The Effect of School Renaissance on TAAS scores in the McKinney ISD

John A. Nunnery, Steven M. Ross, Elizabeth Goldfeder
Center for Research in Educational Policy
The University of Memphis

June, 2003

EXECUTIVE SUMMARY

Overview

The present research is a third-party study of the effects of the School Renaissance (SR) comprehensive school reform (CSR) model on student achievement in 11 elementary and middle schools in Texas. The context for the study was the McKinney Independent School District (MISD). Implementation of Reading Renaissance in MISD elementary schools began in the spring, 2000 and was mandated in all district schools in the fall, 2000. The mathematics program ("Math Renaissance") was subsequently mandated in all schools (grades 3-8) in January 2002. Given the cooperation of MISD, the Texas Education Agency (TEA), and superintendents from 10 other Texas school districts, a rigorous matched treatment-control pretest-posttest design was possible to employ for the present research. The primary research questions were:

1. How do SR schools compare to matched Comparison schools (C) in student achievement on the Texas Assessment of Academic Skills?
2. Do SR effects relative to C schools vary for (a) reading and mathematics; and (b) elementary and middle schools?
3. Do SR effects vary for different types of students:
    a. free or reduced-price lunch subsample,
    b. Limited English Proficient (LEP), and
    c. Low achievers.
4. Do SR effects relative to C schools vary as a function of the degree of SR model implementation?

Method

Participants

Participating in the study were 3rd, 4th, and 5th grade students in 9 SR and 9 C elementary schools, and 6th, 7th, and 8th grade students in 2 SR and 2 C middle/junior high schools from the 1997-98 through the 2001-02 school years.

For the Reading program, the baseline (pre-implementation) period was the 1998-99 school year. The baseline for the Mathematics program was the 1999-00 school year. The TEA's Academic Excellence Indicator System's (AEIS) Comparable Improvement data was used to select the most similar C school to each SR school based on level of SR implementation (i.e., the degree to which any SR components, such as Accelerated Reader were *not* being used by a possible C school), the base year accountability rating (low performing, acceptable, recognized, exemplary), and base year percent of economically-disadvantaged students.

Analyses

Overall analyses of program effects. The primary measures used in the study were the Texas Learning Index (TLI) reading and mathematics scores obtained through administration of the Texas Assessment of Academic Skills (TAAS). TLI scores were also converted to a dichotomous measure indicating whether students performed at or above grade level. Two methods were used to provide an overall view of the performance of SR versus matched C schools. First, median effect size estimates were computed for each grade level and year to provide cohort performance profiles. Secondly, the percentage of students performing at or above grade level in SR and C schools was computed for each year.

An assumption of the multivariate analyses used to determine program effects is that the dependent measures (reading and mathematics achievement scores) have a multivariate normal distribution. An examination of the distributions of the reading and math TLI scores indicated

that TLI scores were strongly negatively skewed (asymmetrical) and highly leptokurtic. Data transformations were conducted to induce the distributions to normality as well as stabilize variances across schools and treatment groups. Repeated-measures analyses were conducted on the transformed variables to test program effects, using 2001 and 2002 transformed scores as dependent measures; 2000 transformed scores and free or reduced-price lunch status as covariates; and program type (SR vs. C) and matched pair as between-subjects factors. Because 1998-99 was the baseline year for the reading program, similar repeated-measures analyses were conducted for the 2001 5^{th} and 8^{th} grade cohorts using transformed reading scores as the within-subjects factor and 1998-99 scores as the achievement covariate. For math, the baseline year was 1999-00, with 2001 and 2002 representing the second and third years of implementation, respectively.

Analyses of subgroup performances. Descriptive analyses were performed on the 2001 and 2002 fifth grade cohorts to provide an overview of how three student subgroups performed in SR versus C schools: students eligible for free or reduced-price lunch; limited English proficient (LEP) students; and students who performed at- or below grade level in third grade (i.e., received a score of 70 or below on the TAAS subtest).

Analyses of implementation effects. The research question regarding implementation impacts was addressed through repeated-measures analysis of transformed TLI reading and mathematics scores for the 2002 5^{th} grade cohort. The latter was the first cohort of students for which Reading Renaissance and Math Renaissance were made available from third grade onward. Three school-level indices were used to classify schools on the basis of intensity of implementation of the program: the overall reading program implementation index, the overall mathematics program implementation index, and the ratio of program-certified teachers to total

student enrollment in grades 3, 4, and 5. Hierarchical cluster analysis using Ward's minimum variance method was used to categorize schools into maximally homogeneous groups based on standardized implementation indicator measures (see Romesburg, 1990, p. 129-135). Results yielded a 2-cluster solution, with 4 schools clustered tightly in the "high implementation" group, and 5 schools clustered in the "low implementation" group.

Results

Major findings in reading and mathematics are summarized below from the (a) descriptive analyses of yearly cohort results, and (b) inferential repeated-measures comparisons between SR and C schools for the longitudinal cohorts. The former analyses examined all available student scores for each grade and subject in the given year, regardless of whether the student attended the same school in any prior years. Because students' prior achievement and socioeconomic status were not taken into account, these achievement profiles need to be viewed cautiously in judging program effects. The second set of analyses (inferential tests) included only students in fifth-grade and eighth-grade cohorts who had three successive years of achievement test data. Accordingly, with prior achievement, mobility, and socioeconomic status controlled a more sensitive picture of SR vs. C outcomes can be obtained.

Reading: Grades 3-5

- For separate cohorts, median effect size estimates indicating SR relative to C outcomes were generally small and stable from 1999 to 2002.
- The percentage of students reading at or above grade level improved substantially at all schools and was nearly equal at SR and C schools.
- Repeated-measures analyses of both 2001 and 2002 fifth-grade cohorts indicated statistically significant differences favoring SR schools after controlling for 3rd grade achievement and socioeconomic status.

- Participation in the SR program accounted for about the same amount of variance in student outcomes as socioeconomic status after controlling for prior achievement.
- Effect size estimates were $d = +0.22$ and $d = +0.17$ for 2001 and 2002 fifth grade cohorts, respectively, indicating that SR students performed about one-fifth of a standard deviation higher than comparison students in matched schools.

Reading: Grades 6-8

- Median effect size estimates indicating SR relative to C outcomes consistently improved from 1999 to 2002.
- The percentage of students reading at or above grade level improved at a greater rate in SR than in C schools.
- Repeated-measures analyses revealed no significant differences between SR and C schools, although results directionally favored SR schools.

Math: Grades 3-5

- Median effect size estimates indicating SR relative to C outcomes for separate cohorts remained stable and near zero from 2000 to 2002.
- Similarly, the percentage of students performing at or above grade level rose at nearly equal rates for SR and C schools.
- Repeated-measures analysis showed a significant program effect favoring SR schools for the 2002 fifth-grade cohort. Program type explained substantially more variance in math outcomes than did socioeconomic status after controlling for prior math achievement.
- Repeated-measures analyses indicated that SR students performed about one-fifth of a standard deviation higher than comparison students ($d = +0.20$).

Math: Grades 6-8

- Median effect size estimates indicating SR relative to C outcomes improved at all grade levels from 2000 to 2002.
- The percentage of students performing at or above grade level rose at a greater rate in SR than in C schools.
- Repeated-measures analysis revealed a significant program effect favoring SR schools for the 2002 eighth-grade cohort after controlling for prior math achievement and socioeconomic status.
- Repeated-measures analyses indicated that SR students performed about one-fifth of a standard deviation higher than comparison students (d = +0.17).

Subgroup Performances

- Descriptive analyses showed trends indicating that SR students who were eligible for free or reduced-price lunch demonstrated larger gains than C students in both Reading and Mathematics in the 2001 and 2002 5^{th} grade cohorts.
- Similar trends, showing larger SR than C gains, were indicated for the LEP subgroup.
- Trends for the subgroup that performed at- or below grade level in third grade were comparable for SR and C students in the 2001 cohort. C students in the 2002 cohort, however, demonstrated a slightly higher gain than SR students in both subjects.

Implementation Effects

- Repeated-measures analyses indicated that students in high-implementation SR schools achieved at significantly higher levels in both reading and mathematics than students in similar comparison schools, after controlling for students' prior achievement and socieoeconomic status.

- Results in low-implementation SR schools were directionally higher than comparison schools in both reading and mathematics, but the differences were not large enough to attain statistical significance.

Conclusions

The major conclusions from the study are highlighted below. In view of the overall positive results obtained in this study, it is recommended that future research examine school environment variables such as school climate and teacher buy-in to better explain varied program effectiveness across schools and identify the optimum conditions for schools to realize benefits from SR adoption. Conclusions from this study are:

- In the elementary grades, SR schools showed a clear trend of improved performance. Effects of the program were small, but statistically significant, for both reading (2001 and 2002 5th grade longitudinal cohorts) and mathematics (2002 longitudinal cohort). However, program effects were generally comparable to or greater than the effects of socioeconomic status after controlling for prior achievement.
- In the middle school grades, SR schools had noticeably improved performance relative to C schools when viewing separate cohorts in both reading and mathematics. However, performance was significantly better only in mathematics in the 2002 8th grade longitudinal cohort.
- Descriptive data showed trends for higher SR than C gains in both reading and mathematics for both 2001 and 2002 cohorts in the free or reduced-price lunch subgroup and in the LEP subgroup. However, the reverse was true for the 2002 cohort in the low-achieving subgroup.
- Implementation intensity was positively related to SR program effects on both reading and mathematics scores.

In a recent meta-analytic study of 29 models, Borman et al. (2002) found an overall effect size of from +0.10 to +0.14, with the range for most successful category, labeled by the authors as "proven models," being +0.17 to +0.21. Only 3 out of the 29 models achieved this high status (Direct Instruction, School Development Program, and Success For All). Turning to the present research, it is noteworthy that four out of six significant inferential comparisons were associated with effect sizes for SR ranging from +0.17 to +0.22. Including the two nonsignificant effects (the two eight-grade cohorts in reading), both of which directionally favored SR, would still yield a relatively high overall median ES of +0.17 for the entire study. Compared to the above three "proven models," SR has been made available to schools for only a short time and has had many fewer years to be researched.

Taken as a whole, the present results are clearly suggestive of its benefits for student achievement, and if consistently replicated in future studies would strongly imply proven effectiveness as CSR model. Not surprisingly, the present results also imply that the program impacts are significantly more positive when implementation intensity is high. It is noteworthy that the program schools participating in this study were relatively high-performing, with some having 80-90% of their students performing at or above grade level. Thus, the achievement effects occurred in a context where potential gains might have been limited by ceiling effects. Further study is needed to determine whether stronger effects might occur with lower-performing populations.

The Effect of School Renaissance on TAAS scores in the McKinney ISD

Overview

The present research is a third-party study of the effects of the School Renaissance (SR) comprehensive school reform (CSR) model on student achievement in 11 elementary and middle schools in Texas. The context for the study was the McKinney Independent School District (MISD). In the fall, 1999, MISD initiated training for the SR reading program ("Reading Renaissance"). Implementation of Reading Renaissance in elementary schools began in the spring, 2000 and was mandated in all MISD schools in the fall, 2000. The mathematics program ("Math Renaissance") was subsequently mandated in all schools (Grades 3-8) in January, 2002.

School Renaissance Program

The development of SR has evolved over the past two decades from several programs created by Terry and Judy Paul, the founders of Renaissance Learning, Inc. The earliest and most widely use of these programs is Accelerated Reader, designed to increase students' motivation and ability to read. The broader philosophy of Reading Renaissance was first introduced to educators through professional development seminars in 1992. Accelerated Math software and Math Renaissance were introduced in 1998, and Accelerated Writer and Writing Renaissance were initiated in 2002. SR incorporates all of these programs, along with other critical CSR elements, such as providing extensive professional development, conducting formative evaluation, ensuring faculty buy-in, aligning curricula with state standards, and specifying and monitoring appropriate resource allocation. Curriculum alignment with state standards, using Standards Master (assessment and web-based reporting software), receives strong emphasis in the model. Currently, there are approximately 60,000 schools nationwide at various stages of implementing SR or one of its major components, Reading Renaissance and Math Renaissance. Close to 250 schools nationwide are implementing the full model.

*Study Rationale and Research Question*s

With the passage in 2002 of the federal legislation, No Child Left Behind (U.S. Congress, 2001), increased focus has developed for identifying "proven" practices that demonstrably raise student achievement. Consistent with this movement is the current national demand for increased rigor in educational research. Standards for high-quality scientific research have recently been proposed from national organizations, such as the National Research Council (Feuer, Towne, & Shavelson, 2002; Shavelson & Towne, 2002), Educational Quality Institute (www.eqireports.org), and What Works Clearinghouse (wwcinfo@w-w-c.org). Although randomized experiments represent the highest standard of validity (U.S. Department of Education, 2002), rigorous matched-control group designs are likely to be far more practical to implement across broad and diverse application sites (see Borman, 2002; Slavin, 2002). In contrast to single-group pre-post designs, which comprise a high percentage of existing CSR evaluations (see Borman et al., 2002), matched-control group studies allow student achievement gains and key outcomes to be compared between schools that use the program of interest and similar schools that do not.

MISD's policy decision that all district schools implement the SR model precluded conducting a randomized field study to assess model effects. However, given the cooperation of MISD, the Texas Education Agency (TEA), and superintendents from 10 other Texas school districts, a rigorous matched treatment-control, pretest-posttest design was possible to conduct. The methodology and instrumentation are described in the immediately following sections below. The primary research questions were:

1. How do SR schools compare to matched Comparison schools (C) in student achievement on the Texas Assessment of Academic Skills?

2. Do SR effects relative to C schools vary for (a) reading and mathematics; and (b) elementary and middle schools?
3. Do SR effects vary for different types of students:
    a. free or reduced-price lunch subsample,
    b. Limited English Proficient (LEP), and
    c. Low achievers.
4. Do SR effects relative to C schools vary as a function of the degree of SR model implementation?

Method

Participants

Participating in the study were 3rd, 4th, and 5th grade students in 9 School Renaissance (SR) and 9 Comparison (C) elementary schools, and 6th, 7th, and 8th grade students in 2 SR and 2 C middle/junior high schools from the 1997-98 through the 2001-02 school years. All SR schools were located in MISD. The specific schools examined included: (1) Finch Elementary, (2) Webb Elementary, (3) Burks Elementary, (4) Caldwell Elementary, (5) Valley Creek Elementary, (6) Glen Oaks Elementary, (7) Slaughter Elementary (8) Reuben Johnson Elementary, (9) C.T. Eddins Elementary, (10) Faubion Middle, and (11) Dowell Middle.

The Texas Education Agency's (TEA) Academic Excellence Indicator System's (AEIS) Comparable Improvement data was used to select the Comparison schools. Each year, for each school in Texas, the AEIS Comparable Improvement report identifies 40 demographically similar schools based on the percent of:

- African American students enrolled,
- Hispanic students enrolled,
- White students enrolled,

- economically disadvantaged students enrolled,
- limited English proficient (LEP) students enrolled, and
- student mobility as determined from cumulative attendance.

For the Reading program, the baseline (pre-implementation) period was the 1998-99 school year. The baseline for the Mathematics program was the 1999-00 school year. From the group of 40 base-year comparison schools, the most similar school to each McKinney school was selected according to level of SR implementation (i.e., the degree to which any SR components, such as Accelerated Reader were *not* being used by the prospective C school), the base-year accountability rating (low performing, acceptable, recognized, exemplary), and base-year percent of economically disadvantaged students.

Eleven schools, representing nine districts, were selected as matched control schools. Eight of the original nine districts granted permission to release the student-level TAAS data. (Given that four schools were represented in two of these districts, ten schools had been approved on the first round.) One district superintendent declined participation. For this district, a replacement selection was made and the approval procedure repeated. The replacement district superintendent gave approval. The day prior to this, however, a representative from one of the school districts (which had already granted approval) communicated that they were not a TAAS school, that Grade 2 was their highest level, and that they needed to withdraw approval. Several days later, the TEA communicated that another school that had given permission was also not a TAAS school. Replacement selections were made for these two districts, and the approval procedure was repeated. Both of the schools chosen on the third round were approved, although one district gave conditional consent, requiring that their school's data be "scrubbed" for categories in which there were five students or less, in order to protect students' confidentiality.

Table 1 provides a breakdown of the percentage of students receiving free or reduced-price lunch, the percentage of students with limited English proficiency, and the percentage of students within various ethnic categories for each matched school pair during the 1999-2000 school year. Demographic profiles indicate that SR schools were quite similar to their matched pair with respect to the percentage of students receiving free or reduced-price lunch. The largest discrepancies on this variable were in Pair 3 (33% SR, 25% C), Pair 9 (48% SR, 67% C), and Pair 11 (23% SR 32% C; see Table 1). In Pair 2 and Pair 3, the SR school had substantially higher percentages of students with limited English proficiency (LEP) than did the C school; 33% versus 5%, and 35% versus 7%, respectively. Otherwise, the percentage of LEP students was equal or nearly equal within matched pairs. The student sample sizes for the study are indicated for Grades 3 to 6 in Tables A.1 to A.4 in Appendix A.

Table 1

Demographic Characteristics of Participating Schools by Matched Pair and Program Type (School Year 1999-2000)

Pair/ Program	Level	Free or reduced-price lunch	LEP[1]	Ethnicity				
				African-American	Asian	Hispanic	Native American	White
1								
School Renaissance	Elementary	32%	1%	20.6%	0.5%	13.8%	0.5%	64.6%
Comparison		31%	1%	12.2%	4.1%	10.6%	2.0%	71.1%
2								
School Renaissance	Elementary	59%	33%	12.7%	0.0%	50.4%	0.4%	36.6%
Comparison		58%	5%	0.0%	0.0%	51.8%	0.0%	48.25
3								
School Renaissance	Elementary	33%	35%	2.1%	1.5%	39.2%	0.6%	45.8%
Comparison		25%	7%	20.4%	2.8%	11.3%	0.0%	65.5%
4								
School Renaissance	Elementary	55%	17%	18.5%	1.3%	34.4%	0.0%	45.8%
Comparison		55%	15%	20.7%	9.5%	43.6%	0.4%	25.8%
5								
School Renaissance	Elementary	0%	0%	1.0%	0.8%	2.4%	0.5%	95.3%
Comparison		1%	0%	0.7%	2.9%	3.7%	0.0%	92.6%

Table 1 (continued)

Demographic Characteristics of Participating Schools by Matched Pair and Program Type (School Year 1999-2000)

Pair/ Program	Level	Free or reduced-price lunch	LEP[1]	Ethnicity				
				African-American	Asian	Hispanic	Native American	White
6								
School Renaissance	Elementary	3%	2%	4.5%	4.0%	7.4%	0.0%	84.1%
Comparison		10%	1%	0.7%	2.9%	3.7%	0.0%	92.6%
7								
School Renaissance	Elementary	18%	2%	14.1%	1.8%	9.6%	0.3%	74.2%
Comparison		19%	7%	3.9%	5.2%	16.7%	0.0%	74.1%
8								
School Renaissance	Elementary	2%	1%	2.7%	2.1%	2.4%	0.3%	92.6%
Comparison		1%	0%	2.0%	3.9%	3.3%	0.0%	90.8%
9								
School Renaissance	Elementary	48%	21%	12.2%	0.5%	39.4%	0.5%	47.3%
Comparison		67%	15%	12.1%	0.0%	64.5%	0.0%	23.4%

Table 1 (continued)

Demographic Characteristics of Participating Schools by Matched Pair and Program Type (School Year 1999-2000)

Pair/ Program	Level	Free or reduced-price lunch	LEP[1]	Ethnicity				
				African-American	Asian	Hispanic	Native American	White
10								
School Renaissance	Middle/Jr. High	20%	4%	8.2%	2.1%	19.5%	0.3%	70.0%
Comparison		15%	5%	6.4%	4.6%	16.8%	0.1%	72.1%
11								
School Renaissance	Middle/Jr. High	23%	4%	10.1%	1.7%	18.6%	0.4%	69.3%
Comparison		32%	3%	4.0%	0.7%	26.6%	0.4%	68.3%

[1]Limited English Proficiency.

Measures

The primary measures used in the study were the Texas Learning Index (TLI) reading and mathematics scores obtained through administration of the Texas Assessment of Academic Skills (TAAS). The TLI has a common interpretation across grades: a score of 70 or above indicates the student performed at or above grade level expectations. A student receiving the same score at consecutive grade levels made one year of academic progress. For example, a student scoring 65 in reading in 3rd grade and a 65 in reading in 4th grade made one year's academic progress, although the score indicates the student is performing below grade level. TLI scores were also converted to a dichotomous measure indicating whether students performed at or above grade level.

