# THE IMPACT OF CLASS SIZE REDUCTION ON STUDENT ACHIEVEMENT

**Penny Fidler, Ph.D.**

**Program Evaluation and Research Branch**
**Los Angeles Unified School District**

**Planning, Assessment and Research Division Publication No. 109**

# THE IMPACT OF CLASS SIZE REDUCTION ON STUDENT ACHIEVEMENT

**Penny Fidler, Ph.D.**

**September 7, 2001**

**Program Evaluation and Research Branch**
**Los Angeles Unified School District**

**Planning, Assessment and Research Division Publication No. 109**

# EXECUTIVE SUMMARY

## Introduction

In July 1996, the California legislature enacted Senate Bill (SB) 1777, which made funds available to all school districts in California to reduce class sizes in kindergarten through 3rd grades to a twenty to one ratio of students to teachers. During the first year of the program, 1st and 2nd grade Los Angeles Unified School District (LAUSD) classrooms were reduced in size to twenty or fewer students followed by kindergarten and 3rd grade classrooms in the 1997-98 school year.

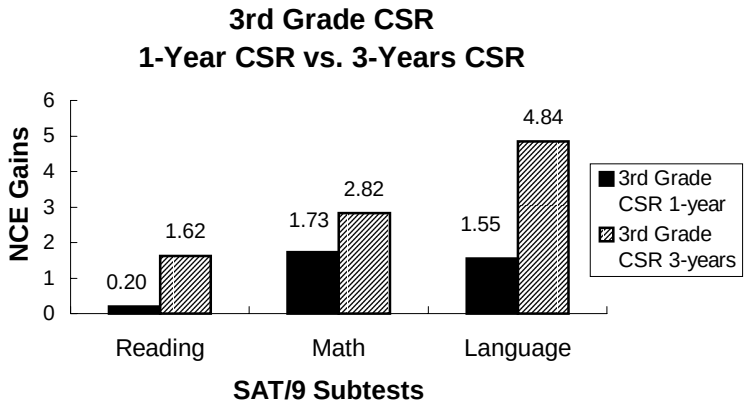
The purpose of this study was to examine the impact of class size reduction (CSR) on achievement among 3rd, 4th, and 5th grade students with different numbers of years of participation in CSR. Single-year matched gains (same students tested spring-to-spring) were calculated for the 1997-98 and 1999-00 school years. Student progress was assessed using scores from the SAT/9 reading, mathematics, and language achievement tests for the 1997-98, and 1999-00 school years.

This study focused on student achievement as measured by the SAT/9 reading, math, and language tests comparing three groups of students who participated in CSR from 0 to 3 years. The design of this study was a quasi-experimental longitudinal design. This design allowed for comparisons between the following groups: a) 3rd grade students with 3-years CSR vs. 3rd grade students with 1-year CSR, b) 4th grade students with 2 prior years CSR vs. 4th grade students with 0-years CSR, and c) 5th grade students with 1 prior year CSR vs. 5th grade students with 0-years CSR.

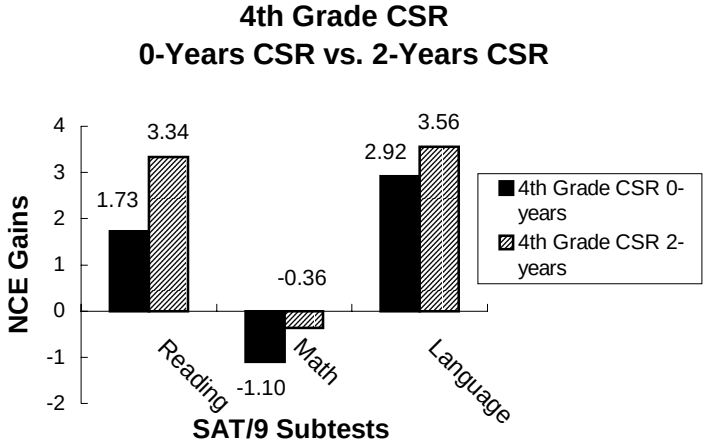
## Findings

The statistically significant findings are as follows: Third grade students with 3-years of CSR had larger gains in reading (matched NCE gain = 1.62) and language (4.84) than 3rd grade

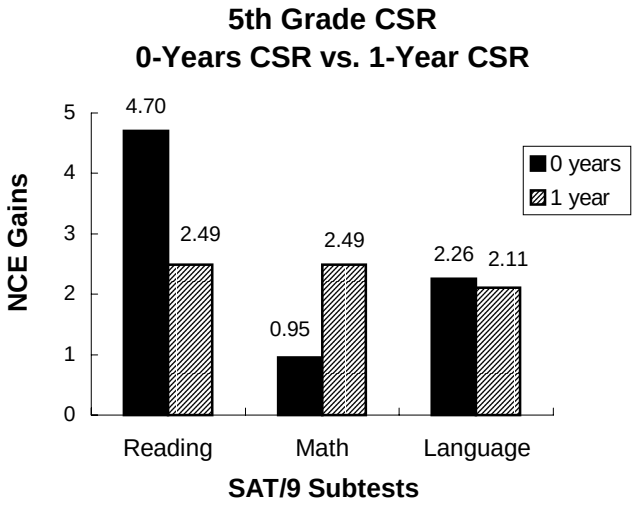
students with 1 year of CSR (.20, 1.55). In math, students with 3 years of CSR out-gained 1-year CSR students (2.82 vs. 1.73).




Fourth grade students with 2 years of CSR scored significantly larger gains in reading (3.34) than the 4th grade students with 0 years of CSR (1.73). They also exhibited somewhat larger gains in language (3.56) than students with 0 years of CSR (2.92). In mathematics, the data show losses for both groups.

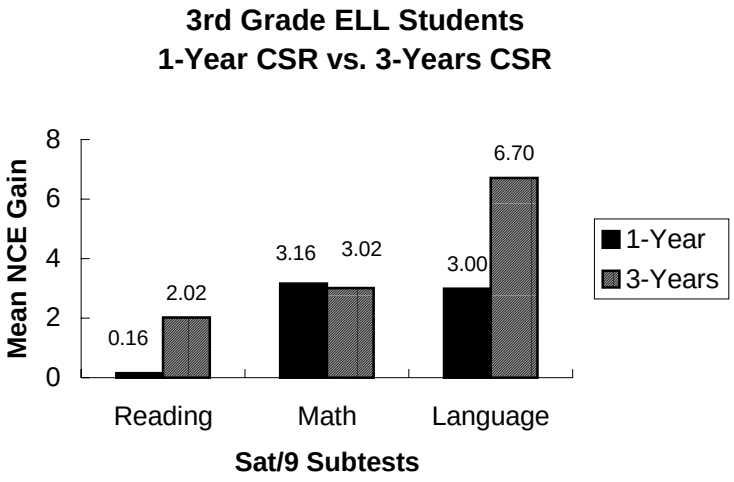


Fifth grade students with 1 year of CSR showed lower gains in reading than students having 0 years of CSR (2.49 vs. 4.7). These 1-year CSR students who had not been in reduced size classes for the past two years also showed lesser gains in language than non-CSR students.




In math, these 1-year CSR students showed larger gains (2.49) than their 0-years CSR comparison group (0.95).

Further analyses were performed by language classification. Third grade English language learners (ELL) with 3 years of CSR had significantly larger gains (2.02) in reading than students with 1 year of CSR (.16). In language, those with 3 years of CSR significantly outperformed (6.70) those with only 1 year of CSR (3.00). The results for 3rd grade ELL students in mathematics were less positive, with 1-year CSR students actually gaining slightly more (3.16) than 3-year CSR students (3.02).




