grade teachers in California public schools. Not only did the employment of third-grade teachers increase by more than 50 percent within two years, there was substantial teacher movement among schools and districts. In particular, there is evidence that inner city schools serving high poverty, high proportion black or Hispanic student bodies lost many teachers to more affluent communities. As we document below, this led to a substantial increase in the proportion of teachers lacking full certification and experience in these schools. Whether this portends a long-term decline in teacher quality depends upon the links between quality, certification, and teacher cohort size. We now draw upon previously developed models of teacher labor supply and demand in order to highlight issues particularly germane to California. Following this discussion is a description of the changes in California third-grade teachers following the implementation of CSR.

CSR and the Supply and Demand for Teachers

The teacher labor market depends upon a number of factors that affect both demand and supply. Recent work by Flyer and Rosen (1997) highlights the impact of the improvement in women's labor force opportunities on both demand and supply. While the effect on teacher labor supply may be ambiguous, the decline in the earnings of women teachers relative to women in other occupations suggests that the expansion of alternative opportunities reduced average teacher quality during the latter part of the 1900s (Hanushek and Rivkin, 1997).

State licensing is another important factor that has an ambiguous effect on teacher quality. Though education and training requirements in combination with any certification examinations may raise the skills of some prospective teachers, they may also discourage or even prevent others from entering the profession by raising the opportunity cost of teaching. Consider the demand for teachers by school s in a state such as California that requires a school to hire a fully certified

teacher if possible. Equation (1) represents demand as a positive function of skill, certification, and other factors such as connections with specific teacher training programs and a negative function of teacher salaries.

(1) D_s = f(skill, certification, other factors, salary)

Schools are assumed to rank order teachers on the basis of expected classroom effectiveness (i.e. skill), certification status, other factors and salary. Because schools value both skill and certification, in order to be preferred to a fully certified teacher, a non-fully certified applicant must be of higher expected quality, other things equal.

Equation (2) describes the supply of teachers to school s as a positive function of salary and non-pecuniary amenities in school s and a negative function of salaries and amenities in nearby districts, distance to desirable housing, earnings opportunities in other occupations, and certification.[4]

(2) S_s = f(salary & amenities$_s$, salary & amenities $_{\text{other schools}}$, salary & amenities $_{\text{other occupations}}$, distance$_s$, certification)

The magnitude and distribution of any changes in teacher characteristics following the implementation of CSR clearly depends upon both the details of the class size reduction, any commensurate increases in salary, and the elasticities of teacher supply for different types of schools. In terms of the first, Appendix Table 1 shows that CSR reduced third-grade class size from roughly 30 to 20 for all types of schools. As a result, the number of CSR-induced openings at a given school depended only on the size of the school, not on the characteristics of its student body. With respect to teacher salaries, Appendix Table 2 shows changes over time in average earnings of

[4] Evidence on non-pecuniary characteristics strongly suggests that teachers tend to prefer schools with lower percentages of poor, non-white and low performing students (Hanushek, Kain and Rivkin, 2001). Recent work by Lankford, Loeb, and Wyckoff also (2002) highlights the very localized nature of teacher labor markets and strong preference for shorter commutes..

young, non-teacher female college graduates in California, probably the best measure of the opportunity cost of teaching. Though the small sample sizes lead to large year-to-year fluctuations in annual earnings, taken as a whole the table strongly suggests that the rise in non-teacher earnings matched if not exceeded the rise in starting teacher salaries during the 1990s, a period of rapid economic growth in California and the nation as a whole. Specifically, between 1993 and 1999 non-teacher earnings grew by three percentage points more than teacher starting salaries (26 to 23 percent). Using three year averages (1993-1995 to 1998-2000) increases the gap to ten percentage points.

Finally measures of supply conditions for different types of schools are not readily available, although Ballou and Podgursky (1997) and others argue that excess supply characterizes most teacher labor markets, especially those in middle class communities. Such schools can meet increased demand by drawing from new teachers, those currently out of teaching, and teachers working in less preferred districts, though the need to fill so many slots in such a short period likely necessitated even these schools to hire less qualified applicants. On the other hand, schools serving academically and economically at risk students likely confront a far more inelastic supply of certified teachers. Lankford, Loeb and Wyckoff (2002) find that the combination of distance to desirable housing and characteristics associated with a high poverty, high proportion black or Hispanic student body severely inhibits the ability of administrators to attract and retain high quality or even fully certified teachers. As the next section documents, these factors led to a dramatic change in the characteristics of elementary school teachers in high poverty, high percent black and Hispanic schools.

Teacher Experience, Certification, and Education

This section describes changes in teacher experience, certification, and education following the implementation of CSR.[5] The tables are constructed in the following way. Schools are divided into four categories according to the percentage of students eligible for a free lunch.[6] Then, the average of the teacher characteristic, such as the school's percentage of new teachers, is calculated from all the schools in that free lunch category. In the case of the final row ("All Students"), schools are weighted according to total enrollment, and the reported percentage reflects the average for all students in that income category. In the other rows the calculations weight each school by the number of students in the specific racial/ethnic group (e.g. Asians in the case of the first row) thereby creating averages for students in each of the four racial/ethnic groups.[7] Differences among racial/ethnic groups within each free lunch category provide information on the degree to which the distributions of teacher characteristics differ by race/ethnicity independent of income.[8]

Experience

Although average experience is not closely linked with achievement gains, recent work suggests that first and second year teachers perform markedly worse than more experienced colleagues, and the distributions of these inexperienced teachers is highlighted.[9] Table 1 describes the percentages of new and second year teachers in third grade. A clear pattern in teacher experience based on percent free lunch exists throughout the time period: the percent of teachers with no years or one year of experience increases with the percent of free lunch for all

[5] Bohrnstedt and Stecher (1999), Stecher and Bohrnstedt (2000), and Betts, Rueben, and Danenberg (2000) also describe the distribution of public school teachers in California, though they do not focus on first and second year teachers or divide schools on the basis of student race/ethnicity and income.
[6] The measure of poverty is actually the percent of students in the school who are eligible for free or reduced price lunch. For simplicity, we refer to this percentage as the percent free lunch.
[7] The calculations assume that there is no systematic variation within schools by race or ethnicity in the probability of having inexperienced or uncertified teachers.
[8] Because the free lunch variable is a crude measure of income, race/ethnic differences conditional on income may also reflect income differences not captured by the four free lunch categories.
[9] See Rivkin, Hanushek, and Kain (2000) for evidence on teacher experience.

race/ethnicities. The gap in teacher experience between low-poverty and high-poverty schools widens for nonwhites following the implementation of CSR. For students in the highest poverty schools (75 percent or more free lunch), the increase between 1995 and 1997 in the percentage of teachers with no experience ranges from over 180 percent for whites to over 220 percent for Hispanics. Although the percentage of teachers in their first year declines for all demographic groups between 1997 and 1999, the rate in 1999 still exceeds that in 1990 for nonwhites in high poverty schools.

The substantial variation in the percentage of new teachers by race/ethnicity and percent free lunch, coupled with the similarity across these groups in the average reductions in class size (see Appendix Table 1), indicate a substantial teacher movement away from high poverty, high percent nonwhite schools. Using data not available for this study, Reichardt (2000) does find evidence of such mobility in the first year following CSR implementation.

Certification

The left half of Table 2 reports the percentage of teachers who lack full certification. In 1990, there is virtually no systematic ordering by income. By 1995, non-certification rates have begun to climb, particularly in higher percent free lunch schools, although some of the increase is likely a result of better reporting of teachers with emergency credentials. There is no doubt, however, that non-certification rates increase dramatically following CSR. In 1997, the non-certification rates were between 3.3 and 6.1 percent for low poverty schools (0 to 25 percent free lunch), compared with a range of 12.0 to 27.9 percent for high-poverty schools (75 percent or more free lunch). Similar differences remain in 1999.

Within each income category, the increases in non-certification rates from 1995 to 1999 for blacks and Hispanics are noticeably higher than for Asians and whites. In 1999, between 23 and 28

of black and Hispanic students in high-poverty schools (75 percent free lunch or more) have teachers who lack full certification.