Analyses

Description of school-level performance. Mean reading and mathematics scores on the TLI were computed for all participating schools. Reading means were computed for the school years 1997-98 through 2001-02, whereas mathematics means were computed for the years 1998-99 through 2001-02. An effect size estimate (*ES*) was computed for each matched pair by subtracting the C school mean from the SR school mean, then dividing this difference by the standard deviation for the entire grade-level sample. Effect size estimates thus express mean differences in terms of standard deviation units, and therefore can be directly compared across years. Descriptive profiles were generated for the 1997-98 to 2001-02 school years for reading, and 1998-99 to 2001-02 school years for mathematics.

Description of program-level performance. In addition to the school-by-school descriptive profiles, two methods were used to provide an overall view of the performance of SR versus matched C schools. First, median effect size estimates were computed for the overall SR

vs. C comparisons for each grade level and year to provide cohort performance profiles. Secondly, the percentage of students performing at or above grade level in SR and C schools was computed for each year. The descriptive profiles illustrate the performance of all tested students in each school in a given year (i.e., students may not have attended the school in previous years).

Data transformations for inferential analyses. An assumption of the multivariate analyses used to determine program effects is that the dependent measures (reading and mathematics achievement scores) have a multivariate normal distribution. An examination of the distributions of the reading and math TLI scores indicated that TLI scores were strongly negatively skewed (asymmetrical) and highly leptokurtic (peaked; see Table 2). Data transformations were conducted to induce the distributions to normality, as well as stabilize variances across schools and treatment groups (Berenson, Levine, & Goldstein, 1983; Tabachnick & Fidell, 2001). First, each variable was reflected by subtracting each score from the maximum score plus one. This procedure reverses the skewness, making the transformed variables positively skewed with a minimum score of 1. Next, the log base 10 was computed for the reflected variables, which induces the variable to normality and stabilizes variances across groups. Finally, the transformed scores were subtracted from the maximum transformed score to aid in interpretation (i.e., so that higher scores on the transformed variable indicate higher levels of achievement). Table 2 shows distributional statistics (skewness and kurtosis) for the original and transformed variables.

Table 2

Distributional Impact of Variance Stabilizing and Normality Inducing Data Transformations

	Before Transformation		After Transformation[1]	
Variable	Skewness	Kurtosis	Skewness	Kurtosis
1997-98 Reading TLI	-1.91	2.67	-0.66	-0.11
1998-99 Reading TLI	-1.87	2.25	-0.05	-0.10
1999-00 Reading TLI	-2.01	2.82	-0.16	0.07
2000-01 Reading TLI	-2.74	7.09	-0.25	0.50
2001-02 Reading TLI	-2.86	7.65	-0.40	0.51
1997-981998 Math TLI	-2.11	3.58	-0.19	-0.22
1998-99 Math TLI	-2.09	3.10	-0.33	-0.27
1999-00 Math TLI	-2.24	3.68	-0.49	-0.09
2000-01 Math TLI	-3.29	10.46	-0.47	0.57
2001-02 Math TLI	-3.37	10.77	-0.62	0.64

[1]Log base 10 of reflected variable. Reflected variable equals the maximum score +1 minus the original score. The transformed variable scores were then subtracted from the maximum transformed variable score to aid in interpretation (i.e., so the transformed variable is positively correlated with the original variable).

As indicated in Table 2, the transformation was quite successful in inducing the variables to normality. For example, the original 2002 Reading TLI scores were strongly negatively skewed (*Sk* = -2.86) and highly leptokurtic (*K* = 7.65), whereas the transformed variable was almost normally distributed (*Sk* = -0.40, *K* = 0.51). Similar improvements were effected for all achievement variables. Likewise, as indicated in Table 3, the transformation stabilized variances across groups. Six of ten of the original 2002 variables violated the homogeneity of variance assumption, whereas only two of the transformed variables violated this assumption (2002 4^{th} grade reading and 2002 4^{th} grade math; see Table 3). Even for the 4th grade scores, heterogeneity of variance was substantially reduced through data transformation.

Table 3

Levene's Test for Equality of Variances Across Schools for Original and Transformed Dependent Variables: 2001-02 TLI Scores by Grade Level

TLI Subtest/ Grade	Original Variable				Transformed Variable			
	F	df_n	df_d	*p*	*F*	df_n	df_d	*p*
Reading								
4th	4.54	17	1,279	.000[a]	2.02	17	1,279	.008[a]
5th	7.17	17	1,281	.000[a]	1.35	17	1,281	.151
6th	0.59	3	1,341	.623	0.83	3	1,341	.476
7th	0.71	3	1,209	.549	0.02	3	1,209	.906
8th	0.62	3	1,391	.602	0.13	3	1,391	.943
Math								
4th	4.85	17	1,279	.000[a]	3.29	17	1,279	.000[a]
5th	12.28	17	1,281	.000[a]	1.82	17	1,281	.022
6th	4.94	3	1,341	.002[a]	3.24	3	1,341	.021
7th	5.60	3	1,209	.001[a]	0.50	3	1,209	.684
8th	1.01	3	1,391	.388	1.57	3	1,391	.194

[a]Significantly heterogeneous variances at $p < .01$.

Repeated-measures analyses. Repeated-measures analyses were conducted on the transformed variables to test program effects, using 2001 and 2002 transformed scores as dependent measures; 2000 transformed scores and free or reduced-price lunch status as covariates; and program type (SR vs. C) and matched pair as between-subjects factors. Repeated-measures analyses were only possible for 8th grade and 5th grade cohorts, because students would have to remain in the same school for three consecutive years to have matched data available across years. Thus, covariate achievement data would be 2000 3rd grade scores for the 5th grade analyses, and 2000 6th grade scores for the 8th grade analyses. Because 1999 was the baseline year for the reading program, similar repeated-measures analyses were conducted for the 2001 5th and 8th grade cohorts using transformed reading scores as the within-subjects factor and 1999 scores as the achievement covariate. For math, the baseline year was 1999-00, with 2001 and 2002 representing the second and third years of implementation, respectively.

Computation of effect size estimates. Cohen's *d* was computed as the measure of program effects in cases where the effect size was significantly different from zero. Cohen's *d* is the difference between the treatment and comparison group means expressed in terms of standard deviation units. Typically, Cohen's *d* is derived by subtracting the comparison group mean from the treatment group mean, then dividing by the pooled estimated of the population standard deviation. However, because inferential analyses were conducted on transformed variables, a different approach to estimating d was used. First η^2 (the proportion of total variance attributable to treatment differences) was computed from repeated-measures analyses of reading and mathematics achievement, then converted to a simple correlation (*r)* by taking the square root. This conversion is meaningful in the present instance because there were only two

treatment levels. The effect size estimate (Cohen's *d*) can then be obtained by the following formula (Rosenthal, 1991, p. 20).

$$d = \frac{2r}{\text{sqrt}\,(1 - r^2)}$$

Supplementary Subgroup Analyses

Descriptive analyses were performed on the 2001 and 2002 fifth grade cohorts to provide an overview of how three student subgroups performed in SR versus comparison (C) schools: students eligible for free or reduced-price lunch; limited English proficient (LEP) students; and students who performed at- or below grade level in third grade (i.e., received a score of 70 or below on the TAAS subtest). Because these student subgroups were relatively small and unevenly distributed across matched pairs of schools, it was not possible to incorporate the random school variable into the analyses. Therefore, inferential tests of program effects were not warranted due to the inability to utilize the matched-schools comparison design.

Supplementary Implementation Analyses

To address the research question regarding the influences of implementation quality on SR program impacts, the SR schools were classified on the basis of quantitative indicators as having either "high intensity" or "low intensity" implementation in reading and mathematics. Repeated-measures analysis of transformed TLI reading and mathematics scores for the high- and low-intensity groups compared to the C group was then conducted for the 2002 5th grade cohort. The 2002 5th grade cohort was the first cohort of students for which Reading Renaissance and Math Renaissance were made available from third grade onward. Because only two middle schools participated in the study, it was not possible to analyze implementation effects for the 8th grade cohort (i.e., due to confounding of school and implementation level).

Scores from 4th and 5th grade were treated as repeated measures, while student SES (eligibility for free or reduced-price lunch) and prior achievement (3rd grade TLI scores) were treated as covariates. Implementation level (high-intensity, low-intensity, Comparison) and matched pair (a random variable) were the independent variables. The analyses were similar to those employed in the main report, where school-matching and data transformation procedures are described in detail.

Implementation categorization. Three school-level indices were used to classify schools on the basis of intensity of implementation of the program: the overall reading program implementation index, the overall mathematics program implementation index, and the ratio of program-certified teachers to total student enrollment in grades 3, 4, and 5. The overall reading implementation index is computed by the SR developers by multiplying (a) the proportion of earned points on the AR reading quizzes to the expected points for the median reading level at that grade by (b) the percentage of students in the grade who average above 85% correct on the quizzes. On AR reports, the developers often present the points ratio as the proportion of 60 minutes during which students are reading based on the assumption that students who read for 60 minutes per day would earn the expected points. The overall reading implementation index, in turn, is informally defined for the entire class or school as the average time spent reading multiplied by the quality of the reading being done. For mathematics, the index is calculated by multiplying the average proportion of the major library completed by the percent of student who average 85% on Accelerated Math tests.

Hierarchical cluster analysis using Ward's minimum variance method was used to categorize schools into maximally homogeneous groups based on standardized implementation indicator measures (see Romesburg, 1990, p. 129-135). With this method, one computes the

average difference or distance between each case and all the others based on the values of the implementation variables, then sequentially combines cases into "clusters" that are maximally homogeneous on all variables. At each step, the distance between combined clusters is computed and reported in an agglomeration schedule. A large value of the agglomeration schedule index indicates that highly dissimilar clusters were combined at that step.

Two methods are generally used to determine the appropriate number of clusters: determining where a large "jump" occurs in the agglomeration schedule index, and visual inspection of a dendrogram. The greatest increase in the agglomeration schedule index occurred between the 2- and 3-cluster solutions (stages 7 and 8), indicating that a 2-cluster solution was appropriate (see Figure 1). The dendrogram shown in Figure 2 indicates the relative distance at which cases were combined into clusters, and clusters were combined into larger clusters. Visual inspection of the cluster solution dendogram confirmed that a 2-cluster solution was appropriate (see Figure 2), with 4 schools clustered tightly in the "high implementation" group, and 5 schools clustered in the "low implementation" group.

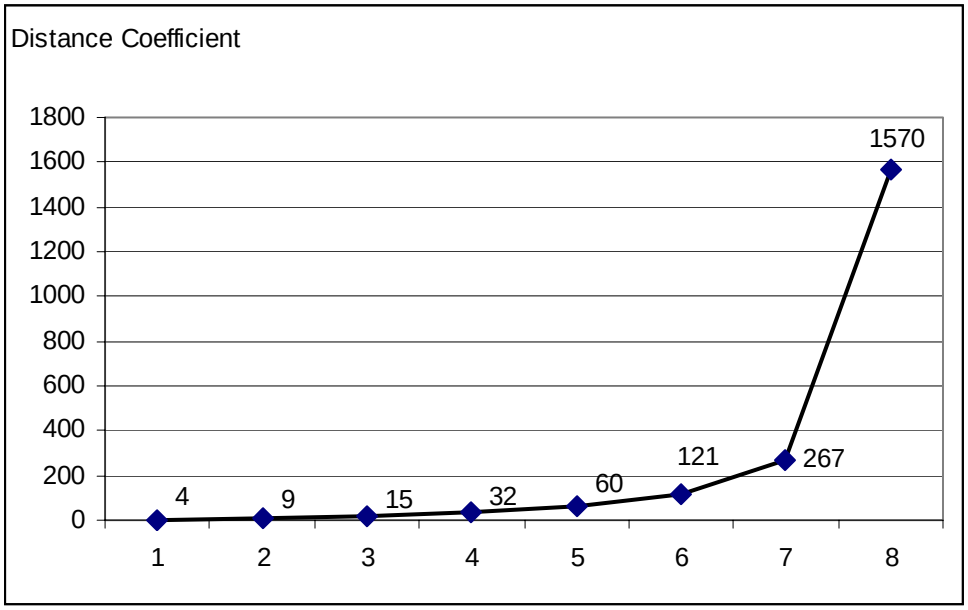

Figure 1. Distance coefficients by stage for implementation clusters.

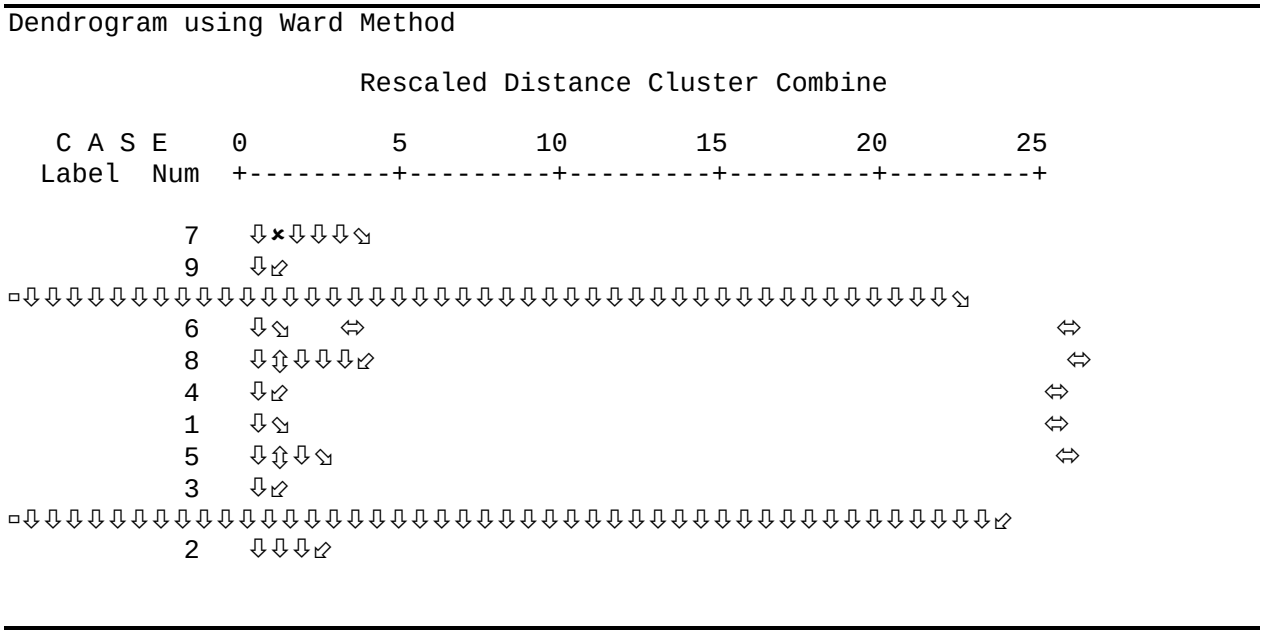

Figure 2. Cluster solution dendrogram for reading implementation indices.

To determine whether mean implementation index scores differed between high- and low-implementation schools, *t*-tests for independent samples were performed. As shown in Table 4, large and statistically significant differences were observed on all implementation indices: number of certified teachers per student enrollment in grades 3-5 (M_H = 8.93 versus M_L = 3.99; $t(7) = 3.41$, $p = .01$); reading composite index (M_H = 74.00 versus M_L = 50.20; $t(7) = 6.13$, $p < .01$); and math composite index (M_H = 36.25 versus M_L = 12.60; $t(7) = 3.45$, $p = .01$).

Table 4

Mean Implementation Index Scores by Implementation Intensity

Intensity	*n*	No. Certified per Student Enrollment[a]	Reading Composite Index	Math Composite Index
High	4	8.93	74.00	36.25
Low	5	3.99[1]	50.20[1]	12.60[1]

[1]Significantly lower than high-intensity implementation mean at $p < .01$.
[a]The number of program certified teachers per 100 students in grades 3, 4, and 5.

Results

Reading

Cohort Performance

This section provides a descriptive profile of school- and program-level reading performance from 1999 (baseline year) through 2002 for SR and C schools. All means and effect sizes are based on data from all students who completed TAAS assessments in reading in a given year. Data for each year within a given grade level represent a different cohort of students. Because these profiles do not take into account individual students' prior performance or socioeconomic status, attributions of any patterns to program performance should be made cautiously.

Third grade. Table 5 provides mean TLI Reading scores and effect size estimates for SR and C schools by matched pair for successive cohorts of 3rd grade students from 1997-98 through 2001-02. From 1999 to 2002, mean reading TLI scores improved at seven of the nine SR schools, and at eight of the nine C schools (see Table 5). During the same time period, cohort effect size estimates increased at five of the nine SR schools, ranging from -0.51 to +0.68 in

2002. From 1999 to 2002, the percentage of SR 3rd graders reading at or above grade level increased from 83% to 90%, versus 81% to 88% at C schools (see Figure 3). Median effect size estimates remained virtually unchanged and nearly equal to zero from 1999 (*ES* = -0.02) to 2002 (*ES* = -0.04), indicating that third-grade reading performance in a typical SR school was virtually equal to that of a matched C school during all years considered (see Figure 4).

Table 5

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Third Grade Reading

	1997-98			1998-99			1999-00			2000-01			2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
1	82.12	70.52	-0.40	78.68	73.85	-0.17	76.50	72.84	-0.14	84.75	83.98	-0.04	85.96	76.73	-0.50
2	81.63	56.22	-0.88	86.12	68.07	-0.62	87.24	72.88	-0.53	88.98	79.52	-0.53	89.89	80.36	-0.51
3	72.92	*	*	66.94	75.72	+0.30	66.79	83.95	+0.64	70.28	87.74	+0.97	83.71	79.57	-0.22
4	58.43	55.80	-0.09	69.95	63.35	-0.23	71.03	63.69	-0.27	75.08	77.60	+0.14	75.15	74.41	-0.04
5	83.29	81.73	-0.05	86.35	86.05	-0.01	84.73	84.53	-0.01	89.52	87.22	-0.13	74.83	87.51	+0.68
6	81.21	81.85	+0.02	87.56	85.98	-0.05	85.66	85.85	+0.01	85.45	86.28	+0.05	83.47	86.87	+0.18
7	78.79	74.55	-0.15	71.34	78.51	+0.25	76.88	75.50	-0.05	86.54	82.32	-0.24	81.65	83.71	+0.11
8	84.28	86.16	+0.07	85.02	87.23	+0.08	84.84	89.23	+0.16	89.49	90.04	+0.03	89.71	88.57	-0.06
9	67.90	64.33	-0.12	58.93	58.30	-0.02	76.50	72.84	-0.14	73.89	86.94	+0.73	76.92	85.13	+0.44

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate.
ES was computed by subtracting the C school mean from the SR school mean within each matched pair, then dividing by the total standard deviation.
*Data not available. School did not open until 1998-99 school year.