Conclusions

Other things being equal, trends in the data support the hypothesis that longer exposure to CSR results in higher achievement in reading and language. These results were especially evident for English learners. In mathematics, however, the results were mixed. Third grade students with greater CSR exposure showed larger gains in math, but this was not true for English learners.

There have been several prior studies of the effects of CSR on achievement. Tennessee's class size reduction experiment showed that after four years of CSR, student achievement in reading and mathematics increased by approximately one quarter of a standard deviation. Rand/AIR's statewide study in California indicates that there are small (.04 - .10 of a standard deviation) effects of smaller class sizes on reading, math, language, and spelling achievement, but not for English language learners. Vital Research conducted a cross-sectional analysis of LAUSD student achievement data for 3rd grade students in CSR for two years as compared to

3rd grade students who never participated in CSR. They found effect sizes of .19 in reading, .25 in mathematics, and .27 in language, but excluded English language learners from the analysis.

It is worth noting that the methodology used in our study differs from the above-mentioned studies in that we used matched NCE gains (same students tested spring-to-spring) to assess the impact of CSR. This design makes it less likely that changes in student composition would account for any observed differences in achievement identified before and after CSR. Consequently, our use of matched gains resulted in smaller effect sizes (approximately .13 in reading, .07 in mathematics, .22 in language) than those previously reported by Vital Research for LAUSD 3rd graders. Our research indicates that CSR is benefiting student achievement in most subjects, but it may take more time to see large and sustained gains from this program among LAUSD students.

## The Impact of Class Size Reduction on Student Achievement

In July 1996, the California legislature enacted Senate Bill (SB) 1777. This legislation made funds available to all school districts in California that reduced class sizes in kindergarten through 3rd grades to a twenty to one ratio of students to certified teachers. On August 5, 1996, the Los Angeles Unified School District's (LAUSD) Board of Education directed that class sizes be reduced to twenty or fewer students with priority given to 1st and 2nd grade the first year of the program. In the 1997-98 school year, class size reduction (CSR) was extended to kindergarten and 3rd grade classrooms. The district reported that it was successful in achieving full compliance (93%) with SB 1777, beginning with the 1997-98 school year.

The implementation of CSR has been a unique challenge for LAUSD. Due to the size of the district, the shortage of certified teachers, and the lack of sufficient numbers of classrooms, the implementation of CSR has been very difficult, and as a result, far from perfect.

## Literature Review

There is a great deal of interest in and enthusiasm for CSR and many states are in the process of implementing CSR as a part of their educational policy. Class size reduction has been promoted as an educational panacea. Does the research support the hope that smaller class sizes will improve learning? While empirical evidence on the relationship between CSR and achievement is mixed, many studies indicate that reduced class sizes leads to improved student achievement (Finn & Voekl, 1992; Glass, Cahen, Smith & Filby, 1982; Robinson, 1990). Class size reduction appears to be more effective for minority, low-achieving, and low SES students (Angrist & Lavy, 1997). The Sage Report illustrated results for CSR in Wisconsin (Halbach, Ehrle, Zahorik, & Molnar, 2001). The researchers reported effect sizes of .14 to .35 (Cohen's-d)

for the achievement scores of African American students who were in reduced class sizes as compared to African American students who were not in reduced class sizes.[1]

Glass and Smith (1978) conducted a comprehensive meta-analysis of CSR studies and found when classes are comprised of twenty or fewer students, achievement scores increase. Ferguson (1991) reported results from a large-scale study in Texas. He found significant and positive relationships between class size and student achievement (effect sizes ranged from .15 to .35). However, Hanushek (1999) reanalyzed Ferguson's data and disputed the large effects that were found. Meanwhile, Greenwald, Hedges, and Caine (1996) have criticized the analytic techniques used in some of the research on CSR. They suggested that future research include controlled, experimental studies on the effect of CSR on student achievement.

Recent experimental studies of CSR have added tremendously to what is known about the effects of CSR on student achievement. Tennessee's Project STAR (Student Teacher Achievement Ratio) was a four-year longitudinal study of K-3 classrooms that employed a controlled research experiment to more reliably study the effects of smaller class sizes on achievement. The researchers found that students participating for four years in the STAR experiment had effect sizes (ES) of approximately 0.25 and the effects for inner-city minority students were twice as large as those for the rest of the population (Finn & Achilles, 1999). The effect sizes found in the STAR experiment are quite large and educationally important.[2] The STAR results revealed that the greatest gains occurred in kindergarten and 1st grade, and that just one-year exposure to CSR in the 3rd grade did not significantly impact test scores (Finn, 1998).[3]

[1] Effect size is a standardized value, which indicates the strength or efficacy of a treatment (e.g. CSR, Open Court Reading Program). A more detailed discussion of ES can be found on page 6 of this report.
[2] STAR effect sizes were .15 to .30.
[3] It is important to note that the students in the Tennessee study have very different demographic characteristics than the students involved in the LAUSD study. "Minority" students comprise only 25% of the population in the Tennessee study as compared to 90% "minority" students in LAUSD.

In a comprehensive analysis of CSR in California, Stecher, McCaffrey, Burroughs, Wiley, and Bohrnstedt (2000) examined the effects of CSR among 3rd grade students who were in reduced class sizes in California for one year (1997-98) and compared them to 3rd grade students who participated in CSR for two years (1997-99). They also assessed the "persistence" of CSR effects on achievement in 4th grade students no longer in CSR. Because implementation of the legislation statewide did not include any attempt to randomize assignment to conditions and the data could not be linked to individual students, Stecher et al. (2000) employed statistical controls (adjustments) in their study.[4] They found that students in CSR for two years had significantly higher achievement scores than students in CSR for one year. There was also evidence to support a "persistence" of CSR effects for 4th grade students one year after these students were no longer in reduced class sizes (Stecher et al., 2000). The ES that were reported are relatively small and not statistically significant (.032 to .067). While they concluded that there was some improvement in these students' achievement due to CSR, Stecher et al. (2000), in contrast to other findings, did not find any significant improvement in achievement for English Language Learners (ELL). They indicated that the lack of evidence for a CSR effect might be due to the fact that ELL students are concentrated in crowded, low SES schools, which lack the resources to adequately implement CSR. Finally, Stecher et al. (2000) reported that 5th grade students, the students with the least exposure to CSR, showed no appreciable increases in achievement.

Vital Research (2001) examined the effects of CSR on achievement in LAUSD students. They employed a quasi-experimental cross-sectional design that compared 3rd grade students' SAT/9 NCE scores from the spring administration in 1997 to 3rd grade students NCE scores in the spring of 1999. The students in the spring 1997 (control group) were compared to same grade

[4] In their regression analyses, Stecher et al. (2000), used 5th grade scores to control for "school effects" because of limitations in their data.

students in the spring 1999 (two years CSR). The districtwide findings are as follows: the effect sizes were .19 in reading, .25 in mathematics, and .27 in language. The ES findings for school calendar are as follows: 1) Year round calendar: .29 in reading, .47 in mathematics, and .51 in language; and 2) Traditional calendar: .17, .24 and .32, in reading, mathematics, and language, respectively. ELL students were not included in this analysis due the unavailability of their SAT/9 scores in 1997-98.

## Measurement Issues in CSR Research

### Matched Gain Scores

This study uses matched gain scores to represent improvements in student achievement. For example, matched yearly scores for 1999-00 are based on students who tested in both the 1999 and 2000 testing periods. A gain score is calculated by subtracting a student's pre-test score from his post-test score. This difference represents a particular student's gain in achievement, as measured by the SAT/9. Gain scores have been criticized as being unreliable (Cronbach & Furby, 1970; Linn & Slinde, 1977). However, Collins (1996) noted that the unreliability of the gain score does not imply that within examinees, change is imprecisely measured. Mellenbergh's (1999) results from examples of Classical Test Theory and Item Response Theory support Collins (1996) research. As Cribbie and Jamieson (2000) report, gain scores most accurately represent achievement when the population being measured is normally distributed. An analysis of LAUSD student data for 1998-99 and 1999-00 did in fact show that SAT/9 NCE scores were normally distributed. Therefore, employing matched gains strengthens the conclusions that can be made in this study. The use of matched gain scores allows for the assessment of longitudinal trends in the data (Linn, 2001).