Education

One certification requirement for experienced teachers is the completion of 30 course hours in addition to a bachelor's degree.[10] The fact that education and certification are tied together is reflected in the right half of Table 2. Similar to experience and certification, the percent of teachers with only a bachelor's degree rises monotonically with the percent free lunch. This pattern is evident for the entire decade. The difference between white and nonwhite grows following the class size reduction, and the percentage of teachers with no postgraduate education is far larger in the highest percent free lunch category for nonwhite students. Over one third of the teachers for black and Hispanic students, along with nearly 30 percent for Asian students, lack any education beyond the completion of a bachelor's degree. The broad based trend towards fewer teachers with only a bachelor's degree observed between 1990 and 1995 for all demographic categories is nowhere to be seen in 1999.

Large Urban Districts

The statewide trends reveal dramatic changes in teacher characteristics, but they almost certainly conceal substantial variation among metropolitan areas, particularly given evidence that teacher labor markets are highly localized (Lankford, Loeb, and Wyckoff, 2002). Therefore the following table illustrates trends in experience and certification for each of the six largest school districts in California. Because of the smaller sample sizes and the high correlation between poverty and racial composition, the tables divide schools only by race/ethnicity.

[10] Currently, new teachers can obtain a preliminary full credential – valid for up to three years – without completing the 30 hour requirement.

The left half of Table 3 describes changes in the percentages of teachers with zero or one year of experience. The share of inexperienced teachers in some of California's largest districts has risen to an alarming level. In 1999-2000, for black and Hispanic students in San Diego, nearly 30 percent of the teachers are in their first or second year. Around 20 percent of the teachers for black and Hispanic students in Los Angeles, Long Beach, Oakland, and San Francisco are in their first or second year. In sharp contrast, Asian and white students in these districts – as well as all students in Fresno – have rates of inexperienced teachers quite comparable to the average California school with a similar demographic composition.

The distribution of the proportion of teachers without full certification presented in the right panel of Table 3 differs somewhat from the experience distributions. In some ways, the certification rates for the six largest districts mirror that of the state as a whole: essentially no uncertified teachers in 1990-1991 and 1995-1996, followed by a large increase in 1997-1998 with the class size reduction. Furthermore, by 1999-2000 the percentage of not fully certified teachers is noticeably higher for nonwhite students than for white students. However, there is also substantial heterogeneity among the districts. In San Diego, only 5.6 percent of the teachers for black students and 8.4 percent of the teachers for Hispanic students do not have full certification despite the large numbers of new teachers.[11] In Long Beach, Los Angeles, and Oakland, on the other hand, over one fourth of black and Hispanic students have uncertified teachers, whereas the proportions for Asian and white students are similar to the statewide averages for these two demographic groups.[12]

[11] An alternative explanation is that the certification data are quite poor in San Diego, raising questions about this aspect of the analysis. Similar concerns exist with the 1997-1998 data for San Francisco, where every single teacher claims to be fully certified. The regression analysis in subsequent sections is not affected by any possible problems with the data in San Francisco or San Diego.

[12] Fresno has very few teachers without full certification in any of its schools.

Summary

These tables document substantial changes in the distributions of teacher characteristics in third grade in California during the 1990s. All of the shifts in experience, certification, and education should not be attributed to CSR, because Stecher and Bohrnstedt (2000) show that certification fell at the high school level as well. Nevertheless, the much larger changes in third grade concentrated around the time of class size reduction provide strong evidence that CSR altered the composition of the third-grade teaching force in important ways. Whether the changes in observable characteristics actually reflect a substantial decline in teacher quality is a key issue that we explore in the next two sections.

4. Empirical Model

California third-grade students, particularly nonwhites from low-income families living in large, urban districts, experienced dramatic changes in both class size and teacher characteristics following the implementation of CSR. The key question is how the reduction in class size and any accompanying changes in teacher quality affected student performance. This section develops the methods used to analyze the inter-relationship among class size, teacher quality, and achievement.

Previous Work

Recent research utilizes social experiments and innovative statistical methods to identify the causal effect of smaller classes *holding teacher quality constant*. The most prominent of these is the Tennessee STAR experiment. Students were randomly assigned to small classes or larger classes. A comparison between achievement in large and in small classes provides an estimate of the benefits of smaller classes but provides no information on changes in teacher quality. Krueger (1999) and Krueger and Whitmore (2001) find that smaller class sizes in kindergarten and first

grade had a significant and lasting impact on achievement. However, Hoxby (2000) discusses potentially serious problems with the randomization of both students and teachers.[13]

Recent studies by Angrist and Lavy (1999), Hoxby (2000), and Rivkin, Hanushek and Kain (2000) identify plausibly exogenous variation in class size in order to learn more about the effects of class size on achievement (in Israel, Connecticut, and Texas, respectively). Except for Hoxby (2000), the studies find that smaller classes significantly increase achievement in the early grades (grade 5 and below), and the effects tend to be larger for lower income students. Because data problems attenuate the estimates in Hoxby (2000), the overall pattern of results suggests that smaller classes raise achievement, other things held constant.[14]

One important feature of these studies is that they go to great lengths to hold all other factors, particularly teacher quality, constant. However, an extensive class size reduction program such as that undertaken by California inevitably alters the composition and quality of the teaching force. Consequently the controlled experiment provides only partial information on the likely effect of reducing class size by ignoring its impact on teacher quality. Depending upon the elasticity of supply of teacher quality and teacher preferences along a number of dimensions, the existing estimates may seriously overstate the benefits of smaller classes, particularly for schools that have a difficult time attracting teachers.

The initial analysis of CSR conducted by the Class Size Reduction Research Consortium does consider the impact of changes in teacher quality.[15] Consortium evaluations of the effects of CSR on achievement compare schools that implemented CSR in the initial year with those that did not (although funding was available for all schools, some did not participate in the first year). In

[13] Also see Ehrenberg et al (2001) for additional discussion of Project STAR.
[14] In the Connecticut data used by Hoxby (2000), tests are administered in the Fall. Therefore, the tests are regressed on class size for the previous school years. If a student moved into the school or district for the current academic year, he or she would receive an erroneous measure of class size. Such measurement error attenuates the estimates of class size effects on achievement.

order to account for differences between schools that did and did not implement CSR in a timely manner, demographic information and fourth or fifth grade test scores are used as controls.

Essentially these difference-in-difference estimators compare the difference between third and fifth grade test scores in schools that adopted CSR to the difference in schools that did not. The validity of this approach rests on the assumption that the timing of CSR implementation did not affect achievement in non-CSR grades. But the CSR Consortium documents large increases in inexperienced teachers in the fourth and fifth grades following CSR even though the program had no direct effect on class sizes in these grades (Stecher and Bohrnstedt, 2000). In combination with estimates presented below showing that changes in fifth grade average achievement are systematically related to changes in third-grade class size, this raises serious doubts that this method produces consistent estimates.

The approach adopted in this paper relaxes some of the more problematic assumptions underlying the Consortium evaluation in an effort to obtain better estimates of the effect of CSR. Moreover, it attempts to gain a more comprehensive understanding of the links between class size reduction, teacher quality, and student achievement.

Empirical Model

The implied model that lies behind this analysis has the general form:

$$A_{isgc} = \sum_{\tau=0}^{g} X_{i\tau c} + \sum_{\tau=0}^{g} CS_{i\tau c} + \sum_{\tau=0}^{g} TC_{i\tau c} + \sum_{\tau=0}^{g} \gamma_i + \sum_{\tau=0}^{g} \theta_\tau + \sum_{\tau=0}^{g} \varepsilon_{i\tau c} \tag{3}$$

Equation (3) describes student achievement (A_{isgc}) for individual i in school s, grade g and cohort c. Achievement is a function of the entire histories of family and community background variables

[15] See Bohrnstedt and Stecher (1999) and Stecher and Bohrnstedt (2000).

(X); of class size (CS); of teacher characteristics (TC); of fixed individual skill and motivation[16] (γ); of a school fixed effect capturing unobserved school differences (θ); and of random errors (ε). For notational ease, each of the inputs is measured in achievement units such that the coefficient is 1.

