What Is the Tradeoff Between Smaller Classes and Teacher Quality?

March 2004

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

and

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

Abstract: This paper investigates the effects of California's class size reduction program on teacher quality and student achievement to gain a better understanding of the impact of a large-scale decrease in class size. It uses year-to-year differences in class size generated by both variation in enrollment and the state's class size reduction program and seeks to identify both the direct effects of smaller classes and any related changes in teacher quality. The results show that smaller classes raise mathematics and reading achievement and that the effects are larger in the earlier grades. The need to hire large numbers of new teachers, many of whom lacked full certification, did reduce the average quality of instruction following the implementation of class size reduction. However, there is not strong evidence of a longer-term decline in the quality of instruction.

---

* Any opinions expressed are those of the authors alone and do not necessarily reflect any position of the Public Policy Institute of California. 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, Society of Labor Economists, 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 Cerdán and Jackie Chou for excellent research assistance.

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), these studies took great pains to hold all other factors constant. Yet as Heckman (2001) emphasizes, it is important to consider both the direct and indirect effects of a policy in order to gain a comprehensive understanding of its impact. In the case of a large-scale reduction in class size, the possibility of a dramatic change in teacher quality at many schools is quite real. In addition to adding many rookie teachers, a class size reduction induced expansion of the teaching force could 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. In fact, Ross (1999) describes just such an 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.

This paper investigates the effects of the California CSR program on student achievement with special attention to changes in teacher quality. We begin by documenting changes in the distribution of elementary school teacher characteristics following the implementation of CSR in 1996. Differences by student demographic composition receive particular attention, given the evidence that schools serving substantial numbers of minority and low socioeconomic status (SES) students appear to

have more difficulty attracting and retaining teachers. Following this description we turn to an analysis of the effects of CSR on achievement, both the direct benefits of smaller classes and any accompanying changes in 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. Evidence strongly suggests that the implementation of CSR also affected the composition of teachers and student achievement in grades not directly affected by CSR. Moreover, there is no attempt to disentangle any adjustment costs from the longer-term benefits of smaller classes.

Our approach removes school by year and school by grade fixed effects and uses year-to-year variations in class size and teacher characteristics for the years 1997 to 2001 to estimate the effects of smaller classes and teacher characteristics on mathematics and

reading achievement. The class size coefficients can be multiplied by the average reduction in class size under CSR to obtain estimates of the direct gains from reducing class size, while coefficients on teacher certification and experience variables can be used to estimate any temporary decline in the quality of instruction related to the hiring of large numbers of inexperienced teachers, many of whom lacked full certification.[2] In addition, differences in the size of entering teacher cohorts can be used to estimate the elasticity of the quality of instruction with respect to the number of teachers hired, providing a means of estimating longer-term changes in the quality of instruction related to the need to hire so many new teachers.

The results show that other things equal, smaller classes raise mathematics and reading achievement and that the effects appear to be larger in the earlier grades. On the other hand, the dramatic increase in the number of rookie teachers, particularly those lacking full certification, offset some of the benefits of smaller classes in the short term. However, there is not strong evidence of a longer-term decline in the quality of instruction caused by the need to hire large numbers of new teachers following the implementation of CSR.

The paper is structured as follows. Section 2 describes the data set and test score instrument used in California. Section 3 documents changes over time in a number of teacher characteristics following the implementation of CSR. Section 4 develops the empirical model used to investigate the achievement effects of CSR. Section 5 reports the

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

results, focusing on any related changes in teacher quality. 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 serving third grade in California with the exception of charter and alternative schools (such as special education schools). Seven years of data are used: 1990-1991, 1995-1996, and 1997-1998 through 2001-2002 (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, published by Harcourt, Brace & Co. Therefore, the analysis of student achievement is limited to the years 1997 through 2001. We use average mathematics and reading test scores in the empirical analysis.

3. California Public Elementary School Teachers: 1990-2001

The implementation of CSR led to dramatic changes in the size and composition of the elementary school teaching force in California public schools. The need to hire large numbers of teachers in a very short period led to a dramatic increase in the share of teachers without experience and full certification. Whether this portends a long-term

decline in teacher quality depends upon the links between quality, certification, and teacher cohort size. This section describes changes in teacher experience and certification following the implementation of CSR.[3] Prior to describing changes in teacher characteristics following the implementation of CSR, we draw upon previously developed models of teacher labor supply and demand in order to highlight issues particularly germane to California.

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. 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 determinant of teacher quality, though there is a serious debate over its impact. Though education and training requirements in combination with any certification examinations may improve skills and eliminate some potentially bad teachers, they may also discourage or even prevent others from entering the profession by raising the opportunity cost of teaching. In addition, some potentially effective teachers may be ruled out by examinations that focus on only a subset of the skills used in teaching.

[3] 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 year teachers or divide schools on the basis of student race/ethnicity and income.

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) $D_s = f(\text{skill, certification, salary, budget, other factors})$

Schools are assumed to rank order teachers on the basis of expected classroom effectiveness (i.e. skill), certification status, and other factors possibly unrelated to instructional effectiveness. 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) $S_s = f(\text{salary \& amenities}_s, \text{salary \& amenities}_{\text{other schools}}, \text{salary \& amenities}_{\text{other occupations}}, \text{distance}_s, \text{certification})$

The magnitude and distribution of any changes in teacher characteristics following the CSR driven increase in demand clearly depends upon the details of class size reduction, any commensurate increases in salary, and the elasticities of teacher supply for different types of schools. In terms of the first, Jepsen and Rivkin (2002) show that CSR reduced

[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, Rivkin, and Kain 2002). Recent

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 1 shows changes over time in average earnings of 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

---

work by Lankford, Loeb, and Wyckoff also (2002) highlights the very localized nature of teacher labor markets and strong preference for shorter commutes.

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 and Certification

This section describes changes over time in teacher experience and certification by student demographic characteristics. The figures for experience and certification are constructed in the following way. Schools are divided into four categories according to the percentage of students eligible for a free lunch.[5] 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. Low poverty schools are defined as schools with less than 25 percent free lunch, whereas high poverty schools are defined as schools with 75 percent or more free lunch. 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.[6] 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.[7]

[5] 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.
[6] The calculations assume that there is no systematic variation within schools by race or ethnicity in the probability of having inexperienced or uncertified teachers.
[7] 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.

Figure 1 describes the percentages of new teachers grades 2 through 5 (elementary school grades with standardized tests).[8] A clear pattern in teacher experience based on percent free lunch exists throughout the time period: the percent of new teachers increases as percent free lunch increases for all 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, the percentage of teachers with no experience increased by over 8 percentage points for nonwhites (increases of 100 percent or more), compared with increases of just over 5 percentage points for nonwhites in low-poverty schools.

Figure 2 reports the percentage of teachers who lack full certification. In 1990, nearly all teachers are certified, so 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 and 6 percent for low poverty schools, compared with a range of 12 to 28 percent for high-poverty schools.