Dose-response Methodology

Dose-response methodology is becoming a popular technique for analyzing program effects on student achievement (Stecher et al., 2000). A dose-response model was used because it posits that students spending more time in reduced class sizes should show higher levels of achievement as measured by the SAT/9 than students spending less time in reduced class sizes (Finn, 1998; Stecher et al. 2000). Linn (2001) has recently suggested that this approach to measuring and reporting achievement is an effective way to illustrate student "improvement over time."

Effect Sizes

How do we judge the magnitude of an effect? The ES index that will be employed in this study is Cohen's-d (Cohen, 1988). Cohen (1988) defines a small ES to be about .2, a medium ES is approximately .5, and a large ES is greater than .8. The ES analysis adds information regarding the educational importance of the calculated gain score (Glass & Smith, 1978). For example, an ES of .25 indicates that a student scored one-fourth of a standard deviation higher than the comparison group. Kirk (1995) reminds us that small and medium effect sizes are common in the social sciences. In educational research, effect sizes of the magnitude .03 to .40 have been reported to be educationally important (Glass & Smith, 1978). An ES of .03 may seem to be very small, but under certain circumstances, it may indicate a promising program effect.

The statistical significance of a set of data is not informative about the practical importance (or substantive significance) of the findings. It has been suggested that an index of substantive significance (educational importance) can be derived from an ES estimate (Rosenthal, 1991; Rosenthal & Rubin, 1978, 1979, 1991, 2001). At this time, the exact magnitude of an educationally significant ES is unknown (APA Task Force on Effect Size Reporting, 2001; Rosenthal & Rubin, 2001). Educational importance must be judged on an ad

hoc basis. It is up to policy makers and program experts to justify whether the magnitude of a program effect warrants the expenditure. We found that the range of "educationally important" effect sizes reported in the CSR research ranged from .04 to .56 using Cohen's-d ES index.

## Purpose of Study

The purpose of this study is to examine the impact of CSR on achievement in 3rd, 4th, and 5th grade students. Matched gains in achievement will be compared for the 1997-98 and 1999-00 school years.[5] The two cohorts from each grade represent a differential amount of CSR exposure. Student progress will be assessed using scores from the SAT/9 reading, mathematics, and language achievement tests for the 1997-98, and 1999-00 school years.

This report focuses on differences in 3rd, 4th, and 5th grade students' SAT/9 NCE achievement gains. Differences in gain scores and effect sizes will be reported for all grades by reading, mathematics, and language SAT/9 subtests. Following the results by grade, each language classification will be presented by subtest by grade. The language classifications are as follows: ELL, Redesignated English Proficient (RFEP), Initially Fluent English Proficient (IFEP), and English Only (EO). It is important to examine achievement gains by language program because ELL students have illustrated larger program effects than those students in the other language classifications (Finn & Achilles, 1990; Salazar, 1998).

## Research Questions

1. Will 3rd grade students who participated for three years in CSR have higher achievement gains than 3rd grade students who have participated for one year in CSR?
2. Will 4th grade students with two prior years of CSR and 5th grade students with one prior year of CSR, have similar gains to those of 3rd grade students?
3. Will there be a "persistence" of CSR effects for 4th and 5th grade students?

[5] The 1997-98 and 1999-00 school years are being compared to provide comparisons for students who had the earliest CSR exposure to those having the most recent CSR exposure.

4. Will participation in CSR have a differential effect on achievement due to language classification?

## Method

Participants

Four cohorts of students were identified for this study.[6] These cohorts are described below.

Cohort 0: Selected students for this cohort were 4th and 5th grade students in 1997-98 who did not have any CSR exposure.

Cohort 1: Selected students for this cohort were 3rd grade students in 1997-98, 4th grade students in 1998-99, and 5th grade students in 1999-00. These students completed one year of CSR in 1997-98.

Cohort 2: Selected students for this cohort were 2nd grade students in 1997-98, 3rd grade students in 1998-99 and 4th grade students in 1999-00. These students participated in two years of CSR from 1997 to 1999.

Cohort 3: Selected students for this cohort were 1st grade students in 1997-98, 2nd grade students in 1998-99, and 3rd grade students in 1999-00. These students participated in three years of CSR from 1997 to 2000.

These four cohorts were chosen so that the dose-response design could be used to examine the effects of different lengths of exposure for 3rd grade students and also extended to 4th and 5th grade students to glean information about the persistence of CSR effects. Comparisons are made throughout the study between these cohorts at different points in their exposure to or distance from CSR. Therefore, valid conclusions can be drawn about both the amount of exposure a student has to CSR and the lasting effects of one and two years of CSR.

[6] Only those students with matched gain scores were employed in the analysis.

## Design and Procedure

The design of this study is a quasi-experimental design. This design allows for comparisons between 3rd grade students (Cohort 3 vs. Cohort 1), between 4th grade students (Cohort 2 vs. Cohort 0), and between 5th grade students (Cohort 1 vs. Cohort 0) having differential exposure (in years) to reduced class sizes. This study focused on the SAT/9 reading, math, and language matched NCE scores for students who are receiving instruction or who have received instruction in the district CSR Program.

Gain scores were calculated for each matched grade cohort. In order to adjust for unusually high or low scores, the following procedure was employed: all absolute gains larger than thirty NCE were recoded to thirty NCE to eliminate the impact of unusually high or low scores. This procedure insures that the distribution of scores falls between plus and minus three standard deviations from the mean and controls for outliers and possible errors in the data.

If reduced class sizes had a large impact on student achievement then it would be expected that the differences in gain scores and effect sizes for the students having more exposure to CSR would be larger than the differences in gain scores and effect sizes for students having less exposure to CSR (Glass & Smith, 1978).

## Threats to the Validity of the Study

Because history is a threat to the validity of the results, achievement scores were compared for other cohorts who completed one and two years of CSR. History means that other programmatic effects or events other than CSR have affected all grades or all SAT/9 subtest scores. If history is not an issue, then any substantial gains scored solely by Cohort 3 is most likely the result of CSR.

## Results

Academic Achievement – SAT/9

Tables 1-3 depict the matched gains, standard deviations, and sample sizes for 3rd, 4th, and 5th grade students in 1997-98 and 1999-00, regardless of language classification.

Table 1 illustrates the findings for 3rd grade students. The matched gains for Cohort 3 were compared to the matched gains for Cohort 1. The differences in matched gains along with their associated effect sizes are shown at the bottom of the table.

Table 2 shows the results for 4th grade students. This table illustrates the matched gains of Cohort 2 and the matched gains for Cohort 0. At the bottom of the table, the differences in matched gains along with their associated effect sizes are shown.

Table 3 presents the findings for 5th grade students. This table contains comparisons of Cohort 1 and Cohort 0. The effect sizes of the differences in matched gains are depicted at the bottom of the table.

3rd Grade Achievement Reading, Mathematics, and Language

Table 1 shows 3rd grade achievement gains, standard deviations, sample sizes, and effect sizes (Cohort 3 minus Cohort 1).