The entire histories of these inputs are not available for this study or any other, and achievement is not reported for individuals. The inability to follow students over time negates the possibility of analyzing the rate of learning with a value added model. As mentioned above, the focus is on third grade due to data limitations for other grades affected by CSR. Consequently, we utilize data from 1997 and 1999 to estimate the following levels model for third grade achievement:

$$(4) \qquad A_{sc} = X_{sc}\alpha_X + CS_{sc}\alpha_{CS} + TC_{sc}\alpha_{TC} + \theta_s + d_c + d_c * I_{late} + \omega_{sc} + \varepsilon_{sc}$$

Equation (4) describes third-grade student achievement (A_{sc}) for school s and cohort c (1997 or 1999). All variables are measured at the third-grade level unless otherwise stated. The vector TC includes the percentage of teachers in their first year, the percentage in their second year, the percentage not fully certified, and the percentage with no graduate education. The vector X includes percentage black, percentage Hispanic, percentage Asian, percentage eligible for a free lunch, and percentage LEP.[17] A dummy variable (d_c) equal to one for the 1999 cohort captures any upward drift in the test score as well as the effects of other state policies and systematic changes over time. Because the set of schools that implemented CSR "late" (in 1998 or 1999) likely differ from other schools, the equation contains a dummy variable for these schools (I_{late}). This variable is interacted with d_c so that it controls for any systematic differences in the rate of test score growth over time between early and late adopters. Unobserved school quality is decomposed into a

[16]This term incorporates any fixed learning differences that are formed before school entry.
[17] Due to data constraints, we use school-level averages for percent free lunch and percent LEP. All other variables are available at the grade level.

permanent (θ) and a time varying component (ω), the latter of which primarily reflects changes over time in teacher quality that are not captured by the included characteristics.

Because the limited set of explanatory variables omits many important current and past determinants of achievement, OLS estimation of equation (4) would almost certainly confound the effects of the specific measured elements with unobserved differences in family and community background as well as individual skills and motivation. In an effort to minimize any bias, we make use of the availability of two years of data and estimate a model of school fixed effects that uses only within school variation over time to identify the coefficients. Not only does the fixed effects approach remove all time invariant school characteristics (θ_s), it also controls for school average family influences that are constant over time. For example, if some school districts tend to attract more highly educated parents than others, the persistent aspect of the education gap will be removed. However, if changes in class size are correlated with unobserved factors that also affect achievement, the fixed effect estimates will be biased.

In the case of California, two potentially confounding factors would appear to be most important. The first is teacher quality. If a reduction in class size that requires additional teachers tends to dilute unobservable teacher quality (i.e. decrease ω), the class size coefficients will conflate the true class size effect with the change in quality, because cov(CS, ω) will not equal zero, conditional on the included explanatory variables. Rather cov(CS, ω)>0, and the estimated class size effects will be biased toward zero.

The second potential impediment to the identification of the benefits of smaller classes relates to the sources of class size differences between 1997 and 1999. One factor that contributes to changes over time in class size is fluctuations in enrollment. CSR participants in both years experience annual fluctuations in class size driven by small enrollment differences among cohorts. Controlling for changes in student characteristics, this type of natural variation first used by Hoxby

(2000) likely provides a valid source of identification, i.e. cov(CS, ω)=0.[18] The second major source of class size variation is the CSR program itself. Nearly 35 percent of schools implemented CSR after 1997, resulting in large reductions in class size during this period. Importantly, these schools contribute the bulk of the variation in class size. If facility constraints, administrative turmoil, difficulties attracting teachers, or other factors that directly affect achievement contributed to the decision to delay program adoption, the class size estimate will be biased downward.

The regressions directly control for any systematic differences associated with the timing of CSR implementation with the inclusion of a dummy variable equal to one for schools that reduced class size in either 1998 or 1999 interacted with the indicator variable for 1999.[19] Moreover, we take advantage of the two different sources of class size variation in order to identify changes in teacher quality that may accompany class size reduction. Defining I_{early} as a dummy variable for schools that had implemented CSR as of 1997, we estimate the following equation:

$$A_{sc} = X_{sc}\alpha_X + CS_{sc} * I_{late}\alpha_{CSlate} + CS_{sc} * I_{early}\alpha_{CSearly} + \theta_s + d_c + d_c * I_{late} + \omega_{sc} + \varepsilon_{sc} \tag{5}$$

Because schools that must reduce class size more in order to qualify for CSR funding must add proportionately more teachers, any decline in teacher quality that accompanied implementation of CSR should reduce the estimate of α_{CSlate}. Consequently the difference between the class size coefficients for late (α_{CSlate}) and early ($\alpha_{CSearly}$) adopters provides an estimate of the decline in teacher quality that accompanied class size reduction.

If the observed teacher variables accounted for all variation in teacher quality, their inclusion into equation (5) would eliminate any difference between α_{CSlate} and $\alpha_{CSearly}$ that was due

[18] Hoxby (2000) argues that enrollment differences that are not de-trended may be correlated with unobserved changes in family background. However, we find that the exclusion of the set of student demographic variables has no effect on the class size coefficients, meaning that any omitted variables bias would have to come from variables orthogonal to these very important predictors of achievement.
[19] Omitted factors would have to be correlated with the change in class size in order to introduce bias.

to teacher quality. Because these three characteristics almost certainly explain only a smaller portion of the quality variation, that is unlikely to be the case. Instead, if schools adopting CSR were forced to hire from further down the teacher quality distribution, the inclusion of the three teacher characteristics would not eliminate difference between α_{Cslate} and $\alpha_{CSearly}$.

$$A_{sc} = X_{sc}\alpha_X + CS_{sc} * I_{late}\alpha_{CSlate} + CS_{sc} * I_{early}\alpha_{CSearly} + TC_{sc} * I_{late}\alpha_{TClate} + TC_{sc} * I_{early}\alpha_{TCearly} + \theta_s + d_c + d_c * I_{late} + \omega_{sc} + \varepsilon_{sc} \quad (6)$$

The magnitude of any quality decline depends upon teacher labor supply conditions for a school, and the previous section showed that there is likely to be substantial variation by student demographic characteristics and other factors. Consequently, we allow the effects of CSR and estimates of the change in teacher quality to vary by student demographics and district.

Although the difference between estimates of α_{CSlate} and $\alpha_{CSearly}$ provides some information on the change in teacher quality that accompanied the implementation of CSR, it does not capture potentially important sources of quality decline. First, CSR may have affected teacher quality in other grades. A decline in class size from thirty to twenty certainly makes teaching third grade relatively more desirable than teaching fourth or fifth grade, and the rise in the share of rookie teachers in these grades indicate that many teachers moved to earlier grades in order to take advantage of the now smaller classes. Consequently any decline in quality was likely spread across all grades, and we examine the link between fifth grade achievement and third grade class size.

Second, teacher quality may also be affected by hiring in neighboring districts or other schools in the same district. Consider two schools in Los Angeles Unified school district, one in a middle class neighborhood in the San Fernando Valley and the other in a high poverty area of South Central Los Angeles. Class size reduction in the San Fernando Valley school opens up additional jobs that may be filled by teachers currently working in South Central if they have the appropriate level of seniority. There may not be any decline in quality in the San Fernando Valley

school, but teacher quality could fall at the school in South Central if that school is unable to find suitable replacements. Because any such decline in teacher quality is not linked to the magnitude of class size change at the affected school, it will not be captured by the methods used in this paper. Again this suggests that a comparison of third grade class size coefficients by the timing of CSR implementation will understate any decline in teacher quality associated with the program.

Finally, any measurement error in the class size variable would almost certainly reduce the gap between estimates of α_{CSlate} and $\alpha_{CSearly}$, because the ratio of signal to noise would be much larger for late adopters.

5. Class Size, Teacher Characteristics, and Student Achievement

This section reports the results from the analysis of mathematics and reading achievement for the state as a whole and for a set of large urban districts. Though micro-data are not available, estimates are permitted to vary by student demographic characteristics in order to investigate possible differences in both teacher labor supply conditions and direct effects of class size. Baseline results are reported prior to the examination of differences by CSR timing.

