How Flexible is Educational Production? Combination Classes and Class Size Reduction in California

David Sims*
November, 2003

In 1996 the State of California implemented an extensive and costly program to reduce class sizes for K-3 students by an average of ten pupils. The program's non-linear reward scheme provided incentives for schools to both reduce class size by creating new classes and to smooth class size across grades by creating combination classes. I use the rules created by the class size reduction policy to generate instruments for both class size and the percentage of students taught in combination classes. I find that the use of combination classes has a negative and significant effect on the test scores of second and third grade students in California. Furthermore, this negative effect counteracts any positive effect of class size reduction for students placed in combination classes due to program incentives. While smaller classes may be beneficial, the negative effects of combination classes are large enough that the net effect of the California Class Size Reduction Program may have been negative.

---

*Graduate Student, MIT Department of Economics E52-351, 50 Memorial Drive, Cambridge, MA. 02142-1347

I would like to thank Joshua Angrist, Daron Acemoglu, David Autor, Alexis Leon, Cindy Perry, James Poterba, Michael Steinberger, and participants in the MIT Bradley Public Economics/Labor Economics Seminar and MIT Labor lunch for helpful comments and suggestions. Remaining errors are the author's sole responsibility.

I. Introduction

The California Class Size Reduction Program, adopted in 1996, was one of the largest state education reforms of the decade. Though a number of states adopted measures to reduce the size of elementary school classes, the California program was the most ambitious in scope, affecting millions of kindergarten through third grade students at a cost of several billion dollars. Instead of providing a scale of rewards based on reductions in average class size, the program rules provided an "all or nothing" payment for schools that met a threshold requirement of fewer than twenty students per class. Though policy makers intended to provide schools with incentives to hire more teachers and create more classes, the non-linear reward structure they created also provided schools with incentives to smooth class size across grades by creating combination classes, thereby reducing maximum class size without increasing the number of teachers or lowering average class size.

The effect of incentive schemes in public education has been of recent interest to economists. Acemoglu, Kremer and Mian (2003) argue that high-powered incentives create distortions in educational production. They claim that the government provides most elementary education because it can avoid the high powered incentives offered by firms. [1] Recent research by Jacob (2002) and Jacob and Levitt (2003), shows that perverse effects can come from adding high-powered incentive programs in public

---

[1] There is a large literature on the perverse effects of non-linear incentives in firms. For example, the multitasking literature beginning with Holmstrom and Milgrom (1991) details the costly nature of providing high powered incentives to an increasingly flexible workforce. Chevalier and Ellison (1997) demonstrate that mutual funds alter their holdings at the end of the year in response to returns earlier in the year, since managers do not have incentives to maximize fund value but to meet targets. Oyer (1998) shows that managers and salespeople respond to quota incentives by varying sales and prices over the fiscal year, particularly in the final quarter.

education. They find that the adoption of high stakes testing in Chicago led to teachers teaching to the test and helping their students cheat.

This paper also contributes to the literature on combination classes. A combination class is an otherwise normal, self-contained classroom in which multiple grades are generally taught by the same instructor. Though there is little work by economists on the subject, there is a large education literature. Despite the large volume of empirical work, the conclusions are far from uniform. Some studies such as Russell, Rowe, and Hill (1998) find negative consequences of combination classes, while others such as Pavan (1992) find that combination classes enhance academic achievement.[2]

Likewise, theoretical arguments over combination classes remain unresolved. The advocates of combination classes insist that they foster cooperation and critical thinking. Opponents argue that such classes breed confusion and resentment among students who have difficulty working together. Teachers often claim that combination classes are more difficult to instruct.

The debate over the desirability of combination classes can be seen as a debate about academic tracking. Mixing students across grade level leads to a wider range of student interest and ability levels within a classroom due to students' differing levels of experience in an academic setting and prior exposure to certain material. Since diversity is a primary aim of non-tracking initiatives, a program that expands combination classes can be thought of as an experiment in a variety of extreme non-tracking.[3]

---

[2] A review of this education literature can be found in Veenman (1995), and Guttierez and Slavin (1992).

[3] The literature on tracking has long considered it a disastrous policy for poor students (eg. Slavin 1990). However, recent empirical findings cast doubt upon this conventional wisdom. For example, Figlio and Page (2002) show that corrections for endogeneity and selection lead to positive estimates of the impact of tracking on the test scores of poor students.

This paper develops a simple model that relates two forms of classroom organization, class size and student homogeneity, to illustrate the effect that creating combination classes may have in the context of class size reduction. I then estimate the impact of class size and the percentage of students in combination classes on student achievement using instruments derived from the non-linear relationship between enrollment and classroom organization.

I find that the use of combination classes reduces student test scores. The negative effects are larger for third graders than second graders. Estimates of class size effects are small and statistically insignificant. Using generous outside estimates for the effect of class size, I conclude that the 4-5% of students placed in combination classes by the program are worse off than in the absence of such a program. If class size effects are small then the program had a net negative effect on student achievement.

The remainder of the paper is organized as follows: Section two describes the institutional background and theoretical framework, section three discusses the data and identification strategy, section four presents the results and interpretation, and section five concludes.

II. Background

A. The California Class Size Reduction

The California Class Size Reduction Program arose from an unexpected political alliance in the summer of 1996. At the time, the state had a budget surplus and there was widespread interest in using the money to improve primary education. A large portion of the legislature favored a program to reduce class size, then about thirty students per class. Governor Wilson, on the other hand, supported a program offering school vouchers to students. Other initiatives were also discussed.

As the 1996-97 school year approached with the prospect of no major reform, the governor gave his support to the class size reduction advocates. The resulting law took effect less than a month before the school year began. Schools scrambled to adopt the program but many did not have time to fully implement it in the first year.

The Class Size Reduction Program provided incentives for schools to voluntarily reduce their class sizes in the early grades. The state committed to pay each school district $650 dollars for every student in a participating program grade. A school grade was considered a participant if it was in a participating district and had all of its students in that grade in a class of twenty students or fewer. This payment was sizeable relative to California's 1995-96 per pupil expenditure of $6,068. The payment amount steadily increased in subsequent years and stood at $906 in 2002-03. Anticipating a lack of classroom space the state also arranged to subsidize the procurement of temporary classrooms with payments of $25,000. After the first year this subsidy rose to $40,000.

Schools were required to reduce class sizes in a particular order. To participate in the program, schools were required to reduce the class size of first graders. Only when first grade class sizes were below twenty could a school receive money for reducing the size of their second grade classes. A school that had reduced class sizes for first and second graders could receive program money for reducing the size of either kindergarten or third grade classes. After the first year, the program was amended to allow reduction of both kindergarten and third grade classes, though schools still had to reduce first and second grade classes.

The large awards offered by the state led to high program participation rates. Table 1 shows the level of participation and overall participation percentage in the first few years of the program. In the first year, two-thirds of first graders were in classes that qualified for a subsidy, though few kindergarteners and third graders were. However, in subsequent years participation became nearly universal in first and second grade and reached considerable levels in the other grades. By year three, all grade levels exceeded eighty percent participation and by the fourth year all grade levels had a participation rate of over ninety percent.

California's Class Size Reduction Program was extraordinarily expensive. In its first year, including payments for classroom space, the program committed the state to over $1.3 billion in payments. This number increased as per student award and participation levels rose so that by 2001 the subsidy payments constituted six percent of the state education budget.

Previous research on the California Class Size Reduction includes a state commissioned evaluation by a consortium of five companies. Their report found modest gains in student achievement associated with the reduction in class size. Since the program had been offered to all school districts, there was not a clear control group for the study. The consortium's primary solution was to use a difference-in-difference estimator based on the difference in fifth versus third grade test scores in adopting and non-adopting schools. This strategy assumes that the program did not affect fifth grade students. This seems unlikely since fifth grade students may have seen their class size increase and teacher characteristics change as teachers with seniority were transferred to lower grades.

The study by Rivkin and Jepsen (2003) uses variation in the timing of program adoption to identify the effects of smaller class size on test scores. They find large and significant effects, especially for students in poorer districts. They also investigate a potentially perverse effect mentioned in the consortium report. The Class Size Reduction Program forced many districts to hire new teachers with little experience and incomplete credentials. Hanushek, Rivkin and Kain (1998) suggest that inexperienced teachers reduce student achievement. Rivkin and Jepsen argue that the influx of inexperienced teachers in California reduced student test scores, especially in heavily African-American schools. They suggest that the CSR program had net positive effects but increased educational inequality. Nevertheless, the influx of inexperienced teachers represents a short run adjustment rather than a lasting problem.

A feature of the Class Size Reduction program which has not drawn attention is the incentives it provided to use combination classes. Students from eligible grades in combination classes qualified for program money as long as the size of the combination class was below twenty students. This applied even when some of the students in the class were not from eligible grades. For example, a class of eighteen third graders and two fourth graders received a payment for the eighteen third graders. In practice schools were far more likely to combine classes within eligible grades as this reduced the inefficiency of putting students that did not qualify for the subsidy in a smaller class.

The remainder of the paper demonstrates that these incentives led to reduced educational achievement and smaller class size reductions.

B. Theoretical Background

This section outlines an education production function that relates classroom structure to student outcomes. Following Lazear (2001), classroom instruction time is considered a public good and the amount of classroom time available to the teacher is assumed to be scarce and fixed. Any time that a teacher must spend working with an individual student, whether on material the rest of the class understands, individual questions, or discipline, is time not producing classroom instruction. This is a type of congestion effect. If students in combination classes are more likely to require individual teacher time than students in a single-grade class, then combination classes will have less time for learning and lower achievement. In addition to its simplicity, this model agrees with anecdotal evidence provided by teachers that combination classes are harder to control and short on time.