Non-certification rates remain at high levels in 2001, especially for nonwhites attending high-poverty schools. 21 percent of Hispanic students and 25 percent of black students in high-poverty schools have teachers who lack full certification, compared with rates of only 10 percent for white students and 13 for Asian students.

[8] Rivkin, Hanushek, and Kain (2000) find that gains to experience are concentrated in the first few years of teaching.

4. Empirical Model

California elementary school 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. Though there are concerns about some aspects of the randomization of students and teachers (c.f. Hoxby (2000)), the findings are generally consistent with other recent studies using non-experimental methods.

Work 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 the data structure attenuates the estimates in Hoxby (2000), the overall pattern of results suggests that smaller classes raise achievement, other things held constant.[9]

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 and may face difficult facility constraints. Consequently the controlled experiment provides only partial information on the likely effect of reducing class size. Depending upon the elasticity of supply of teacher quality, facility characteristics and other conditions at a school, the existing estimates may seriously overstate the likely benefits of CSR, 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.[10] 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 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.

[9] 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

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 in Jepsen and Rivkin (2002) 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 and the transactions costs experienced during the implementation period.

Basic Model

The empirical model used in this paper is quite similar to that used by Rivkin, Hanushek and Kain (2000) with two main exceptions. First, microdata are not available, therefore the data are aggregated to the school by grade by year level. Second, students are not followed as they progress through school, so it is not possible to investigate directly the rate of learning. However, because grade and year specific random variation in class size

---

attenuates the estimates of class size effects on achievement by reducing the covariance between achievement and class size.

and teacher characteristics is used to identify the variable effects, the regression coefficients should capture the impacts on learning over a single school year.

Equation (3) describes average student achievement (A_{sgy}) for in grade *g* in school *s* in year *y* as a function of demographic variables (*X*); class size (*CS*); teacher characteristics (*TC*); and a composite error term that includes a school by grade fixed effect (θ_{sg}), a school by year fixed effect (θ_{sy}), a year by grade fixed effect (d_{gy}), and a stochastic term (ε_{gsy}).

$$(3) \qquad A_{gsy} = X_{gsy}\alpha_X + CS_{gsy}\alpha_{CS} + TC_{gsy}\alpha_{TC} + d_{gy} + \theta_{sg} + \theta_{sy} + \varepsilon_{gsy}$$

The vector TC includes the percentage of teachers in their first year, the percentage in their second year, and the percentage not fully certified, and the vector X includes percentage black, percentage Hispanic, percentage Asian, percentage eligible for a free lunch, and percentage LEP.[11] Year by grade dummy variables for each cohort (d_{gy}) capture any upward drift in the test score as well as the effects of other state policies and systematic changes over time. Finally, the error term contains three components: school by year and school by grade fixed effects and a random error.

Because the limited set of explanatory variables omits many important current and past determinants of achievement, OLS estimation of equation (3) 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 the multiple years and grades to estimate a set of successively more comprehensive fixed effects models. The first controls

[10] See Bohrnstedt and Stecher (1999) and Stecher and Bohrnstedt (2000).
[11] Due to data constraints, we use the school-level average for percent free lunch. All other variables are available at the grade level.

only for school fixed effects, which removes any time invariant school characteristics and school average family influences that are constant over time.

However, any changes over time in school or neighborhood not captured by the time varying explanatory variables will contaminate the results. Therefore, we include school by year fixed effects in order to account for such changes. Yet even school by year fixed effects may not generate consistent estimates if there are systematic achievement patterns as students progress through school which are related to grade differences in class size or teacher characteristics. Consequently, we add a full set of school by grade fixed effects in order to account for such differences. This leaves only differences by school, grade, and year to identify the class size and teacher characteristic effects, and there is little reason to believe that such differences are systematically related to other factors that determine achievement. Nevertheless, we also include a set of student demographic characteristics that vary by grade and year to control for any remaining confounding factors.

The full fixed effects model should generate consistent estimates of class size and average teacher characteristic effects. However, interpretation of the teacher characteristic effects is complicated by the substantial changes in the size of the California teaching force during this period. Consider first the experience parameters. In a steady state environment of a fixed number of new entrants annually, the teacher experience parameters would combine the effects of on the job learning with composition changes due to non-random exits from the profession and would be stable over time. A large change in the size of the entering teacher cohort, however, could cause the composition of new teachers to diverge from that in prior years as long as salaries do not adjust to maintain constant quality or the

supply of teacher quality is not perfectly elastic. In this case the experience parameters would be expected to change over time in inverse relationship to cohort size. Notice that this same line of reasoning holds for the certification variable.

We will estimate separate experience and certification coefficients for each year in order to investigate the sensitivity of cohort quality to cohort size. Comparisons of these estimates with year-to-year changes in cohort size will provide information on the supply elasticity of teacher quality with respect to cohort size.

5. Class Size, Teacher Characteristics, and Student Achievement

This section reports the results from the analysis of average mathematics and reading achievement. Unless otherwise indicated, all regression specifications include the percentages of free lunch, black, Hispanic, Asian, and LEP students and full sets of grade by year dummies. Descriptive statistics are reported in Appendix Table 2. Prior research showing larger class size effects in earlier grades leads us to fully interact class size and grade.[12]

Baseline Results

Table 1 reports the estimated effects of class size and teacher characteristics for six specifications that differ according to the included covariates and fixed effects. Column 1 includes no fixed effects, Column 2 includes school fixed effects, Column 3 includes school by grade fixed effects, Column 4 includes school by year fixed effects, and

[12] Krueger (1999) and Rivkin, Hanushek, and Kain (2000) are two examples of papers that report such a pattern.

Columns 5 and 6 include both school by grade and school by year fixed effects. All specifications except for Column 6 include a full set of student demographic variables.

The estimated effects of class size on mathematics achievement are negative and highly significant across the board, though effect magnitudes do vary by specification. Importantly, the class size estimates are virtually unchanged by the exclusion of the highly significant student demographic variables (see Appendix Table 3), indicating that the school by grade and school by year fixed effects remove any confounding factors.

Table 1 reveals smaller effects for 4th and 5th grade than for 2nd and 3rd grade, though the magnitude of the class size effect does not decline monotonically as students age. One possible explanation for the smaller 2nd grade effect is downward bias from measurement error. Because virtually all schools participated in the CSR program in 2nd grade in all the years, there is far less class size variation in 2nd grade than in any of the other grades (see Appendix Table 4), and the lack of variation may reduce the signal to noise ratio and exacerbate any measurement error.

The right panel results for reading suggest that class size exerts a much smaller effect on reading achievement, particularly beyond grade 2.[13] In fact the hypothesis that class size does not affect reading achievement in the 5th grade cannot be rejected at any conventional significance level once school by grade fixed effects are included.