Baseline Results

Table 4 reports the estimated effects of class size and teacher characteristics on the percentages of third grade non-LEP students in a school who exceed the 50th percentile (median) mathematics and reading test scores calculated on a national basis. The odd numbered columns do not include any controls for teacher characteristics, and therefore the class size coefficients capture the effects of any associated changes in these variables. The even numbered columns control for differences in experience, certification, and education. All regression specifications include the percentages of free lunch, black, Hispanic, Asian, and LEP students, an indicator variable for late

(1998 or 1999) CSR adopters interacted with a 1999 cohort dummy variable and the 1999 cohort dummy variable. Descriptive statistics are reported in Appendix Table 3.

All class size estimates are negative and highly significant. Consistent with recent research on Texas, the effects on reading are roughly seventy percent as large as the effects on mathematics.[20] The class size estimates are quite insensitive to the inclusion of the other teacher characteristics, probably because most schools had already implemented CSR by 1997 and the change in class size is not highly correlated with changes in experience, percent not fully certified, or education. Importantly, the class size and teacher characteristic estimates are also quite insensitive to the inclusion of the student characteristics. Therefore any omitted variables bias must be orthogonal to student income and race/ethnicity, significant predictors of achievement.[21]

Each one student decrease in class size raised the percentage who exceeded the median mathematics score by approximately 0.4 points and the percentage who exceeded the median reading score by roughly 0.3 percentage points. Therefore, the ten student average decline due to CSR raised the percentages that exceeded the median by approximately 4 percentage points in mathematics and 3 percentage points in reading. Further interpretation of the magnitude of these effects is difficult for two reasons. First, the dependent variable measures the percentage that crosses a threshold rather than the number or percentage correct. Second, the link between these test scores and ultimate effects on educational attainment, earnings, and other important outcomes is not known.

In terms of the teacher characteristics, the findings in Table 4 contradict some of the established views on the determinants of quality. First, consistent with other studies, the table

[20] See Rivkin, Hanushek, and Kain (2000).
[21] Similarly, Altonji, Elder, and Taber (2000) find that their coefficient of interest (Catholic schooling) is not sensitive to the omission of subsets of observable characteristics. More generally, they demonstrate that selection on unobservables must be greater than selection on observables in order to eliminate the causal effect of Catholic schooling on high school graduation and college attendance rates.

provides little or no evidence that postgraduate education improves teacher quality. None of the coefficients on percent B.A. only are significant at any conventional level, and all are extremely small. Further dividing teachers on the basis of a master's degree (not reported) does not change the results. Second, none of the specifications produce a positive and significant relationship between achievement and the percentage of teachers not fully certified. As discussed earlier, this result is not that unexpected if uncertified teachers must be of higher quality in order to be hired in districts that could choose a fully certified teachers instead. Appendix Table 4 shows that the results for certification are sensitive to the model specification. In particular, the failure to account for the highly nonlinear gains to experience that are concentrated early in the career can lead to upwardly biased estimates of the gains to certification (see columns 1, 2, 5, and 6). If such gains are controlled, however, the results are not sensitive to the taxonomy used to classify teachers lacking full certification.

On the other hand, the results for experience in Table 4 are consistent with prior expectations. For both mathematics and reading, the percentage of students who exceeded the 50th percentile in a class with a new teacher was roughly 3 percentage points lower than in classrooms with teachers with at least two years of experience. All of these estimates are highly significant. Although all coefficients on percentage of teachers with one year of experience are also negative, none are significant.

Differences by Student Demographic Composition

A number of studies find that class size and teacher experience effects are larger for low-income or nonwhite students. In order to examine whether this holds in California we interact class size and the teacher characteristics with the percentages of free lunch and black students. The most striking pattern in Table 5 is that the effect of class size increases as percent low income increases but not as percent black increases. The larger and significant positive interaction term with percent

black indicates that an additional black reduces the benefit from smaller classes even if that student is poor.[22] The effects for teacher experience in Table 5 are less precise, but they show the same pattern as class size. In contrast, there is no evidence that having a teacher without full certification or without any postgraduate training imposes a larger cost on low-income or nonwhite students.

Did CSR Reduce Teacher Quality?

The results in Tables 4 and 5 provide only partial information on the total impact of CSR. Most schools had already reduced class size by 1997, the first year of the sample (and the second year of CSR). Any changes in teacher quality not captured by the included characteristics are ignored because the tests were not administered prior to the implementation of CSR at the majority of schools. Therefore we split the sample on the basis of CSR timing in order to gain a better understanding of the relationship among class size, teacher quality, and achievement. As previously discussed, the class size coefficients estimated from the sample of schools that adopted CSR as of 1997 provide estimates of the pure effect of class size, whereas those estimated from the late adopters combine this direct effect of class size with accompanying changes in teacher quality.

Table 6 reports mathematics and reading class size coefficients from specifications based on equation (5) that divide schools on the basis of CSR timing. Three regressions are estimated for both mathematics and reading. For each regression, "Class Size/Late Adopter" and "Class Size/Early Adopter" capture the baseline class size effects for the students in late adopting and early adopting schools, respectively, while "Class Size Interaction…/Late Adopter" and "Class Size Interaction…/Early Adopter" capture any interactions with percent free lunch or percent black for students in the two groups of schools. Schools that did not adopt CSR as of 1999 were excluded from the analysis that underlies Tables 6 and 7.

[22] An alternative explanation is that blacks benefit less than their non-black classmates. In the absence of student level data the competing explanations cannot be tested, though past evidence provides little reason to believe that blacks benefit less from smaller classes.

The similarity of the late adopter and early adopter class size coefficients in Column 1 reveal little or no evidence that the quality of third-grade math instruction systematically declined following the implementation of CSR. In the case of reading the early adopter coefficient in Column 4 is roughly 50 percent larger, though not significantly different at the 10 percent level. However, the previous section suggested that the dynamics of the teacher labor market are likely to vary systematically by school demographic composition and other factors, and the direct benefits of smaller classes may also vary by these same factors. Therefore the remaining columns report results from regressions that permit class size effects to vary by both demographic composition and the timing of CSR implementation.

Despite the limited variation in class size among the early adopting schools, Columns 2 and 5 reveal a clear contrast between the effects of changes in percent low income and changes in percent black; this difference is particularly striking in mathematics. Although the class size interaction with percent free lunch is much larger in magnitude for early adopters, there is very little difference in the comprehensive class size effects on the basis of CSR timing in predominantly low-income schools. For example, the estimated gain in the percentage exceeding the nationally norm referenced median mathematics score in schools in which 75 percent of the students are eligible for a free lunch and 0 percent are black is 6.9 percentage points for late adopters and 7.0 percentage points for early adopters; the corresponding estimates for reading are 2.2 and 3.2 percentage points for early and late adopters respectively.

An increase in the percentage of students who are black, on the other hand, significantly reduces the benefits of CSR for the late adopting schools. In the case of mathematics, the magnitude of the interaction coefficient is roughly twice as large as the corresponding coefficient on low income, implying that an additional poor black student tended to reduce the benefits of smaller class in late adopting schools. Although the coefficient for the class size/percent black interaction is also positive for the early adopting schools, it is much smaller both in magnitude and

in comparison to the class size/percent low income interaction. Consider schools with a 60 percent black and a 75 percent low income student body (the average for schools with greater than 50 percent black enrollment). The estimated gain in math from reducing class size by 10 for such schools is 4.6 percentage points if the schools adopted early and 0.3 percentage points (essentially zero) if the schools adopted late. The results for reading are not nearly as strong, but there remains a substantial difference between early and late adopting schools.

The results for reading also indicate somewhat lower benefits in late adopting schools, but the gap is not as large and even late adopters enjoyed a significant positive effect in reading. The coefficient on the percent black interaction term for schools that implemented CSR early is much smaller (and negative in the reading regressions) and statistically insignificant, refuting the notion that the direct benefits of smaller classes are simply much smaller in schools with a large black enrollment share.

An important question is whether the decline in teacher quality suffered by schools with a high percentage black enrollment is temporary or longer term. Given the difficulties faced by new teachers, teacher quality would be expected to fall in most schools just implementing CSR because of the need to hire large numbers of new teachers (schools in more affluent communities may be an exception if they tend to fill new positions with experienced teachers). This decline, which would occur even if these schools continued to hire from the same pool of applicants, is an unavoidable short-term cost of reducing class size that would disappear as these new teachers acquired more experience.