Consider the following setup: With probability q a student does not require individual attention from the teacher in a one unit period of classroom time. If each student's behavior is independently determined, the probability that a classroom of n such students has no interruptions and that learning takes place is q^n. In this hypothetical classroom the learning of one time period has a value of w. The per student value of classroom educational production is:

$$wq^n \qquad (1)$$

This simple formulation highlights two features of classroom production. First, since $q<1$, the marginal effect of increasing class size is negative. Second, the marginal effect of increasing q is positive. Decreasing class size has a positive effect on per student educational production while increasing the need for teacher time spent with individual students reduces learning. Now envision two hypothetical classrooms, both the same size, one of which has only students from grade g, while the other combines a percentage α students from g with students from another grade j. This combination class has lower per student educational production if:

$$q_g^{n} > q_g^{\alpha n} q_j^{(1-\alpha)n} \qquad (2)$$

This inequality depends on the relative magnitudes of q_g and q_j. One assumption that conforms to common ideas about childhood behavior is that disruptiveness is a function of age. If the need for teacher time is strictly a function of age then the older grade will have a larger q than the younger one. This means that the inequality in (2) will be true if $j<g$ but false in the opposite case. In this situation, the older students lose by being in a combination class, while the younger students benefit.

Another plausible assumption is that the need for teacher attention may be higher when a student is in a classroom where the teacher covers multiple curricula than in a classroom with a common curriculum. In this case, combination classes are disruptive. The inequality in (2) becomes:

$$q_g^{n} > q_{gc}^{\alpha n} q_{jc}^{(1-\alpha)n} \qquad (3)$$

where $q_{gc} < q_g$ is the probability that a student in grade g does not require teacher attention.

This inequality depends both on the relative magnitudes of student disruptiveness and on the previous assumption about student age. If age is not a factor in determining q, then the inequality in (3) will be true and the combination class will have lower educational output. If age is a factor, students in grade g will have lower per student output when combined with students from lower grades. However, if they are combined with students from a higher grade, the effect is ambiguous and depends on the ordering of q_g, q_j, q_{gc}, and q_{jc}. The model predicts that students combined with those in lower grades will always be worse off, while students combined with those in higher grades may fare better or worse depending on how disruptive the various groups are.

The model can also show the class size reduction level required to offset the effects on grade g students of their class becoming a combination class. Assuming that $q_g > q_j$ this quantity can be found by solving a slight variation on the inequality in (2), namely:

$$q_g^{n*} = q_g^{\alpha n} q_j^{(1-\alpha)n} \qquad (2')$$

where n* represents the original class size that would make the students in grade g indifferent to moving into a smaller combination class. After some algebra this becomes:

$$n^* = (1-\alpha)n\,[(\log q_j)/(\log q_g)] + \alpha n \qquad (2'')$$

III. Data and Identification

A. Data

This paper draws upon two data sources. Data from the Standardized Testing and Reporting (STAR) program were provided by the assessment division of the California Department of Education. The STAR program began with the administration of standardized tests to students in grades two and above in the spring of 1998.[4] In 1998-

---

[4] Immediately before 1998 there is no reliable statewide testing data for the early elementary grades. This makes it impossible to estimate preprogram test scores, discussed later.

2000, the test years used in this paper, the elementary STAR included the Stanford 9 norm-referenced test.

I use scores for second and third graders from both the mathematics and language sections of the test to measure educational achievement. These scores are available on a school by grade level basis rather than a classroom by classroom basis. I use the National Percentile Rank (NPR) of a hypothetical mean student in a particular grade for a specific school in math or language as dependent variables.

The rest of the data came from the Educational Demographics Office of the California Department of Education. These included detailed reports from schools and teachers about their classes, and contained information on a variety of teacher characteristics such as experience, education level, class sizes, and demographics. The data also provided demographic information including the number and ethnicity of students in each grade, the number of English learners, and the number of students receiving free or subsidized meals. I aggregate this data to the school-grade level where necessary and match it to test scores.

The dataset used in my analysis consists of observations on second graders from the 1998-2000 test years and third graders from the 1999-2000 test years.[5] I eliminate observations for which the necessary demographic and testing information is unavailable and observations for which average class size cannot accurately be figured.[6] The bulk of

---

[5] Corresponding to the 1997-98 through 1999-2000 school years and 1998-99 to 1999-2000 school years respectively. Third graders from test year 1998 were omitted because of their lower participation rate, and the inability to classify all of them as participants on non-participants in the program.

[6] Neither of these seems to be a systematic error. I also eliminate the few schools that had more than 240 students in a grade. Because my approach relies on non-linear variations in enrollment, it requires a sufficient density of observations to be effective.

the analysis also excludes approximately 1,500 observations of grades for which the school did not participate in Class Size Reduction that year.

The dataset's size and detail is greater than that of the data used in most previous studies of the effects of combination classes on student achievement. However, an important limitation in the data is the inability to measure outcomes on the classroom level. Largely because of this, I am not able to look at the detailed workings of combination classes, but rather look at the percentage of students in a school and grade that are in combination classes.

Another limitation is the lack of outcome data for pre-program years. Pre-program data would provide a valuable check on the identification strategy and allow estimation of "value-added" models. Finally, for confidentiality reasons, test scores are unavailable for any school and grade where ten or fewer students were tested. Thus, extremely small schools are excluded from the sample. Fortunately, the vast majority of the schools in California are larger than this cutoff.

Descriptive statistics are found in Table 2, reported separately by grade. Program participants scored close to the national average on standardized tests, though they scored slightly above average in math but below average in reading. The two grades also had fairly similar characteristics. However, a higher percentage of second grade students (15.05%) than third grade students (12.23%) were in combination classes. Also, second graders in program schools appear to have slightly less experienced teachers and slightly higher poverty and English learner percentages than third graders.

B. Graphical Analysis

The California Class Size Reduction Program provided schools with an incentive to create combination classes. Figures 1 and 2 illustrate two effects of these incentives. Figure 1 plots the average class size for a school and grade against the number of students in that grade. The figure also plots a predicted class size function similar to that of Angrist and Lavy (1999). To obtain this function, I begin by estimating the smallest number of equal sized classes under twenty students that would accommodate the enrollment of each school grade. The predicted number of classes is:

$$CLN_{sgt} = (\text{int}[(STUDENTS_{sgt}\text{-}1)/20]\ ) \qquad (4)$$

where s indexes school, g indexes grade and t indexes time. Int(.) represents the integer function, meaning that int(n) in the largest integer less than or equal to n. Using this variable, the predicted class size can be defined as:

$$PCS_{sgt} = [(ENROLLMENT_{sgt}/\ CLN_{sgt}] \qquad (5)$$

where $ENROLLMENT_{sgt}$ is the total number of students in that school and grade observation.

This predictive function indicates what the average class size would be if students were actually divided into the predicted number of classes. Figure 1 shows that actual class size is generally greater than the predicted class size. The two come closest to matching at twenty student intervals.

Figure 2 graphs the percentage of students in a grade in combination classes against the total school enrollment in that grade. The pattern is striking. The proportion of students in a combination class decreases markedly whenever the size of that grade approaches a multiple of twenty. It is easy two see how these two patterns are related.

Instead of lowering class size the predicted amount, administrators kept class size under the twenty student maximum without opening as many new classes by putting the excess students into combination classes. The closer a grade is to having natural multiples of twenty students, the less administrators can employ this type of shifting.

To illustrate the process of choosing class sizes and combination class levels, consider a school that had thirty students in a single class, in each of two first and second grade classes prior to the program. Without combination classes this school would have to implement the program for first and second graders by hiring two new teachers and providing four classes with fifteen students each. However, the additional money that would be paid to the school for implementing the program might be as little as $39,000.[7] Obviously, it would be impossible to hire two additional teachers with this amount. To adopt, this school must shift money from other areas to pay for the extra cost. However, if the school is allowed to count combination classes toward its goal it can hire only one new teacher and have three classes of twenty students. One of these classes is a combination class with equal numbers of students from each grade. This way, the school covers more of its costs from the program bonus.

A negative effect of combination classes on achievement may explain a puzzling pattern in the Class Size Reduction adopters' test scores. Figures 3 and 4 present this pattern in two different formats. Figure 3 plots math scores against the number of students enrolled in a grade. For convenience, the plot shows only schools with between 16 and 124 students in a grade, but the same pattern continues at higher enrollment levels. The vertical lines show multiples of twenty students that would require the

---

[7] This is figured at the initial payment rate of $650. Even at the 2002 rate of $906, it is hard to imagine paying the salaries and benefits of two new teachers for $54,360.

creation of an additional class if combination classes were not available. If the only effect of the program were a reduction in class size, and if the reduction in class size from forming a new class led to higher test scores, we would observe test scores rising after crossing a twenty student threshold.

In contrast, test scores appear to rise as they approach an enrollment threshold and drop off immediately afterward. Figure 4 isolates this pattern by plotting test scores against the distance from an enrollment threshold. This figure plots test scores against the distance in students from an enrollment threshold. Two facts emerge: First, test scores rise as enrollment levels approach a threshold. Second, test scores fall discontinuously as the threshold is crossed, and continue to fall thereafter. This suggests that something other than class size is driving the variation in test scores with enrollment.