Table 1 also reveals significant effects of teacher experience and certification on achievement that unlike class size are quite similar in magnitude in the two subjects. Grade interactions are not included based upon other work and preliminary regressions that did

not reveal a systematic pattern of effects by grade. In terms of effect magnitudes, the penalty for having a first year teacher in terms of mathematics achievement is slightly higher than the gain from the ten student reduction in class size in 3rd grade, roughly double the gain in 2nd grade, and three or four times the gain in 5th grade. The much smaller benefits of class size reduction in reading lead to a smaller benefit of CSR relative to the cost of having an inexperienced teacher. The costs of having a second year teacher are much smaller and roughly comparable to having a teacher who lacked full certification. Thus it appears that the learning curve is quite steep in the first year and flattens out rather quickly.

An important issue concerns the source of the certification effect. Teachers lacking full certification may be less talented teachers on average than those with full certification, but they also have less coursework and classroom experience as student teachers prior to entering the classroom for the first time as a rookie teacher. If the inferior performance comes primarily from the lack of coursework and experience, experience should dissipate the fully certified/not fully certified differential.

In order to investigate the source of the certification differential we redefine the teacher characteristics by interacting certification and experience. Table 2 reports the estimated effects of having a new and a second year teacher with full certification and a new and a second year teacher without full certification (the class size estimates are virtually unchanged by the reparameterization of the certification and experience variables). The estimates provide some support for the belief that lack of experience and coursework constitutes the major source of disadvantage for teachers lacking full

[13] In contrast, Rivkin, Hanushek, and Kain (2000) find that class size impacts on reading are roughly 70

certification, at least for those teachers who return for their second year of teaching. The large gap between certified and not fully certified teachers in the first year declines substantially by the second year, and this holds despite the fact that many teachers become certified between their first and second years. Of course the gap would also be reduced by the departure of the worst of the non-certified teachers prior to their second year, and more information is needed to know for certain. Nonetheless, the certification gap basically disappears by the second year.

Cohort Size and Teacher Quality

This section uses differences in the size of entering teacher cohorts to estimate the elasticity of the quality of instruction with respect to the number of teachers hired. In most labor market analyses it is quite difficult to separate any change in cohort quality from changes in labor supply, because both lead to changes in average wages. However, in this case we have a direct measure of cohort quality that can be used to identify any changes.

Table 3 reports changes over time in the percentages of fully certified and not fully certified first and second year teachers. Note that all specifications also include full sets of certification by year variables for those with one year of experience, class size by grade variables and student demographic variables. The tables also report the sizes of entering cohorts in each year.

The results reveal little systematic relationship between cohort size and teacher quality (as measured by student achievement). If anything, both the mathematics and reading estimates suggest that the quality of not fully certified first-year teachers is worse

---

percent as large as those on mathematics.

in 2001 (the smallest teacher cohort) than in 1997 (the largest). And there is no systematic pattern whatever for new teachers with full certification.

One potential problem with the analysis presented in Table 3 is that schools exert some discretion in hiring regarding certification. Consequently Table 4 groups both certified and non-certified teachers together and ignores differences in certification. The results in Table 4 again show no strong relationship between cohort size and teacher quality, especially for first-year teachers. In fact the coefficient on percent first-year teachers is larger (in absolute terms) in 2001 than in 1997.

6. Summary and Policy Implications

The implementation of CSR brought dramatic changes to California elementary school teachers and classrooms. Average class size fell by roughly one third, from 30 to 20. This necessitated a huge expansion of the teaching force following the implementation of CSR in 1996 including the hiring of large numbers of teachers without full certification. This expansion affected all elementary grades, not just those with smaller classes. Although the increase in the number of teachers with little or no experience was a temporary phenomenon, the share of teachers without full certification today remains far above its pre-CSR level. Moreover, these teachers are concentrated in schools serving primarily economically disadvantaged and nonwhite students.

The rapid expansion in the number of teachers and consequent increase in the share lacking full certification raise the possibility of a deterioration in teacher quality that could have offset some of the benefits of the now smaller classes. This decline in teacher quality

is of particular concern for non-white students in high-poverty schools, where new, not-certified teachers were concentrated.

Figure 3 simulates the CSR induced decline in teacher quality using the coefficients on the teacher characteristics from Table 2 and the changes following the implementation of CSR in the shares of teachers falling into the various experience/certification categories for four subsets of students: Asians attending schools with fewer than 25% of students classified as economically disadvantaged; whites attending schools with fewer than 25% of students classified as economically disadvantaged; blacks attending schools with greater than 75% of students classified as economically disadvantaged; and Hispanics attending schools with greater than 75% of students classified as economically disadvantaged (changes are calculated relative to 1995). Notice that the effects of experience and certification are fixed over time due to the lack of systematic changes; therefore the changes over time result entirely from changes in the experience/certification distribution. Moreover, since preliminary work revealed no systematic differences by demographic group in the benefits of smaller classes and the student income variable did not vary by grade, we assign all students the same benefit of smaller classes.

The Figure reveals quite clearly the disproportionate burden borne by Black and Hispanic and economically disadvantaged students in terms of the decline in teacher quality. The CSR induced decline in teacher quality was roughly one sixth as large in mathematics and one fourth as large in reading as the benefit of reducing class size by ten students for blacks and Hispanics attending high poverty schools in 1997. By comparison, the teacher quality decline for whites in 1997 was roughly half as large. Although the decline in the share of rookie teachers and those lacking full certification reduces the

impact of the teacher quality decline as of 2001, quality remains substantially below the pre-CSR level even as late as 2001, particularly for nonwhite students. Moreover, the trends suggest that the share of new/uncertified teachers in high poverty schools primarily serving blacks and Hispanics may remain above the 1995 levels for the foreseeable future.

It is clear that the need to hire large numbers of new teachers reduced the average quality of instruction and that the effects were concentrated in schools serving minority and economically disadvantaged students. If these schools continue to require large numbers of new teachers lacking full certification the gains to smaller classes the decline in teacher quality will continue to offset benefits of smaller classes. However, there is little or no evidence that the average quality of instruction is linked closely with cohort size, meaning that it is unlikely that one of the legacies of CSR will be cohorts of teachers who perform markedly worse than their predecessors.

The unintended consequences of class-size reduction could probably be reduced if the program were implemented more gradually. 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 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.

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

Hanushek, Eric (1986), "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24(3): 1141-1177.

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.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff (2002), "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis," *Educational Evaluation and Policy Analysis,* 24(1), 37-62.