**Table 1**
**3rd Grade**
**Matched Gain Scores in Reading, Mathematics, and Language**
**Cohort 3 minus Cohort 1**

| **Descriptive Statistics** | | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **Cohort 1** | | | |
| NCE Gain | .20 | 1.73 | 1.55 |
| SD | 11.39 | 14.13 | 14.30 |
| N | 15,025 | 16,899 | 16,512 |
| **Cohort 3** | | | |
| NCE Gain | 1.62 | 2.82 | 4.84 |
| SD | 11.28 | 14.80 | 14.42 |
| N | 45,761 | 49,143 | 47,592 |
| **Effect Sizes** **Cohort 3 minus Cohort 1** | | | |
| Difference | 1.43 | 1.09 | 3.29 |
| **Effect Size** | **.13**[a] | **.07** | **.22** |

[a.] Effect Sizes Calculated with Cohen's-d

The results in Table 1 illustrate mixed findings. The difference in gain scores was positive on all three subtests for Cohort 3 vs. Cohort 1. However, the effect sizes were much smaller in mathematics than in language or reading. The ES for the difference in language gains is educationally and statistically significant (d = .22). An ES of .22 means that Cohort 3 scored almost one-fourth of a standard deviation higher than Cohort 1 in language. The ES for reading is also educationally important (d = .13). The ES for mathematics (d = .07) is much smaller than the ES for language or reading and is not considered to be educationally important.[7]

4th Grade Achievement in Reading, Mathematics, and Language

[7] In this study, an educationally important effect size is considered to be .10 or greater.

Table 2 depicts 4th grade achievement in reading, mathematics, and language (Cohort 2 minus Cohort 0).

**Table 2**
**4th Grade**
**Matched Gain Scores in Reading, Mathematics, and Language**
**Cohort 2 minus Cohort 0**

| **Descriptive Statistics** | | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **Cohort 0** | | | |
| NCE Gain | 1.73 | -1.10 | 2.92 |
| SD | 10.81 | 12.27 | 12.35 |
| N | 16,734 | 17,626 | 17,163 |
| **Cohort 2** | | | |
| NCE Gain | 3.34 | -.36 | 3.56 |
| SD | 10.44 | 12.30 | 12.73 |
| N | 46,834 | 48,581 | 46,878 |
| **Effect Sizes** **Cohort 2 minus Cohort 0** | | | |
| Difference in Gains | 1.61 | .74 | .64 |
| **Effect Size** | **.15** | **.06** | **.05** |

The results depicted in Table 2 show that in reading, the difference between Cohort 2 and Cohort 0 resulted in an educationally important ES ($\underline{d}$ = .15). However, the effect sizes were fairly small in mathematics ($\underline{d}$ = .06) and in language ($\underline{d}$ = .05).

5th Grade Achievement in Reading, Mathematics, and Language

Table 3 shows 5th grade achievement gains, standard deviations, sample sizes, and effect sizes (Cohort 1 minus Cohort 0).

**Table 3**
**5th Grade**
**Matched Gain Scores in Reading, Mathematics, and Language**
**Cohort 1 minus Cohort 0**

| **Descriptive Statistics** | | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **Cohort 0** | | | |
| NCE Gain | 4.70 | .95 | 2.26 |
| SD | 10.07 | 11.47 | 12.54 |
| N | 17,683 | 19,043 | 18,755 |
| **Cohort 1** | | | |
| NCE Gain | 2.49 | 2.49 | 2.11 |
| SD | 10.09 | 11.28 | 12.53 |
| N | 42,590 | 44,597 | 43,923 |
| **Effect Sizes** | | | |
| Cohort 1 minus Cohort 0 | | | |
| Difference in Gains | -2.21 | 1.54 | -.15 |
| **Effect Size** | **-.22** | **.14** | **-.01** |

The comparison of Cohort 1 and Cohort 0 showed a significant decrease in reading (d = -.22). The ES in language (d = -.01) indicates that there was virtually no change in language achievement. However, the ES in mathematics was educationally significant (d = .14).

Summary of 3rd, 4th, and 5th Grade Achievement in Reading, Mathematics, and Language

Table 4 provides a summary of the results for Tables 1 – 3 presented above.

**Table 4**
**Summary of Tables 1 - 3**
**Differences in Gains and Effect Sizes for SAT/9 Subtests by Grade**

| | Reading | Mathematics | Language |
|---|---|---|---|
| **3rd Grade** (Cohort 3 minus Cohort 1) | | | |
| Difference | 1.43 | 1.09 | 3.29 |
| **Effect Size** | **.13** | **.07** | **.22** |
| **4th Grade** (Cohort 2 minus Cohort 0) | | | |
| Difference | 1.61 | .74 | .64 |
| **Effect Size** | **.15** | **.06** | **.05** |
| **5th Grade** (Cohort 1 minus Cohort 0) | | | |
| Difference | -2.21 | 1.54 | -.15 |
| **Effect Size** | **-.22** | **.14** | **-.01** |

The ES for 3rd grade students in reading is educationally important (d = .13). Forth grade students also had an educationally important ES (d = .15). However, for 5th grade students, the ES in reading shows a large decrement (d = -.22). The reading ES for 3rd grade students is similar to the ES for 4th grade students, which indicates some persistence of CSR reading effects.

The results found for mathematics are in the opposite direction of those found for reading and language. The effect sizes for 3rd (d = .07) and 4th grade students (d = .06) were much smaller than the ES for 5th grade students (d = .14).

In language, 3rd grade students had the largest ES among all three grades (d = .22). This ES is educationally and statistically significant. The small effect sizes for 4th and 5th grade students suggest that there is no persistence of CSR effects in language.

Tables 5-7 illustrate the results presented above broken down by language classification.[8] These analyses are critical because of the underrepresentation of tested ELL students in spring 1997. At that time, the district had a policy of not testing ELL students in the English Language. We were interested in finding out whether the results would have been different from what we found had all ELL students tested in spring 1997.

The individual tables showing the gains, standard deviations, and effect sizes in reading, mathematics, and language by language classification for 3rd, 4th, and 5th grade students are included in the Appendix (Tables A1-A9).

Reading, Mathematics, and Language Achievement by Language Classification

3rd Grade Results in Reading, Mathematics, and Language by Language Classification.

**Table 5**
**3rd Grade**
**Summary of 3rd Grade Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Language Classification**
**Cohort 3 minus Cohort 1**

| | ELL | RFEP | IFEP | EO |
|---|---|---|---|---|
| **Reading** | | | | |
| Difference | 1.86 | 1.66 | .78 | 1.60 |
| **Effect Size** | **.16** | **.15** | **.06** | **.12** |
| **Mathematics** | | | | |
| Difference | -.14 | .66 | .96 | 1.22 |
| **Effect Size** | **-.01** | **.05** | **.07** | **.08** |
| **Language** | | | | |
| Difference | 3.70 | .85 | .70 | 1.29 |
| **Effect Size** | **.25** | **.05** | **.05** | **.09** |

Table 5 illustrates the results for Cohort 3 minus Cohort 1 and the associated effect sizes by language classification.[8] Third grade students in Cohort 3 had the largest dose of CSR. In language, ELL students in this cohort had the largest ES (d = .25). This ES represents the most

[8] The language classifications are as follows: ELL (English Language Learner), RFEP (Fully Designated English Proficient), IFEP (Initially Fluent English Proficient), and EO (English Only).

educationally important finding in Table 5. The other three language classifications had fairly small effect sizes in language. RFEP students had an educationally important ES in reading (d = .15). The effect sizes for IFEP students were very small for all three subtests. The largest ES for EO students was in reading and is educationally important (d = .12). These students had smaller effect sizes in the other 2 subtests.