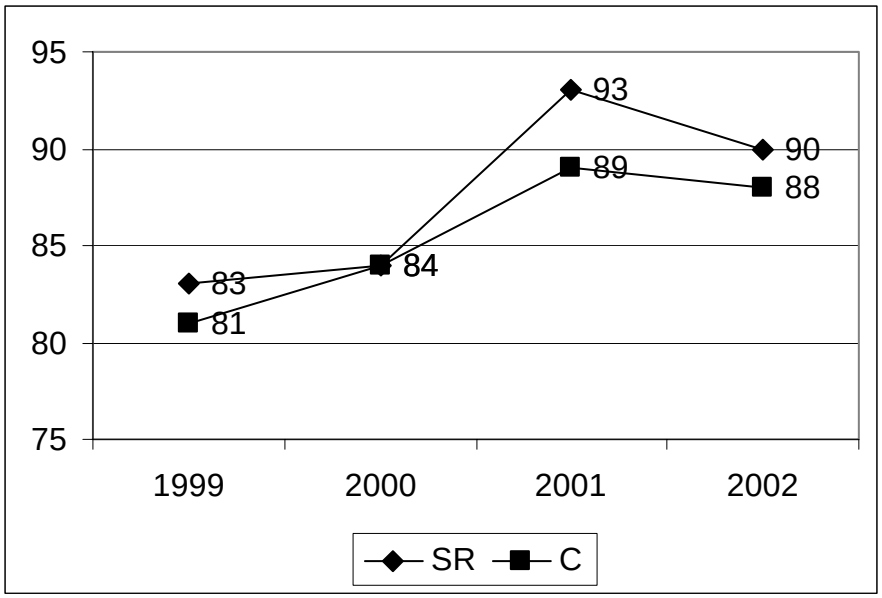

Figure 3. Percentage of 3rd Grade Students Reading At or Above Grade Level in School Renaissance (SR) and Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

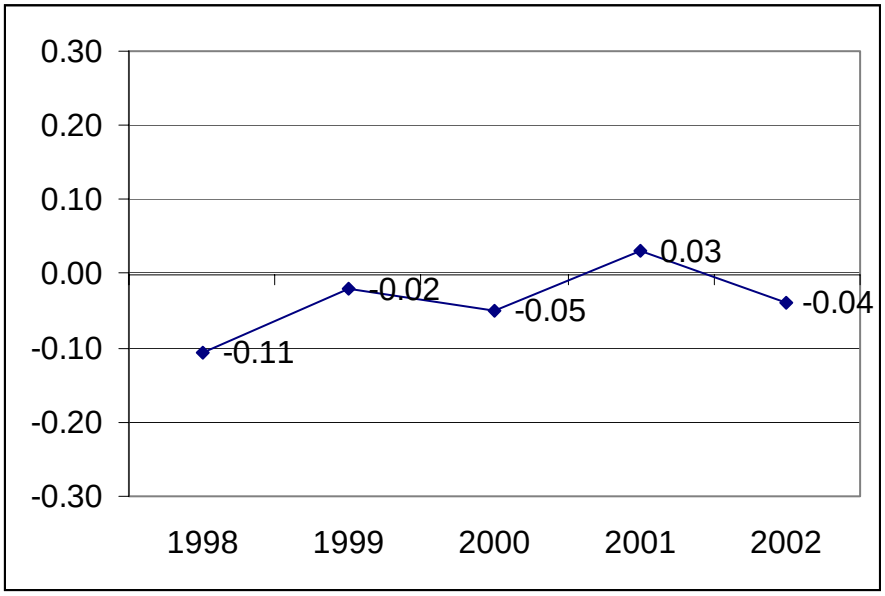

Figure 4. Median Effect Size Estimates by Year: Third Grade Reading.
*Year represents spring testing (1999 = 1998-99 school year).

Fourth grade. From 1999-2000, mean cohort reading scores increased at seven of nine SR schools (one remained equal) and all C schools from 1998-99 to 2001-02 (see Table 6). Cohort effect size estimates increased for five of nine matched pairs, and remained equal in one pair, ranging from -0.37 to +0.34 in 2002 (see Table 6). As Figure 5 illustrates, the percentage of students reading at or above grade level improved and was nearly equal for SR and C schools between 1999 and 2002; from 81% versus 80%, respectively, in 1999; to 92% versus 91% in 2002. As with third grade, median effect size estimates were equal to or near zero for all four years (see Figure 6).

Table 6

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Fourth Grade Reading

	1997-98			1998-99			1999-00			2000-01			2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
1	80.52	72.29	-0.27	84.73	70.09	-0.54	79.88	79.98	0.00	89.07	83.74	-0.28	90.19	86.20	-0.22
2	81.77	62.47	-0.64	80.63	68.38	-0.45	88.68	78.81	-0.37	87.77	74.83	-0.68	91.55	85.64	-0.32
3	74.81	*	*	70.58	87.30	0.61	75.88	68.13	-0.29	80.38	88.84	0.44	88.36	90.52	0.12
4	64.83	66.74	0.06	66.11	66.23	0.00	70.92	69.58	-0.05	79.26	84.70	0.28	75.52	81.88	0.34
5	86.21	85.19	-0.03	89.45	86.71	-0.10	89.28	89.45	0.01	90.90	87.79	-0.16	90.36	90.91	0.03
6	87.16	90.99	0.13	86.02	85.11	-0.03	87.83	90.08	0.08	90.33	87.12	-0.17	92.02	85.11	-0.37
7	77.61	75.12	-0.08	76.13	76.62	0.02	77.56	75.98	-0.06	85.26	89.40	0.22	84.66	86.80	0.12
8	86.24	86.25	0.00	89.93	90.15	0.01	89.98	92.07	0.08	90.63	93.06	0.13	91.08	89.07	-0.11
9	72.00	64.53	-0.25	65.63	67.20	0.06	64.89	74.48	0.36	81.20	79.02	-0.11	82.95	83.98	0.06

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate.
ES computed by subtracting comparison school mean from Renaissance school mean within each matched pair, then dividing by the total standard deviation.
*Data not available. School did not open until 1998-99 school year.

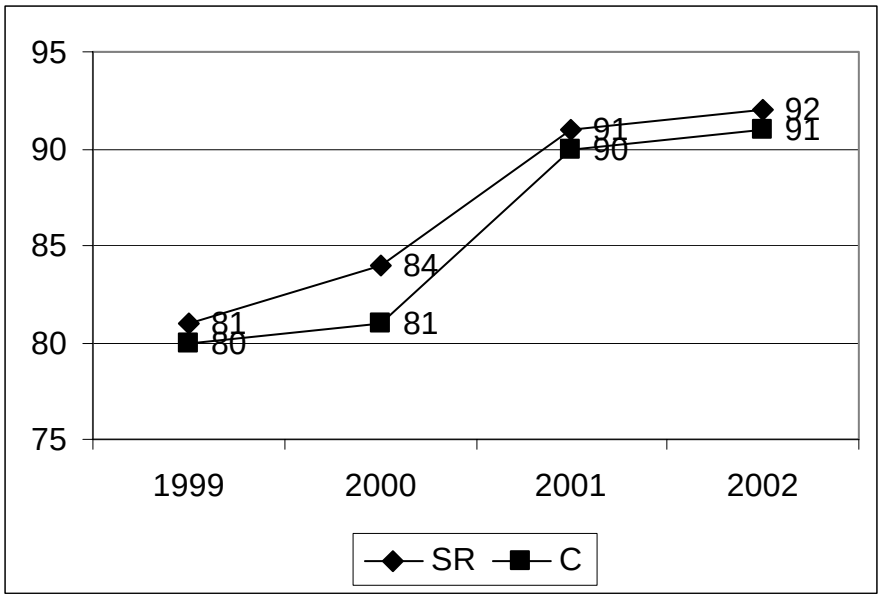

Figure 5. Percentage of 4th Grade Students Reading At or Above Grade Level in School Renaissance (SR) and Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

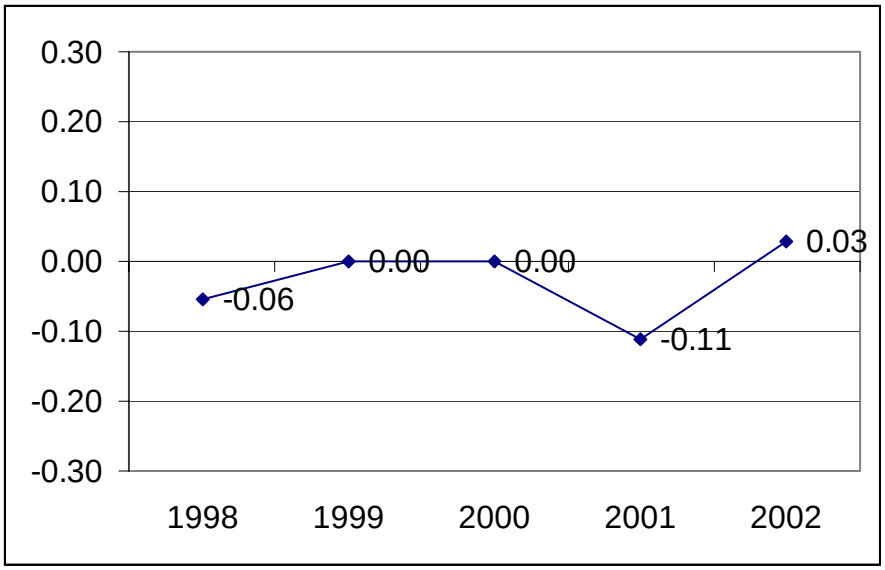

Figure 6. Median Effect Size Estimates by Year: Fourth Grade Reading.
*Year represents spring testing (1999 = 1998-99 school year).

Fifth grade. Mean reading scores increased at all SR and C schools from 1999 to 2002 (see Table 7). Effect size estimates increased in four of nine pairs from 1999 to 2002, ranging from -0.29 to +0.28 in 2002 (see Table 7). The percentage of students reading at or above grade level mirrored the fourth-grade pattern, with 82% versus 80% in 1999, and 93% versus 92% in 2002, for SR and C schools, respectively (see Figure 7). Effect size estimates remained stable and near zero throughout the same period (see Figure 8).

Table 7

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Fifth Grade Reading

	1997-98			1998-99			1999-00			2000-01			2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
1	79.49	70.31	-0.37	83.34	77.89	-0.18	87.63	80.16	-0.27	92.85	83.98	-0.44	89.60	85.80	-0.20
2	85.57	70.19	-0.62	77.48	65.45	-0.39	78.93	79.61	+0.02	90.58	72.69	-0.88	88.93	88.49	-0.02
3	81.67	*	*	73.63	78.40	+0.16	75.26	69.87	-0.20	88.00	91.97	+0.20	90.37	88.42	-0.10
4	72.37	66.33	-0.24	67.15	66.53	-0.02	73.04	68.41	-0.17	79.02	75.52	-0.17	80.97	82.94	+0.10
5	92.93	87.07	-0.24	90.35	90.82	+0.02	91.57	91.34	-0.01	91.50	94.73	+0.16	96.00	95.61	-0.02
6	88.99	91.73	+0.11	86.19	89.81	+0.12	86.41	85.78	-0.02	89.56	93.74	+0.21	94.02	91.57	-0.13
7	83.55	76.06	-0.30	70.81	76.97	+0.20	80.54	80.26	-0.01	89.57	86.51	-0.15	88.55	83.05	-0.29
8	87.46	90.57	+0.12	88.02	84.98	-0.10	91.66	92.77	+0.04	92.63	90.37	-0.11	93.21	95.03	+0.10
9	69.56	68.49	-0.04	67.09	61.19	-0.19	66.98	76.83	+0.36	76.67	78.38	+0.08	78.12	83.38	+0.28

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate.
ES computed by subtracting comparison school mean from Renaissance school mean within each matched pair, then dividing by the total standard deviation.
*Data not available. School did not open until 1998-99 school year.

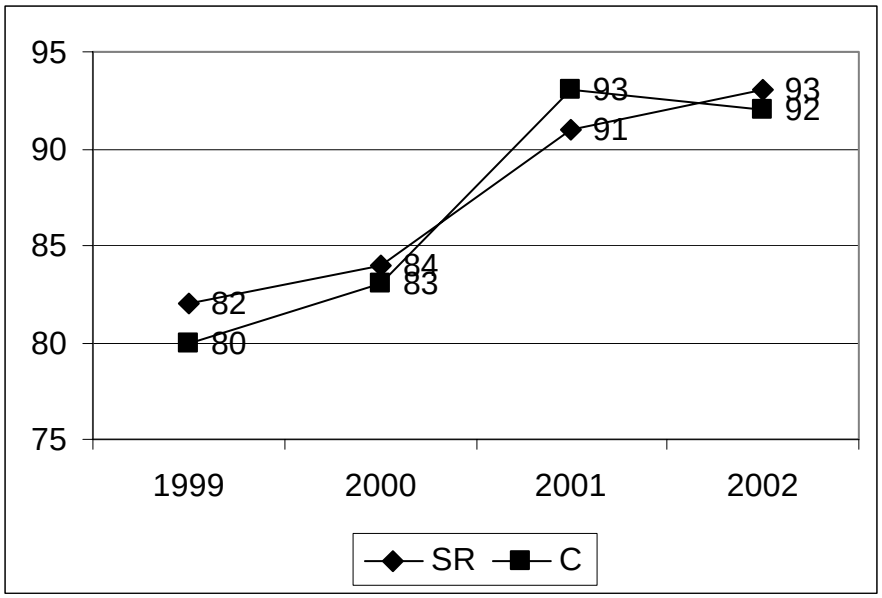

Figure 7. Percentage of 5th Grade Students Reading At or Above Grade Level in School Renaissance (SR) and Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

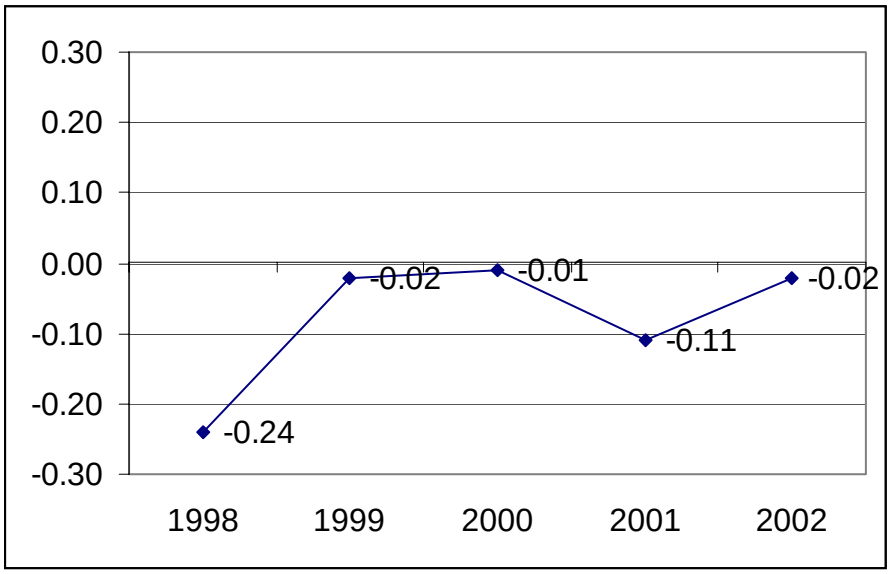

Figure 8. Median Effect Size Estimates by Year: Fifth Grade Reading.
*Year represents spring testing (1999 = 1998-99 school year).

Sixth grade. Mean reading scores increased in SR and C schools from 1999 to 2002 (see Table 8). Effect size estimates increased at both SR schools in these years, from -0.18 to +0.09 for Pair 10, and from -0.17 to -0.06 for Pair 11 (see Table 8). The percentage of students reading at or above grade level showed marked improvement in SR schools, from 76% versus 84% (for the C schools) in 1999, to 90% versus 89% in 2002 (see Figure 9). The C schools appeared to have a higher percent of students reading at or above grade level than the SR schools in the baseline year of 1999 (prior to SR implementation). By 2002, these percentages were nearly identical to both SR and C schools. Likewise, median effect size estimates improved substantially between 1999 and 2002, from -0.17 to +0.01 (see Figure10).

Table 8

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Sixth, Seventh, and Eighth Grade Reading

	6th Grade 1997-98			6th Grade 1998-99			6th Grade 1999-00			6th Grade 2000-01			6th Grade 2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
10	81.80	82.00	+0.01	83.69	79.61	-0.18	82.26	81.64	-0.03	87.17	86.34	-0.05	86.29	88.00	+0.09
11	82.42	75.65	-0.29	81.39	77.64	-0.17	81.99	80.90	-0.05	81.79	84.55	+0.16	87.88	86.67	-0.06
	7th Grade 1997-98			**7th Grade 1998-99**			**7th Grade 1999-00**			**7th Grade 2000-01**			**7th Grade 2001-02**		
10	80.30	78.75	-0.07	81.13	78.73	-0.10	81.76	77.55	-0.18	87.21	86.01	-0.06	86.38	88.40	+0.10
11	81.22	74.90	-0.28	81.23	73.59	-0.32	77.93	78.41	+0.02	86.99	88.34	+0.07	86.92	87.17	+0.01
	8th Grade 1997-98			**8th Grade 1998-99**			**8th Grade 1999-00**			**8th Grade 2000-01**			**8th Grade 2001-022**		
10	78.00	79.28	+0.05	81.09	80.69	-0.02	84.18	80.19	-0.16	83.25	79.07	-0.15	85.13	85.80	+0.02
11	85.28	77.33	-0.33	79.49	76.17	-0.14	81.30	78.30	-0.12	79.70	81.76	+0.08	86.60	85.24	-0.05

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate.
ES computed by subtracting the C school mean from the SR school mean within each matched pair, then dividing by the total standard deviation.

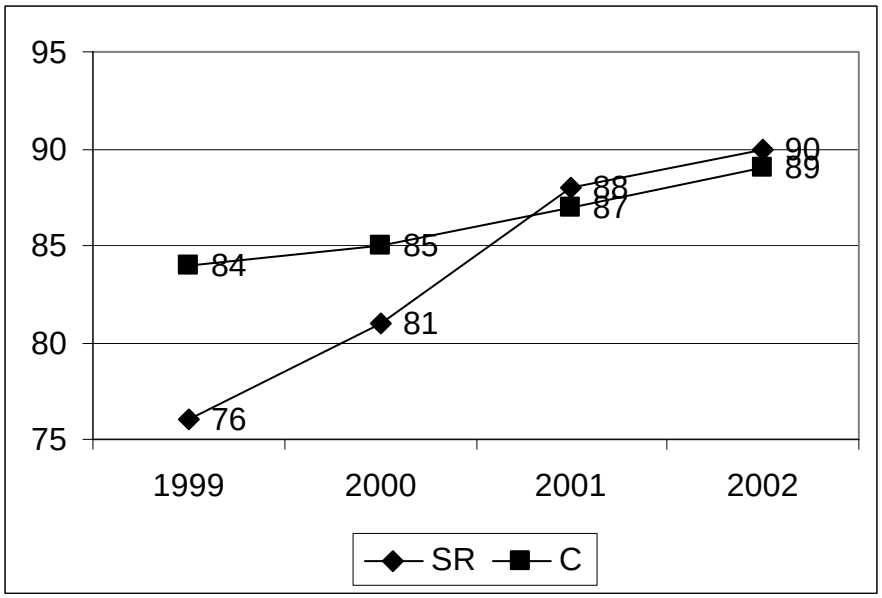

Figure 9. Percentage of 6th Grade Students Reading At or Above Grade Level in School Renaissance (SR) and Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

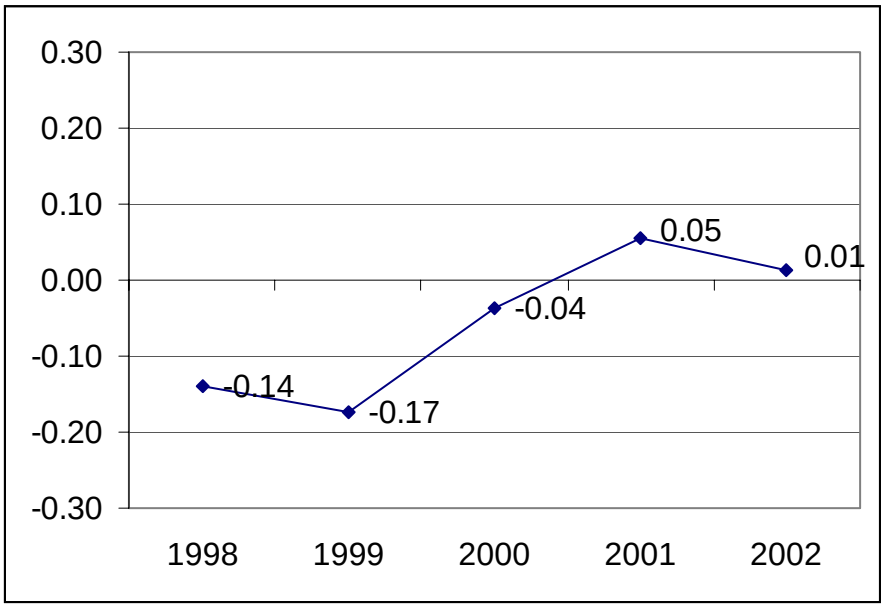

Figure 10. Median Effect Size Estimates by Year: Sixth Grade Reading.
*Year represents spring testing (1999 = 1998-99 school year).

Seventh grade. Mean reading scores increased at all four schools from 1999 to 2002 (see Table 8). Effect size estimates increased substantially at both SR schools, from -0.10 to +0.10 in Pair 10, and -0.32 to +0.01 in Pair 11 (see Table 8). The percentage of students reading at or above grade level improved substantially in SR schools versus C schools, from 77% versus 84% in 1999, to 92% versus 90% in 2002 (see Figure 11). The C schools appeared to have a higher percent of students reading at or above grade level than the SR schools in the baseline year of 1999 (prior to SR implementation). By 2002, these percentages were similar in both SR and C schools. Median effect size estimates showed strong and consistent improvement, from -0.21 in 1999 to +0.06 in 2002 (see Figure 12).