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: Effects of Class Size and Teacher Experience on Student Achievement

	Mathematics						Reading					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Grade 2 class size	-0.174	-0.259	-0.248	-0.254	-0.224	-0.212	-0.213	-0.212	-0.253	-0.19	-0.214	-0.201
	(3.41)	(7.02)	(6.39)	(6.43)	(5.25)	(4.92)	(4.74)	(6.62)	(7.46)	(5.58)	(5.66)	(5.24)
Grade 3 class size	-0.396	-0.391	-0.392	-0.352	-0.306	-0.302	-0.275	-0.228	-0.156	-0.232	-0.132	-0.127
	(16.26)	(21.86)	(19.07)	(18.85)	(13.67)	(13.31)	(12.83)	(14.70)	(8.68)	(14.44)	(6.65)	(6.29)
Grade 4 class size	-0.037	-0.16	-0.115	-0.188	-0.136	-0.127	-0.114	-0.107	-0.068	-0.146	-0.095	-0.084
	(1.43)	(8.36)	(5.16)	(9.33)	(5.56)	(5.15)	(5.03)	(6.47)	(3.47)	(8.42)	(4.40)	(3.84)
Grade 5 class size	-0.084	-0.198	-0.096	-0.220	-0.094	-0.089	-0.134	-0.116	-0.03	-0.133	-0.027	-0.021
	(3.14)	(9.86)	(4.13)	(10.39)	(3.67)	(3.44)	(5.66)	(6.65)	(1.47)	(7.33)	(1.20)	(0.91)
Percent first-year teachers	-0.046	-0.034	-0.032	-0.030	-0.027	-0.027	-0.042	-0.030	-0.027	-0.026	-0.020	-0.021
	(17.83)	(17.50)	(16.37)	(14.28)	(12.19)	(12.22)	(18.56)	(17.76)	(15.51)	(13.99)	(10.37)	(10.38)
Percent second-year	-0.024	-0.014	-0.011	-0.011	-0.006	-0.006	-0.029	-0.015	-0.011	-0.015	-0.009	-0.008
teachers	(9.42)	(7.36)	(5.69)	(5.38)	(3.00)	(2.98)	(12.49)	(9.38)	(6.65)	(8.18)	(4.60)	(4.45)
Percent teachers not	0.01	-0.009	-0.009	-0.009	-0.009	-0.009	0.002	-0.007	-0.008	-0.006	-0.007	-0.008
certified	(4.33)	(4.41)	(4.08)	(4.30)	(3.71)	(3.93)	(1.13)	(4.16)	(4.37)	(3.56)	(3.44)	(3.80)
Observations	90,503	90,503	90,503	90,503	90,503	90,503	90,492	90,492	90,492	90,492	90,492	90,492
School FE	no	yes	no	no	no	no	no	yes	no	no	no	no
School by grade FE	no	no	yes	no	yes	yes	no	no	yes	no	yes	yes
School by year FE	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes
Student characteristics	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	no

Notes: Absolute value of t-statistics is in parentheses. The dependent variable is the average test score in that subject. All regressions include grade by year fixed effects. Student characteristics are percent black, percent Hispanic, percent Asian, percent LEP and school percent free lunch. Coefficients and t-statistics for student demographics are in Appendix Table 3. Regressions are weighted by average school enrollment for 1997-2001. Each column represents a separate regression.

Table 2: Effects of Class Size, and Teacher Experience * Teacher Certification on Student Achievement

	Mathematics						Reading					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Grade 2 class size	-0.171	-0.256	-0.244	-0.254	-0.223	-0.211	-0.209	-0.209	-0.248	-0.190	-0.213	-0.201
	(3.37)	(6.94)	(6.28)	(6.42)	(5.23)	(4.90)	(4.67)	(6.52)	(7.32)	(5.58)	(5.65)	(5.23)
Grade 3 class size	-0.395	-0.390	-0.390	-0.352	-0.306	-0.301	-0.274	-0.227	-0.154	-0.232	-0.131	-0.126
	(16.22)	(21.83)	(18.97)	(18.87)	(13.63)	(13.27)	(12.79)	(14.67)	(8.57)	(14.45)	(6.63)	(6.27)
Grade 4 class size	-0.035	-0.160	-0.115	-0.188	-0.136	-0.127	-0.113	-0.107	-0.067	-0.146	-0.095	-0.084
	(1.37)	(8.36)	(5.16)	(9.33)	(5.57)	(5.16)	(4.98)	(6.46)	(3.46)	(8.40)	(4.40)	(3.84)
Grade 5 class size	-0.082	-0.197	-0.094	-0.219	-0.092	-0.087	-0.132	-0.115	-0.028	-0.133	-0.026	-0.020
	(3.07)	(9.81)	(4.04)	(10.37)	(3.60)	(3.37)	(5.59)	(6.60)	(1.37)	(7.31)	(1.16)	(0.86)
Percent of not-certified	-0.045	-0.044	-0.044	-0.040	-0.037	-0.038	-0.050	-0.041	-0.040	-0.031	-0.027	-0.027
first-year teachers	(12.72)	(16.87)	(16.51)	(13.75)	(12.44)	(12.53)	(15.95)	(18.00)	(17.53)	(12.36)	(10.14)	(10.27)
Percent of certified	-0.041	-0.030	-0.027	-0.028	-0.023	-0.024	-0.036	-0.025	-0.020	-0.025	-0.019	-0.019
first-year teachers	(12.71)	(12.40)	(10.96)	(10.62)	(8.60)	(8.68)	(12.64)	(11.96)	(9.36)	(11.14)	(7.88)	(7.95)
Percent of not-certified	-0.013	-0.020	-0.016	-0.016	-0.009	-0.009	-0.026	-0.023	-0.020	-0.017	-0.011	-0.012
second-year teachers	(3.06)	(6.63)	(5.33)	(4.81)	(2.62)	(2.73)	(7.34)	(8.83)	(7.77)	(6.10)	(3.81)	(3.95)
Percent of certified	-0.027	-0.014	-0.010	-0.012	-0.008	-0.008	-0.030	-0.014	-0.009	-0.016	-0.009	-0.009
second-year teachers	(8.76)	(6.32)	(4.68)	(5.16)	(3.16)	(3.09)	(11.11)	(7.51)	(4.68)	(7.51)	(4.28)	(4.08)
Observations	90,503	90,503	90,503	90,503	90,503	90,503	90,492	90,492	90,492	90,492	90,492	90,492
School FE	no	yes	no	no	no	no	no	yes	no	no	no	no
School by grade FE	no	no	yes	no	yes	yes	no	no	yes	no	yes	yes
School by year FE	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes
Student characteristics	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	no

Notes: Absolute value of t-statistics is in parentheses. The dependent variable is the average test score in that subject. All regressions include grade by year fixed effects. Student characteristics are percent black, percent Hispanic, percent Asian, percent LEP and school percent free lunch. Regressions are weighted by average school enrollment for 1997-2001. Each column represents a separate regression.