One possible explanation for this pattern is the use of combination classes. Figure 2 shows that the percentage of combination students in a grade follows a pattern opposite to test scores. Combination class percentage falls as it approaches the threshold and rises discontinuously afterward.

An alternative explanation involves the Rivkin and Jepsen findings on teacher experience. Schools forming an extra class when crossing a threshold have to hire a new teacher. The new teachers may have had less experience. However, there is little graphical evidence that this explains the test score pattern. Figure 5 plots average teacher experience against grade level enrollment. The drop in teacher experience appears to be a smooth function of enrollment. Teacher experience rises as often as it falls when a threshold is crossed. Intuitively this may happen because some experienced teachers from other grades are willing to switch grades to teach smaller classes. Figure 6 shows a similar pattern for another measure of teacher experience, the percentage of novice teachers in a grade. Although the Class Size Reduction Program may have led to an overall experience decrease, teacher experience does not vary in a non-linear pattern with enrollment in participating grades.

The Class Size Reduction Program did not reduce class sizes by as much as grade level enrollments would predict. It also led to the use of combination classes and a decrease in teacher experience. The use of combination classes may have had a negative effect on test scores that would help explain the unexpected pattern observed in the data.

C. Identification

I use the data described above to capture the effects of combination classes and class size on achievement by exploiting the non-linear relationship between these variables and enrollment. The idea of using a program induced discontinuity as a source of identification is not new. Campbell (1969) discusses the use of regression-discontinuity designs in empirical research. More recently Hoxby (2001), Angrist & Lavy (1999), and Guryan(2001) make use of regression discontinuities to form instruments for instrumental variables estimation in education related investigations.

The causal relationship of interest is:

$$TESTSCORE_{sgt} = X_{sgt}'\alpha + \phi CLASSIZE_{sgt} + \delta COMBINATIONPCT_{sgt} + \gamma_g + \tau_t + \eta_{sgt} \quad (6)$$

where s indexes school, g indexes grade and t indexes year. X is a vector of demographic controls including grade level enrollment, percentage black and hispanic students, percentage of English learners, and percentage of students that qualify for free or subsidized meals. Also, γ is a grade effect, τ a time effect and η_{sgt} is the error term.

OLS estimates of equation (6) are unlikely to have a causal interpretation because the demographic variables included in the regression are unlikely to completely control for all the factors that relate classroom organization to test scores. For example, parental incomes and levels of involvement are likely to be negatively correlated with combination classes and class size and positively correlated with test scores. Omitting these factors from the regression biases the OLS estimates toward zero.

The presence of two variables with potential causal interpretations in the regression is also a concern. Consider estimation of a two stage least squares model which uses a non-linear enrollment function such as Predicted Class Size to instrument for combination class percentage but allows class size to be exogenous. The first stage relationship is:

$$\mathrm{COMBINATIONPCT}_{sgt} = X_{sgt}'\pi_1 + \pi_2 Z_{sgt} + \pi_3 E_{sgt} + \pi_4 E^2_{sgt} + \varepsilon_{sgt} \qquad (7)$$

where X now includes demographic controls, year effects and class size , E is enrollment in grade g, and Z is the instrument, in this case predicted class size. This leads to the second stage:

$$\mathrm{TESTSCORE}_{sgt} = X_{sgt}'\varphi + \rho\,\mathrm{COMBINATIONPCT}^*_{sgt} + \mu E_{sgt} + \theta E^2_{sgt} + \omega_{sgt} \qquad (8)$$

where COMBINATIONPCT* is the predicted combination percentage produced by the first stage.

This specification assumes the instrument Z does not affect test scores except through its effect on combination class percentage. The assumption is likely violated in this case, since Figure 1 shows an apparent correlation between predicted class size and actual class size. An administrator who crosses an enrollment threshold does not have to put the extra students in a combination class, since she may opt instead to create a new class and reduce class size.

A potential solution to this "two causes" problem exploits the non-linear relationship between combination class percentage, class size, and enrollment to construct instruments for both class organization variables. Each of these instruments is a different non-linear function of enrollment. I have already introduced two potential instruments, Predicted Class Size and Predicted Number of Classes, in the graphical analysis. In principle these two variables can be used to instrument for both endogenous regressors. In practice, both variables are different functions of the same underlying enrollment variable, and this strategy is unlikely to provide precise estimates of both coefficients.

Another approach is to generate an instrument that is correlated with combination classes but uncorrelated with class size, conditional on enrollment. The predicted class size of a lower grade is a candidate. If a school reduced class size at the second or third grade level, the grade immediately below the observed grade must also have participated. To form a combination class there must have been students at two grade levels available to combine. Since schools were far less likely to combine students with non-participating grades, the ability of a school to form combination classes depended on the predicted class size of the immediately lower grade. The Lower Grade Class Size Predictor is:

$$PCS_{s(g-1)t} = [(STUDENTS_{s(g-1)t} / CLN_{s(g-1)t}] \ . \qquad (9)$$

I also construct a Combination Classes Predictor (CSP) that by design is purged of correlation with class size. To do this I calculate the number of classes of fewer than twenty students required for the students in a grade and the grade below. Then I subtract this number from the predicted number of classes of fewer than twenty students the school would require to avoid mixing these two grades. Intuitively the predictor counts additional classes a school would need to form to participate in both grades and avoid combination classes. The formula is:

$$CSP_{sgt} = [\ (CLN_{sgt} + CLN_{s(g-1)t}) - (CLN_{s(g+(g-1))t}] \qquad (10)$$

I present results that use a variety of methods to deal with the potential confounding effects of class size, including instrumenting for both potentially endogenous variables, and using the Combination Classes Predictor as an instrument. In all cases the estimates of the effects of combination students on achievement are similar. The effects of class size are generally small and always statistically insignificant.

IV. Results

A. OLS

Table 3 presents ordinary least squares estimates of equation (6) for second graders. The dependent variable is the national percentile rank of the hypothetical average student in mathematics. Panel A provides results for the sample of participating schools discussed in the data section. Column (1) presents a specification comparable to many previous studies of class size. The class size coefficient is small and insignificant. However, some measures of teacher experience are correlated with student achievement.

Both the percentage of first year teachers and the percentage of teachers with credential waivers have negative significant coefficients.[8]

Columns (2) – (5) add the percentage of students in combination classes to the specification. Estimate of this coefficient are consistently in the neighborhood of -.073, implying that a five percentage point change in students in combination classes would leads to a drop of about one-third of a percentile in math scores. This seems like a modest effect. This result is robust to the level of control and addition of smooth enrollment controls. Teacher inexperience continues to play a negative role, with estimates of a similar magnitude. The class size coefficient falls toward zero once higher order demographic controls are added to the model.[9] The small coefficient on class size differs from recent research documenting large class size effects[10]

These results do not conclusively demonstrate that the class size changes caused by the Class Size Reduction Program had no effect on test scores. There is no pre-treatment versus post-treatment element in any of these estimates. Since the sample is composed of schools that have all implemented the reduction program, the variation in class size is smaller than in most populations. Most schools in the sample lie within a 2.5 student range of class size. This may account for the failure to find large class size

---

[8] To get a teaching credential in California, a candidate must take 30 credit hours beyond a bachelors degree in a recognized education program. This is often referred to as the “fifth year”. Teachers with a bachelors degree who pass other certification requirements such as the competency test can get an emergency credential which allows them to teach for a few years under the understanding they will use the time to complete the other requirements. A credential waiver, is more radical and releases the teacher from even more requirements of the credentialing process. Because a large part of credentialing is gaining classroom understanding, experience, and performing student teaching, these variables can still be thought of as a type of experience measure.

[9] Throughout the following tables the estimated models contain the full set of demographic controls up to third order terms, except where specified otherwise, as well as year effects and grade effects when relevant. These coefficients are not reported because they are not a primary object of interest in this paper and because the follow a pattern that previous research predicts. Namely, test scores drift upward over time, and schools with a high percentage of disadvantaged students perform perform poorly.

[10] For example Angrist and Lavy (1999), Krueger (1999), Finn and Achilles (1990).

effects in this paper. This result is similar to Hoxby (2001), which finds no class size effects when examining natural population variation.

If the lack of variation in class size is partially responsible for the small class size coefficient estimates, then OLS estimation using a sample with more variation should yield larger coefficients. Panel B of Table 3 confirms this. It re-estimates the specifications used in panel A on a sample which adds the 445 second grade classes that did not participate in the Class Size Reduction Program to the previous data. The coefficient on class size is larger and consistently negative. However, the class size coefficient is still insignificant in all but one of the specifications. Panel B also shows slightly attenuated teacher experience and combination class effects when compared to the CSR sample estimates.

Table 4 demonstrates that the percentage of students in combination classes has larger effects on second grade language scores than on math scores. The coefficient estimates of combination class effects are -.094, about thirty percent greater than the math coefficients. Increased class size effects are also larger, about half the magnitude of the combination class effect, but still statistically insignificant. Teacher experience remains important, as do credential waivers.

The results in Table 5 show that OLS regressions using third grade test scores produce similar patterns. Columns (1)-(3) show that the effect of combination class percentage on math scores is about the same as for second graders. Class size effects are now positive but are still small and imprecisely estimated. In contrast with the second grade results the percentage of first year and credential waiver teachers have no significant relationship with test scores. Columns (4)-(6) show similar findings for class size and combination class percentage when the dependent variable is language scores.

B. IV

First Stage Estimation

The OLS results provide a reference point but are unlikely to have a clear causal interpretation. Any omitted variable, such as parental education or involvement that is positively correlated with test scores and negatively correlated with percentage of students in combination classes, will bias the OLS estimates toward zero.