4th Grade Results in Reading, Mathematics, and Language by Language Classification

**Table 6**
**4th Grade**
**Summary of 4th Grade Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Language Classification**
**Cohort 2 minus Cohort 0**

| | ELL | RFEP | IFEP | EO |
|---|---|---|---|---|
| **Reading** | | | | |
| Difference | 2.10 | 2.97 | 2.14 | .64 |
| **Effect Size** | **.20** | **.30** | **.21** | **.06** |
| **Mathematics** | | | | |
| Difference | .18 | .33 | .88 | 1.08 |
| **Effect Size** | **.02** | **.03** | **.07** | **.09** |
| **Language** | | | | |
| Difference | .44 | -.39 | 1.01 | .42 |
| **Effect Size** | **.03** | **-.03** | **.08** | **.03** |

Table 6 depicts the results for Cohort 2 minus Cohort 0 and the associated effect sizes by language classification. The effect sizes in reading were large and educationally significant for ELL students (d = .20), RFEP students (d = .30), and IFEP students (d = .21). In contrast, EO students had a very small ES in reading. Table 6 shows that there were no educationally important effect sizes in mathematics or language for 4th grade students.

5th Grade Results in Reading, Mathematics, and Language by Language Classification

**Table 7**
**5th Grade**
**Summary of 5th Grade Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Language Classification**
**Cohort 1 minus Cohort 0**

| | ELL | RFEP | IFEP | EO |
|---|---|---|---|---|
| **Reading** | | | | |
| Difference | -1.92 | -2.33 | -2.00 | -2.59 |
| **Effect Size** | **-.19** | **-.25** | **-.19** | **-.25** |
| **Mathematics** | | | | |
| Difference | 1.21 | 1.55 | 1.43 | 1.59 |
| **Effect Size** | **.12** | **.13** | **.12** | **.12** |
| **Language** | | | | |
| Difference | -1.04 | .25 | .56 | -.23 |
| **Effect Size** | **-.08** | **.02** | **.04** | **-.02** |

Table 7 illustrates the differences in gains between Cohort 1 and Cohort 0 and the associated effect sizes by language classification. The effect sizes in reading for all language classifications reflect extreme decrements in achievement. Mathematics effect sizes were fairly strong and very similar across all language classifications. In language, there was little gain or loss.

Academic Achievement for Language Classification by Reading, Mathematics, and Language by Cohort

Tables 8-11 present the results for ELL, IFEP, RFEP, and EO, respectively by cohort and SAT/9 subtests.

ELL Language Classification

**Table 8**
**Summary of ELL Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Grade**

| ELL | | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **3rd Grade** (Cohort 3 minus Cohort 1) | | | |
| Difference | 1.86 | -.14 | 3.70 |
| **Effect Size** | **.16** | **-.01** | **.25** |
| **4th Grade** (Cohort 2 minus Cohort 0) | | | |
| Difference | 2.10 | .18 | .44 |
| **Effect Size** | **.20** | **.02** | **.03** |
| **5th Grade** (Cohort 1 minus Cohort 0) | | | |
| Difference | -1.92 | 1.21 | -1.04 |
| **Effect Size** | **-.19** | **.12** | **-.08** |

The findings in Table 8 illustrate a similar pattern to the results shown in Table 4. Third grade ELL students had the largest ES in language (d = .25), followed by reading (d = .16). Fourth grade ELL students had the largest ES in reading (d = .20). The ES results for ELL students in mathematics are in the opposite direction than were expected. Table 8 shows that there was very little change in mathematics' achievement for either 3rd or 4th grade students. However, 5th grade ELL students had an educationally important ES in mathematics (d = .12).

IFEP Language Classification

**Table 9**
**Summary of IFEP Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Grade**

| IFEP | | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **3rd Grade** (Cohort 3 minus Cohort 1) | | | |
| Difference | .78 | .96 | .70 |
| **Effect Size** | **.06** | **.07** | **.05** |
| **4th Grade** (Cohort 2 minus Cohort 0) | | | |
| Difference | 2.14 | .88 | 1.01 |
| **Effect Size** | **.21** | **.07** | **.08** |
| **5th Grade** (Cohort 1 minus Cohort 0) | | | |
| Difference | -2.00 | 1.43 | .56 |
| **Effect Size** | **-.19** | **.12** | **.04** |

Table 9 does not illustrate any significant achievement effects for 3rd grade IFEP students. The 4th grade IFEP students had a large and educationally significant ES in reading (d = .21), but no appreciable achievement effects in mathematics or in language. The results presented in Table 9 indicate a large decline in reading achievement (d = -.19) for 5th grade IFEP students and a very small ES in language (d = .04). However, 5th grade IFEP students had an educationally significant ES in mathematics (d = .12).

.

RFEP Language Classification

**Table 10**
**Summary of RFEP Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Grade**

| RFEP | | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **3rd Grade** (Cohort 3 minus Cohort 1) | | | |
| Difference | 1.66 | .66 | .85 |
| **Effect Size** | **.15** | **.05** | **.05** |
| **4th Grade** (Cohort 2 minus Cohort 0) | | | |
| Difference | 2.97 | .33 | -.39 |
| **Effect Size** | **.30** | **.03** | **-.03** |
| **5th Grade** (Cohort 1 minus Cohort 0) | | | |
| Difference | -2.33 | 1.55 | .25 |
| **Effect Size** | **-.25** | **.13** | **.02** |

Table 10 shows that the 3rd grade RFEP students had an educationally important ES in reading (d = .15), but had very small effect sizes in mathematics and language (d = .05). The 4th grade ES in reading is very large and educationally significant (d = .30). However, the effect sizes for 4th grade mathematics and language illustrate very little improvement in achievement. The ES for 5th grade reading indicates a steep decline in achievement (d = -.25). The trend in mathematics is similar to the trend for ELL and IFEP students, where the smallest amount of CSR resulted in the largest ES. RFEP students showed little change in language achievement.

EO Language Classification

**Table 11**
**Summary of EO Differences in Gains and Effect Sizes for Reading, Mathematics, and Language by Cohort**

| | EO | | |
|---|---|---|---|
| | **Reading** | **Mathematics** | **Language** |
| **3rd Grade** (Cohort 3 minus Cohort 1) | | | |
| Difference | 1.60 | 1.22 | 1.29 |
| **Effect Size** | **.12** | **.12** | **.09** |
| **4th Grade** (Cohort 2 minus Cohort 0) | | | |
| Difference | .64 | 1.08 | .42 |
| **Effect Size** | **.06** | **.09** | **.03** |
| **5th Grade** (Cohort 1 minus Cohort 0) | | | |
| **Difference** | -2.59 | 1.59 | -.23 |
| **Effect Size** | **-.25** | **.12** | **-.02** |

Table 11 indicates that 3rd grade EO students had educationally significant effect sizes in reading (d = .12) and mathematics (d = .12). The 3rd grade ES for language (d = .09) is somewhat smaller than the others just listed. The effect sizes for 4th grade EO students were generally small across subtest. Table 11 also indicates that the 5th grade EO students had a significant decline in reading (d = -.25), a significant ES in mathematics (d = .12), and a very small ES in language (d = -.02).

## Post Hoc Results

These next results are presented to address possible questions arising from the underrepresentation of ELL students in spring 1997. Tables 12-14 show the composition of LAUSD 3rd grade students who took the SAT/9 reading subtest in spring 1997 and spring 2000 with respect to language classification, ethnicity, and meal program participation. Table 15 augments the results shown in Tables 12 and 14 with the addition of gain scores for Cohort 1 and Cohort 3 and resulting effect sizes for language program classification by meal program participation.

Proportion of 3rd Grade Students by Language Classification

**Table 12**
**Reading Matched Scores**
**Grade 3 by**
**Language Classification**

| | ELL | RFEP | IFEP | EO |
|---|---|---|---|---|
| **Cohort 1 (1997-98)** | | | | |
| **%**[a] | 29.6% | 4.0% | 15.5% | 51.0% |
| **N** | 4,350 | 584 | 2,273 | 7,491 |
| **Cohort 3 (1999-00)** | | | | |
| **%** | 64.6% | 4.4% | 7.6% | 23.2% |
| **N** | 25,135 | 1,746 | 2,988 | 9,010 |

[a] Discrepancies in total percent due to missing values.