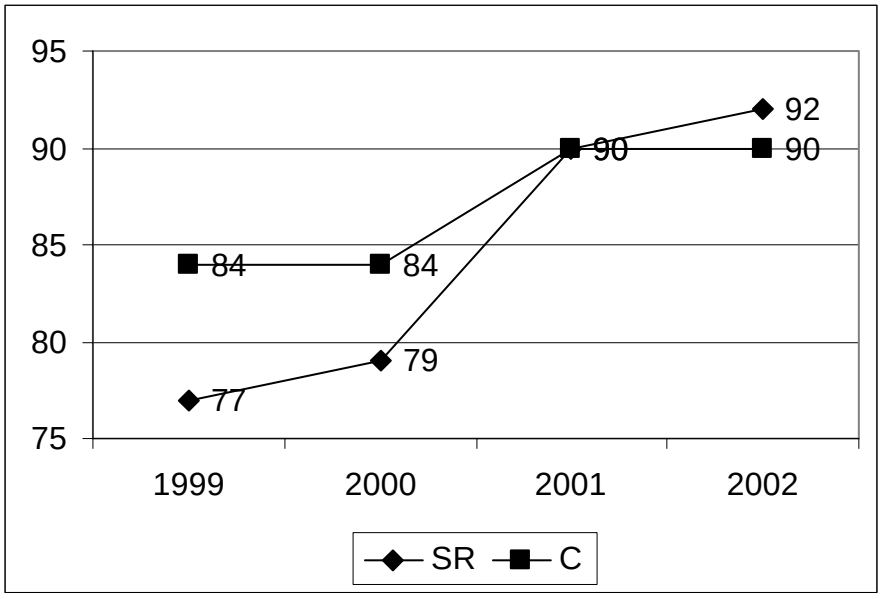

Figure 11. Percentage of 7th Grade Students Reading At or Above Grade Level in School Renaissance (SR) and Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

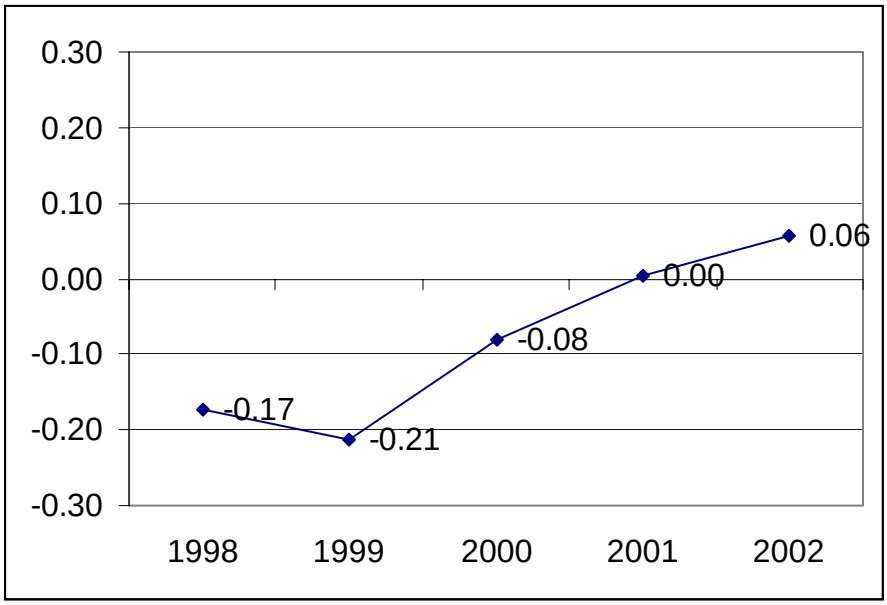

Figure 12. Median Effect Size Estimates by Year: Seventh Grade Reading.
*Year represents spring testing (1999 = 1998-99 school year).

Eighth grade. For all schools, mean reading achievement increased between 1999 and 2002 (see Table 8). Effect size estimates improved modestly for both SR schools, from -0.02 to +0.02 in Pair 10, to -0.14 to -0.05 in Pair 11 (see Table 8). The percentage of students reading at or above grade level increased at a greater rate in SR schools, from 79% (SR) versus 84% (C) in 1999, to 89% versus 88% in 2002 (see Figure 13). The C schools appeared to have a higher percent of students reading at or above grade level than the SR schools in the baseline year of 1999 (prior to SR implementation). By 2002, these percentages were nearly identical in both SR and C schools. Median effect size estimates improved modestly, from -0.08 in 1999 to -0.01 in 2002 (see Figure 14).

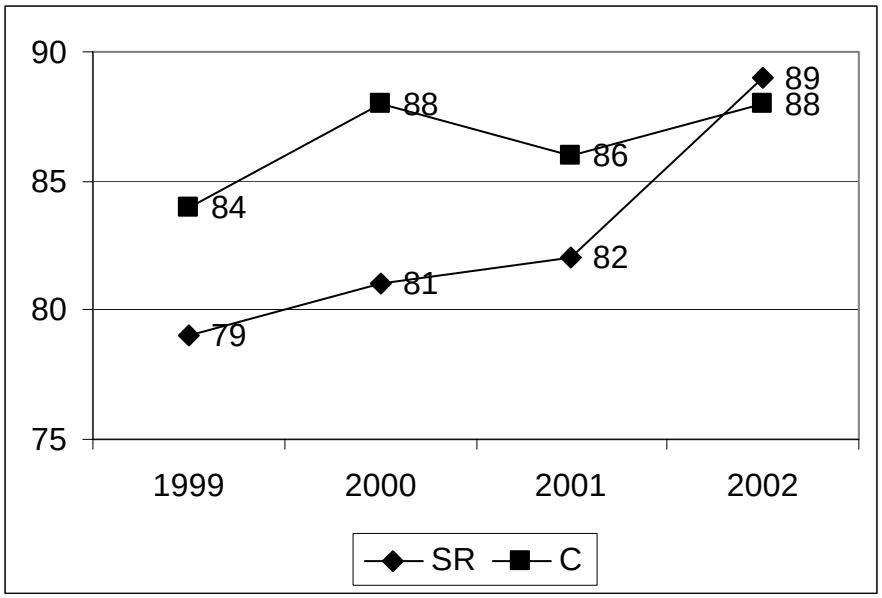

Figure 13. Percentage of 8th Grade Students Reading At or Above Grade Level in School Renaissance (SR) and Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

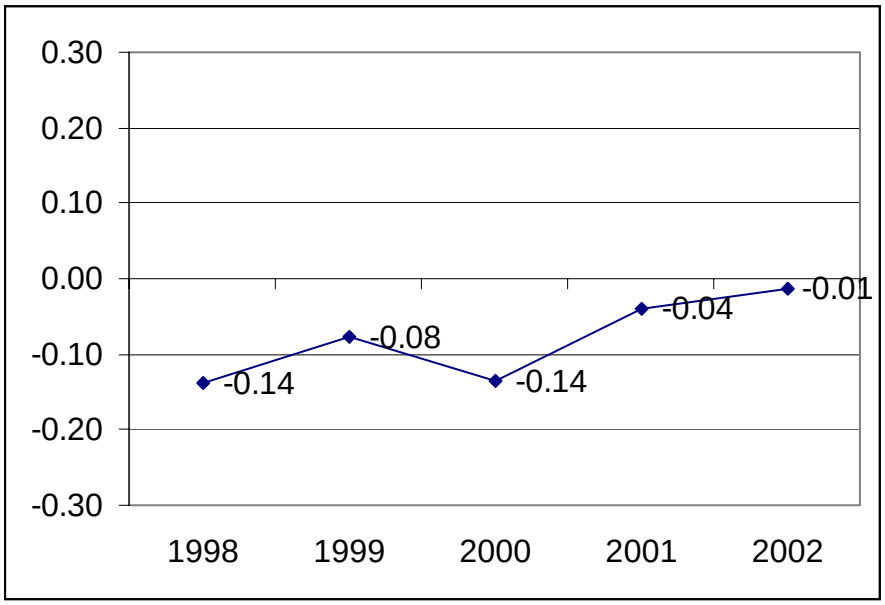

Figure 14. Median Effect Size Estimates by Year: Eighth Grade Reading.
*Year represents spring testing (1999 = 1998-99 school year).

Repeated-Measures Analyses

The repeated-measures analyses presented below test program effects on reading achievement while controlling for student baseline performance, student socioeconomic status, and the school-level random variable representing membership in a matched pair. All analyses were performed on the transformed achievement variables described above. Effect size estimates (Cohen's *d*) for significant treatment effects were computed by converting η^2 to *r*, then converting *r* to *d*.

2001 5th grade cohort: Reading. Three consecutive years of matched data were available for a total of 462 participants in the C schools and 450 participants in the SR schools. In the omnibus multivariate test, the treatment by reading (repeated measures) interaction effect was not significant (Wilk's lambda = 0.98, $F_{1,892} = 2.17$, $p = .141$), indicating that the effect of treatment was constant across years. Program type was found to have a significant main effect ($F_{1,846} = 10.45$, $p = .001$), as were 1999 transformed reading scores ($F_{1,892} = 292.02$, $p < .001$) and free or reduced-price lunch status ($F_{1,892} = 12.87$, $p < .001$). Program type accounted for about 1.2% of the variance in 2001 and 2002 reading scores ($\eta^2 = 0.012$), after controlling for free or reduced-price lunch status (partial $\eta^2 = 0.014$) and 2000 reading scores (partial $\eta^2 = 0.247$). Covariate-adjusted means on the averaged transformed variables (i.e., 2000 and 2001 scores averaged together) were $M = 1.038$ for the C group students, and $M = 1.105$ for SR students. As indicated in Figure 15, the mean difference was somewhat smaller in 2000 (4th grade) than in 2001 (5th grade), suggesting increased program effects with continued participation in the program (although the interaction was not significant). The effect size estimate was $d = +0.22$.

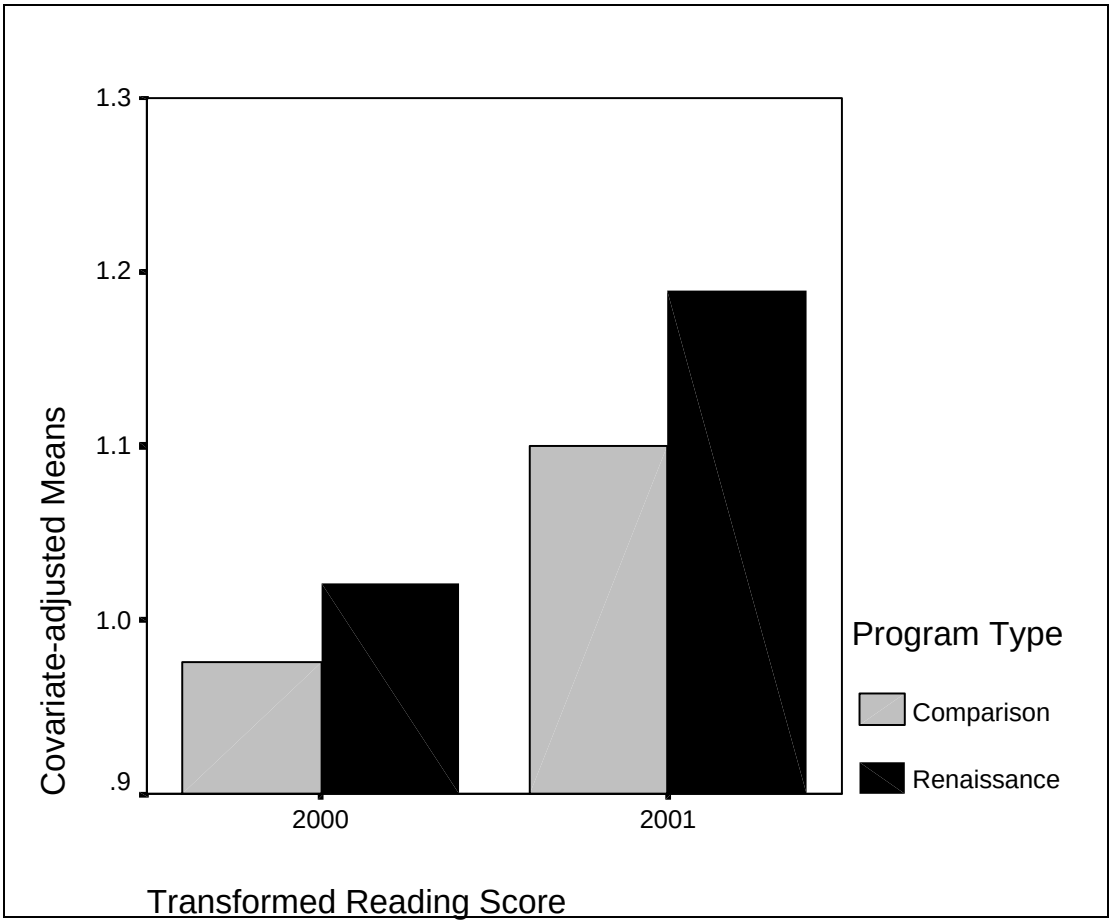

Figure 15. Covariate-adjusted Reading Means by Program Type: 2001 5th Grade Cohort Repeated-measures Analysis.
*Year represents spring testing (2000 = 1999-00 school year).

Note: Test of the average effect indicated a significant difference in means by program type.

2002 5th grade cohort: Reading. Three consecutive years of matched data were available for a total of 449 participants in the C schools and 441 participants in the SR schools. In the omnibus multivariate test, the treatment by reading (repeated measures) interaction effect was not significant (Wilk's lambda = 0.99, $F_{1,870} = 3.18$, $p = 0.08$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (4th and 5th grade) transformed scores. Program type was found to have a significant main effect ($F_{1,870} = 7.00$, $p = .008$), as were 2000 transformed reading score ($F_{1,870} = 363.61$, $p < .001$) and free or reduced-price lunch status ($F_{1,870} = 5.62$, $p = .018$). Program type accounted for about

0.8% of the variance in 2001 and 2002 reading scores ($\eta^2 = 0.008$), after controlling for free or reduced-price lunch status (partial $\eta^2 = 0.006$) and 2000 reading scores (partial $\eta^2 = 0.295$). Covariate-adjusted means on the averaged transformed variables were $M = 1.096$ for C group students and $M = 1.146$ for SR students. As indicated in Figure 16, the mean difference was much larger in 2001 (5th grade) than in 2000 (4th grade), suggesting that the achievement gains accelerated with exposure to the program over multiple years. (As indicated above, the program X year interaction effect approached significance.) The effect size estimate was $d = +0.17$. It should be noted that since third-grade scores were used as the covariate in this analysis (second grade scores were not available), this effect size might underestimate actual program effects (assuming outcomes comparable to other cohort analyses). That is, because third-grade was actually the first program implementation year for the 2002 cohort, to the extent that the SR scores reflected early treatment effects, the covariate would overestimate students' prior achievement (ability) relative to Comparison students. Fifth-grade program effects would therefore be underestimated accordingly.

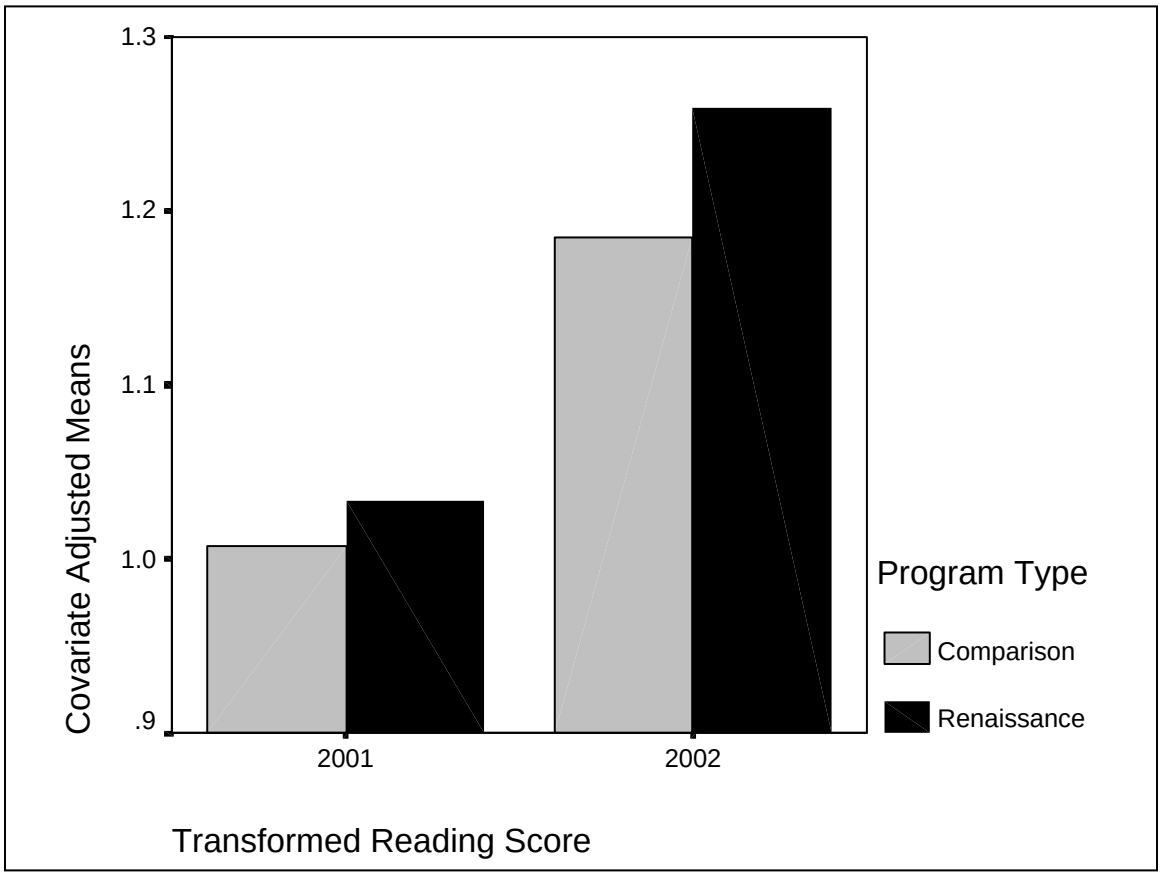

Figure 16. Covariate-adjusted Reading Means by Program Type: 2002 5th Grade Cohort Repeated-measures Analysis.
*Year represents spring testing (2001 = 2000-01 school year).

Note: Test of the average effect indicated a significant difference in means by program type.

2001 8th grade cohort: Reading. Three consecutive years of matched data were available for a total of 448 participants in the C schools and 400 participants in the SR schools. In the omnibus multivariate test, the treatment by reading (repeated measures) interaction effect was not significant (Wilk's lambda = 1.00, $F_{1,842} = 0.01$, $p = 0.94$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (7th and 8th grade) transformed scores. Program type did not have a significant main effect ($F_{1,842} = 0.12$, $p = .913$), although 1999 transformed reading score ($F_{1,842} = 759.81$, $p < .001$) and free or reduced-price lunch status ($F_{1,842} = 31.97$, $p < .001$) did have a significant effect on averaged

transformed scores. Covariate-adjusted means on the transformed variable were $M = 0.931$ for C group students and $M = 0.933$ for SR students.

2002 8th grade cohort: Reading. Three consecutive years of matched data were available for a total of 510 participants in the C schools and 482 participants in the SR schools. In the omnibus multivariate test, the treatment by reading (repeated measures) interaction effect was not significant (Wilk's lambda = 1.00, $F_{1,986} = 0.49$, $p = 0.48$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (7th and 8th grade) transformed scores. Program type did not have a significant main effect ($F_{1,986} = 0.73$, $p = .393$), although 2000 transformed reading score ($F_{1,986} = 639.85$, $p < .001$) and free or reduced-price lunch status ($F_{1,986} = 56.36$, $p < .001$) did have a significant effect on averaged transformed scores. Covariate-adjusted means on the transformed variable were $M = 1.085$ for C group students, and $M = 1.098$ for SR students.

Mathematics

Cohort Performance

This section provides a descriptive profile of school- and program-level mathematics performance from 2000 (baseline year) through 2002 for SR and C schools. All means and effect sizes are based on data from all students who completed TAAS assessments in mathematics in a given year. Data for each year within a given grade level represent a different cohort of students. These profiles do not take into account individual students' prior performance or socioeconomic status. Thus, as for the parallel analyses of reading scores, it may not be appropriate to attribute any patterns to program performance.

Third grade. Mean TLI mathematics scores improved in six of eight SR schools for which 3rd grade data were available from 2000 to 2002; effect size estimates improved for four

out of eight SR schools. In 2002, effect size estimates ranged from -1.01 for Pair 2, to +0.56 for Pair 9 (see Table 9). The percentage of third graders performing at or above grade level in mathematics rose from 84% to 90% between 2000 and 2002 in SR schools, and from 79% to 89% in C schools (see Figure 17). Median effect size (*ES*) estimates increased modestly during this time, from -0.03 in 2000 to +0.13 in 2002 (see Figure 18).