The first five columns of Table 6 present estimates of the first stage relationship described in equation (7) for various instruments. Column (1) shows a significant positive correlation between the Combination Class Predictor and the percentage of students in a combination class, conditional on enrollment and demographic controls.[11] Columns (2) and (3) show that Predicted Class Size and Predicted Number of Classes are also correlated with combination class percentage. This confirms the graphical evidence of Figure 2.

Columns (4) and (5) present regression results from specifications that contain two non-linear functions of enrollment. All these functions, including the Lower Grade Class Size Predictor have significant coefficients. However the use of two instruments does not substantially improve the fit of the prediction. Also, the presence of Predicted Class Size in the regression attenuates the coefficient on the Combination Class Predictor by half. This relationship makes sense as both are functions of the same underlying enrollment variable.

---

[11] Though the tables only report results of quadratic enrollment controls, regressions using quartic enrollment controls yield essentially the same results, with the higher enrollment terms having insignificant coefficients.

The last five columns of Table 6 repeat these regressions, with average class size as the dependent variable. Columns (6) and (9) demonstrate that the Combination Classes Predictor is not a significant predictor of class size, even in specifications which include Predicted Class Size as a regressor. The results also indicate that class size is correlated with the other non-linear enrollment functions, including Lower Grade Predicted Class Size.

Additionally, the table shows that the non-linear enrollment functions are better predictors of combination class percentage than of class size. The coefficients for these three instruments are smaller (in an absolute value sense) and less precisely estimated than their counterparts in the first half of the table.

The non-linear enrollment functions are also poor predictors of teacher experience. Columns (1) – (4) of Table 7 present estimates of the relationship between the non-linear enrollment instruments and average years of teacher experience. In all specifications the quadratic enrollment controls are significant predictors of teacher experience, while the non-linear enrollment functions have coefficients that are not statistically distinguishable from zero. Columns (5) – (8) show the same pattern using a different teacher experience measure, the percentage of novice teachers in a grade and school. Despite a positive and significant OLS relationship with test scores, teacher experience does not vary in a non-linear fashion with enrollment. While teacher experience may have predictive power for test scores on average, it cannot explain the pattern of scores shown in Figure 4.

2SLS

Table 8 presents the results of a two stage least squares estimation of equation (8) for second graders. The first six columns treat class size as exogenous. The final three impose a zero coefficient on class size. The first three columns present results for language achievement. Instrumenting the percentage of combination classes with the Combination Class Predictor yields a significant coefficient estimate of -.195. This implies that a five percentage point increase in students in combination classes leads to a one percentile fall in test scores. This estimate is about 2.5 times the magnitude of the OLS estimates, suggesting that the OLS estimates suffer from omitted variables bias. Columns (2) and (3) report results from estimation using the predicted class size instruments. The coefficient estimates are slightly smaller than those in column (1) but are highly significant.

Columns (4) – (6) show an effect of similar magnitude on math scores. Instrumenting with the Combination Class Predictor, Column (4) gives coefficient estimates of -.180 for the effect of combination classes on math scores. This implies a five percentage point increase in combination class students again results in a one percentile drop in average test scores. Estimation using the predicted class size instruments provides similar results.

In all of these specifications, the class size estimates range from one-third to one-half the magnitude of the combination class estimates. The class size estimates are also very imprecise. In addition to the insignificant class size estimates the table reveals that the coefficients on the smooth enrollment controls are not significantly different from zero in any specification.

As a specification check, the final three columns of Table 8 repeat the estimation of columns (4) – (6), imposing a coefficient of zero on class size. These regressions yield almost identical results. Estimates of the effect of combination classes on second grade math scores yield coefficients of -.18 to -.20 whether class size is treated as exogenous, zero, or an instrument uncorrelated with class size is used.

Table 9 presents similar two stage least squares regressions results for third graders. The estimated coefficient for percentage of students in a combination class is consistent across different instruments and larger than the second grade estimates. Estimates for math and language scores are about -.36 and highly significant. This implies that a five percentage point increase in combination class students corresponds to a one and a half percentile drop in average test scores for the entire grade. The larger third grade estimates may be due to third graders greater propensity to be placed in combination classes with lower graded students. Estimates of average class size effects are positive but extremely imprecise.

As a further specification check, Columns (7) – (9) estimate the effect of classroom organization on the math scores of third graders with the class size coefficient constrained to equal the Rivkin and Jepsen estimates.[12] The resulting coefficient estimates of the effect of combination class percentage on math scores are very similar to to the estimates in columns (4) – (6) which treat class size as exogenous.

Table 10 presents the results of two stage least squares estimation with both class size and combination class percentage treated as endogenous regressors. The combination of instruments used in each specification is shown at the bottom of the

---

[12] See Rivkin and Jepsen (2002) Table 4. p36. The coefficient is adjusted to reflect the different scale of the test score measure used by Rivkin and Jepsen.

column. The second grade results for combination class percentage presented in Panel A are similar to my previous two stage least squares estimates, -.18 for math and -.19 for language. The pattern holds for third graders with estimated coefficients about -.36 for math and -.38 for language.

At first the class size coefficient estimates in this table might seem implausibly large. These coefficients are much larger than the estimates shown in previous tables. However, all the instruments rely on the variation in the same underlying enrollment variable. Because of their colinearity they are unlikely to provide precise estimates for both coefficients. The instruments have a stronger relationship with combination classes than with class size. Thus, the large class size coefficients come with very large standard errors. Such large standard errors make it impossible to rule out any of the earlier class size estimates or a zero effect.

Two stage least squares estimates provide consistent evidence that combination class students explain the perverse effect seen in Figure 4. Furthermore, the effect of combination classes on test scores is larger than OLS estimates suggest. The coefficient estimates are robust across different approaches to the potential confounding effects of class size. Class size effects on the other hand are small or zero and very imprecisely estimated.[13]

C. Comparison and Interpretation

Why do the results show unambiguous negative effects of combination classes while previous research has generated mixed results? Two factors seem important. The first is the variety of organizational structures that might be considered combination

---

[13] This is not surprising. Class size had very small OLS coefficients and a weaker relationship with the instruments than combination classes. This does not necessarily mean there was no effect on test scores from the class size reduction, but rather there is no discernable effect in this sample.

classes. Combination classes in California might be different in some important respects from those studied in other contexts. California exercises an unusual amount of centralized control over school curriculum. In addition, since 1998, the state has required grade-specific standardized tests be administered to all students in grade two and above. These tests reinforce the need to teach distinct skills to students at different grade levels. This structure and testing mean that a combination class teacher in California is less likely to rely on thematic or common curriculum elements than teachers in other settings. In effect, the teacher must teach two separate classes within one classroom. The education literature suggests that a prime source of benefit in combination classes is the ability of students to work together in accomplishing mutual tasks, an advantage that is lost if the students are involved in different tasks.

Though it might seem that this rigid structure limits the applicability of this study to other combination class contexts, education policy trends indicate otherwise. These emerging trends involve a shift toward centralized standards and curriculum and greater grade-specific testing. This makes the California model of the combination class a good approximation for what many states might choose in the future.

A second consideration distinguishing this study from previous work is study design. Many studies rely on small samples of classrooms and are limited by the lack of important data on school characteristics. Additionally, there is no clear source of exogenous variation in the use of combination classes. In contrast, this study uses a clearly defined source of exogenous variation and a relatively large sample of schools.

Did the use of combination classes make the California Class Size Reduction Program at net loss in academic achievement terms? To answer this question, I provide some estimates of the net effect of the program. These estimates are illustrative, requiring assumptions about the effect the program had on test scores through changing class size in the absence of combination classes.

In order to estimate a net effect, I first translate my coefficient estimates into "effect size" (i.e. standard deviation) units. Let β be the estimate of the effect of combination students on test scores, X_{pre} and X_{post} measures of combination percentage before and after the program, and σ the standard deviation of student math scores, then the effect size of the California Class Size Reduction, working through combination classes is:

$$[(X_{post} - X_{pre})\beta] / \sigma \qquad (11)$$

Often, researchers have a choice of σ when calculating effect sizes. Effect sizes calculated using the standard deviation among student test scores will always be smaller than effect sizes that use the standard deviation between groups of students, because the latter has a smaller variance. Because within-student test score information is not available, I calculate effect sizes using the between-grade variation. Because they are presented in standard deviations, effect sizes can be compared across outcome measures.

Table 11 presents estimates of the effect sizes of the increase in combination class percentage due to the Class Size Reduction Program. The first row shows the effect size of a five percentage point increase in combination class students. These results seem modest, representing only 4-10% of a standard deviation decrease in test scores. Unlike other policies that affect all the students in a grade, combination classes may only affect the test scores of combination class students. This is reflected in the second row of the table, which presents effect size estimates scaled by the proportion of combination students in that grade. These adjusted figures indicate that achievement losses for students placed in combination classes are between .24 and .36 of a standard deviation for second graders and .58 to .66 of a standard deviation for third graders.

To figure the net effect of the California Class Size Reduction Program I combine its effects on test scores through three channels, combination classes, class size and teacher experience. I measure changes in these variables from 1995-96 to 1999-2000. During this time average class size dropped by about 10 students for affected grades, average teacher experience decreased by a year and the population of novice teachers grew by seven percentage points. Using the effect sizes implied by my estimates of class size and teacher experience I find that the net effects of the program for second graders are slightly negative with magnitudes of 2-4 percent of a standard deviation. The net effects on third graders are also negative, but more substantial, 10-13 percent of a standard deviation.