Table 12 illustrates the dramatic underrepresentation of ELL students who took the spring 1997 SAT/9 reading test. The numbers of students were similar in the other three language classifications.

Proportion of 3rd Grade Students by Ethnicity

**Table 13**
**Reading Matched Scores**
**Grade 3 by**
**Ethnicity**

| | Black | Asian | Hispanic | White |
|---|---|---|---|---|
| **Cohort 1 (1997-98)** | | | | |
| **%**[a] | 26.1% | 5.3% | 56.29% | 8.50% |
| **N** | 3,828 | 780 | 8,258 | 1,249 |
| **Cohort 3 (1999-00)** | | | | |
| **%** | 11.6% | 6.9% | 80.0% | 4.0% |
| **N** | 4,515 | 2,698 | 31,122 | 1,557 |

[a] Discrepancies in total percent due to missing values.

Table 13 shows the difference in ethnic composition of the students who took the reading subtest in spring 1997 and spring 2000. The most noteworthy results are the differences in the number of Asian and Hispanic students tested in Cohort 1 as compared to Cohort 3.

Proportion of 3rd Grade Students by Meal Program

Table 14 depicts the difference in the number of students taking part in the free/reduced lunch program as reflected by the SAT/9 reading subtest between spring 1997 and spring 2000. Meal program participation has been an important predictor of SES in LAUSD students.

**Table 14**
**Reading Matched Scores**
**Grade 3 by**
**Meal Program**

| | YES | NO |
|---|---|---|
| **Cohort 1 (1997-98)** | | |
| **%**[a] | 67.6% | 32.4% |
| **N** | 10,361 | 4,963 |
| **Cohort 3 (1999-00)** | | |
| **%** | 85.0% | 14.8% |
| **N** | 38,900 | 6,769 |

[a] Discrepancies in total percent due to missing values.

The results shown in Table 14 indicate that there were significantly more students participating in the free/reduced lunch program in Cohort 3 than in Cohort 1.

Gain Scores and Resulting Effect Sizes for Language Program Classification by Meal Program Participation for Cohort 3 minus Cohort 1

Table 15 illustrates the cross-tabulation of language classification by meal program participation.

**Table 15**
**Reading Gains**
**Grade 3**
**Language Classification by Meal Program Participation**
**Cohort 3 minus Cohort 1**

**Descriptive Statistics**

| | ELL | | RFEP | | IFEP | | EO | |
|---|---|---|---|---|---|---|---|---|
| **Meal Program** | **Yes** | **No** | **Yes** | **No** | **Yes** | **No** | **Yes** | **No** |
| **Cohort 1 (1997-98)** | | | | | | | | |
| NCE Gain | .28 | -.60 | -2.73 | .29 | .30 | -.75 | -.63 | .48 |
| SD | 11.20 | 11.27 | 10.09 | 10.69 | 11.23 | 11.07 | 11.75 | 12.22 |
| N[a] | 3,079 | 455 | 434 | 162 | 1,660 | 660 | 5,186 | 3,441 |
| **Cohort 3 (1999-00)** | | | | | | | | |
| NCE Gain | 2.05 | 1.22 | -.19 | -.85 | .83 | .56 | .66 | 2.84 |
| SD | 10.93 | 11.77 | 10.23 | 10.50 | 11.54 | 10.98 | 11.80 | 11.94 |
| N | 25,135 | 891 | 1,446 | 229 | 2,988 | 829 | 8,168 | 4,818 |

**Effect Sizes**
Cohort 3 minus Cohort 1

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Difference | 1.77 | 1.82 | 2.54 | -1.14 | .53 | 1.31 | 1.23 | 2.36 |
| **Effect Size** | **.16** | **.17** | **.25** | **-.11** | **.05** | **.12** | **.10** | **.20** |

[a] Cross-tabulation procedures tend to exacerbate the number of missing cases.

English language learners who participated in the free/reduced lunch program (low SES) had larger gains than higher SES ELL students. The ELL students who comprised Cohort 1 were level 5 English language development (ELD) students. ELD level 5 represents the most English-fluent ELL students. Cohort 3 ELL students were composed of all five levels of ELD. Thus, the difference between Cohort 1 and Cohort 3 matched gains may not correctly reflect the effect of reduced class sizes on reading achievement.

In order to address the difference in composition for the ELL students in Cohort 1 and Cohort 3, analyses were conducted by ELD level for the students in Cohort 3. The results indicated that ELD level 5 students had a slightly smaller gain (M=1.44) than was presented in Table A1 for all ELL students. Thus, the corrected ES for the difference between Cohort 1 ELL students (ELD level 5) and Cohort 3 (ELD level 5) students is .11. This estimate is slightly lower than the effect size that was presented in Table 5 (.16). This estimate also corrects for meal program participation/SES composition between cohorts.

## Discussion

All other things being equal, trends in the data support the hypothesis that longer exposure to CSR results in higher achievement. Our findings are supported by previous research (Finn & Voekl, 1992; Glass, Cahen, Smith & Filby, 1982; Lipsey & Wilson, 1993; Robinson, 1990). Studies conducted on the effects of CSR on achievement, after students were no longer in smaller classes (Stecher et al., 2000), indicate that there are much smaller gains and effect sizes just one year after being returned to larger classes. This trend was also evident in our results for reading and language. In mathematics, however, our results were in the opposite direction than expected. We found that 5th grade students had larger effect sizes than 3rd or 4th grade students. Our findings regarding the impact of CSR on mathematics achievement are not supported by previous CSR research (Finn & Achilles, 1999; Robinson, 1990; Stecher et al., 2000; Vital Research, 2001).

It is interesting that the relationship between CSR exposure and mathematics achievement is very different from the one found for reading and language achievement (see Table 4). Perhaps the lag in mathematics' achievement in reduced-sized classes is due to the fact that mathematics has been under-emphasized by LAUSD in grades K-3. As a result, it takes more years in school for students to accumulate the knowledge in mathematics to effectively

illustrate achievement gains. It is also possible that the curriculum is more rigorous in the upper elementary grades, regardless of CSR. Because so many new teachers had to be hired when CSR was implemented in LAUSD, many less experienced teachers were placed in K-3 classrooms. The expectation of larger achievement gains in mathematics due to CSR may not be met at this time for K-3 students in LAUSD.

Future research will probably illustrate more positive mathematics achievement for students in reduced size classes, especially, in light of the new District Math Plan that is currently being initiated in LAUSD. Stasz and Stecher (2000) reported positive mathematics achievement results for students in reduced size classes. They found significant differences in several instructional practices between CSR and non-CSR classrooms, which have significantly affected mathematics achievement. Some of the differences that Stasz and Stecher (2000) observed between CSR and non-CSR classrooms are as follows: 1) there are more teachers with Master's degrees in reduced sized classes; 2) students spend more time playing mathematics games in reduced sized classes; and 3) students spend more time using patterns to find relationships in mathematics in reduced sized classes. The "reform-oriented" type of instruction they found in smaller classes included collecting and analyzing data; and writing about mathematics.

In order to be able to explain the mathematics results in LAUSD students, we conducted several post hoc analyses. In one analysis, we examined the relationship between student achievement and teacher certification. The preliminary results indicate that 3rd grade teachers, who were certified, had students with larger gains than non-certified teachers. The other post hoc analyses are explicit to language classification and are discussed at the end of the next section.