Table 7 reports mathematics and reading class size coefficients from specifications based on equation (6) that include teacher characteristics. The results for mathematics indicate that the addition of the teacher characteristics, most importantly experience, reduces the percent black/class size interaction coefficient in Column 2 from 0.0109 (in Table 6) to 0.0088 (in Table 7) for

students in schools that implemented CSR in 1998 or 1999; there is little change for the early adopters. Roughly twenty percent of the decline in class size effects for these students is eliminated by the inclusion of teacher experience, indicating that a portion of the decline in teacher quality is temporary. However, differences in teacher experience do not account for most of the apparent decline in teacher quality that accompanied the adoption of CSR. This finding suggests that some schools were forced to draw from further down the quality distribution and that lower average instructional quality is likely to persist for students in high percentage black schools.

As noted earlier, this approach potentially underestimates any decline in teacher quality for a number of reasons including the possibility that any decline in teacher quality was spread across both CSR and non-CSR grades. In order to provide additional evidence, we consider the relationship between fifth-grade achievement and third-grade class size. The results in Table 8 from specifications that exclude 5th grade teacher characteristics but include 5th grade class size show that a ten student reduction in third-grade class size led to roughly 2.0 to 2.4 percentage point decline in the percentages of fifth-grade students who exceeded the national medians in mathematics and reading in schools with 60 percent black enrollment. In addition, it appears that changes in the percentages of fifth grade teaches with zero and one year of experience account for roughly one fourth of the decline in fifth grade teacher quality associated with the reduction in third grade class size; the inclusion of fifth-grade teacher experience reduces the size of the coefficient on third grade class size by roughly 25 percent. Similar to the evidence for third grade, the findings for fifth grade suggest that in higher percentage black schools, CSR led to both the need to hire a number of inexperienced teachers and a decline in average quality conditional on experience.

Labor Market Differences Across California

Because of potential differences in teacher labor market conditions, the findings for the state as a whole may conceal substantial variation among districts in the change in teacher quality and benefits of CSR. It would be very interesting to compute separate estimates for each of the

large districts across the state, but small sample sizes would make this an uninformative exercise with the exception of Los Angeles. Therefore Table 9 reports results for third-grade achievement in the state as a whole, Los Angeles Unified school district, the next five largest districts combined, and all districts other than Los Angeles. Note that at least one third of schools in teach of the six largest districts adopted CSR late, providing the necessary conditions for a decline in teacher quality between 1997 and 1999. However, because school-to-school teacher transfers are commonplace in large districts, the link between CSR timing and any decline in teacher quality within district is much weaker. In combination with the very small sample sizes, this limits the value of dividing the sample for this part of the analysis.

The results in Table 9 suggest that high percentage black *and* high percentage low income schools in LA Unified suffered large declines in mathematics achievement in response to CSR regardless of whether teacher characteristics are included in the regressions. Specifically, the interaction term between percent black and class size is nearly twice as large in LA Unified as in the state as a whole (see Table 5), and the low income interaction is also positive and statistically significant. By comparison, the interaction between class size and percent black for the next five largest districts is much smaller and imprecisely estimated. For all districts other than Los Angeles, the final column reveals a much smaller and less significant positive interaction with percent black that is of the same magnitude as the interaction with percent low income. Finally, the results for reading are much less precisely estimated but suggest similar disparities between LA Unified and either the state as a whole or the next 5 largest districts.

6. Summary and Policy Implications

The characteristics of teachers and classrooms changed dramatically during the 1990s for all demographic groups. The now standard class size of 20 students for grades K through 3 is roughly one third smaller than the average just six years ago, and the results suggest that smaller

classes increased mathematics and reading achievement, other things equal. Yet the rapid expansion of the teaching force needed to staff the additional classes has led to a dramatic increase in the percentage of teachers who lack full certification, who have no postgraduate education, and who are in their first or second year of teaching. These changes have been concentrated in high-percentage nonwhite and free lunch schools, particularly those in Los Angeles. For example, in 1999, nearly 30 percent of the teachers in these schools were not fully certified, while over 10 percent were in the first year of teaching. Overall, the income and race/ethnicity gaps in teacher quality widened along virtually all measured characteristics between 1995 and 1999.

Although neither teacher education nor certification appear to be strongly related to teacher quality as measured by student performance, the pattern of estimated class size effects suggests that teacher quality has deteriorated for students attending schools in which a substantial percentage of the student body is black; this effect is particularly striking in Los Angeles Unified School District. The rise in the incidence of rookie teachers explains a portion of the decline, but it appears that these schools were hiring and retaining teachers from lower down the quality distribution following the implementation of CSR.

This conclusion that a decline in teacher quality offset the benefits of smaller classes is not based on observable measures of quality. Rather it rests upon the assumption of sizeable differences in teacher effectiveness that are not well captured by the included variables. In other words, we believe that the decline in observable teacher characteristics (i.e. a higher proportion of uncertified teachers with little or no experience) is accompanied by a similar decline in unobservable teacher quality for both certified and non-certified teachers. There may be other explanations for the observed pattern of class size estimates, such as differences in other school resources that are correlated with the change in class size, but we believe that the most plausible explanation is a decrease in teacher quality following the implementation of CSR.

These findings highlight the importance of the elasticity of the supply of teacher quality in the determination of the benefits from class size reduction. Although that supply might be highly elastic in schools serving middle- or upper-income students, the results are consistent with the view that schools serving a predominantly black student body face severe difficulties in attracting and retaining high-quality teachers. Persistent teacher shortages in New York City illustrate the challenge of staffing schools that serve disadvantaged children, particularly those in urban districts in which senior teachers are able to choose their preferred school.[23] Similar teacher shortages in California's largest cities, the impending retirement of a large share of teachers, and the expanding wage opportunities for women outside of teaching all suggest that this phenomenon is neither short term nor limited to California.

There are potential policy levers for reducing the adverse costs of class-size reduction by making the supply of teacher quality more elastic. Evidence that certification is only weakly related to quality, particularly for low- income and nonwhite students, and the lack of association between teacher education and student outcomes strongly suggests that personnel policies should be reexamined. Removing the link between postgraduate education and pay and relaxing or eliminating many certification requirements would lower the cost of becoming a teacher with arguably little cost in terms of preserving quality. Higher pay, particularly for teachers who serve disadvantaged students, also may be necessary in order to attract and retain higher-quality teachers.

The unintended consequences of class-size reduction could probably be reduced if the program were targeted toward schools serving disadvantaged populations. Evidence consistently shows that non-disadvantaged students obtain quite small gains from smaller classes, particularly in comparison to those with low family income. In a more fundamental way, the targeting of large sums of money for statewide or nationwide class-size reduction would appear to be a faulty approach to improving school quality. School and district circumstances including personnel

[23] See Lankford, Loeb, and Wyckoff (2002) for a description of the teacher labor market in New York.

considerations are far too disparate for such a blunt, “one-size-fits-all” program. Rather, policies should accommodate the fact that schools and districts face very different constraints and have very different needs.

References

Altonji, Joseph, Todd Elder, and Christopher Taber (2000), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," NBER working paper 7831.

Angrist, Joshua and Victor Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114(2), 533-576.

Ballou, Dale, and Michael Podrusky (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, MI: The Upjohn Institute.

Betts, Julian, Kim Rueben, and Anne Danenberg (2000). *Equal Resources, Equal Outcome? The Distribution of School Resources and Student Achievement in California*. San Francisco, CA: Public Policy Institute of California.

Bohrnstedt, George and Brian Stecher, eds (1999). *Class Size Reduction in California: Early Evaluation Findings, 1996-1998*. Palo Alto, CA: American Institutes for Research.

Boyd, Don, Hamilton Lankford, Susanna Loeb, and James Wyckoff (2002), "Analyzing the Determinants of the Matching of Public School Teachers to Jobs," unpublished manuscript.

Ehrenberg, Ronald, Dominic Brewer, Adam Gamoran, and J. Douglas Willms (2001), "Class Size and Student Achievement," *Psychological Science in the Public Interest*, 2(1), 1-30.