I next figure net effects using the Rivkin and Jepsen estimates of the class size effects of the program[14]. Their estimates imply a math score effect size of .199 standard deviations for a ten student decrease in class size among third graders. They do not estimate models using language scores, but find a .1167 standard deviation effect for reading scores.[15] These larger class size estimates imply a positive program net achievement effect. Third graders experienced a .10σ increase in math scores and a .03σ increase in language scores. Second graders experienced a similar positive effect, assuming they faced the same class size effects.

---

[14] See Rivkin and Jepsen (2002) Table 4. p36.

[15] I use reading scores and language scores interchangeably in this comparison. In fact student scores on the two tests are similar but not identical.

Finally, I calculate the net effect of a hypothetical policy that implements all the Class Size Reduction Rules but does not allow combination classes. In this scenario, there is no perverse combination class effect, class size equals predicted class size, on average more than a student lower than under the real program, and teachers are slightly less experienced.

This hypothetical program has a larger positive net effect than the actual program.. Third grade students experience an increase in math scores of .2σ, almost double the net effect of a program that allows combination classes. The effect on language scores of third graders quadruples, to .12σ. This calculation assumes that all schools that participated in the real program would participate in the hypothetical one. In practice, some schools might not join the program if they were unable to use combination classes. This would diminish the number of students that received the increasd benefits.

Though the program may well have had net positive effects, the effect on the students put into combination classes by the program was almost certainly negative. Angrist and Lavy (1999) and Krueger (1999) consider classroom settings outside California and find larger effect sizes than Rivkin and Jepsen (approximately .3σ). Even if class size effects this large occurred in California, an increase in second grade combination students of five percentage points would roughly offset a contemporaneous class size decrease of eight students. It would take a much larger class size decrease to offset the negative impact of a five percentage point increase in combination students at the third grade level. .

V. Conclusion

The California Class Size Reduction Program spent billions of dollars to reduce class sizes for early elementary school children. However, the program used a non-linear incentive scheme that rewarded schools for meeting a target threshold. These incentives led schools to shift students into combination classes as well as add classes to meet program requirements. This study offers strong evidence against the use of combination classes. Combination classes have an unambiguously negative effect on student achievement and the effect is greater for third graders than second graders. Students placed in combination classes by the program were almost certainly worse off in achievement terms.

The sign of the overall net effect of the Class Size Reduction Program depends crucially on the actual class size effect of the program. My estimates of class size effects were small and not significantly different from zero. These small class size effects lead to the conclusion that the program had negative net effects in the first few years. However, other studies have found larger class size effects that would imply a positive net effect. Further research might examine cross state variation in class size reduction policies to better estimate the magnitude of the class size effect.

References

Acemoglu, Daron, Michael Kremer, and Atif Mian, Incentives in Markets, Firms and Governments. *NBER Working Paper # 9802.* June 2003.

Angrist, Joshua D. and Victor Lavy, Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement,. *Quarterly Journal of Economics* 114. (1999), 533-575.

California Department of Education. *Standardized Testing and Reporting Data*. Years 1998-2000.

______, *Class Size Reduction: Final Participation and Funding Data*. Years 1997-2003

California Department of Education – Educational Demographics Office. *Demographic Data Files.* Years 1995-2001.

Campbell, Donald T., Reforms as Experiements. *American Psychologist* 24. (1969) 409-429.

Card, David and Alan Krueger, Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States,. *Journal of Political Economy* 100. (February 1992) 1-40.

Chevalier, Judith and Glenn Ellison, Risk Taking by Mutual Funds as a Response to Incentives. Journal of Political Economy 105. (December 1997), 1167-1200

CSR Research Consortium, *Class Size Reduction in California: Early Evaluation Findings, 1996-98*. (1999).

______. *Class Size Reduction in California: The 1998-99 Evaluation Findings*. (2000).
.

Figlio, David N. and Marianne E. Page, School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? *Journal of Urban Economics 51*. (May 2002), 497-514

Finn, Jeremy D. and Charles M. Achilles, Answers and Questions About Class Size: A Statewide Experiment, *American Educational Research Journal* 27. (Fall 1990), 557-77.

Guryan, Jonathan, Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts. *NBER Working Paper #8269*. (May 2001).

Gutierrez, R., Slavin, R.E. (1992). Achievement Effects Of The Non-Graded Elementary School: A Best Evidence Synthesis. *Review of Educational Research* 62. (1992), 333-376.

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

______., John F. Kain and Steven G. Rivkin. Teachers, Schools and Academic Achievement. *NBER Working Paper #6691*. (August 1998).

Holmstrom, Bengt, and Paul Milgrom, Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics and Organization 7. (1991), 24-52.

Hoxby, Caroline M, The Effects of Class Size on Student Achievement: New Evidence from Natural Population Variation. *Quarterly Journal of Economics* 115. (November 2000), 1239-85.

Jacob, Brian, and Steven Levitt, Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics 118*. (August 2003), 843-77.

Jacob, Brian, Accountability, Incentives, and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools. *NBER Working Paper # 8968* (2002).

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

Lazear, Edward P., Educational Production. *Quarterly Journal of Economics* 116 (August 2001), 777-803

Oyer, Paul, Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality. Quarterly Journal of Economics 113. (February 1998), 149-85.

Pavan, B. N., The benefits of non-graded schools. *Educational Leadership* 50. (1992) 22-25.

Rivkin, Steven G. and Christopher Jepsen, What is the Tradeoff Between Smaller Classes and Teacher Quality? *NBER Working Paper #9205*. (September 2002).

Russell, V.J., Rowe, K.J., Hill, P.W. Effects of Multi-Grade Classes on Student Progress in *Literacy and Numeracy: Quantitative Evidence and Perceptions of Teachers and School Leaders*. Melbourne, Australia: University of Melbourne (1998).

Slavin. R. E., Achievement Effect of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review of Education Research*. (Fall 1990), 471-499.

Slavin, R. E.. Synthesis of research on grouping in elementary and secondary schools *Educational Leadership* 46.(1988), 67-77.

U.S. Department of Education, National Center for Education Statistics. *Common Core of Data (CCD)*, online version. http://nces.ed.gov/ccd/

Veenman, S. Cognitive and Noncognitive Effects of Multi-grade and Multi-age Classes: A Best Evidence Synthesis. Review of Educational Research 65. (1995), 319-381.

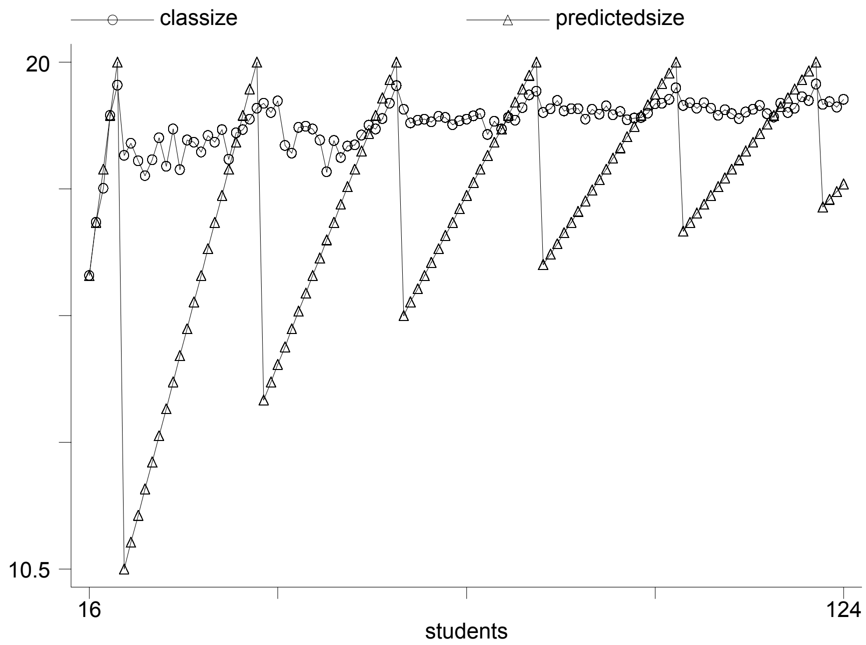
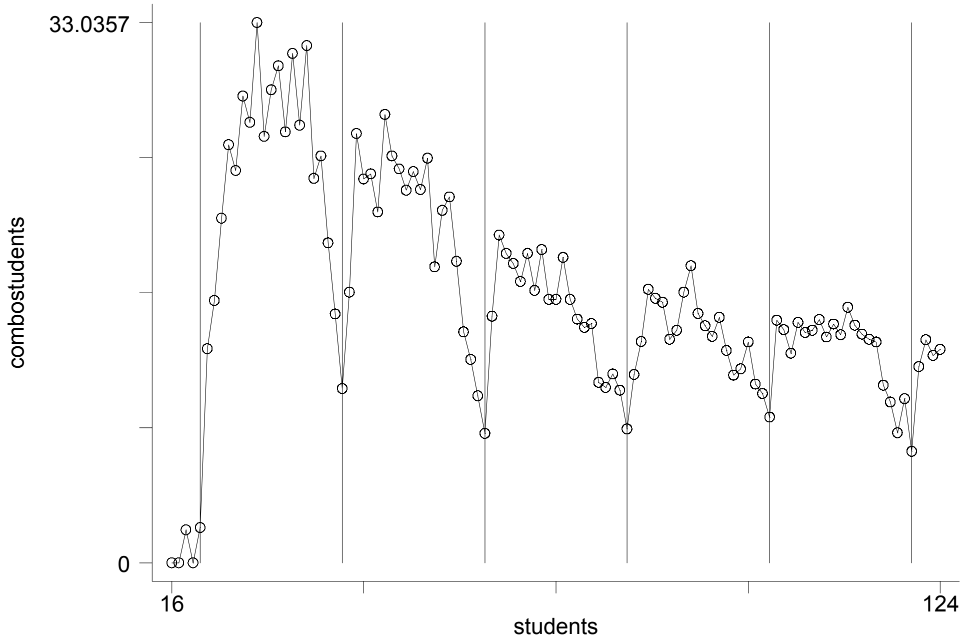


Figure 1: A Plot of the predicted class size function versus actual class



size

Figure 2: A plot of the percentage of students in combination classes versus total grade enrollment. Vertical lines are at 20 student intervals.