## Language Classification

### ELL

Third grade ELL students had educationally important effect sizes in language and reading. The ES for reading was sustained into the 4th grade. This trend seems typical of the manner in which these groups have historically performed on the SAT/9 (Finn & Achilles, 1990; Salazar, 1998). The 3rd grade ELL students' achievement in reading and language performance was remarkable. Because of the CSR initiative, many newly BCLAD certified teachers were placed in schools with higher proportions of ELL students. While these teachers may not be as experienced as the other certified teachers in the district, they appeared to be trained in effective teaching practices for the students they were serving. As a result, ELL students' achievement benefited greatly from CSR.

### IFEP

The results presented in Table 9 of this study provide little support for a CSR effect on achievement for IFEP students**.**

### RFEP

The results presented in Table 10 lend some support for a CSR effect in reading for RFEP students. The findings do not illustrate any support for a CSR effect in mathematics or language achievement. The lower level of achievement gains for students in this language classification may be due to the fact that these students traditionally score the highest on tests of achievement. Because the RFEP students have high pre-test scores, any subsequent gains on the post-test are minimized by an inherent "ceiling" effect in the scores.

### EO Results

The results for EO students indicate that there may be a CSR effect on achievement for all three subtests of the SAT/9. The effect sizes presented in Table 11 may also be the result of

teaching experience. We found some interesting preliminary results on the effects of teaching experience on student achievement in LAUSD. The early findings indicate that EO students have more experienced teachers (M=13.08 years) as compared to ELL students (M=10.52 years). Recent research (Finn & Achilles, 1999; Stecher et al., 2000) supports these preliminary conclusions.

The findings on teacher certification and teaching experience will be part of a larger analysis, which will be conducted to link teacher and classroom effects to individual student achievement. This report is in process and will be released soon. Because of the unclear findings in the relationship between CSR exposure and achievement, we are reporting some general trends that were found in our preliminary results. We examined the change in the percent of students taught by certified teachers in 1998-99 to those taught by certified teachers in 1999-00 by language classification. The findings illustrate that for all language classifications, the percent of students taught by certified teachers was smaller in 1999-00 as compared with 1998-99. Because of CSR, the teacher resources are being stretched to the limit; however, students are still showing increases in achievement.

The early results indicate that the more experienced teachers in LAUSD seem to choose schools that are less crowded and are located in areas that are more affluent. These schools tend to have a majority of EO students. More experienced teachers also seem to choose schools having "traditional" calendars, which are less crowded than schools on year-round calendars (Vital Research, 2001).

## Conclusions

The results presented in this paper regarding the impact of CSR on achievement are mixed. We believe that CSR will help to increase student achievement, especially for students who need it the most: low SES students, limited English-speaking students, and those students in

inner-city schools (Angrist & Lavy, 1997; Finn; 1998; Finn & Achilles, 1999; Greenwald et al., 1996; Grissmer; 1999; Vital Research, 2001).

According to Stecher et al. (2000), there have been small, but consistent achievement gains in students of all backgrounds. The findings presented in our study agree with those presented by Stecher et al. (2000). Furthermore, the ES reported in our study were quite similar to those reported by the majority of CSR research studies (Finn & Achilles, 1999; Glass & Smith, 1978; Stecher et al., 2000).

The 5th grade mathematics effect sizes were strong and very similar across all language classifications. These results do not lend support for a persistence of CSR effects, because the highest dose group (3rd grade students) had very small increments in mathematics achievement. Thus, there appears to be other factors that are impacting achievement, as well as smaller class sizes.

It can be concluded from the results of this study that CSR does help to increase language achievement gains, especially for ELL students. The implications of CSR will be more evident next year, with an additional year of CSR exposure reflected by 2001 SAT/9 scores. It is likely that the long-term effect of CSR will result in improved student achievement, especially for ELL students.

## Assumptions and Limitations

### Threats to Internal Validity

There are several threats to internal validity in the analyses presented in this paper. One threat to the validity of this study is the underrepresentation of ELL students taking the SAT/9 in spring 1997. The second threat is history and was previously discussed. Thus, any inferences must be made with these possible threats in mind.

## Future Research

Future research will examine the impact of teacher characteristics on student achievement. In order to make correct inferences about the relationships between teacher characteristics and individual student achievement, Hierarchical Linear Modeling (HLM; Bryk & Raudenbush, 1992) and Multilevel Structural Equation Modeling (MLSEM) will be employed. The proposed analyses will link results from classroom observations to individual teachers and their student's actual achievement results.

References

Angrist, J.D. & Lavy, V. (1997). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Unpublished manuscript.

Ferguson, R.F. (1991). Paying for public education: New evidence on how and why money matters. Harvard Journal on Legislation, 28, 465-498.

Finn, J.D. (1998). Class size and students at risk: What is known? What is next? Washington D.C.: U.S. Department of Education-OERI.

Finn, J.D. & Achilles, C.M. (1999). Tennessee's class size study: Findings, implications, and misconceptions. Educational Evaluation and Policy Analysis, 21, 97-110.

Finn, J.D. & Voekl, K.E. (1992). Class size: An overview of research. Occasional Paper No. 92-1. Buffalo, NY: State University of New York at Buffalo-Department of Counseling and Educational Psychology.

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

Glass, G. & Smith, M.L. (1978). Meta-analysis of the relationship of class size and student achievement. San Francisco, CA: Far West Laboratory.

Greenwald, R, Hedges, L.V., & Laine, R.D. (1996). The effect of school resources on student achievement. Review of Educational Research, 66, 361-396.

Grissmer, D. (1999). Introduction (Special Issue-Class size: Issues and new findings). Educational Evaluation and Policy Analysis, 21, 93-95.

Halbach, A., Ehrle, K., Zahorik, J., & Molnar, A. (2001). Class size reduction: From promise to practice. Educational Leadership, 29, 32-35.

Hanushek, E.A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects, Educational Evaluation and Policy Analysis, 21, 143-163.

Hedges, L.V. & Olkin, I. (1985). Statistical methods for meta-analysis. San Diego, CA: Academic Press.

Hunter, J.E. & Schmidt, F.L. ((1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park. CA: Sage.

Linn, R.L. (2001). Reporting school quality in standards-based accountability systems. (CRESST Policy Brief 3). Los Angeles: University of California, Center for Evaluation Standards, and Student Testing.

Lipsey, M.W. & Wilson, S. (1993). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.

Molnar, A. Smith, P., Zahorik, P., Palmer, A. Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. Educational Evaluation and Policy Analysis, 21, 165-177.

Robinson, G.E. (1990). Synthesis of research on the effects of class size. Educational Leadership, 13, 80-90.

Rosenthal, R. (1991). Meta-analytic procedures for social research. Newbury Park, CA: Sage.

Salazar, J. J. (1998). A longitudinal model for thirty years of bilingual research. Bilingual Research Journal, 22, 19-30.

Smith, M.L. & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32, 752-760.

Stasz, C. & Stecher, B.M. (2000). Teaching mathematics and language in reduced size and non-reduced size classrooms. Educational Evaluation and Policy Analysis, 22, 313-329.

Stecher, B.M., McCaffrey, D.F., Burroughs, D, Wiley, E., & Bohrnstedt, G.W. (2000). Achievement. In B.M. Stecher & G.W. Bohrnstedt (Eds.), Class size reduction in California: The 1998-99 evaluation findings, (pp. 91-120). Palo Alto: American Institutes of Research.