Table 9

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Third Grade Mathematics

	1998-99			1999-00			2000-01			2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
1	73.70	70.67	-0.11	72.30	68.48	-0.16	84.93	79.80	-0.35	85.07	82.25	-0.18
2	81.12	67.78	-0.50	82.00	71.50	-0.44	86.02	77.25	-0.60	85.58	69.91	-1.01
3	62.62	71.96	+0.35	58.94	79.88	+0.87	67.43	88.43	+1.45	79.95	78.50	-0.09
4	66.13	60.91	-0.20	71.27	N.A.	N.A.	72.82	77.77	+0.34	76.30	78.27	+0.13
5	78.97	83.67	+0.18	80.97	80.96	0.00	86.55	84.12	-0.17	79.40	86.76	+0.47
6	82.22	82.47	+0.01	80.31	85.65	+0.22	82.88	87.66	+0.33	83.81	87.97	+0.27
7	68.54	76.28	+0.29	75.34	73.99	-0.06	84.35	84.74	+0.03	82.24	82.49	+0.02
8	81.77	83.84	+0.08	82.17	86.24	+0.17	83.14	87.44	+0.30	86.51	88.78	+0.15
9	57.58	53.84	-0.14	62.29	60.19	-0.09	75.19	85.96	+0.74	77.56	86.22	+0.56

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate. *ES* computed by subtracting the C school mean from the SR school mean within each matched pair, then dividing by the total standard deviation.
N.A.: data not available.

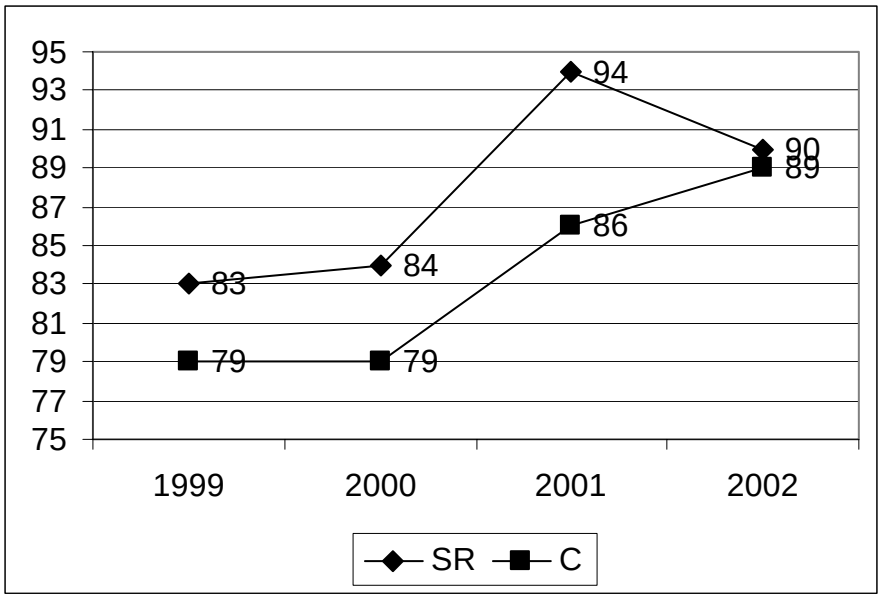

Figure 17. Percentage of 3rd Grade Students Performing At or Above Grade Level in Mathematics: School Renaissance (SR) versus Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

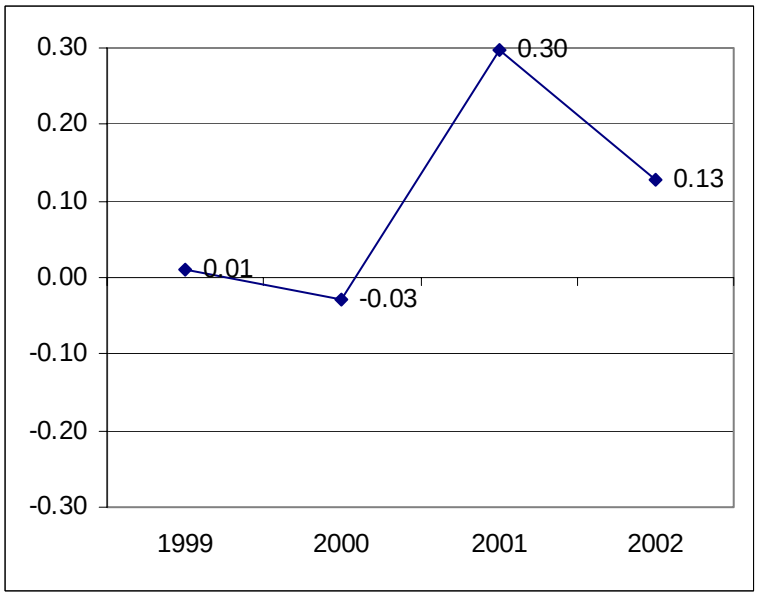

Figure 18. Median Effect Size Estimates by Year: Third Grade Mathematics.
*Year represents spring testing (1999 = 1998-99 school year).

Fourth grade. In fourth grade, average math scores improved at six of eight SR schools from 2000 to 2002 and at eight of nine C schools (see Table 10). Effect size estimates improved for two matched pairs. The percentage of students scoring at or above grade level improved in parallel fashion at SR and C schools, from 84% and 81% to 94% and 92%, respectively (see Figure 19). Median effect size estimates remained equal to or near zero across all years under consideration (see Figure 20). Individual ES estimates ranged from -0.79 to +0.69 in 2002 (see Table 10).

Table 10

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Fourth Grade Mathematics

	1998-99			1999-00			2000-01			2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
1	75.67	67.22	-0.29	70.76	75.55	+0.19	83.09	75.79	-0.44	85.57	77.94	-0.47
2	77.16	66.06	-0.38	82.81	71.25	-0.45	82.18	70.16	-0.73	86.16	73.31	-0.79
3	62.69	80.70	+0.62	66.29	60.15	-0.24	75.60	85.01	+0.57	79.47	84.47	+0.31
4	62.94	63.00	0.00	65.22	N.A.	N.A.	75.23	82.82	+0.46	73.52	84.62	+0.69
5	82.90	83.38	+0.02	84.26	84.59	+0.01	85.06	80.56	-0.27	85.97	85.37	-0.04
6	82.47	82.43	0.00	85.68	84.61	-0.04	84.53	85.28	+0.05	85.52	83.79	-0.11
7	69.11	70.17	+0.04	73.97	73.49	-0.02	81.41	83.09	+0.10	82.58	83.32	+0.05
8	82.67	84.07	+0.05	83.26	85.85	+0.10	85.45	86.22	+0.05	86.48	83.50	-0.18
9	60.94	63.40	+0.08	60.50	65.80	+0.21	74.62	82.00	+0.45	79.83	79.36	-0.03

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate. *ES* computed by subtracting the C school mean from the SR school mean within each matched pair, then dividing by the total standard deviation.
N.A. = data not available.

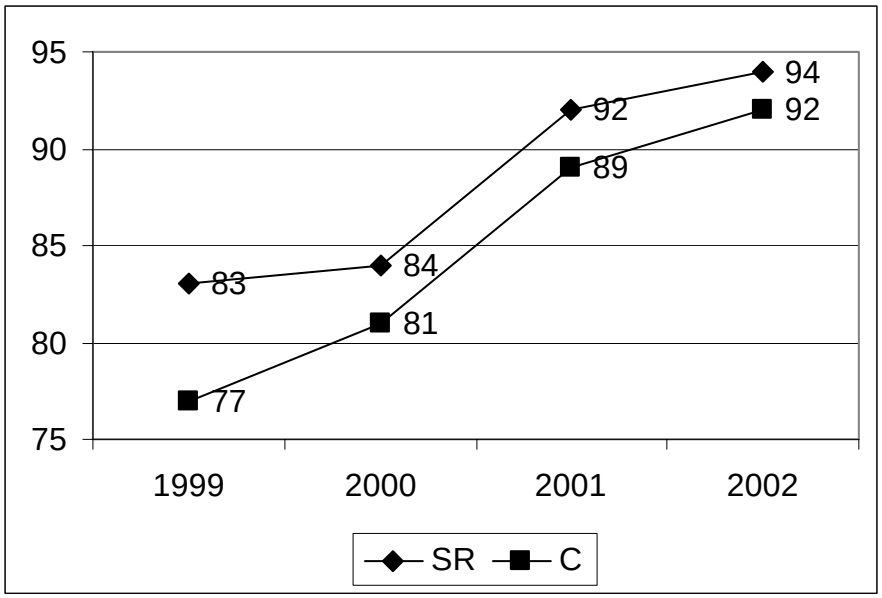

Figure 19. Percentage of 4th Grade Students Performing At or Above Grade Level in Mathematics: School Renaissance (SR) versus Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

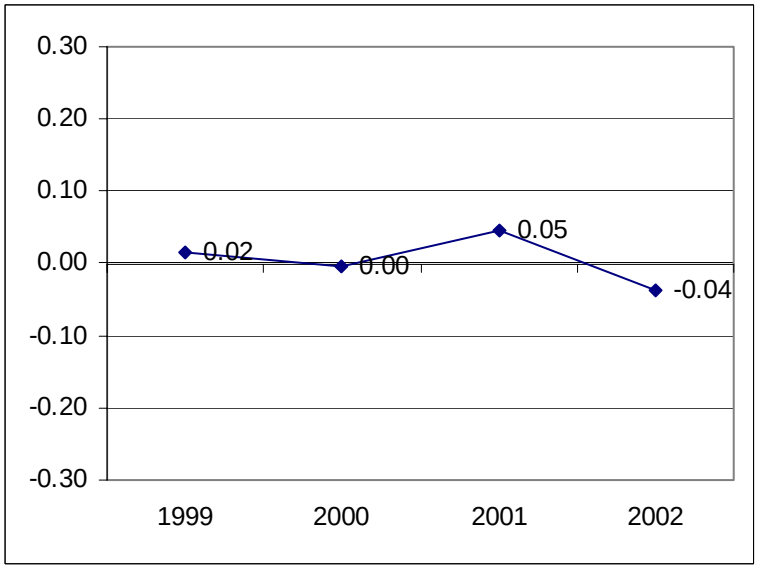

Figure 20. Median Effect Size Estimates by Year: Fourth Grade Mathematics.
*Year represents spring testing (1999 = 1998-99 school year).

Fifth grade. Mean TLI mathematics scores improved in all schools (SR and C) between 2000 and 2002, while *ES* estimates ranged from -0.56 to +0.41 in 2002 (see Table 11). As with fourth grade, the percentage of students performing at or above grade level rose in parallel fashion, although in fifth grade the percentages were virtually equal for SR and C schools each year (see Figure 21). Median effect size estimates declined from -0.07 to -0.14 from 2000 to 2001, but increased to +0.01 in 2002 (see Figure 22).

Table 11

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Fifth Grade Mathematics

	1998-99			1999-00			2000-01			2001-02		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
1	85.60	75.89	-0.36	86.13	78.56	-0.31	88.20	80.75	-0.49	88.98	83.71	-0.34
2	76.52	62.67	-0.52	80.49	76.97	-0.14	87.79	74.33	-0.89	84.45	84.58	+0.01
3	69.49	74.22	+0.18	69.65	57.65	-0.49	82.66	85.96	+0.22	83.96	81.83	-0.14
4	66.78	68.86	+0.08	72.69	N.A.	N.A.	81.18	72.98	-0.54	80.43	86.88	+0.41
5	89.27	84.93	-0.16	86.93	85.51	-0.06	85.37	87.89	+0.17	88.90	87.70	-0.08
6	83.46	87.99	+0.17	86.00	83.96	-0.08	87.52	88.20	+0.04	88.62	90.33	+0.11
7	67.30	73.34	+0.23	77.79	76.76	-0.04	86.99	84.82	-0.14	85.63	76.82	-0.56
8	83.22	82.95	-0.01	86.82	87.18	+0.01	89.06	88.06	-0.07	89.10	89.37	+0.02
9	70.31	59.56	-0.40	67.45	74.17	+0.28	83.50	75.74	-0.51	77.79	82.10	+0.28

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate. *ES* computed by subtracting the C school mean from the SR school mean within each matched pair, then dividing by the total standard deviation.
N.A. = data not available.

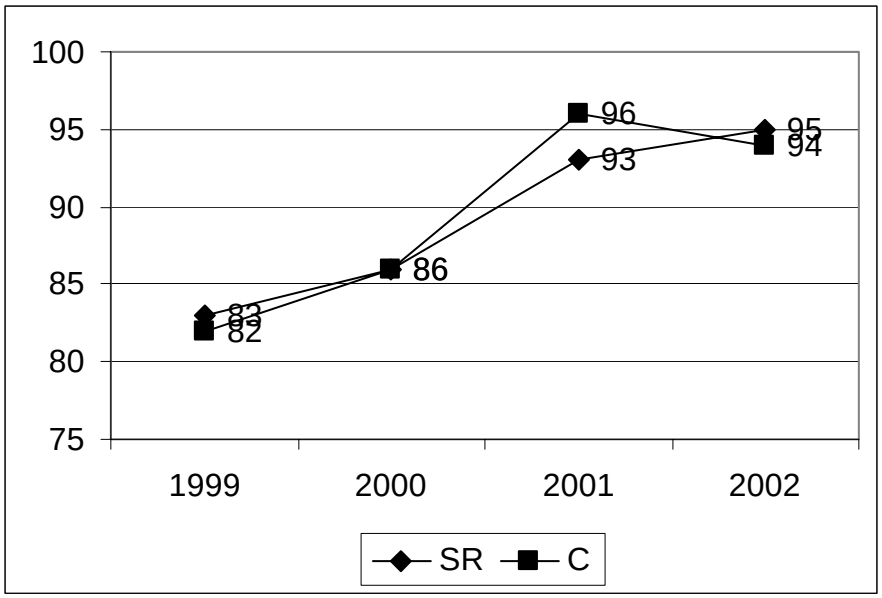

Figure 21. Percentage of 5th Grade Students Performing At or Above Grade Level in Mathematics: School Renaissance (SR) versus Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

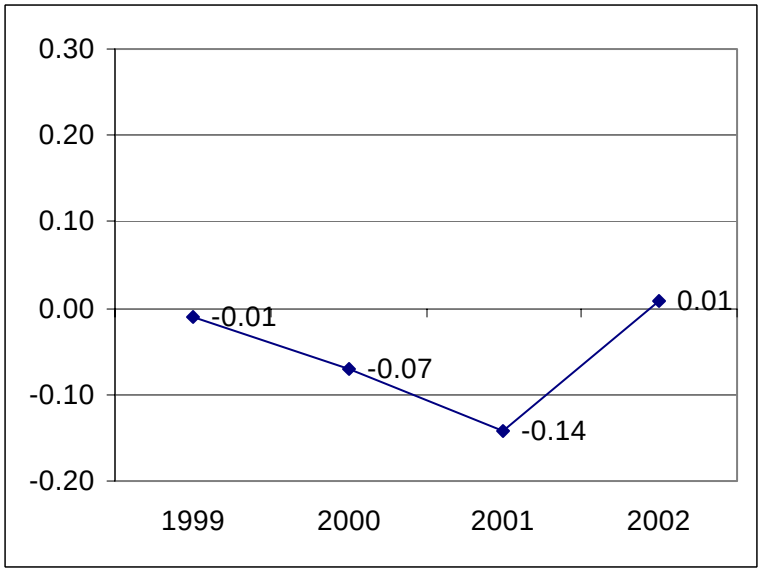

Figure 22. Median Effect Size Estimates by Year: Fifth Grade Mathematics.
*Year represents spring testing (1999 = 1998-99 school year).

Sixth grade. Mean math achievement improved at all schools (see Table 12). The percentage of students achieving at or above grade level improved at an accelerated rate in SR schools, from 82% (SR) versus 87% (C) in 2000, to 94% for both SR and C in 2002 (see Figure 23). The C schools appeared to have a higher percent of students performing at or above grade level than the SR schools in the baseline year of 2000 (prior to SR implementation). By 2002, these percentages were identical in both SR and C schools. Individual effect size estimates were +0.10 in Pair 10 in 2002, and -0.18 in Pair 11, both of which were improvements (see Table 12). The median *ES* improved slightly, from -0.10 to -0.04 (see Figure 24).

Table 12

Mean Texas Learning Index Means and Effect Size Estimates by Matched Pair, Program, and Year: Sixth, Seventh, and Eighth Grade Mathematics

	6th Grade 1998-99			**6th Grade 1999-00**			**6th Grade 2000-01**			**6th Grade 2001-02**		
Pair	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**	**C**	**SR**	**ES**
10	80.03	77.22	-0.13	77.29	78.54	+0.06	83.28	83.70	+0.03	82.84	84.52	+0.10
11	79.69	72.25	-0.34	80.58	74.67	-0.26	81.65	81.36	-0.02	86.00	82.93	-0.18
	7th Grade 1998-99			**7th Grade 1999-00**			**7th Grade 2000-01**			**7th Grade 2001-02**		
10	79.51	76.93	-0.12	79.59	77.85	-0.08	82.12	80.58	-0.09	81.49	86.06	+0.25
11	83.63	68.97	-0.66	79.57	74.75	-0.21	82.92	82.38	-0.03	84.98	82.97	-0.11
	8th Grade 1998-99			**8th Grade 1999-00**			**8th Grade 2000-01**			**8th Grade 2001-02**		
10	76.74	75.72	-0.04	79.15	76.67	-0.11	77.24	74.64	-0.10	78.51	79.45	+0.04
11	78.46	70.67	-0.34	80.73	73.26	-0.33	77.85	78.82	+0.04	80.96	79.25	-0.07

Note: C = Comparison school. SR = School Renaissance school. *ES* = effect size estimate. *ES* computed by subtracting the C school mean from the SR school mean within each matched pair, then dividing by the total standard deviation.

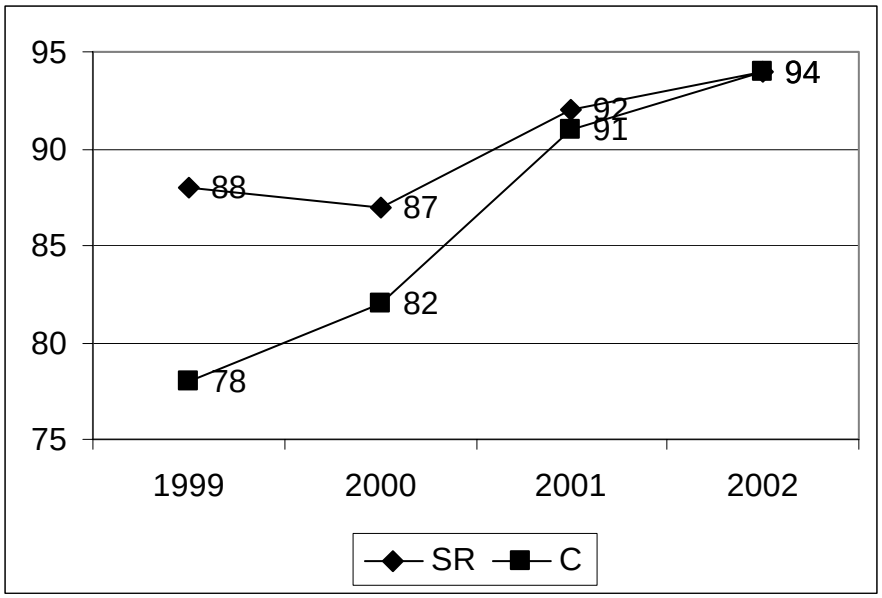

Figure 23. Percentage of 6th Grade Students Performing At or Above Grade Level in Mathematics: School Renaissance (SR) versus Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

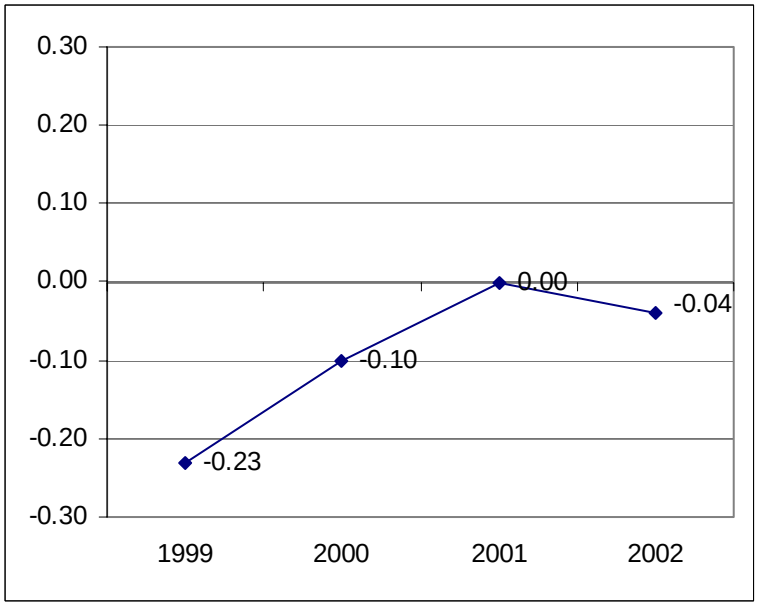

Figure 24. Median Effect Size Estimates by Year: Sixth Grade Mathematics.
*Year represents spring testing (1999 = 1998-99 school year).