Table 3: Estimated Effects of Teacher Experience * Certification by Year

	Mathematics			Reading			Number of
	(1)	(2)	(3)	(1)	(2)	(3)	Teachers
Pct 1st year, not certified - 1997	-0.039	-0.034	-0.033	-0.059	-0.032	-0.027	3,849
	(8.68)	(6.40)	(6.11)	(15.02)	(6.92)	(5.62)	
Pct 1st year, not certified - 1998	-0.053	-0.035	-0.034	-0.057	-0.027	-0.019	3,421
	(10.24)	(5.76)	(5.48)	(12.56)	(5.21)	(3.48)	
Pct 1st year, not certified - 1999	-0.069	-0.047	-0.044	-0.041	-0.032	-0.025	2,900
	(12.20)	(7.11)	(6.55)	(8.24)	(5.66)	(4.26)	
Pct 1st year, not certified - 2000	-0.029	-0.04	-0.036	-0.008	-0.02	-0.024	2,379
	(4.58)	(5.52)	(4.91)	(1.55)	(3.23)	(3.66)	
Pct 1st year, not certified - 2001	-0.016	-0.047	-0.043	0.002	-0.044	-0.045	2,150
	(2.29)	(5.99)	(5.41)	(0.42)	(6.48)	(6.33)	
Pct 1st year, certified - 1997	-0.028	-0.024	-0.017	-0.019	-0.026	-0.017	3,563
	(5.99)	(4.55)	(3.19)	(4.75)	(5.69)	(3.51)	
Pct 1st year, certified - 1998	-0.025	-0.023	-0.025	-0.023	-0.027	-0.022	3,174
	(5.05)	(4.09)	(4.34)	(5.17)	(5.53)	(4.28)	
Pct 1st year, certified - 1999	-0.024	-0.026	-0.022	-0.024	-0.028	-0.025	2,772
	(4.42)	(4.22)	(3.54)	(5.04)	(5.17)	(4.51)	
Pct 1st year, certified - 2000	-0.026	-0.035	-0.023	-0.013	-0.025	-0.009	3,263
	(5.10)	(6.01)	(3.80)	(3.04)	(4.91)	(1.75)	
Pct 1st year, certified - 2001	-0.03	-0.032	-0.032	-0.024	-0.019	-0.021	2,785
	(5.34)	(4.96)	(4.83)	(4.80)	(3.43)	(3.70)	
Observations	90,503	90,503	90,503	90,492	90,492	90,492	
School by grade fixed effects	yes	no	yes	yes	no	yes	
School by year fixed effects	no	yes	yes	no	yes	yes	

Notes: Absolute value of t-statistics is in parentheses. The dependent variable is the average test score in that subject. All regressions include class size by grade, percent of second-year teachers by certification status and year, percent black, percent Hispanic, percent Asian, percent LEP, school percent free lunch, and grade by year fixed effects. Regressions are weighted by average school enrollment for 1997-2001. Each column represents a separate regression. Number of teachers refers to the number of 2nd through 5th grade teachers in that experience and certification category for that year.

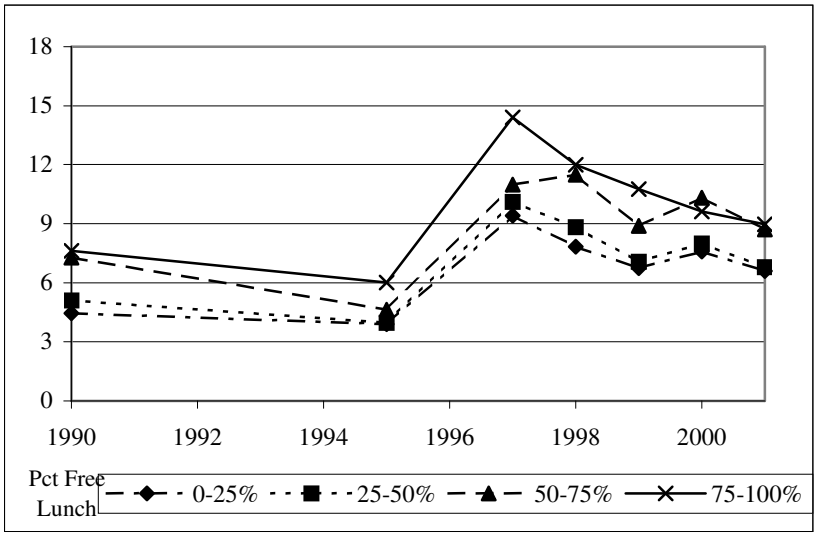
Table 4: Effect of Teacher Experience by Year

	Mathematics			Reading			Number of
	(1)	(2)	(3)	(1)	(2)	(3)	Teachers
Pct 1st year - 1997	-0.035	-0.031	-0.027	-0.042	-0.029	-0.023	7,412
	(10.55)	(7.93)	(6.84)	(14.38)	(8.88)	(6.63)	
Pct 1st year - 1998	-0.039	-0.029	-0.029	-0.04	-0.027	-0.020	6,595
	(10.56)	(6.73)	(6.74)	(12.29)	(7.30)	(5.27)	
Pct 1st year - 1999	-0.047	-0.036	-0.032	-0.031	-0.03	-0.025	5,672
	(11.53)	(7.72)	(6.84)	(8.89)	(7.44)	(6.02)	
Pct 1st year - 2000	-0.028	-0.038	-0.027	-0.01	-0.024	-0.015	5,643
	(6.91)	(8.16)	(5.86)	(2.73)	(6.14)	(3.71)	
Pct 1st year - 2001	-0.024	-0.038	-0.036	-0.011	-0.029	-0.030	4,935
	(5.39)	(7.47)	(7.00)	(3.00)	(6.60)	(6.71)	
Pct 2nd year - 1997	-0.018	-0.013	-0.009	-0.023	-0.019	-0.015	5,906
	(4.63)	(2.99)	(2.09)	(6.94)	(5.11)	(3.96)	
Pct 2nd year - 1998	-0.019	-0.012	-0.008	-0.027	-0.015	-0.010	7,422
	(5.16)	(2.98)	(2.01)	(8.41)	(4.22)	(2.64)	
Pct 2nd year - 1999	-0.019	-0.015	-0.009	-0.019	-0.02	-0.012	6,407
	(4.99)	(3.35)	(2.05)	(5.54)	(5.26)	(3.06)	
Pct 2nd year - 2000	-0.007	-0.012	-0.005	0.004	-0.012	-0.005	5,276
	(1.60)	(2.48)	(1.12)	(1.01)	(2.86)	(1.16)	
Pct 2nd year - 2001	0.003	-0.016	-0.008	0.008	-0.015	-0.006	5,563
	(0.77)	(3.30)	(1.69)	(2.24)	(3.56)	(1.47)	
Observations	90,503	90,503	90,503	90,492	90,492	90,492	
School by grade fixed effects	yes	no	yes	yes	no	yes	
School by year fixed effects	no	yes	yes	no	yes	yes	