# APPENDIX

# Table A1
## 3rd Grade Reading
### Matched Gain Scores in Reading by Language Classification
### Cohort 3 minus Cohort 1

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 1** | | | | |
| NCE Gain | .16 | -1.94 | -.01 | -.18 |
| SD | 11.21 | 10.33 | 11.19 | 12.54 |
| N | 3,534 | 596 | 2,320 | 8,627 |
| **Cohort 3** | | | | |
| NCE Gain | 2.02 | -.28 | .77 | 1.42 |
| SD | 10.96 | 10.26 | 11.42 | 11.90 |
| N | 42,590 | 44,597 | 3,819 | 13,835 |
| **Effect Sizes** | | | | |
| Cohort 3 minus Cohort 1 | | | | |
| Difference | 1.86 | 1.66 | .78 | 1.60 |
| **Effect Size** | **.16** | **.15** | **.06** | **.12** |

# Table A2
# 3rd Grade Mathematics
## Matched Gain Scores in Mathematics by Language Classification
## Cohort 3 minus Cohort 1

### Descriptive Statistics

| | ELL | RFEP | IFEP | EO |
|---|---|---|---|---|
| **Cohort 1** | | | | |
| NCE Gain | 3.16 | 2.70 | 2.32 | 1.02 |
| SD | 13.76 | 14.51 | 14.27 | 14.13 |
| N | 3,971 | 614 | 2,495 | 9,533 |
| **Cohort 3** | | | | |
| NCE Gain | 3.02 | 3.36 | 3.28 | 2.24 |
| SD | 14.91 | 14.84 | 14.33 | 14.69 |
| N | 28,441 | 2,025 | 3,991 | 14,581 |

### Effect Sizes
Cohort 3 minus Cohort 1

| | | | | |
|---|---|---|---|---|
| Difference | -.14 | .66 | .96 | 1.22 |
| **Effect Size** | **-.01** | **.05** | **.07** | **.08** |

## Table A3
## 3rd Grade Language
### Matched Gain Scores in Language by Language Classification
### Cohort 3 minus Cohort 1

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 1** | | | | |
| NCE Gain | 3.00 | 2.65 | 1.84 | .81 |
| SD | 14.21 | 15.01 | 14.03 | 14.31 |
| N | 3,874 | 611 | 2,481 | 9,272 |
| **Cohort 3** | | | | |
| NCE Gain | 6.70 | 3.50 | 2.54 | 2.10 |
| SD | 15.13 | 15.25 | 15.28 | 15.53 |
| N | 27,361 | 1,994 | 3,903 | 14,240 |
| **Effect Sizes** | | | | |
| Cohort 3 minus Cohort 1 | | | | |
| Difference | 3.70 | .85 | .70 | 1.29 |
| **Effect Size** | **.25** | **.06** | **.05** | **.09** |

## Table A4
## 4th Grade Reading
### Matched Gain Scores in Reading by Language Classification
### Cohort 2 minus Cohort 0

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 0** | | | | |
| NCE Gain | 1.71 | .93 | 1.55 | 1.63 |
| SD | 10.90 | 10.51 | 10.64 | 11.07 |
| N | 3,778 | 1,671 | 2,513 | 8,650 |
| **Cohort 2** | | | | |
| NCE Gain | 3.81 | 3.90 | 3,69 | 2.27 |
| SD | 10.53 | 9.53 | 10.19 | 10.59 |
| N | 23,652 | 5,076 | 3,855 | 14,173 |
| **Effect Sizes** | | | | |
| Cohort 2 minus Cohort 0 | | | | |
| Difference | 2.10 | 2.97 | 2.14 | .64 |
| **Effect Size** | **.20** | **.30** | **.21** | **.06** |

## Table A5
## 4th Grade Mathematics
### Matched Gain Scores in Mathematics by Language Classification
### Cohort 2 minus Cohort 0

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 0** | | | | |
| NCE Gain | -.61 | -.70 | -.69 | -1.50 |
| SD | 12.26 | 12.50 | 11.93 | 11.07 |
| N | 4,038 | 1,704 | 2,578 | 9,007 |
| **Cohort 2** | | | | |
| NCE Gain | -.43 | -.37 | .19 | -.42 |
| SD | 12.06 | 12.62 | 12.31 | 12.58 |
| N | 24,920 | 5,140 | 3,931 | 14,501 |
| **Effect Sizes** | | | | |
| Cohort 2 minus Cohort 0 | | | | |
| Difference | .18 | .33 | .88 | 1.08 |
| **Effect Size** | **.02** | **.03** | **.07** | **.09** |

## Table A6
## 4th Grade Language
### Matched Gain Scores in Language by Language Classification
### Cohort 2 minus Cohort 0

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 0** | | | | |
| NCE Gain | 4.58 | 4.02 | 3.23 | 1.87 |
| SD | 13.12 | 12.48 | 12.57 | 12.60 |
| N | 3,939 | 1,679 | 2,523 | 8,734 |
| **Cohort 2** | | | | |
| NCE Gain | 5.02 | 3.63 | 2.22 | 1.45 |
| SD | 12.63 | 12.30 | 12.63 | 12.76 |
| N | 24,920 | 5,140 | 3,931 | 14,501 |
| **Effect Sizes** | | | | |
| Cohort 2 minus Cohort 0 | | | | |
| Difference | .44 | -.39 | 1.01 | .42 |
| **Effect Size** | **.03** | **-.03** | **.08** | **.03** |

## Table A7
## 5th Grade Reading
### Matched Gain Scores in Reading by Language Classification
### Cohort 1 minus Cohort 0

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 0** | | | | |
| NCE Gain | 5.15 | 4.85 | 4.19 | 4.14 |
| SD | 9.99 | 9.51 | 10.19 | 10.49 |
| N | 4,495 | 3,394 | 2.244 | 7,499 |
| **Cohort 1** | | | | |
| NCE Gain | 3.23 | 2.52 | 2.19 | 1.55 |
| SD | 10.06 | 9.36 | 10.05 | 10.49 |
| N | 17,570 | 8,453 | 3,340 | 13,165 |
| **Effect Sizes** | | | | |
| Cohort 1 minus Cohort 0 | | | | |
| Difference | -1.92 | -2.33 | -2.00 | -2.59 |
| **Effect Size** | **-.19** | **-.25** | **-.19** | **-.25** |

## Table A8
## 5th Grade Mathematics
### Matched Gain Scores in Mathematics by Language Classification
### Cohort 1 minus Cohort 0

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **ELL** | **RFEP** | **IFEP** | **EO** |
| **Cohort 0** | | | | |
| NCE Gain | 1.31 | .95 | .99 | .73 |
| SD | 10.98 | 11.48 | 11.56 | 11.67 |
| N | 4,968 | 3,521 | 2,350 | 7,916 |
| **Cohort 1** | | | | |
| NCE Gain | 2.62 | 2.50 | 2.42 | 2.32 |
| SD | 10.71 | 11.49 | 11.75 | 11.79 |
| N | 18,905 | 8,608 | 3,436 | 13,583 |
| **Effect Sizes** | | | | |
| Cohort 1 minus Cohort 0 | | | | |
| Difference | 1.31 | 1.55 | 1.43 | 1.59 |
| **Effect Size** | **.12** | **.13** | **.12** | **.12** |

## Table A9
## 5th Grade Language
### Matched Gain Scores in Language by Language Classification
### Cohort 1 minus Cohort 0

**Descriptive Statistics**

| | ELL | RFEP | IFEP | EO |
|---|---|---|---|---|
| **Cohort 0** | | | | |
| NCE Gain | 3.10 | 3.34 | 1.99 | 1.36 |
| SD | 12.62 | 12.32 | 12.31 | 12.64 |
| N | 4,973 | 3,477 | 2,377 | 7,819 |
| **Cohort 1** | | | | |
| NCE Gain | 2.06 | 3.59 | 2.55 | 1.13 |
| SD | 12.52 | 12.28 | 12.62 | 12.59 |
| N | 18,530 | 8,535 | 3,389 | 13,401 |

**Effect Sizes**

Cohort 1 minus Cohort 0

| | | | | |
|---|---|---|---|---|
| Difference | -1.04 | .25 | .56 | -.23 |
| **Effect Size** | **-.08** | **.02** | **.04** | **-.02** |