Flyer, Frederick and Sherwin Rosen (1997), "The New Economics of Teachers and Education," *Journal of Labor Economics*, 15(1): S104-S139.

Hanushek, Eric and Steven Rivkin (1997), "Understanding the 20th Century Explosion in U.S. School Costs," *Journal of Human Resources*, Winter: 35-68.

Hanushek, Eric, Steven Rivkin, and John Kain (2002), "Why Public Schools Lose Teachers," unpublished manuscript.

Heckman, James (2001), "Accounting for Heterogeneity, Diversity and General Equilibrium in Evaluating Social Programmes," *The Economic Journal*, 111(4), F654-F699.

Hoxby, Caroline Minter (2000), "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115(4), 1239-1286.

Jepsen, Christopher and Steven Rivkin (2002). *Class Size Reduction, Teacher Quality, and Academic Achievement in California Elementary Public Schools*. San Francisco, CA: Public Policy Institute of California.

Krueger, Alan (1999), "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114(2), 497-532.

Krueger, Alan and Diane Whitmore (2001), "The Effect of Attending a Small Class in the Early Grades on College Test-Taking and Middle School Test Results: Evidence from Project STAR," *The Economic Journal*, 111(1), 1-28.

Reichardt, Robert (2000), *The Cost of Class Size Reduction: Advice for Policymakers, Ph.D. Dissertation*, The RAND Graduate School.

Rivkin, Steven, Eric Hanushek, and John Kain, (2000), "Teachers, Schools, and Academic Achievement," Working Paper.

Ross, Randy (1999), "How Class-Size Reduction Harms Kids in Poor Neighborhoods," *Education Week*, May 26, 1999.

Stecher, Brian and George Bohrnstedt, eds (2000). *Class Size Reduction in California: The 1998-1999 Evaluation Findings*. Sacramento, CA: California Department of Education.

Table 1. Teacher Experience by Student Income and Race/Ethnicity

	Percent	Percent 0 Years Experience				Percent 1 Year Experience			
	Free Lunch	1990	1995	1997	1999	1990	1995	1997	1999
Asian	0-25	4.5	3.4	10.6	4.5	6.0	3.4	7.8	5.9
	25-50	4.9	3.1	9.0	4.8	6.7	3.1	9.1	6.4
	50-75	6.6	3.8	12.1	6.1	7.3	5.7	10.0	8.3
	75-100	8.0	5.2	14.9	8.3	8.8	6.7	9.6	11.0
	All	5.7	3.9	11.7	5.8	6.9	4.8	9.0	7.7
Black	0-25	4.4	3.6	9.9	4.9	7.3	3.8	8.2	5.6
	25-50	5.3	3.1	8.7	5.4	7.0	3.8	7.5	5.6
	50-75	10.0	5.0	11.6	6.4	7.6	6.0	10.3	7.9
	75-100	9.2	9.4	20.6	11.3	8.8	8.4	11.7	13.8
	All	7.6	6.4	15.2	8.3	7.7	6.5	10.4	9.9
Hispanic	0-25	5.0	3.7	10.5	4.6	8.4	4.1	7.9	6.4
	25-50	5.7	4.2	9.2	4.7	7.7	4.1	8.2	6.2
	50-75	7.6	5.2	13.6	6.3	8.5	5.2	9.2	8.4
	75-100	7.4	7.6	17.6	8.8	7.6	7.0	10.6	10.7
	All	6.8	6.3	15.1	7.4	7.9	5.9	9.8	9.3
White	0-25	4.5	3.7	9.3	4.2	7.1	4.3	8.2	5.5
	25-50	5.7	3.2	7.6	3.3	7.1	3.5	7.5	5.1
	50-75	6.5	4.2	10.7	4.4	8.0	4.8	8.1	6.2
	75-100	8.3	6.2	12.7	6.2	8.0	6.1	9.7	9.1
	All	5.3	3.9	9.5	4.2	7.3	4.3	8.1	5.9
All Students		6.1	5.1	12.6	6.2	7.5	5.2	9.1	8.0

Notes: Schools are divided into four categories according to the percentage of students eligible for a free lunch. Then, the average of each teacher characteristic is calculated from all the schools in that free lunch category. In the case of the final row ("All Students"), schools are weighted according to total enrollment, and the reported percentage reflects the average for all students in that income category. In the other rows the calculations weight each school by the number of students in the specific racial/ethnic group (e.g. Asians in the case of the first row) thereby creating averages for students in each of the four racial/ethnic groups.

Table 2. Teacher Certification and Education by Student Income and Race/Ethnicity

	Percent	Percent Not Fully Certified				Percent with B.A. Only			
	Free Lunch	1990	1995	1997	1999	1990	1995	1997	1999
Asian	0-25	0	0.1	5.0	4.6	12.6	8.4	15.2	11.2
	25-50	0.6	0.3	5.9	5.8	13.8	12.2	16.4	13.4
	50-75	0.4	1.1	10.6	10.2	20.2	13.9	20.1	19.2
	75-100	0.5	2.5	14.2	16.0	28.0	19.7	24.6	28.5
	All	0.3	1.0	8.9	8.7	17.4	13.5	19.0	17.4
Black	0-25	0	0.2	6.1	6.3	12.9	11.9	16.5	11.3
	25-50	0.5	0.5	6.7	6.0	16.3	14.3	18.1	14.7
	50-75	1.4	2.4	12.1	13.5	23.5	15.0	23.1	22.5
	75-100	1.1	5.8	27.9	28.1	33.6	27.6	35.7	38.8
	All	0.9	3.3	18.2	18.0	23.1	20.1	27.7	27.3
Hispanic	0-25	0.1	0.5	5.3	5.2	13.7	10.1	15.1	12.9
	25-50	0.7	0.7	6.7	6.3	14.8	11.9	16.3	13.2
	50-75	0.8	2.0	13.6	11.6	20.0	12.7	21.8	20.3
	75-100	0.6	4.0	24.8	23.2	32.6	25.9	34.4	33.9
	All	0.6	2.8	18.4	17.2	22.8	19.4	27.7	26.7
White	0-25	0.1	0.3	3.3	3.4	12.3	9.8	12.4	11.0
	25-50	0.4	0.6	4.7	4.4	14.7	11.1	13.0	11.5
	50-75	0.4	1.0	8.2	7.3	19.5	11.8	16.7	16.7
	75-100	0.2	1.6	12.0	12.0	25.2	17.5	22.3	21.9
	All	0.2	0.6	5.6	5.2	14.5	11.3	14.5	13.3
All Students		0.4	1.8	12.6	12.2	18.5	15.6	21.8	21.1

Notes: Schools are divided into four categories according to the percentage of students eligible for a free lunch. Then, the average of each teacher characteristic is calculated from all the schools in that free lunch category. In the case of the final row ("All Students"), schools are weighted according to total enrollment, and the reported percentage reflects the average for all students in that income category. In the other rows the calculations weight each school by the number of students in the specific racial/ethnic group (e.g. Asians in the case of the first row) thereby creating averages for students in each of the four racial/ethnic groups.