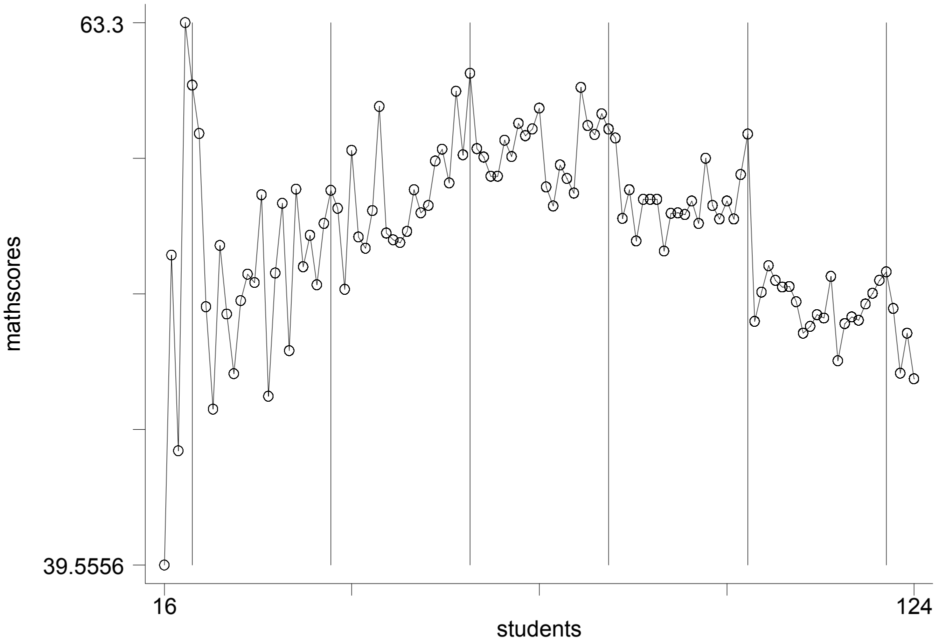


Figure 3: A plot of math test scores by grade level enrollment. Vertical lines are at 20 student intervals.

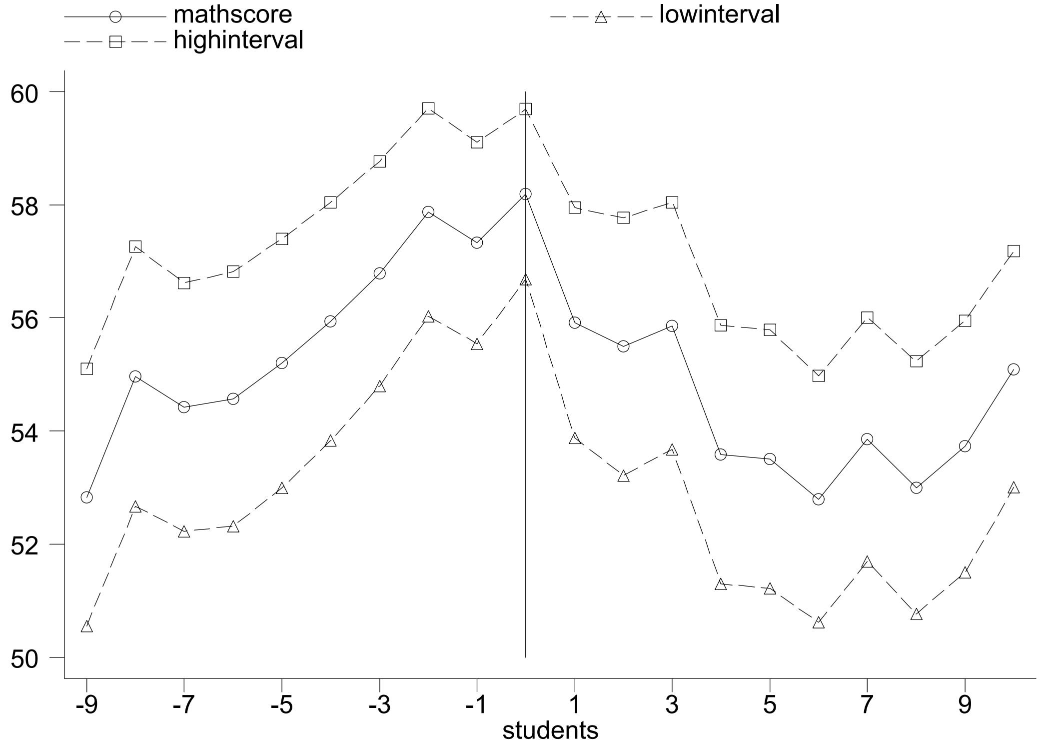


Figure 4: Plot of math test scores versus school grade enrollment measured as distance from a multiple of twenty students. Dashed lines give a 95% confidence interval.

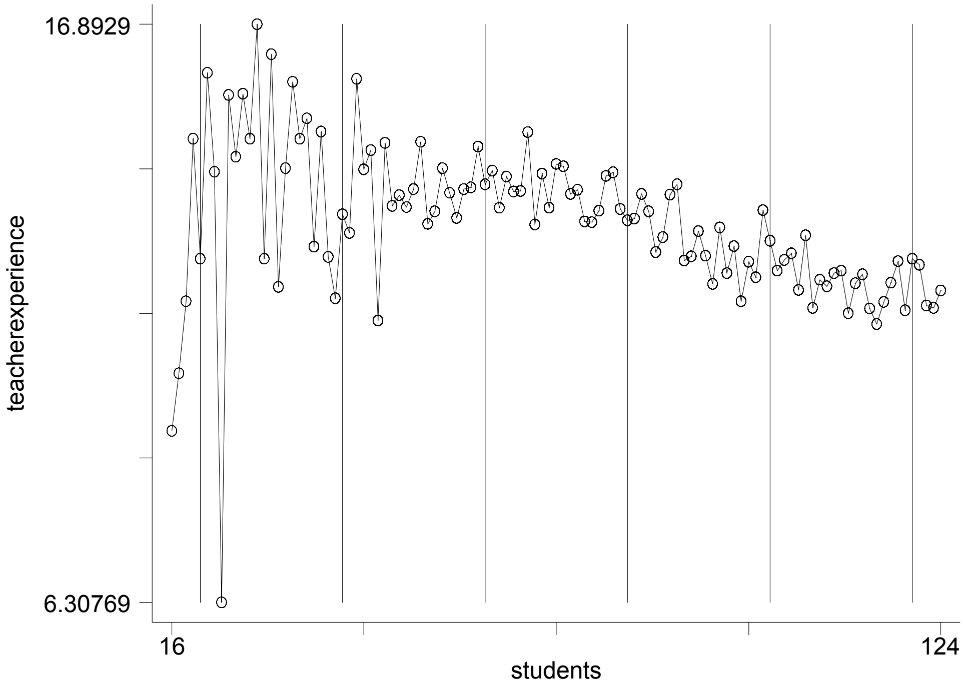


Figure 5: plot of years of average years of teacher experience by grade level enrollment. Vertical lines are at 20 student intervals.

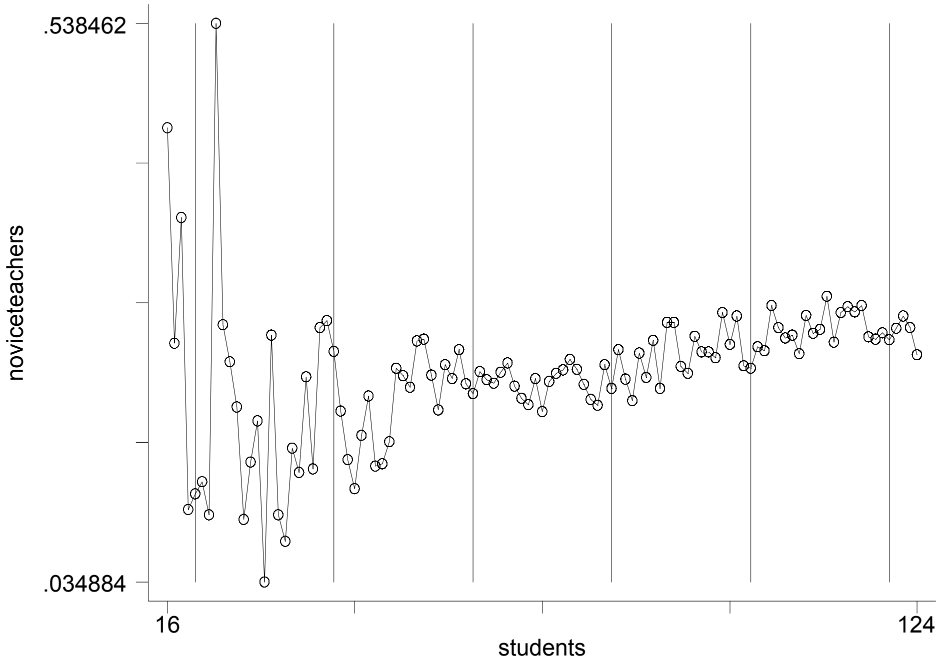


Figure 6 : A plot of the percentage of novice teachers by grade level enrollment. Vertical lines are at 20 student intervals.