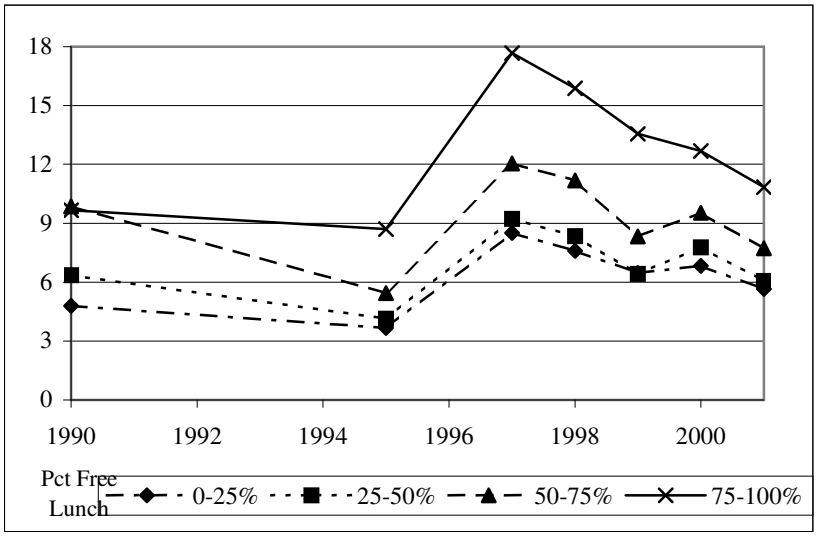
Notes: Absolute value of t-statistics is in parentheses. The dependent variable is the average test score in that subject. All regressions include class size by grade, percent black, percent Hispanic, percent Asian, percent LEP, school percent free lunch, and grade by year fixed effects. Regressions are weighted by average school enrollment for 1997-2001. Each column represents a separate regression. Number of teachers refers to the number of 2nd through 5th grade teachers in that experience category for that year.

Figure 1: Percent New Teachers by Race/Ethnicity, School Income and Year

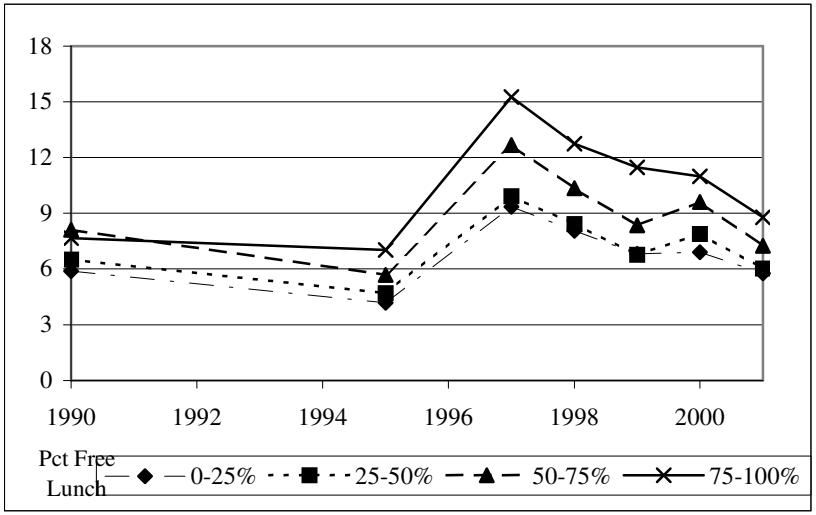
Asian



Black



Hispanic



White

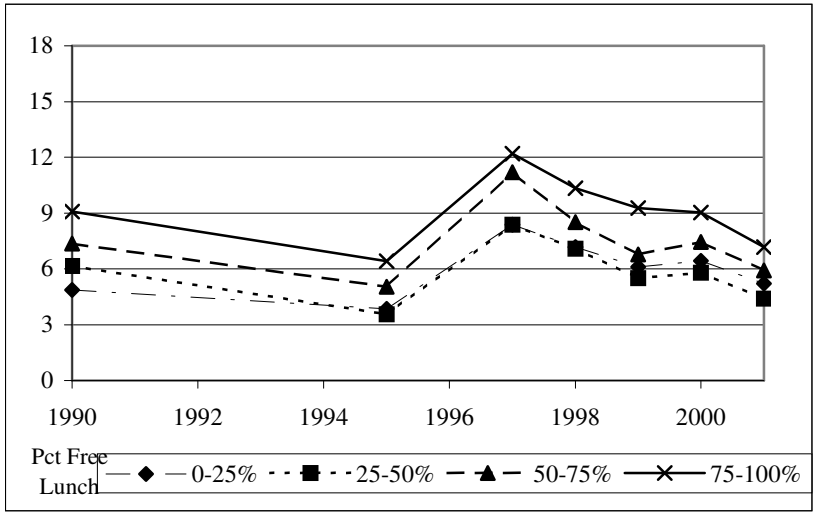
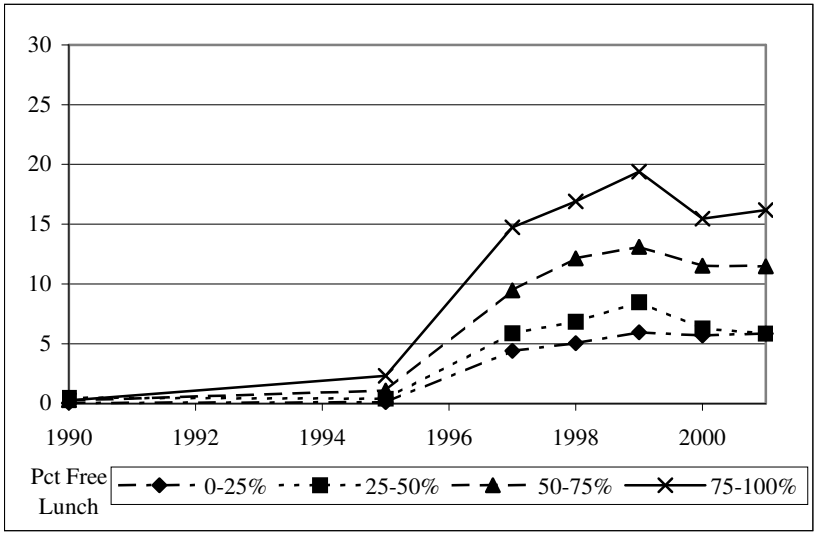
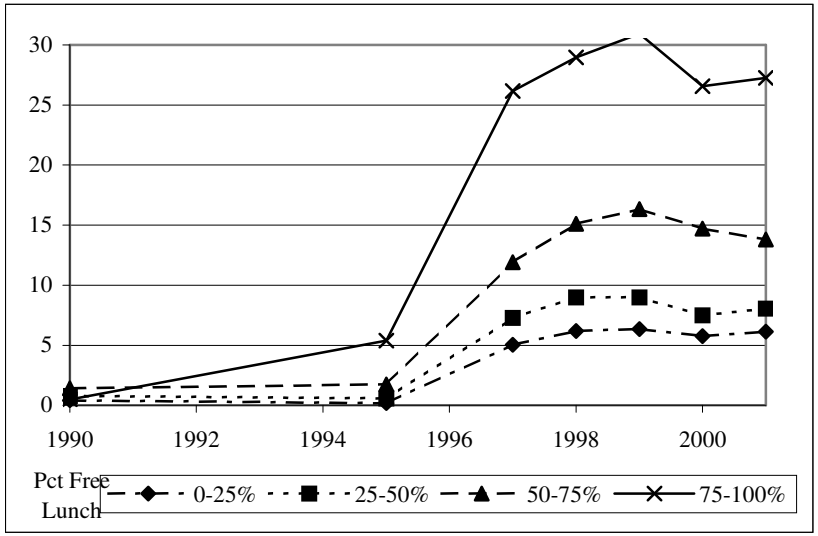