Table 3. Teacher Experience and Certification by District and by Student Race/Ethnicity in the Six Largest School Districts

		Percent 0 or 1 Years Experience				Percent Not Fully Certified			
		1990-91	1995-96	1997-98	1999-00	1990-91	1995-96	1997-98	1999-00
Fresno	Asian	14.3	5.7	13.7	10.9	0	0	1.8	1.2
	Black	11.9	6.3	18.6	6.2	0	0	3.2	1.1
	Hispanic	13.0	3.7	17.6	8.7	0	0	2.0	1.8
	White	9.0	1.5	8.2	2.7	0	0	0.4	1.4
	All	12.0	4.0	15.0	7.7	0	0	1.8	1.5
Long Beach	Asian	18.0	15.4	29.2	19.7	0	11.1	27.1	30.1
	Black	20.2	16.1	28.2	20.4	0	10.0	28.6	30.0
	Hispanic	20.9	16.0	30.8	22.7	0	13.0	30.0	30.7
	White	9.8	7.1	20.7	12.8	0	3.6	17.7	14.4
	All	17.1	14.2	28.1	20.0	0	10.2	26.9	27.6
Los Angeles	Asian	11.3	9.0	30.6	12.6	0	2.8	28.2	15.6
	Black	13.6	17.5	35.8	21.2	0	6.0	37.1	29.7
	Hispanic	12.4	14.2	31.5	18.6	0	4.6	36.9	25.8
	White	11.8	13.0	29.3	12.4	0	3.1	22.9	13.1
	All	12.4	14.2	31.9	18.0	0	4.5	34.9	24.5
Oakland	Asian	15.8	7.4	16.1	8.8	0	2.4	15.7	16.7
	Black	14.0	7.1	22.5	19.6	0	3.9	25.9	29.1
	Hispanic	20.8	7.7	24.4	21.0	0	7.3	24.4	28.1
	White	5.7	2.0	7.6	3.0	0	1.3	4.7	4.9
	All	14.8	7.0	20.9	17.3	0	4.2	22.4	25.4
San Diego	Asian	19.9	7.7	18.8	14.5	0	0	1.3	2.1
	Black	25.4	11.4	29.4	31.4	0	0	1.6	5.6
	Hispanic	20.9	16.4	30.4	29.6	0	0	3.1	8.4
	White	10.5	5.0	16.3	11.8	0	0	0.3	1.6
	All	17.6	10.3	24.3	22.9	0	0	1.8	5.2
San Francisco	Asian	7.0	4.6	19.4	20.4	0	0	0	15.9
	Black	10.3	9.6	19.1	21.7	0	0	0	25.5
	Hispanic	10.6	7.9	24.6	17.8	0	0	0	22.2
	White	3.6	2.1	13.7	9.2	0	0	0	14.8
	All	8.1	6.0	19.7	18.7	0	0	0	18.9

Note: The percentages are weighted by the number of students in each racial/ethnic category.

Table 4. Estimated Effects of Class Size and Teacher Characteristics on Third Grade Test Scores

	Mathematics				Reading			
Include Student Characteristics?	no	no	yes	yes	no	no	yes	yes
Class Size	-0.37	-0.39	-0.40	-0.41	-0.22	-0.23	-0.26	-0.26
	(4.56)	(4.71)	(4.85)	(5.01)	(3.37)	(3.51)	(3.85)	(3.99)
Teacher Characteristics								
Percent 0 Years Experience		-0.028		-0.027		-0.031		-0.030
		(2.51)		(2.43)		(3.37)		(3.24)
Percent 1 Year Experience		-0.007		-0.006		-0.004		-0.003
		(0.61)		(0.56)		(0.42)		(0.34)
Percent Not Fully Certified		-0.018		-0.018		-0.011		-0.011
		(1.36)		(1.38)		(1.05)		(1.07)
Percent B.A Only		-0.002		-0.003		0.010		0.009
		(0.20)		(0.33)		(1.26)		(1.14)
Observations	7,626	7,626	7,626	7,626	7,612	7,612	7,612	7,612

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the school percentage of non-LEP students who exceed the 50th percentile on a nationally determined basis. Other regressors include percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, a dummy variable for the year 1999, and the 1999 dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. All regressions are weighted by the number of test takers. Each column represents a separate regression.

Table 5. Estimated Effects of Class Size and Teacher Characteristics on Third Grade Test Scores, by Student Demographic Composition

	Mathematics	Reading
Class Size	-0.24	-0.13
	(2.37)	(1.52)
Class Size Interaction	-0.0052	-0.0028
with Percent Free Lunch	(4.02)	(2.71)
Class Size Interaction	0.0092	-0.0004
with Percent Black	(3.12)	(0.17)
Teacher Characteristics		
Percent 0 Years Experience	0.037	0.0018
	(1.63)	(0.10)
0 Years Experience Interaction	-0.0014	-0.0007
with Percent Free Lunch	(3.62)	(2.08)
0 Years Experience Interaction	0.0011	0.0004
with Percent Black	(1.46)	(0.60)
Percent 1 Year Experience	0.035	-0.0007
	(1.55)	(0.04)
1 Year Experience Interaction	-0.0009	-0.0001
with Percent Free Lunch	(2.22)	(0.34)
1 Year Experience Interaction	0.0003	0.0003
with Percent Black	(0.37)	(0.44)
Percent Not Fully Certified	-0.029	0.003
	(0.91)	(0.13)
Certification Interaction	0.0002	-0.0003
with Percent Free Lunch	(0.39)	(0.70)
Certification Interaction	0.0004	0.0003
with Percent Black	(0.59)	(0.57)
Percent B.A. Only	-0.003	0.004
	(0.15)	(0.24)
B.A. Only Interaction	-0.00001	-0.00003
with Percent Free Lunch	(0.03)	(0.10)
B.A. Only Interaction	-0.00001	0.0005
with Percent Black	(0.01)	(0.96)
Observations	7,626	7,612

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the school percentage of non-LEP students who exceed the 50th percentile on a nationally determined basis. All regressions include school fixed effects, percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, a dummy variable for the year 1999, and the 1999 dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. Regressions are weighted by the number of test takers. Each column represents a separate regression.

Table 6. Estimated Effects of Class Size on Third Grade Test Scores, by Timing of CSR Implementation and Student Demographic Composition
Teacher Characteristics Excluded

	Mathematics			Reading		
Interaction Group	None	% Black, % Free Lunch	% Black	None	% Black, % Free Lunch	% Black
Class Size						
Late Adopter	-0.422	-0.233	-0.479	-0.215	-0.074	-0.198
	(3.95)	(1.90)	(4.34)	(2.50)	(0.74)	(2.22)
Early Adopter	-0.365	0.122	-0.380	-0.321	0.014	-0.245
	(1.68)	(0.46)	(1.68)	(1.83)	(0.06)	(1.34)
Class Size Interaction with Percent Black						
Late Adopter		0.0109	0.0070		0.0009	-0.0011
		(3.60)	(2.39)		(0.37)	(0.46)
Early Adopter		0.0041	0.0004		-0.0062	-0.0081
		(0.63)	(0.06)		(1.19)	(1.57)
Class Size Interaction with Percent Free Lunch						
Late Adopter		-0.0061			-0.0030	
		(4.60)			(2.85)	
Early Adopter		-0.0110			-0.0053	
		(3.57)			(2.65)	
Observations	6,818	6,818	6,818	6,802	6,802	6,802

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the school percentage of non-LEP students who exceed the 50th percentile on a nationally determined basis. All regressions include school fixed effects, percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, a dummy variable for the year 1999, and the 1999 year dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. Regressions are weighted by the number of test takers. Each column represents a separate regression.

Table 7. Estimated Effects of Class Size on Third Grade Test Scores, by Timing of CSR Implementation and Student Demographic Composition
Teacher Characteristics Included

	Mathematics			Reading		
Interaction Group	None	% Black, % Free Lunch	% Black	None	% Black, % Free Lunch	% Black
Class Size						
Late Adopter	-0.434	-0.207	-0.462	-0.225	-0.087	-0.199
	(4.07)	(1.66)	(4.17)	(2.61)	(0.87)	(2.23)
Early Adopter	-0.382	0.010	-0.401	-0.359	-0.111	-0.280
	(1.75)	(0.04)	(1.77)	(2.04)	(0.50)	(1.53)
Class Size Interaction with Percent Black						
Late Adopter		0.0088	0.0049		0.0003	-0.0018
		(2.80)	(1.63)		(0.12)	(0.74)
Early Adopter		0.0047	0.0015		-0.0068	-0.0084
		(0.71)	(0.23)		(1.28)	(1.59)
Class Size Interaction with Percent Free Lunch						
Late Adopter		-0.0062			-0.0029	
		(4.53)			(2.66)	
Early Adopter		-0.0105			-0.0046	
		(3.28)			(1.78)	
Observations	6,818	6,818	6,818	6,802	6,802	6,802

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the school percentage of non-LEP students who exceed the 50th percentile on a nationally determined basis. All regressions include school fixed effects, percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, percent new teachers, percent with one year of experience, percent without full certification, percent with only a bachelor's degree, a dummy variable for the year 1999, and the 1999 year dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. Regressions are weighted by the number of test takers. Each column represents a separate regression.