Table 1 - Participation in the California Class Size Reduction Program

Year:		1997	1998	1999	2000	2001
Kindergarten:						
	Students	64,779	321,209	393,036	421,943	439,439
	Percentage	14%	69%	86%	92%	96%
First Grade:						
	Students	428,242	484,518	483,714	477,150	480,307
	Percentage	88%	99%	99%	98%	99%
Second Grade						
	Students	262,074	468,103	475,477	472,842	475,702
	Percentage	57%	96%	98%	97%	97%
Third Grade						
	Students	79,062	309,828	410,089	444,136	458,040
	Percentage	18%	67%	84%	91%	91%

Participants are those students in a school that applies for program funding for the student's grade and has all class sizes in that grade at twenty or fewer pupils. The year listed corresponds to the spring portion of the school year. Thus 1997 is the first program year.
Source: California Department of Education

Table 2 - Descriptive Statistics

	Grade 2	Grade 3
Math NPR	51.80 (19.56)	55.30 (18.52)
Language Arts NPR	47.94 (20.80)	49.63 (18.63)
Percentage of students in the grade in combination classes	15.05 (15.66)	12.23 (13.76)
Number of combination classes in the school	2.68 (2.90)	2.55 (2.90)
Average Class Size	18.03 (1.22)	18.15 (1.42)
Average Teacher Experience	12.58 (6.38)	12.90 (6.37)
Percentage novice teachers	25.15 (25.66)	23.66 (25.22)
Percentage first year teachers	7.42 (14.81)	6.42 (13.61)
Percentage second year teachers	9.56 (16.51)	8.90 (15.94)
Percentage third year teachers	8.17 (15.12)	8.35 (14.69)
Percentage teachers with emergency credential	9.32 (17.52)	9.27 (17.39)
Percentage teachers with credential waiver	0.57 (4.81)	0.49 (4.26)
n=	9974	6079

Table continues on next page. Standard Errors are in parentheses below means. Unit of observation is the school grade year. NPR is the National Percentile Rank of the hypothetical average student.

Table 2 - Descriptive Statistics - continued

	Grade 2	Grade 3
Percentage of free/reduced Meal students in school	52.41 (29.71)	52.01 (29.70)
Percentage African-American students in grade	9.16 (13.50)	9.65 (14.34)
Percentage Hispanic students in grade	39.83 (28.04)	38.69 (27.74)
Percentage English Learner students in grade	28.77 (23.92)	27.15 (23.44)
Grade enrollment	98.87 (39.51)	97.84 (39.35)
School enrollment	621.70 (229.79)	617.00 (236.03)
n=	9974	6079

Standard errors are in parentheses below means. Throughout the paper and remaining tables percentage of students with subsidized meals, percentage of minority students and percentage english learners are used as the demographic controls.

Table 3 - OLS Estimates of the Effect of School Characteristics on Second Grade Math Scores

	(1)	(2)	(3)	(4)	(5)
A. Class Size Reduction Schools Only (n=9974)					
Percentage of students in combination classes		-0.073***	-0.075***	-0.070***	-0.069***
		(0.010)	(0.010)	(0.010)	(0.010)
Average Class Size	0.028	-0.022	-0.013	0.001	0.002
	(0.115)	(0.115)	(0.115)	(0.114)	(0.114)
Average Teacher Experience	0.054	0.066**	0.063*	0.079**	0.078**
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Percentage of first year Teachers	-0.038***	-0.036***	-0.036***	-0.038***	-0.038***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Percentage of second year teachers	-0.011	-0.009	-0.009	-0.009	-0.009
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of teachers with emergency credentials	0.021*	0.018	0.018	0.008	0.008
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Percentage of teachers with credential waivers	-0.117***	-0.108***	-0.109***	-0.104***	-0.104***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Enrollment in grade			-0.003	-0.006	0.019
			(0.005)	(0.005)	(0.054)
Enrollment in grade squared*100					-0.024
					(0.045)
B. Class Size Reduction Schools and Non-Participants (n=10419)					
Percentage of students in combination classes		-0.070***	-0.073***	-0.068***	-0.068***
		(0.010)	(0.010)	(0.010)	(0.010)
Average Class Size	-0.120	-0.142*	-0.135	-0.111	-0.112
	(0.084)	(0.084)	(0.084)	(0.084)	(0.084)
Average Teacher Experience	0.050	0.062*	0.058*	0.074**	0.074**
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Percentage of first year Teachers	-0.036***	-0.034***	-0.034***	-0.036***	-0.036***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Percentage of teachers with emergency credentials	0.015	0.011	0.013	0.004	0.004
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Percentage of teachers with credential waivers	-0.114***	-0.105***	-0.107***	-0.102***	-0.102***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Demographic controls	Y	Y	Y	Y	Y
higher order controls				Y	Y

Estimates are of Equation (6) in the text. *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. All regressions are weighted by the number of test takers.

Table 4 - OLS Estimates of the Effect of School Characteristics on Second Grade Language Scores

	(1)	(2)	(3)	(4)	(5)
Percentage of students in combination classes		-0.098***	-0.099***	-0.094***	-0.094***
		(0.008)	(0.008)	(0.008)	(0.008)
Average Class Size	-0.002	-0.069	-0.058	-0.050	-0.050
	(0.104)	(0.104)	(0.104)	(0.103)	(0.104)
Average Teacher Experience	0.073**	0.090***	0.086***	0.099***	0.099***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Percentage of first year Teachers	-0.037***	-0.035***	-0.035***	-0.036***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of second year teachers	-0.010	-0.007	-0.007	-0.007	-0.007
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Percentage of teachers with emergency credentials	0.022**	0.017*	0.018*	0.008	0.008
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of teachers with credential waivers	-0.142***	-0.130***	-0.132***	-0.125***	-0.125***
	(0.026)	(0.027)	(0.027)	(0.026)	(0.026)
Enrollment in grade			-0.004	-0.007*	0.022
			(0.004)	(0.004)	(0.048)
Enrollment in grade squared*100					-0.027
					(0.041)
n	9974	9974	9974	9974	9974
level demographic controls	Y	Y	Y	Y	Y
higher order controls				Y	Y

The table mirrors the estimates of Table 3 Panel A using language scores as the dependant variable. *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. All regressions are weighted by the number of test takers. Level controls refer to controls for percent minority, percent subsidized lunch and percent English learners as well as year effects. Higher order controls add quadratic and cubic terms as well as interactions for the demographic control variables.

Table 5 - OLS Estimates of the Effect of School Characteristics on Third Grade Test scores

Dependent Variable:	**math scores**			**language scores**		
	(1)	(2)	(3)	(4)	(5)	(6)
Percentage of students in combination classes	-0.077*** (0.013)	-0.073*** (0.013)	-0.072*** (0.013)	-0.085*** (0.011)	-0.079*** (0.011)	-0.079*** (0.011)
Average Class Size	0.016 (0.121)	0.023 (0.119)	0.024 (0.119)	0.022 (0.101)	0.026 (0.099)	0.039 (0.099)
Average Teacher Experience	0.073** (0.036)	0.083** (0.036)	0.081** (0.036)	0.121*** (0.031)	0.130*** (0.031)	0.126*** (0.030)
Percentage of first year Teachers	0.000 (0.014)	-0.002 (0.014)	-0.002 (0.014)	0.014 (0.012)	0.012 (0.012)	0.013 (0.011)
Percentage of second year teachers	-0.005 (0.011)	-0.010 (0.011)	-0.011 (0.011)	0.004 (0.010)	-0.002 (0.009)	-0.003 (0.009)
Percentage of teachers with emergency credentials	0.016 (0.013)	0.014 (0.013)	0.014 (0.013)	0.016 (0.010)	0.011 (0.010)	0.010 (0.010)
Percentage of teachers with credential waivers	-0.033 (0.036)	-0.029 (0.036)	-0.029 (0.037)	-0.071** (0.030)	-0.065** (0.031)	-0.064** (0.031)
Enrollment in grade	-0.002 (0.005)	-0.004 (0.005)	0.084 (0.055)	-0.005 0.004	-0.007* 0.004	0.064 0.048
Enrollment in grade squared*100			-0.083* (0.047)			-0.080** 0.041
n=	6079	6079	6079	6079	6079	6079
level controls	Y	Y	Y	Y	Y	Y
higher order controls		Y	Y		Y	Y

The table estimates equation (6) in the text for third graders. *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. All regressions are weighted by the number of test takers. Level controls refer to controls for percent minority, percent subsidized lunch and percent English learners as well as year effects. Higher order controls add quadratic and cubic terms as well as interactions for the demographic control variables.