Seventh grade. Average TLI scores increased at all schools between 2000 and 2002 (see Table 12). Effect size estimates improved for both SR schools, from -0.08 to +0.25 in Pair 10, and -0.21 to -0.11 in Pair 11 (see Table 12). Median ES estimates indicated steady improvement in SR schools, from -0.14 (2000) to -0.06 (2001) to +0.07 (2002; see Figure 25). Likewise, the percentage of students performing at or above grade level improved 10 points between 2000 and 2002, versus 4 points for comparison schools (see Figure 26). The C schools appeared to have a higher percent of students performing at or above grade level than the SR schools in the baseline year of 2000 (prior to SR implementation). By 2002, these percentages were nearly identical in both SR and C schools.

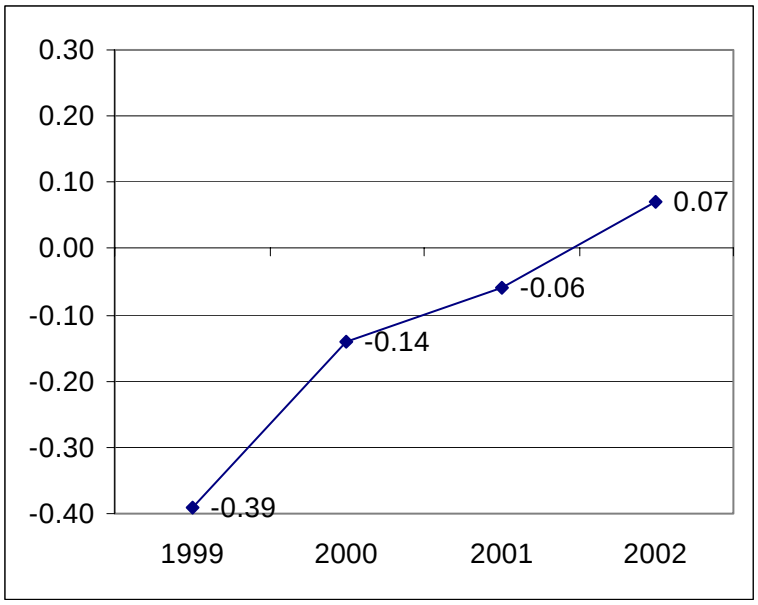

Figure 25. Median Effect Size Estimates by Year: Seventh Grade Mathematics.
*Year represents spring testing (1999 = 1998-99 school year).

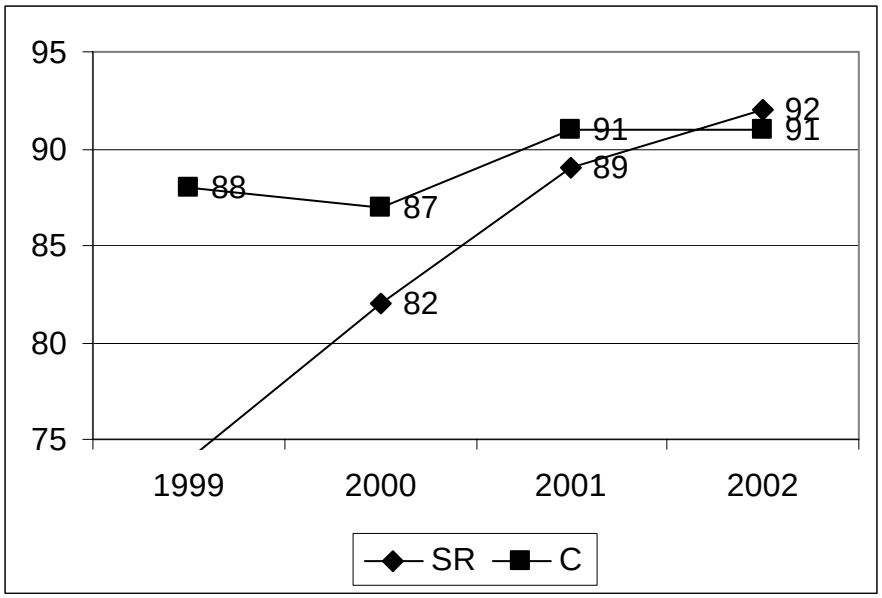

Figure 26. Percentage of 7th Grade Students Performing At or Above Grade Level in Mathematics: School Renaissance (SR) versus Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

Eighth grade. Mean math achievement improved at both SR schools, but declined in one C school and remained about the same in the other (see Table 12). SR schools showed consistent improvement in the percentage of students performing at or above grade level, from 82% in 2000 to 88% in 2002 (see Figure 27). C schools demonstrated a slight decline over these years, from 90% to 89%. The C schools appeared to have a higher percent of students performing at or above grade level than the SR schools in the baseline year of 2000 (prior to SR implementation). By 2002, these percentages were nearly identical in both SR and C schools. Median ES estimates improved somewhat, from -0.22 in 2000 to -0.02 in 2002 (see Figure 28).

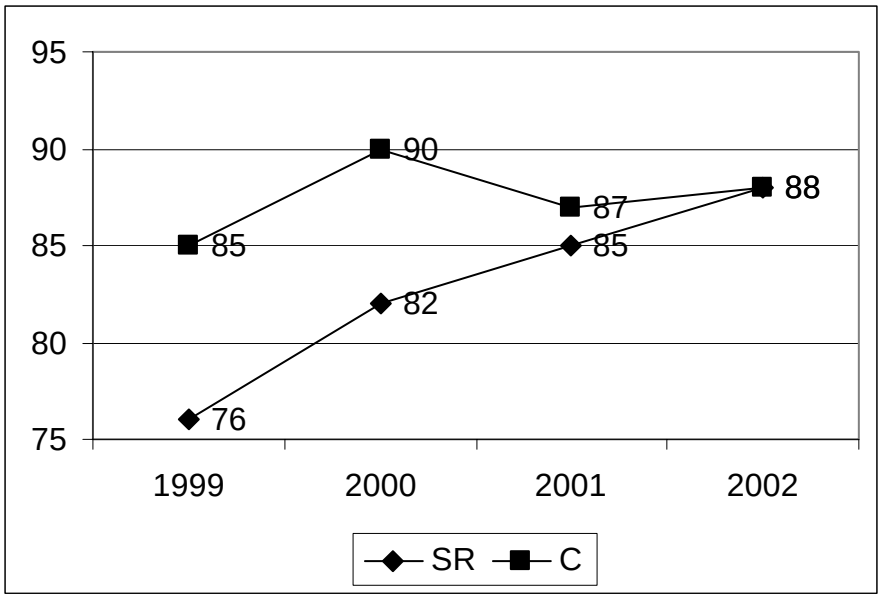

Figure 27. Percentage of 8th Grade Students Performing At or Above Grade Level in Mathematics: School Renaissance (SR) versus Comparison (C) Schools by Year.
*Year represents spring testing (1999 = 1998-99 school year).

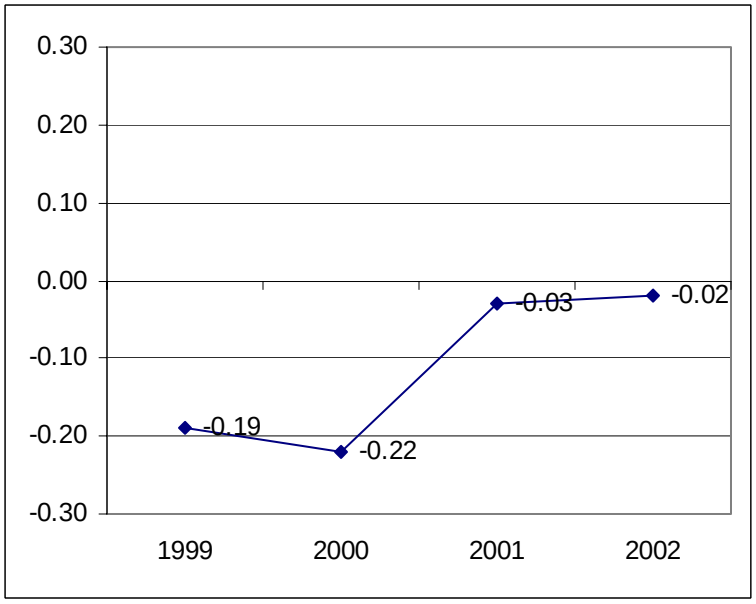

Figure 28. Median Effect Size Estimates by Year: Eighth Grade Mathematics.
*Year represents spring testing (1999 = 1998-99 school year).

Repeated-Measures Analyses

The repeated-measures analyses presented below test program effects on mathematics achievement while controlling for student baseline performance, student socioeconomic status, and the school-level random variable representing membership in a matched pair. All analyses were performed on the transformed achievement variables as described above. Effect size estimates (Cohen's *d*) for significant treatment effects were computed by converting η^2 to *r*, then converting *r* to *d*.

2002 5th grade cohort: Mathematics. Three consecutive years of matched data were available for a total of 449 participants in the C schools and 416 participants in the SR schools. In the omnibus multivariate test, the treatment by mathematics (repeated measures) interaction effect was not significant (Wilk's lambda = 1.00, $F_{1,845} = 0.13$, $p = 0.72$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (4th and 5th grade) transformed scores. Program type was found to have a significant main effect ($F_{1,845} = 8.42$, $p = .004$), as was 2000 transformed math score ($F_{1,845} = 381.94$, $p < .001$). Free or reduced-price lunch status did not have a significant effect on averaged transformed scores ($F_{1,845} = 0.025$, $p = .875$). Program type accounted for about 1.0% of the variance in 2001 and 2002 reading scores ($\eta^2 = 0.010$). Covariate-adjusted means on the averaged transformed variables were $M = 1.206$ for C group students and $M = 1.257$ for SR students. As indicated in Figure 29, the mean difference was consistent across 4th and 5th grades. The effect size estimate was $d = +0.20$.

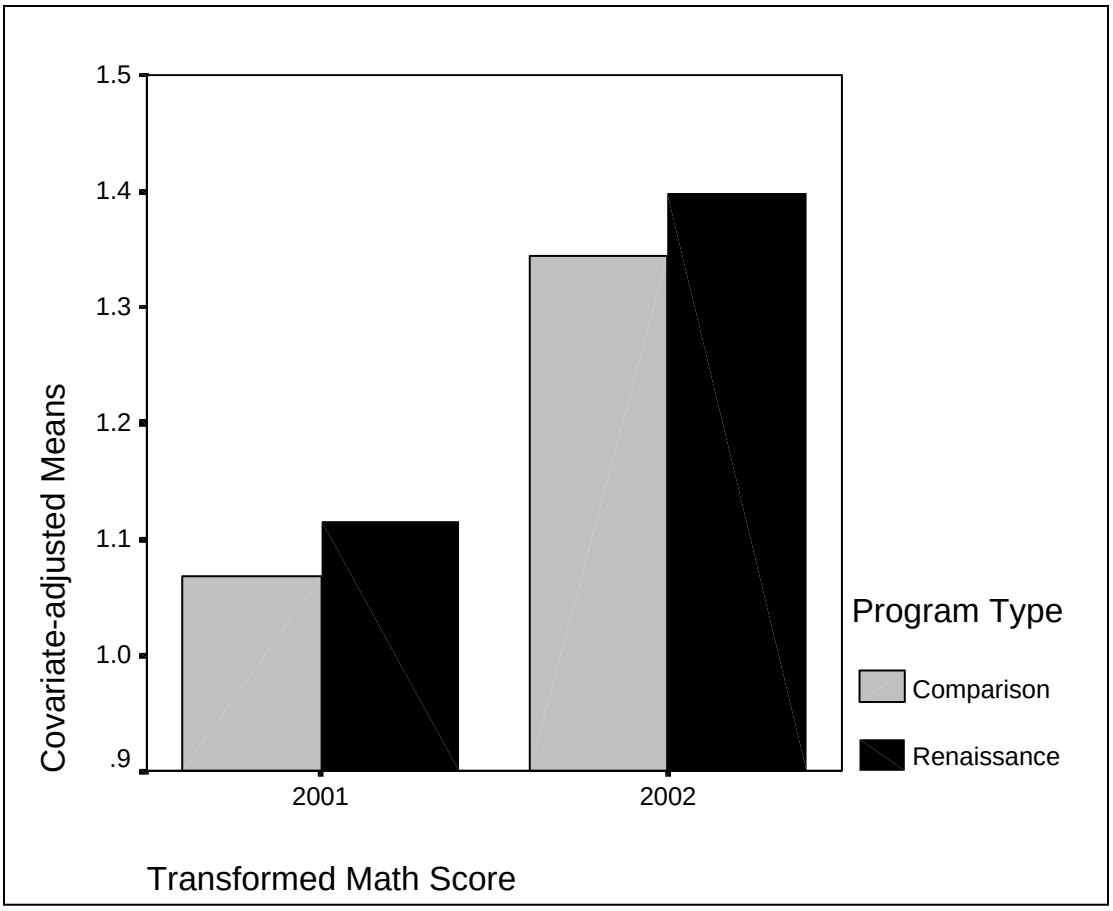

Figure 29. Covariate-adjusted Mathematics Means by Program Type: 2002 5th Grade Cohort Repeated-measures Analysis.
*Year represents spring testing (2001 = 2000-01 school year).

Note: Test of the average effect indicated a significant difference in means by program type.

2002 8th grade cohort: Math. Three consecutive years of matched data were available for a total of 510 participants in the C schools and 482 participants in the SR schools. In the omnibus multivariate test, the treatment by reading (repeated measures) interaction effect was not significant (Wilk's lambda = 0.99, $F_{1,986} = 0.64$, $p = 0.42$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (7th and 8th grade) transformed scores. Program type was found to have a significant main effect ($F_{1,986} = 7.68$, $p = .006$), as was 2000 transformed math score ($F_{1,986} = 715.60$, $p < .001$). Free or

reduced-price lunch status did not have a significant effect on averaged transformed scores ($F_{1,986}$ = 12.99, $p < .001$). Program type accounted for about 0.8% of the variance in 2001 and 2002 reading scores ($\eta^2 = 0.008$). Covariate-adjusted means on the transformed variable were $M = 1.211$ for SR students, and $M = 1.160$ for C group students. As indicated in Figure 30, the mean difference was somewhat larger in 8th grade than in 7th grade, suggesting that the achievement gains accelerated with exposure to the program over multiple years (however, the program X year interaction was not significant). The effect size estimate was $d = +0.17$.

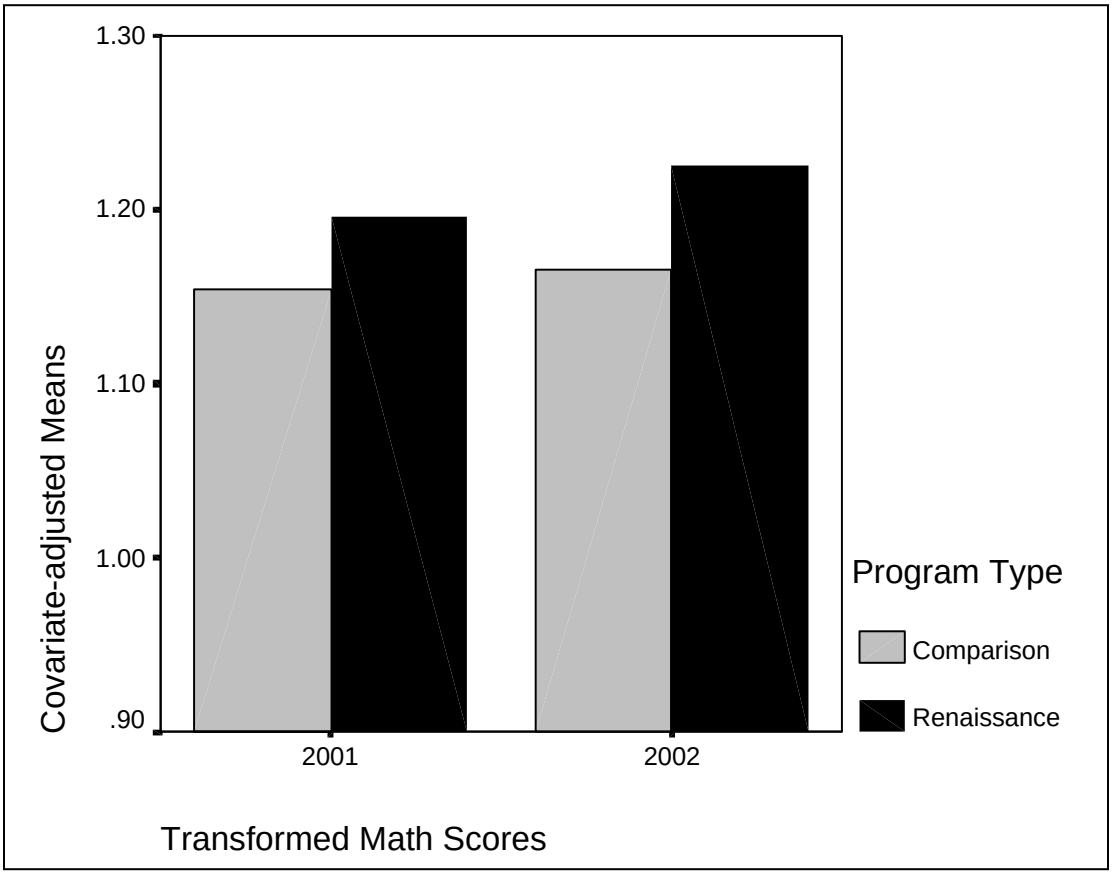

Figure 30. Covariate-adjusted Mathematics Means by Program Type: 2002 8th Grade Cohort Repeated-measures Analysis.
*Year represents spring testing (2001 = 2000-01 school year).

Note: Test of the average effect indicated a significant difference in means by program type.

Analyses of Student Subgroups

Descriptive analyses were performed on the 2001 and 2002 fifth grade cohorts to provide an overview of how three student subgroups performed in SR versus comparison (C) schools: students eligible for free or reduced-price lunch; limited English proficient (LEP) students; and students who performed at- or below grade level in third grade (i.e., received a score of 70 or below on the TAAS subtest). Because these student subgroups were relatively small and unevenly distributed across matched pairs of schools, it was not possible to incorporate the random school variable into the analyses. Therefore, inferential tests of program effects were not warranted due to the inability to utilize the matched-schools comparison design. Consequently, results should be interpreted as suggestive rather than definitive.

Students Eligible for Free or Reduced-Price Lunch

SR students who were eligible for free or reduced-price lunch posted larger gains than their counterparts in both Reading and Mathematics in the 2001 and 2002 5th grade cohorts (see Figures 31-36). The largest difference in gains was observed for the 2001 cohort in Reading, in which the SR student mean increased from 62.7 in 3rd grade to 77.6 in 5th grade (+14.9; see Table 13), while the C student mean increased from 73.0 to 80.8 (+7.8; see Table 13). The 2002 cohort was observed to have large differential gains in Mathematics favoring SR students (72.1 to 85.5; + 13.4) over C students (74.4 to 82.3; + 7.9; see Table 13). Trends also favored the SR 2001 cohort in Mathematics, and the SR 2002 cohort in Reading, although differences were smaller.