Table 8. Estimated Effects of Third-Grade and Fifth-Grade Class Size on Fifth-Grade Test Scores, by Teacher Characteristics Included and Student Demographic Composition

	Mathematics				Reading			
Teacher Characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Fifth Grade Class Size	-0.24	-0.26	-0.26	-0.27	-0.07	-0.08	-0.08	-0.08
	(3.99)	(4.23)	(3.39)	(3.44)	(1.34)	(1.53)	(1.19)	(1.20)
Fifth Grade Class Size Interaction with Percent Black			0.0018	0.0003			0.0008	-0.0008
			(0.38)	(0.07)			(0.19)	(0.20)
Third Grade Class Size	0.002	-0.006	-0.031	-0.028	0.035	0.030	0.002	0.004
	(0.04)	(0.15)	(0.71)	(0.64)	(0.97)	(0.84)	(0.04)	(0.10)
Third Grade Class Size Interaction with Percent Black			0.0039	0.0029			0.0039	0.0034
			(2.13)	(1.57)			(2.51)	(2.15)
Observations	6,748	6,748	6,748	6,748	6,716	6,716	6,716	6,716

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the school percentage of non-LEP students who exceed the 50th percentile on a nationally determined basis. All regressions include school fixed effects, percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, a dummy variable for the year 1999, and the 1999 year dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. Regressions are weighted by the number of test takers. Models controlling for teacher characteristics also include percent new teachers, percent with one year of experience, percent without full certification, and percent with only a bachelor's degree. Each column represents a separate regression.

Table 9. Estimated Class Size Effects by Student Demographic Composition and Community Type

	Mathematics				Reading			
	Entire State	Los Angeles Unified	Next 5 Largest Districts	All Districts Except LA Unified	Entire State	Los Angeles Unified	Next 5 Largest Districts	All Districts Except LA Unified
Class Size	-0.21	-1.53	-0.10	-0.13	-0.10	-1.12	-0.53	-0.12
	(2.07)	(2.55)	(0.18)	(1.20)	(1.25)	(2.12)	(1.08)	(1.39)
Class Size Interaction with Pct Black	0.0103	0.0217	0.0023	0.0071	0.0005	0.0085	-0.0130	-0.0007
	(3.52)	(3.66)	(0.21)	(2.10)	(0.22)	(1.64)	(1.46)	(0.25)
Class Size Interaction with Pct Free Lunch	-0.0057	0.0130	-0.0143	-0.0066	-0.0032	0.0106	-0.0001	-0.0031
	(4.47)	(1.93)	(2.04)	(4.91)	(3.06)	(1.80)	(0.01)	(2.87)
Observations	7,626	816	644	6,810	7,612	816	646	6,796

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the percentage of third grade students who exceed the 50th percentile on a nationally determined basis. All regressions include school fixed effects, use data from 1997 and 1999, and are weighted by the number of test takers. Other regressors include percent new teachers, percent with one year of experience, percent without full certification, percent with only a bachelor's degree, percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, a dummy variable for the year 1999, and the 1999 dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. Each column represents a separate regression.

Appendix Table 1. Average Class Size for Grade 3,
by Income and Race/Ethnicity

	Percent Free Lunch	Average Class Size 90	95	97	99
Asian	0-25	29.0	29.4	22.9	20.0
	25-50	28.8	29.3	23.4	19.6
	50-75	28.6	29.2	24.1	19.6
	75-100	28.4	28.9	23.5	19.9
	All	28.8	29.2	23.4	19.8
Black	0-25	28.9	29.4	23.2	19.8
	25-50	28.3	29.5	23.1	19.6
	50-75	28.5	29.3	23.2	19.7
	75-100	27.7	28.9	23.1	19.6
	All	28.3	29.2	23.2	19.7
Hispanic	0-25	28.9	29.4	23.1	19.7
	25-50	28.8	29.4	23.6	19.8
	50-75	28.6	29.2	23.9	20.1
	75-100	27.5	28.9	24.0	20.2
	All	28.3	29.1	23.9	20.1
White	0-25	28.6	29.0	22.1	19.4
	25-50	28.4	29.1	22.9	19.6
	50-75	28.4	28.9	23.2	19.7
	75-100	28.4	29.1	23.6	19.9
	All	28.5	29.0	22.7	19.6
All Students		28.4	29.1	23.3	19.8

Notes: Schools are divided into four categories according to the percentage of students eligible for a free lunch. Then, the average class size is calculated from all the schools in that free lunch category. In the case of the final row (“All Students”), schools are weighted according to total enrollment, and the reported percentage reflects the average for all students in that income category. In the other rows the calculations weight each school by the number of students in the specific racial/ethnic group (e.g. Asians in the case of the first row) thereby creating averages for students in each of the four racial/ethnic groups.

Appendix Table 2. Salaries for Beginning Teachers and Recent Female College Graduates

	Average Salary	
Year	Beginning Teachers	Recent Female College Graduates
1993	24,598	30,400
1994	25,085	28,091
1995	25,711	31,936
1996	26,684	35,147
1997	27,852	34,078
1998	28,798	34,959
1999	30,214	38,356
2000	33,121	44,037

Note: Teacher salaries are the average starting teacher salaries, as reported by the California Teachers Association. Salaries for recent female college graduates are based on the authors' calculations from the March Current Population Survey data. Recent female college graduates are defined as women who are age 22 to 30, have a bachelor's degree or more, and work 35 or more hours a week and 40 or more weeks a year. There are roughly 100 such women in each year of the CPS. All data are for California only.

Appendix Table 3. Variable Means and Standard Deviations for Third Grade

	Mean	Std Dev
Math Test Score	50.2	23.2
Reading Test Score	44.6	22.0
Class Size	21.7	4.5
	-3.6	5.0
Teacher Characteristics		
Percent 0 Years	-6.9	21.6
Experience	9.5	16.2
Percent 1 Year	8.6	14.5
Experience	-1.0	20.3
Percent Not Fully	13.1	20.1
Certified	-0.9	20.0
Percent B.A. Only	22.4	26.3
	-1.2	23.6
Number of Observations	7,626	

Note: Bottom number for each class size and teacher characteristics refers to the school average change from 1997 to 1999. All statistics are weighted by the number of students in third grade.

Appendix Table 4. Robustness Analysis for Certification and Experience: School Fixed Effect Regression Results by Specification of Teacher Experience, Certification and Education

	Mathematics				Reading			
Class Size	-0.38	-0.41	-0.41	-0.41	-0.16	-0.27	-0.26	-0.26
	(7.50)	(5.02)	(4.99)	(5.00)	(3.92)	(4.04)	(3.97)	(3.97)
Teacher Characteristics								
Percent 0 Years			-0.030	-0.027			-0.032	-0.029
Experience			(2.72)	(2.41)			(3.58)	(3.21)
Percent 1 Year			-0.008	-0.006			-0.005	-0.003
Experience			(0.70)	(0.55)			(0.51)	(0.33)
Average		0.067				0.101		
experience		(1.54)				(2.88)		
Percent Not Fully	-0.029	-0.025			-0.022	-0.017		
Certified	(2.27)	(1.96)			(2.15)	(1.61)		
Percent Intern				-0.029				-0.028
				(1.24)				(1.48)
Percent Emergency			-0.011	-0.016			-0.003	-0.007
or Waiver			(0.82)	(1.10)			(0.28)	(0.63)
Percent B.A. Only	-0.008	-0.005	-0.005	-0.003	0.005	0.009	0.007	0.009
	(0.82)	(0.49)	(0.54)	(0.33)	(0.60)	(1.15)	(0.91)	(1.14)
Observations	7,626	7,626	7,626	7,626	7,612	7,612	7,612	7,612

Notes: Absolute value of t-statistics are in parentheses. The dependent variable is the school percentage of non-LEP students who exceed the 50th percentile on a nationally determined basis. All regressions include school fixed effects, percent black, percent Hispanic, percent Asian, percent free lunch, percent LEP, a dummy variable for the year 1999, and the 1999 dummy variable interacted with a dummy variable for schools adopting CSR in 1998 or 1999. Regressions are weighted by the number of test takers. Each column represents a separate regression.