Table 6 - First Stage Estimates for pooled second and third graders

Dependent Variable:	**Percentage of Students in Combination Classes**					**Average Class Size**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Instruments:										
Combination Classes	1.720***			0.734***		0.001			0.034	
Predictor	(0.174)			(0.174)		(0.016)			(0.018)	
Predicted Class Size		-1.794***		-1.599***	-1.632***		0.061***		0.070***	0.057***
		(0.102)		(0.110)	(0.100)		(0.009)		(0.010)	(0.009)
Predicted Number of Class			5.515***					-0.158***		
Groups			(0.393)					(0.033)		
Predicted Class size of					-1.172***					0.019**
lower grade					(0.111)					(0.008)
Enrollment Controls:										
Own Grade Enrollment	-0.115***	-0.026	-0.368***	-0.048**	-0.070***	0.013***	0.010***	0.020***	0.010***	0.012***
	-0.023	(0.017)	(0.026)	(0.024)	(0.025)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Own Grade Enrollment	0.018**	-0.003	0.015**	0.002	0.006	-0.003***	-0.003***	-0.004***	-0.002**	-0.002**
Squared*100	(0.009)	(0.007)	(0.006)	(0.009)	(0.010)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Lower Grade Enrollment	0.028			0.015	0.080***	0.001			-0.001	-0.003
	(0.021)			(0.021)	(0.023)	(0.002)			(0.002)	(0.002)
Lower Grade Enrollment	-0.005			-0.002	-0.019**	-0.001			-0.001	-0.001
Squared*100	(0.008)			(0.008)	(0.008)	(0.001)			(0.001)	(0.001)
Root MSE	13.797	13.677	13.715	13.705	13.64	1.2105	1.2076	1.2085	1.2085	1.2083
N	15,976	16,053	16,053	15,976	15,976	15,976	16,053	16,053	15,976	15,976

The table estimates equation (7) in the text for both second and third graders *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 7 - Pooled Second and Third Grade Predictions of Teacher Experience Using Non-linear Enrollment Instruments

Dependent Variable:	Average Years Teacher Experience				Percent Novice teachers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Instruments:								
Combination Class	0.034				-0.0071			
Predictor	(0.099)				(0.0041)			
Predicted Class		0.003		0.002		-0.0009		-0.0001
Size		(0.044)		(0.0432)		(0.001)		(0.0017)
Predicted Number of			0.038				0.0025	
Class Groups			(0.155)				(0.0065)	
Predicted Class size				0.012				0.0000
Of Lower Grade				(0.0398)				(0.0015)
Enrollment controls:								
Own Grade	-0.023**	-0.036***	-0.037***	-0.0359***	0.0008**	0.0003***	0.0008*	0.0012***
Enrollment	(0.011)	(0.008)	(0.011)	(0.0078)	(0.0004)	(0.0001)	(0.0004)	(0.0003)
Own Grade Enrol.	0.002	0.008***	0.008***	0.0075**	-0.0001	-0.0002**	-0.0002**	-0.0002*
Squared*100	(0.004)	(0.003)	(0.003)	(0.0029)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Lower Grade	-0.011			0.002	0.0001			-0.0004***
Enrollment	(0.009)			(0.0030)	(0.0004)			(0.0001)
N	15,976	16,053	16,053	15,976	15,976	16,053	16,053	15,976

The table shows the relationship between teacher experience and my instruments for pooled second and third graders *** Indicates 1% significance level and ** 5%. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 8 - 2SLS Estimates of the Determinants of Second Grade Test Scores

Dependent Variable:	**Language Scores**			**Math Scores**					
Class Size Control:	**Class Size Exogenous**						**Class Size Effect Zero**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classroom Variables:									
Percentage of students in combination classes	-0.195***	-0.163***	-0.177***	-0.180***	-0.199***	-0.167***	-0.179***	-0.197***	-0.166***
	(0.060)	(0.053)	(0.045)	(0.070)	(0.059)	(0.049)	(0.070)	(0.058)	(0.048)
Average Class Size	-0.095	-0.083	-0.086	-0.051	-0.062	-0.044			
	(0.110)	(0.110)	(0.109)	(0.122)	(0.124)	(0.121)			
Enrollment Controls:									
Enrollment in grade	-0.034	-0.019	-0.032	-0.021	-0.022	-0.019	-0.020	-0.022	-0.020
	(0.027)	(0.017)	(0.027)	(0.029)	(0.019)	(0.029)	(0.029)	(0.019)	(0.029)
Enrollment in grade squared*100	0.005	0.003	0.005	-0.001	0.003	-0.009	-0.001	0.004	-0.001
	(0.011)	(0.007)	(0.011)	(0.012)	(0.008)	(0.012)	(0.012)	(0.008)	(0.012)
Enrollment in lower Grade	0.016		0.016	0.004		0.003	0.005		0.003
	(0.021)		(0.021)	(0.024)		(0.024)	(0.024)		(0.024)
Enrollment in lower grade squared*100	-0.003		-0.003	0.003		0.003	0.004		0.003
	(0.008)		(0.008)	(0.009)		(0.009)	(0.009)		(0.009)
N	9950	9974	9950	9950	9974	9950	9950	9974	9950
Instruments	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1

This table presents the results of 2SLS regressions for the second graders Combination class percentage is the endogenous regressor. CSP stands for the Combination Class Predictor, PCS for Predicted Class Size and PCS-1 for the Predicted Class Size of the lower grade. All these instruments are defined in the text. Columns (1) –(6) treat class size as an exogenous regressor while columns (7) –(9) impose a zero coefficient. *** Indicates 1% significance level and ** 5%. All regressions are weighted by the number of test takers. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 9 - 2SLS Estimates of the Determinants of Third Grade Test Scores

Dependent Variable:	**Language Scores**			**Math Scores**					
Class Size Control:	**Class Size Exogenous**						**Class Size Effect = -.37**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classroom Variables:									
Percentage of students in combination classes	-0.368*** (0.102)	-0.301*** (0.066)	-0.328*** (0.060)	-0.360*** (0.121)	-0.302*** (0.076)	-0.323*** (0.069)	-0.362*** (0.125)	-0.322*** (0.077)	-0.331*** (0.074)
Average Class Size	0.102 (0.120)	0.098 (0.112)	0.093 (0.115)	0.078 (0.141)	0.084 (0.133)	0.069 (0.135)	-0.370	-0.370	-0.370
Enrollment Controls:									
Enrollment in grade	-0.078** (0.031)	-0.072*** (0.018)	-0.075** (0.030)	-0.054 (0.036)	-0.045** (0.021)	-0.051 (0.035)	-0.025 (0.078)	-0.040 (0.021)	-0.015 (0.069)
Enrollment in grade squared*100	0.026** (0.012)	0.021*** (0.007)	0.026** (0.012)	0.017 (0.014)	0.011 (0.008)	0.017 (0.014)	0.011 (0.010)	0.010 (0.008)	0.011 (0.010)
Enrollment in lower grade	-0.008 (0.028)		-0.007 (0.028)	-0.001 (0.033)		-0.001 (0.033)	-0.001 (0.033)		-0.001 (0.033)
Enrollment in lower grade squared*100	-0.002 (0.011)		-0.007 (0.011)	-0.002 (0.014)		-0.002 (0.013)	-0.003 (0.014)		-0.003 (0.013)
N	6026	6079	6026	6026	6079	6026	6026	6079	6026
Instruments	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1

This table presents the results of 2SLS regressions for the third graders. Combination class percentage is the endogenous regressor. CSP stands for the Combination Class Predictor, PCS for Predicted Class Size and PCS-1 for the Predicted Class Size of the lower grade. All these instruments are defined in the text. All regressions treat class size as exogenous. *** Indicates 1% significance level and ** 5%. All regressions are weighted by the number of test takers. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions.

Table 10 - 2SLS Estimates for Test Scores with Double Instrumenting

	math		language	
	(1)	(2)	(3)	(4)
A. Second Graders				
Percentage of students in combination classes	-0.174*	-0.180***	-0.202**	-0.193***
	(0.096)	(0.068)	(0.087)	(0.059)
Class Size	-0.383	-0.158	-1.085	-1.223
	(2.829)	(1.902)	(2.571)	(1.683)
N	9950	9950	9950	9950
Instruments	PCS PCS-1 NCL	PCS CLN CSP	PCS PCS-1 NCL	PCS CLN CSP
B. Third Graders				
Percentage of students in combination classes	-0.368**	-0.355***	-0.395**	-0.369***
	(0.193)	(0.139)	(0.177)	(0.118)
Class Size	-1.263	-1.315	-1.844	-1.638
	(5.186)	(3.017)	(4.709)	(2.546)
N	6026	6026	6026	6026
Instruments	PCS PCS-1	PCS CLN CSP	PCS PCS-1	PCS CLN CSP

This table presents the results of 2SLS regressions where both class size and combination class percentage are treated as endogenous regressors CSP stands for the Combination Class Predictor, CLN for the predicted number of class groups, PCS for Predicted Class Size and PCS-1 for the Predicted Class Size of the lower grade. All these instruments are defined in the text. *** Indicates 1% significance level and ** 5%. All regressions are weighted by the number of test takers. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 11 - Effect Sizes of a five percentage point increase in combination class percentage

Grade Level:	**second grade**		**third grade**	
Dependent Variable:	math	language	math	language
Effect on all students	.035-.051σ	.039-.049σ	.081-.093σ	.082-.097σ
Effect on only combination class students	.24-.36σ	.25-.35σ	.58-.63σ	.59-.66σ

Effect sizes are calculated by dividing the change in test scores implied by a five percentage point increase in combination class students by the standard deviation of test scores. In this case the between class standard deviation is used since the within student standard deviation is unavailable. The second row is obtained through scaling the first row by the reciprocal of the proportion of students in combination classes.