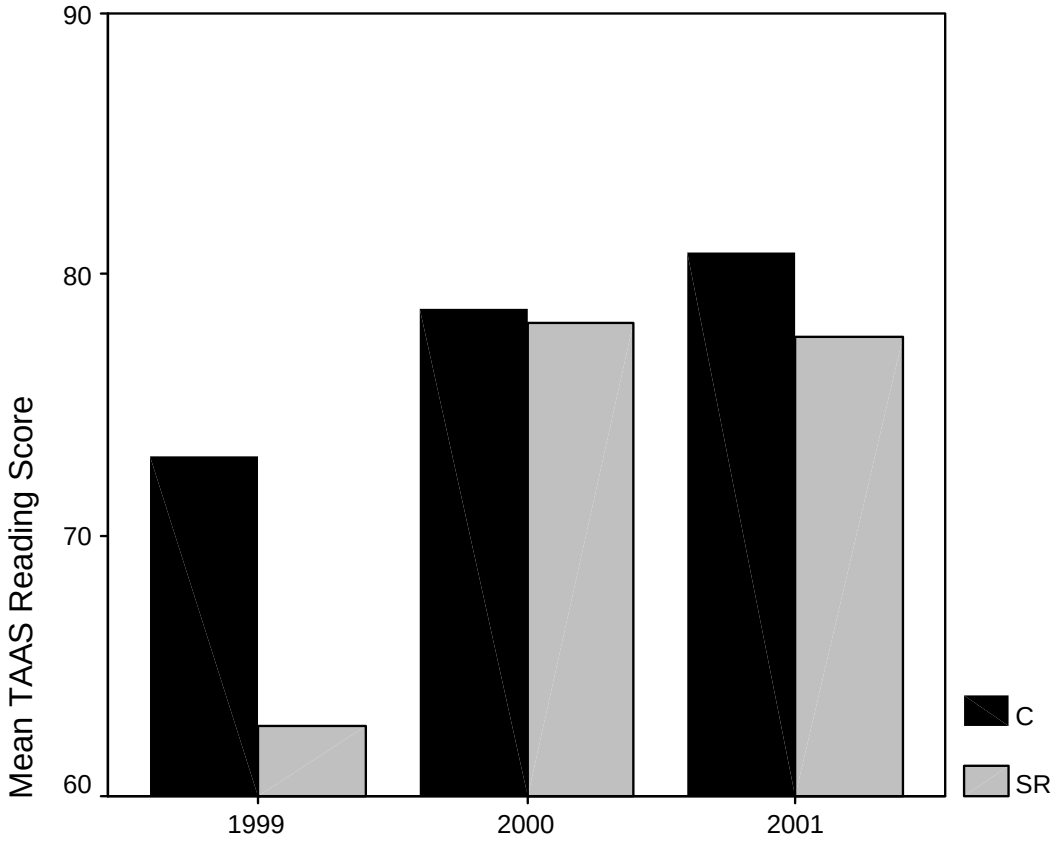

Figure 31. Mean TAAS Reading Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

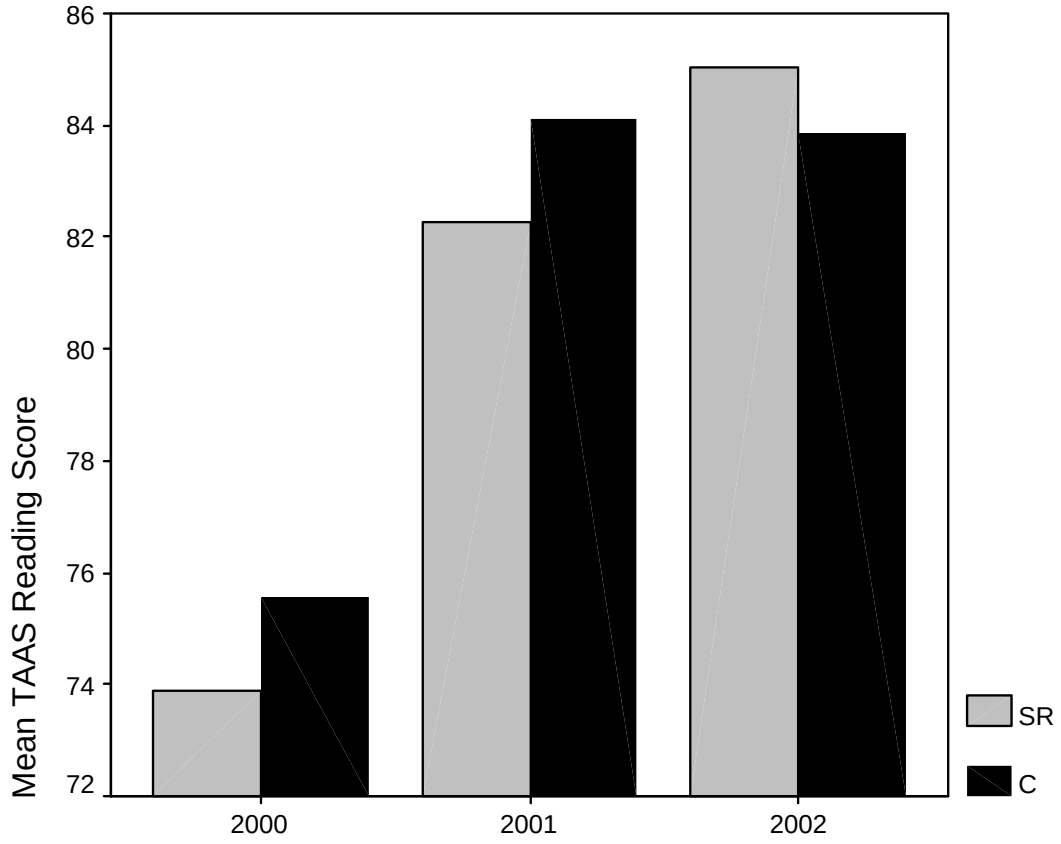

Figure 32. Mean TAAS Reading Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

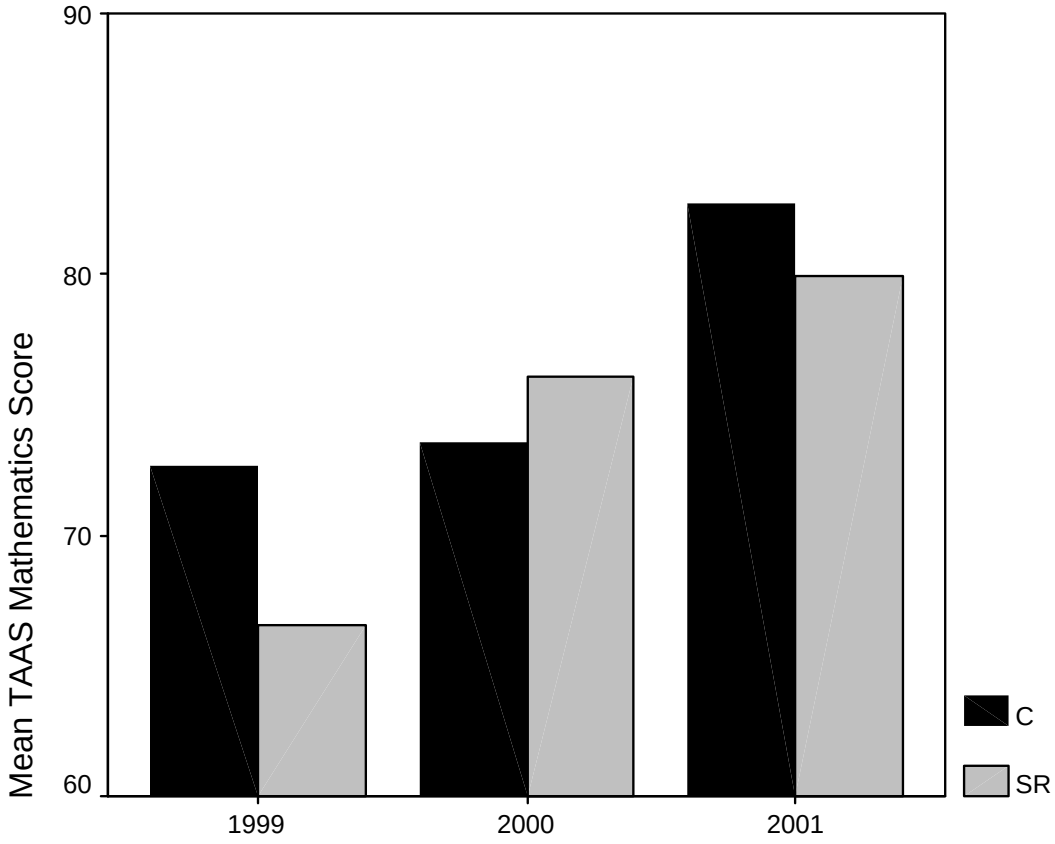

Figure 33. Mean TAAS Mathematics Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

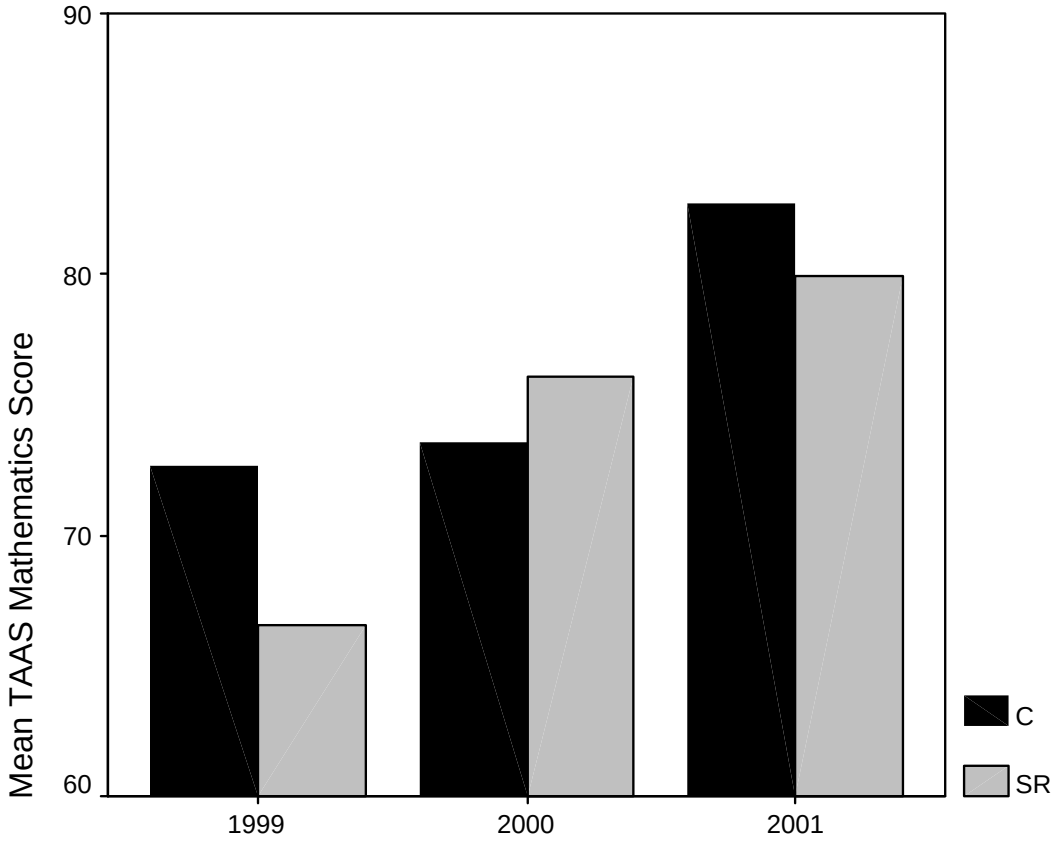

Figure 34. Mean TAAS Mathematics Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

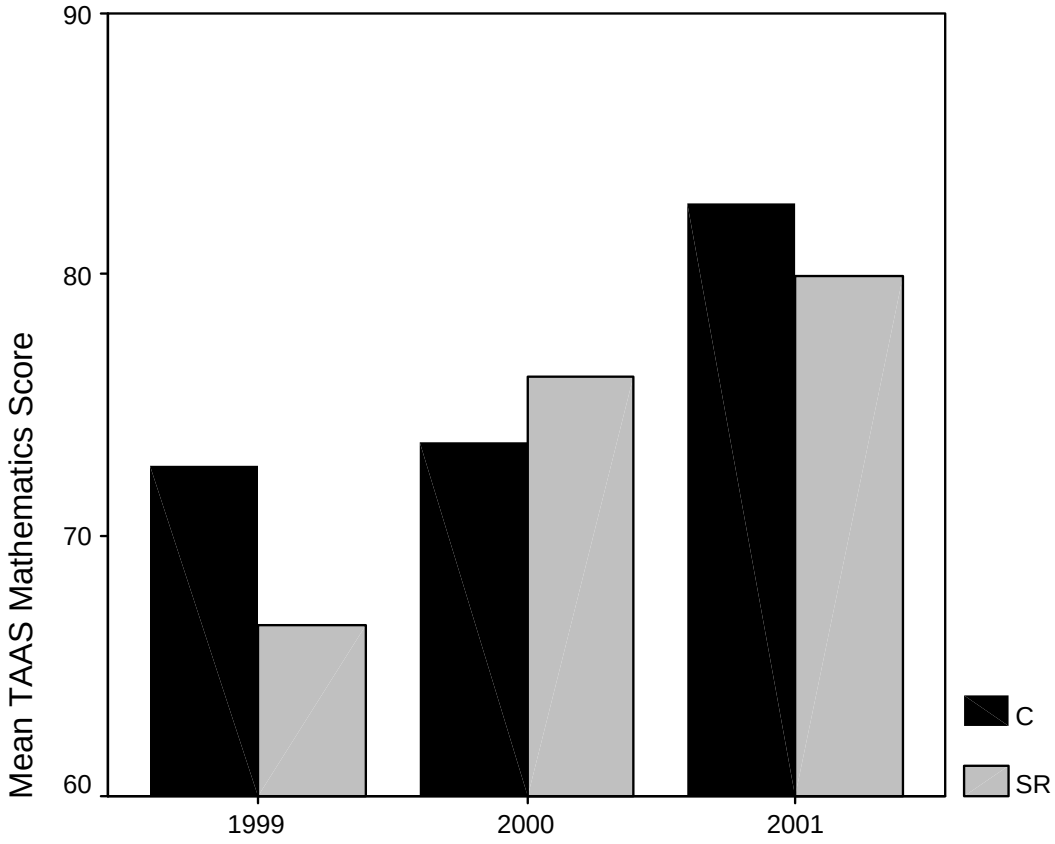

Figure 35. Mean TAAS Mathematics Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

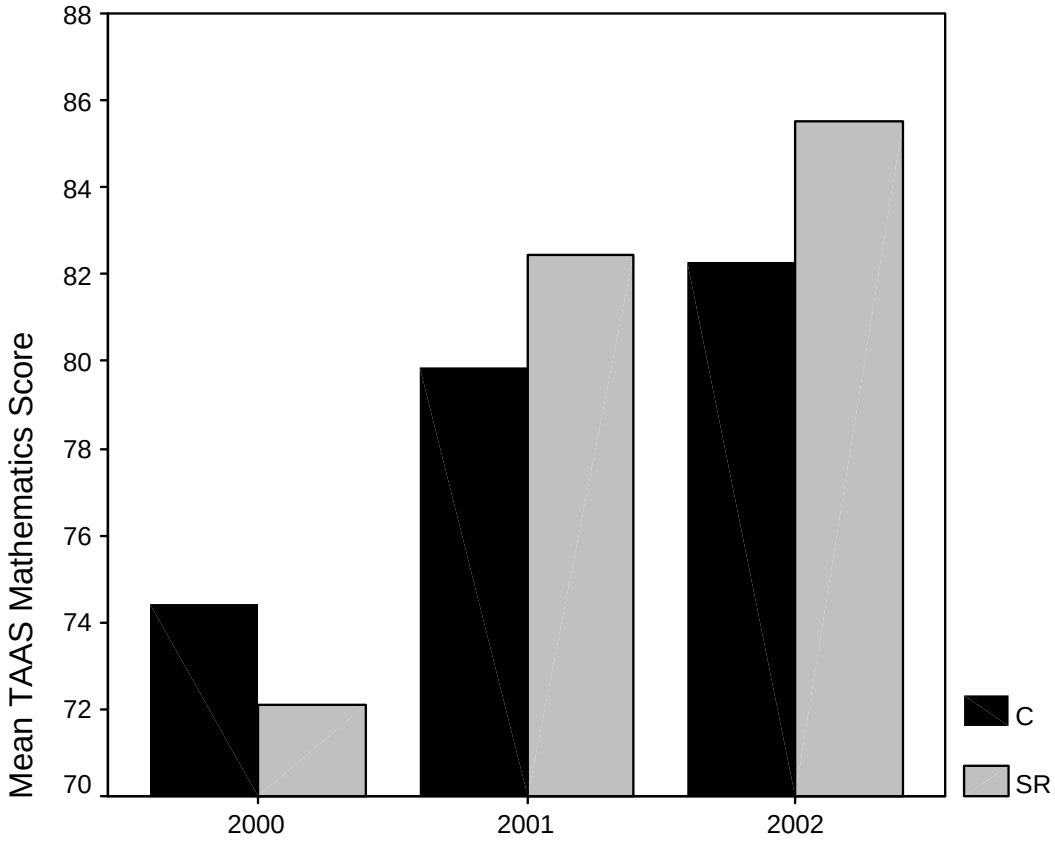

Figure 36. Mean TAAS Mathematics Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

Table 13

Longitudinal Cohort Performance of Student Subgroups by Cohort, Subgroup, Program, and Year: Mean TAAS Reading and Mathematics Scores

	Free or reduced-price lunch				Limited English Proficient				At or Below Grade Level			
Cohort / Subtest	3rd	4th	5th	n	3rd	4th	5th	n	3rd	4th	5th	n
2001 5th Grade Reading												
SR	62.7	78.1	77.6	58	23.4	84.6	79.4	11	24.8	66.2	69.6	49
C	73.0	78.7	80.8	110	39.4	44.6	45.7	10	26.7	66.1	71.2	51
2002 5th Grade Reading												
SR	73.9	82.3	85.0	75	50.6	65.7	70.3	22	38.6	68.4	73.8	44
C	75.6	84.1	83.9	98	58.1	79.7	71.3	15	37.2	62.1	78.8	57
2001 5th Grade Math												
SR	66.5	76.1	79.9	48	0.0	76.4	79.0	7	37.1	70.2	74.3	42
C	72.7	73.6	82.8	110	57.0	47.1	53.2	10	30.0	65.6	67.5	36
2002 5th Grade Math												
SR	72.1	82.5	85.5	58	59.4	77.0	81.1	20	47.0	72.7	74.1	44
C	74.4	80.0	82.3	98	65.4	81.6	68.9	15	45.8	73.0	79.1	57

Limited English Proficient Students

As with students who were eligible for free or reduced-price lunch, LEP students in SR schools were observed to have larger gains in mean TAAS scores for both cohorts and both subject areas examined (see Figures 37-40). Sample sizes were particularly small for these analyses, however, ranging from $n = 7$ (SR 2001 cohort in Mathematics) to $n = 22$ (SR 2002 cohort in Reading). Thus, estimated means could be strongly influenced by a few extreme scores. The largest difference in gains was observed for the 2001 cohort in Mathematics, in which the SR student mean increased from 0.0 in 3rd grade to 79.0 in 5th grade (+79.0; see Table 13), while the C student mean decreased from 57.0 to 53.2 (-3.8). This large extremely large increase in Mathematics scores may be partially attributable to a concomitant rise in reading scores for the SR group from M=23.4 in 3rd grade to M=79.4 in 5th grade—it is possible that the 2001 SR students were unable to successfully complete the 3rd grade math test due to very low English reading achievement. The 2002 cohort also was observed to have large differential gains favoring SR students over C students in both Reading (50.6 to 70.3; + 19.7) and Mathematics (59.4 to 81.1; +21.7; see Table 13).