Figure 2: Percent Not Certified, by Race/Ethnicity, School Income, and Year

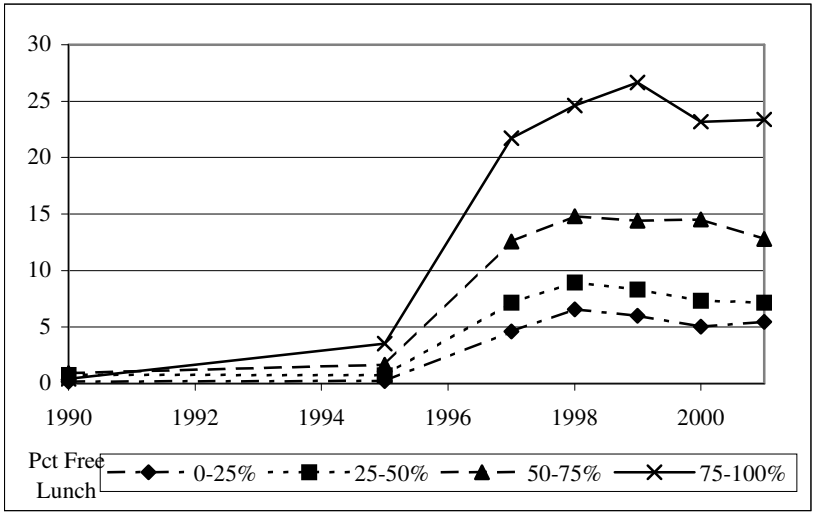
Asian



Black



Hispanic



White

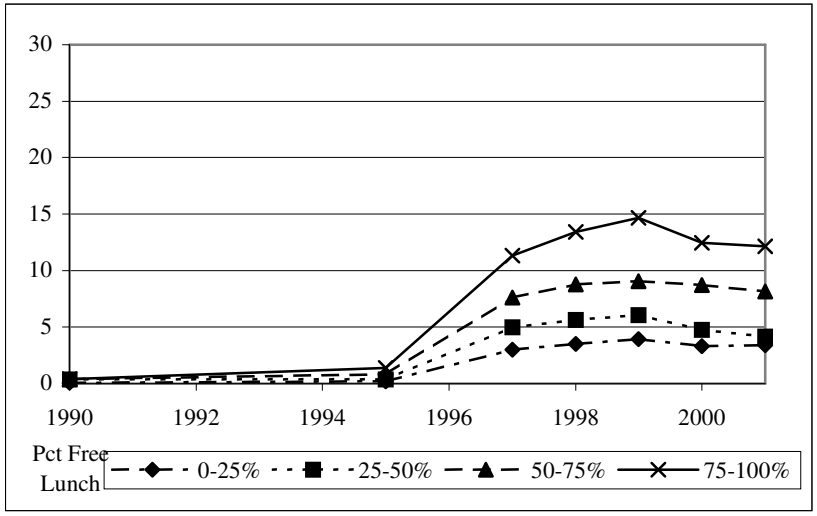
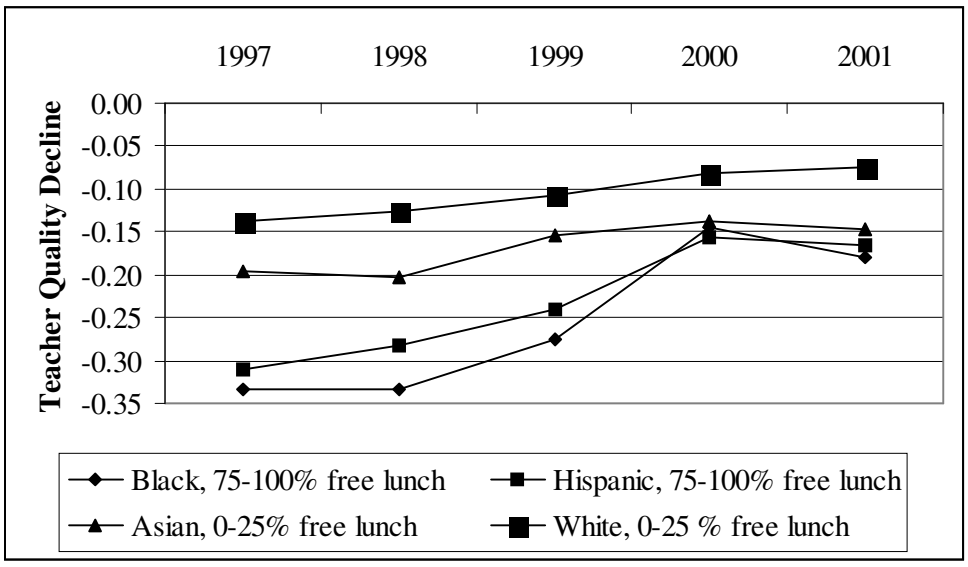
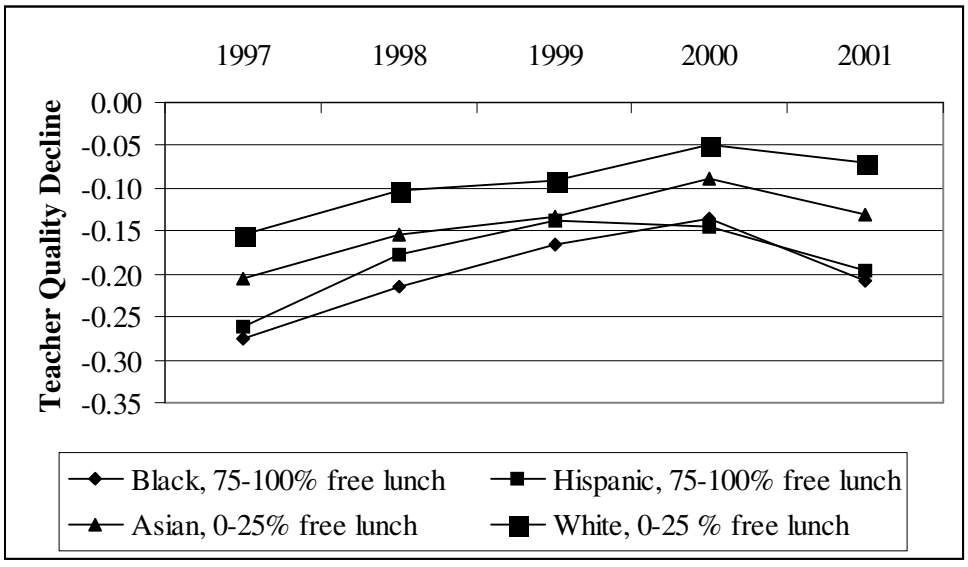


Figure 3: Simulated Decline in Teacher Quality, by Race/Ethnicity, School Income, and Year

Math



Reading



Appendix Table 1. Salaries for Beginning Teachers and Recent Female College Graduates

Year	Average Salary: Beginning Teachers	Average Salary: 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 2: Descriptive Statistics

	All years	1997	1998	1999	2000	2001
Average test scores						
Mathematics test	612.5	600.4	607.0	614.1	618.6	621.9
	(34.11)	(34.49)	(33.19)	(33.73)	(32.66)	(32.04)
Reading test	615.5	607.9	612.2	616.1	619.6	621.4
	(34.72)	(36.43)	(34.91)	(35.02)	(33.01)	(32.46)
Class size						
Overall	24.52	25.39	24.62	24.40	24.20	24.04
	(5.64)	(5.54)	(5.62)	(5.67)	(5.61)	(5.63)
2nd grade	19.06	19.38	19.08	19.04	18.91	18.88
	(1.62)	(2.26)	(1.34)	(1.33)	(1.38)	(1.58)
3rd grade	20.29	22.93	20.39	19.70	19.33	19.21
	(3.53)	(5.00)	(3.58)	(2.78)	(2.08)	(1.96)
4th grade	28.89	29.10	29.04	29.04	28.75	28.54
	(3.41)	(3.30)	(3.23)	(3.51)	(3.35)	(3.63)
5th grade	29.26	29.38	29.35	29.31	29.25	29.01
	(3.28)	(3.23)	(3.17)	(3.36)	(3.11)	(3.52)
Teacher characteristics						
Percent 1st year not certified	4.36	6.15	5.13	4.30	3.35	2.99
	(11.99)	(14.76)	(12.80)	(11.54)	(10.44)	(9.57)
Percent 1st year certified	4.68	5.66	4.89	4.14	4.73	4.00
	(12.78)	(14.40)	(13.12)	(11.84)	(12.80)	(11.54)
Percent 2nd year not certified	3.41	2.67	4.47	4.05	2.83	3.03
	(10.30)	(9.38)	(11.77)	(11.12)	(9.49)	(9.39)
Percent 2nd year certified	5.49	6.39	6.35	5.26	4.76	4.75
	(13.48)	(14.92)	(14.44)	(13.08)	(12.45)	(12.30)
Percent 1st year	9.18	11.94	10.07	8.44	8.61	7.00
	(17.24)	(19.95)	(17.87)	(16.31)	(16.45)	(14.98)
Percent 2nd year	8.98	9.12	10.84	9.33	7.89	7.78
	(16.55)	(17.13)	(18.04)	(16.76)	(15.37)	(15.15)
Percent 3+ years	81.84	78.95	79.08	82.23	83.50	85.22
	(23.44)	(25.19)	(24.70)	(23.09)	(22.36)	(21.08)
Not certified	12.99	11.78	13.58	14.42	12.57	12.57
	(21.45)	(20.81)	(21.76)	(22.44)	(21.11)	(20.99)
Student demographics						
Percent black	8.77	9.27	9.11	8.76	8.48	8.29
	(12.62)	(13.35)	(13.15)	(12.54)	(12.18)	(11.86)
Percent Hispanic	45.40	43.31	44.10	45.22	46.53	47.71
	(30.08)	(29.58)	(29.83)	(30.06)	(30.31)	(30.39)
Percent Asian	10.77	10.57	10.62	10.75	10.85	11.06
	(13.82)	(13.39)	(13.53)	(13.81)	(14.03)	(14.31)
Percent LEP	29.90	30.81	25.11	31.21	31.01	31.31
	(25.09)	(25.79)	(23.86)	(25.49)	(25.07)	(24.62)
School percent free lunch	55.90	56.80	56.26	55.71	55.23	55.55
	(30.80)	(30.44)	(30.75)	(30.87)	(30.99)	(30.90)
Observations	90,506	17,462	17,873	18,167	18,404	18,600

Notes: Standard errors are in parentheses. Statistics are weighted by average school enrollment for 1997-2001.

Appendix Table 3: Student Demographic Coefficients for Table 1

	Mathematics						Reading					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Percent black	-0.248	-0.262	-0.24	-0.269	-0.235		-0.226	-0.315	-0.294	-0.256	-0.216	
	(63.15)	(24.22)	(22.95)	(21.79)	(19.43)		(65.36)	(33.56)	(32.30)	(24.08)	(20.19)	
Percent Hispanic	-0.132	-0.148	-0.139	-0.156	-0.146		-0.213	-0.233	-0.218	-0.222	-0.197	
	(42.33)	(20.36)	(19.74)	(18.96)	(17.97)		(77.60)	(36.99)	(35.46)	(31.18)	(27.45)	
Percent Asian	0.157	0.054	0.06	0.077	0.090		0.046	-0.076	-0.084	-0.022	-0.019	
	(45.24)	(5.52)	(6.30)	(6.88)	(8.25)		(15.22)	(8.91)	(10.15)	(2.31)	(2.01)	
Percent LEP	-0.064	-0.046	-0.059	-0.04	-0.061		-0.126	-0.088	-0.087	-0.09	-0.089	
	(20.21)	(12.97)	(16.06)	(9.93)	(13.55)		(45.27)	(28.93)	(27.12)	(25.81)	(22.25)	
School percent free lunch	-0.366	-0.017	-0.018				-0.385	-0.022	-0.023			
	(140.01)	(3.33)	(3.78)				(167.46)	(4.89)	(5.65)			
Observations	90,503	90,503	90,503	90,503	90,503	90,503	90,492	90,492	90,492	90,492	90,492	90,492
School FE	no	yes	no	no	no	no	no	yes	no	no	no	no
School by grade FE	no	no	yes	no	yes	yes	no	no	yes	no	yes	yes
School by year FE	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes
Student characteristics	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	no

Notes: Absolute value of t-statistics is in parentheses. The dependent variable is the average test score in that subject. All regressions include class size by grade, teacher experience, teacher certification, and grade by year fixed effects. Regressions are weighted by average school enrollment for 1997-2001. Each column represents a separate regression.

Appendix Table 4: Detailed Analysis of Variance in Class Size

Variance	Grade 2	Grade 3	Grade 4	Grade 5
Overall	2.753	3.555	11.767	10.881
De-mean by school-by-grade	1.808	7.647	5.838	5.359
De-mean by school-by-grade and by school-by-year	0.956	4.153	3.175	2.911