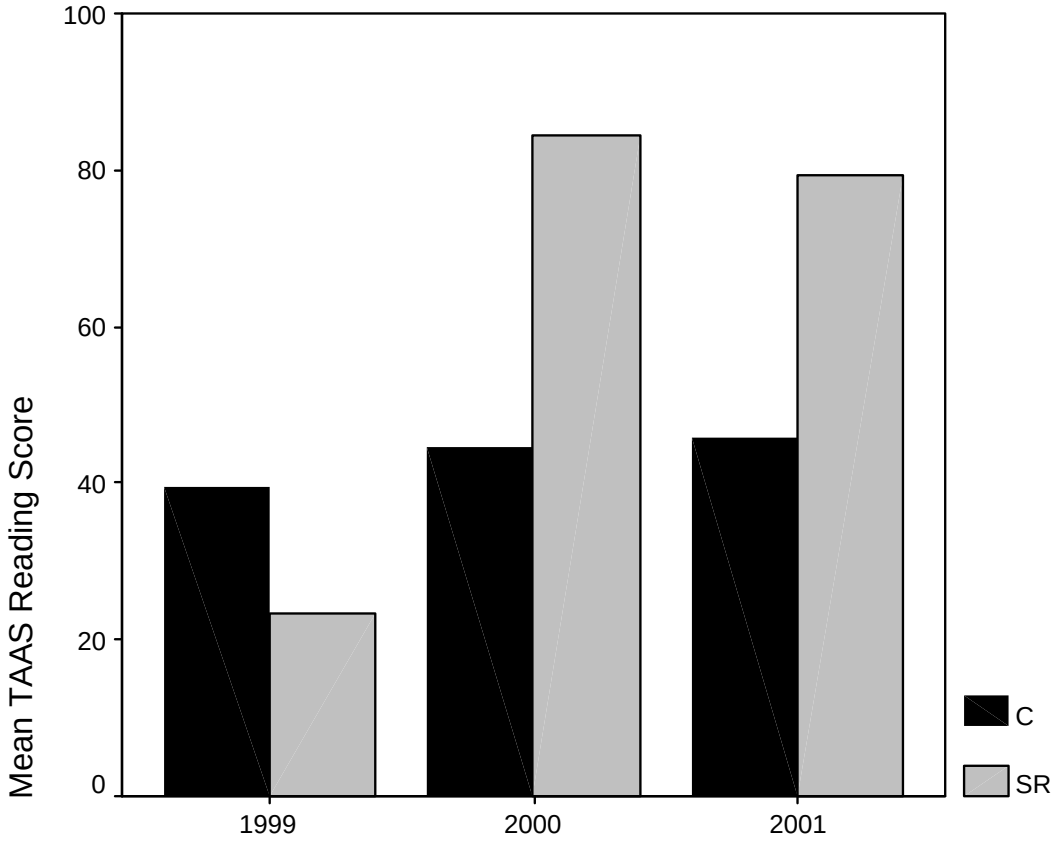

Figure 37. Mean TAAS Reading Scores by Year and Program for Limited English Proficient Students: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

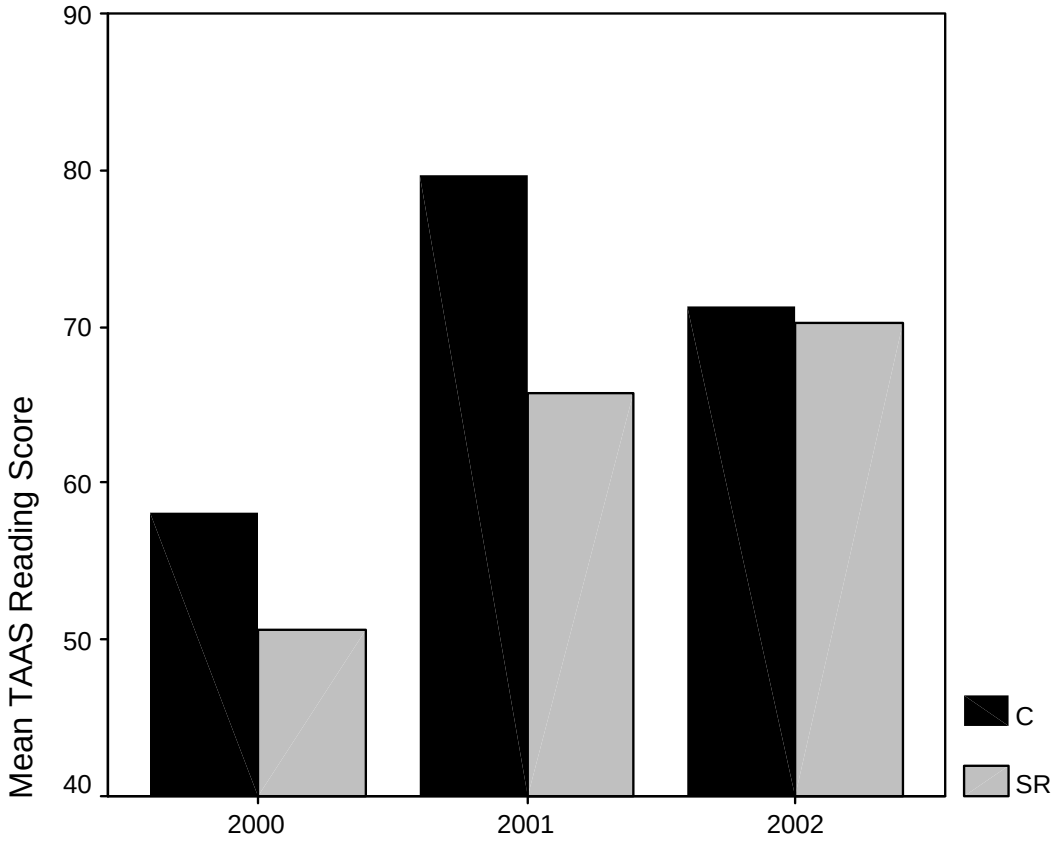

Figure 38. Mean TAAS Reading Scores by Year and Program for Limited English Proficient Students: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

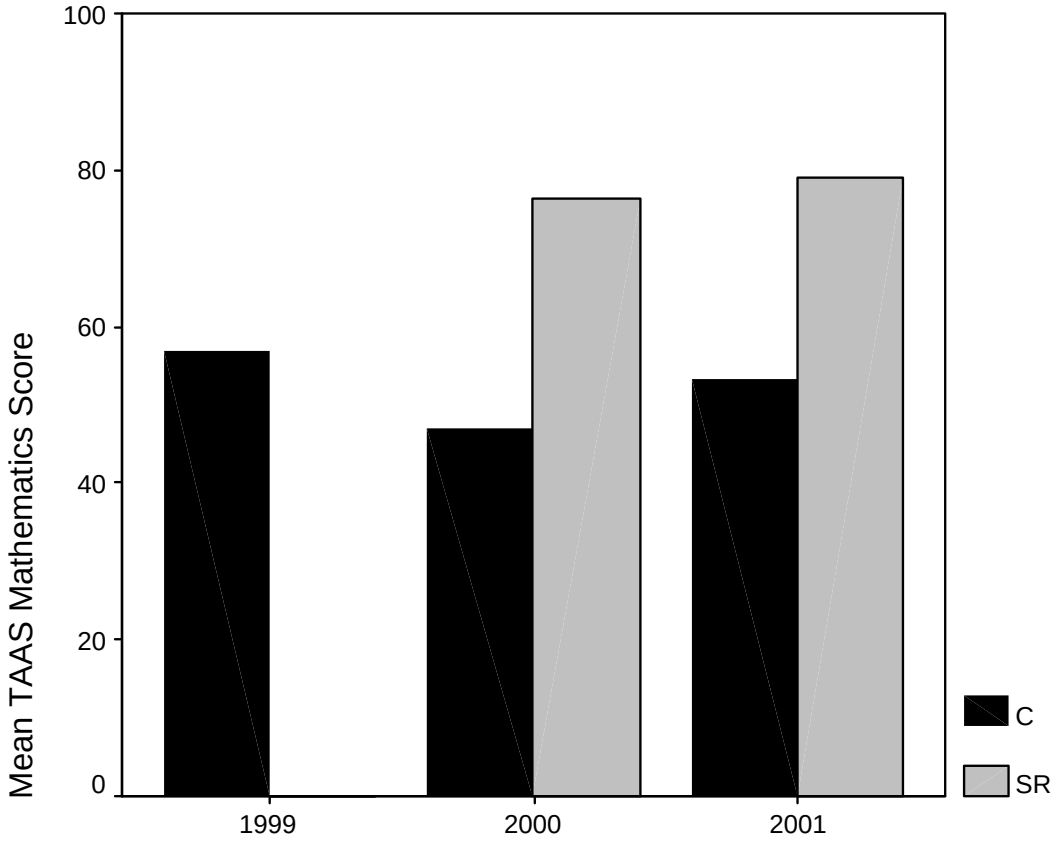

Figure 39. Mean TAAS Mathematics Scores by Year and Program for Limited English Proficient Students: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

Note: 1999 SR Mean = 0.

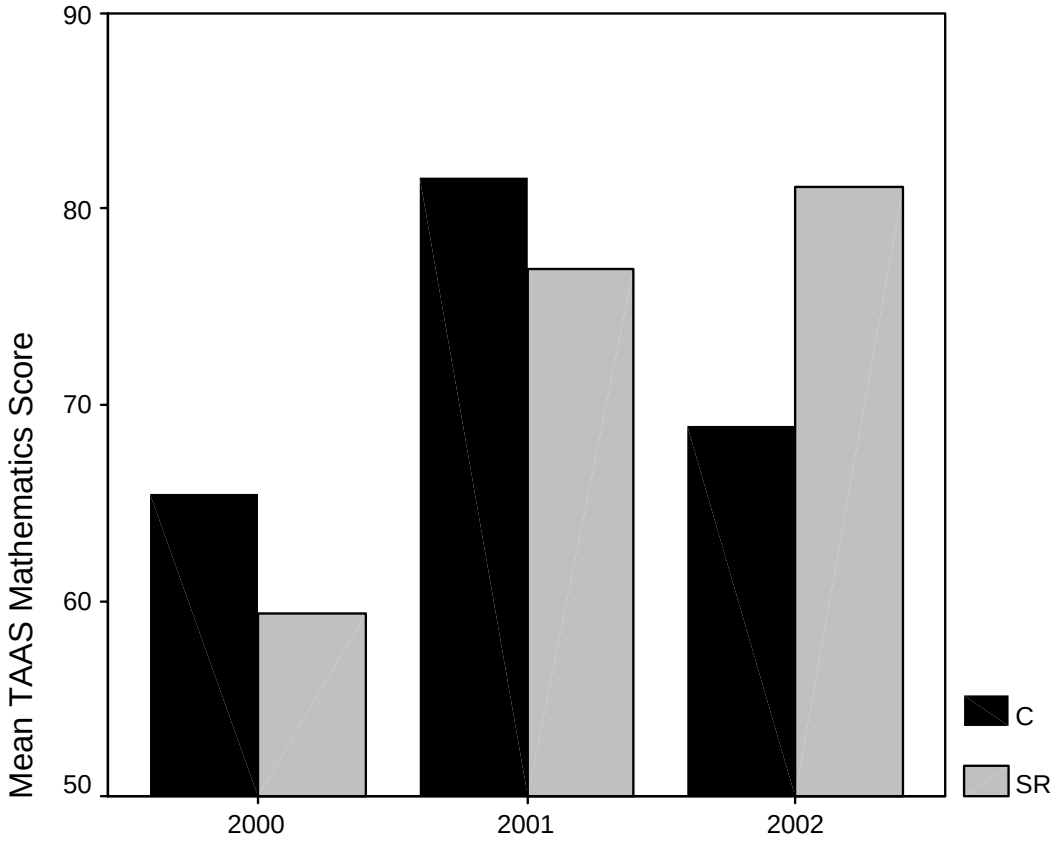

Figure 40. Mean TAAS Mathematics Scores by Year and Program for Limited English Proficient Students: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

Students Scoring At or Below Grade Level in Third Grade

An examination of trends in mean performance for students who scored a 70 or below on the respective TAAS subtest in 3rd grade revealed virtually identical gains for the 2001 cohort in both subject areas, whereas trends for the 2002 cohort slightly favored students in C schools (see Figures 41-45). In both Reading and Mathematics, 2002 5th grade cohort C group students gained about 6 points more than their SR counterparts: from 37.2 to 78.8 versus 38.6 to 73.8 in Reading, and from 45.8 to 79.1 versus 47.0 to 74.1 in Mathematics (see Table 13).

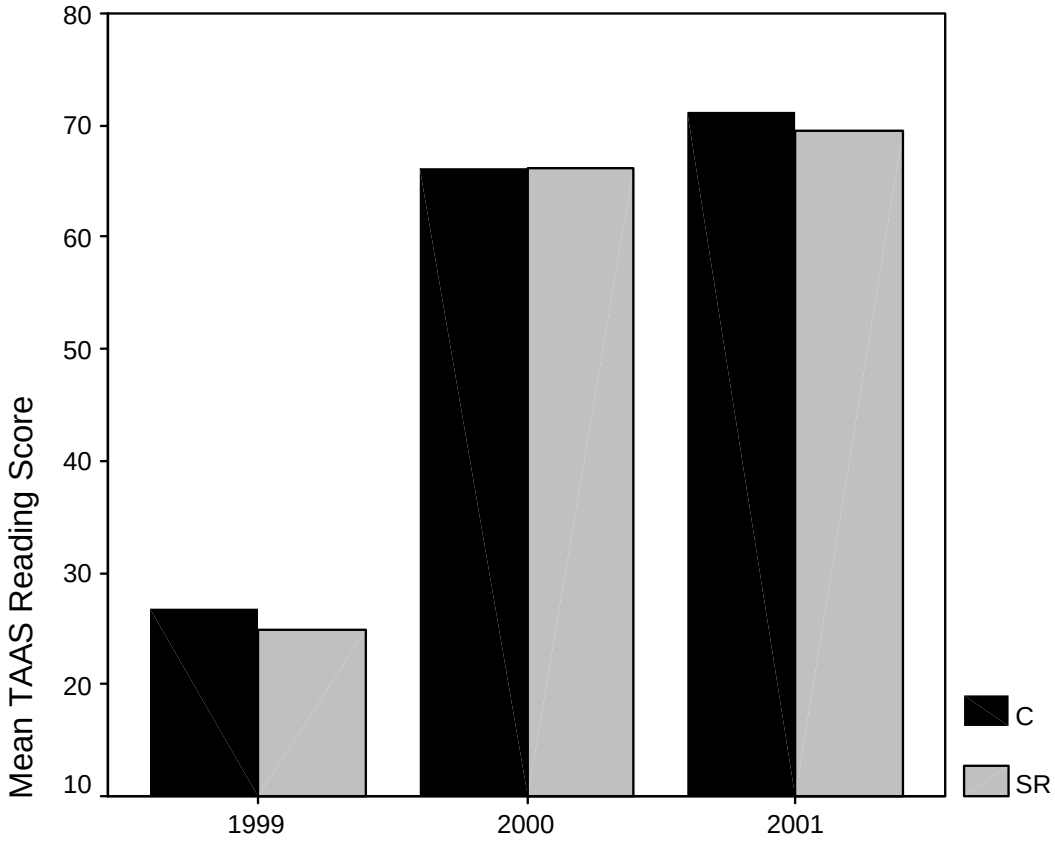

Figure 41. Mean TAAS Reading Scores by Year and Program Students who Scored At or Below Grade Level in 3rd Grade: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

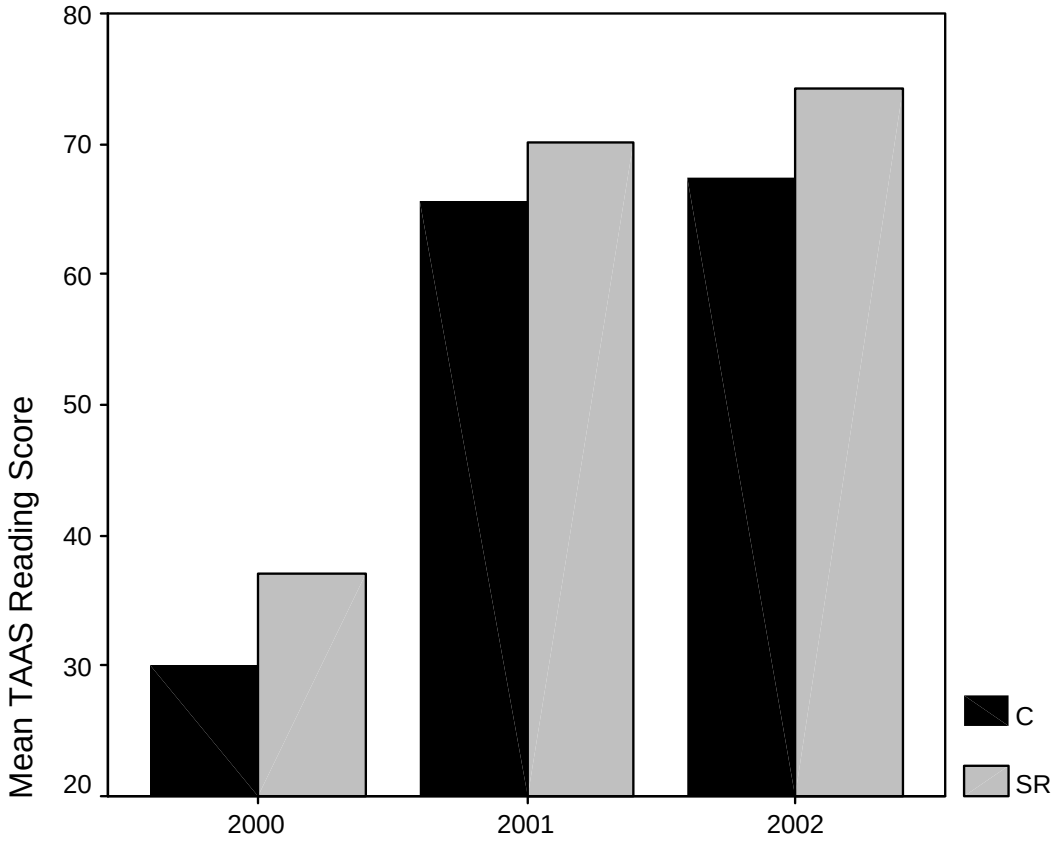

Figure 42. Mean TAAS Reading Scores by Year and Program Students who Scored At or Below Grade Level in 3rd Grade: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

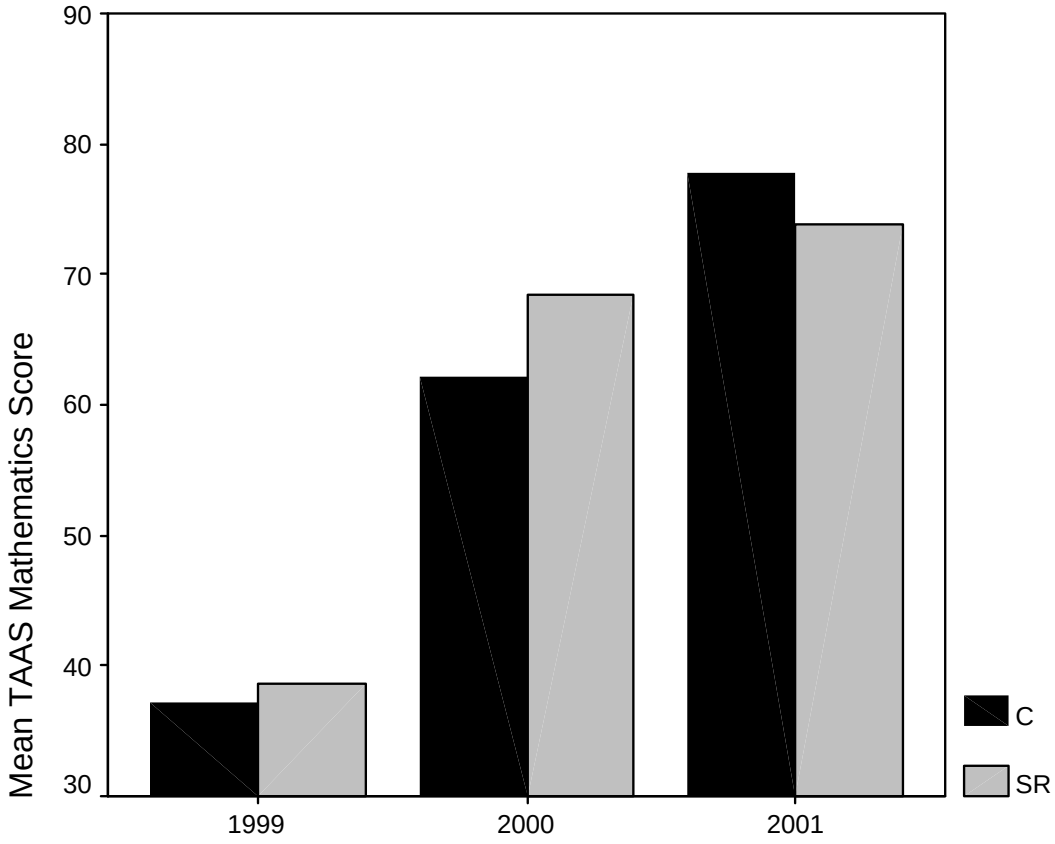

Figure 43. Mean TAAS Mathematics Scores by Year and Program Students who Scored At or Below Grade Level in 3rd Grade: 2001 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

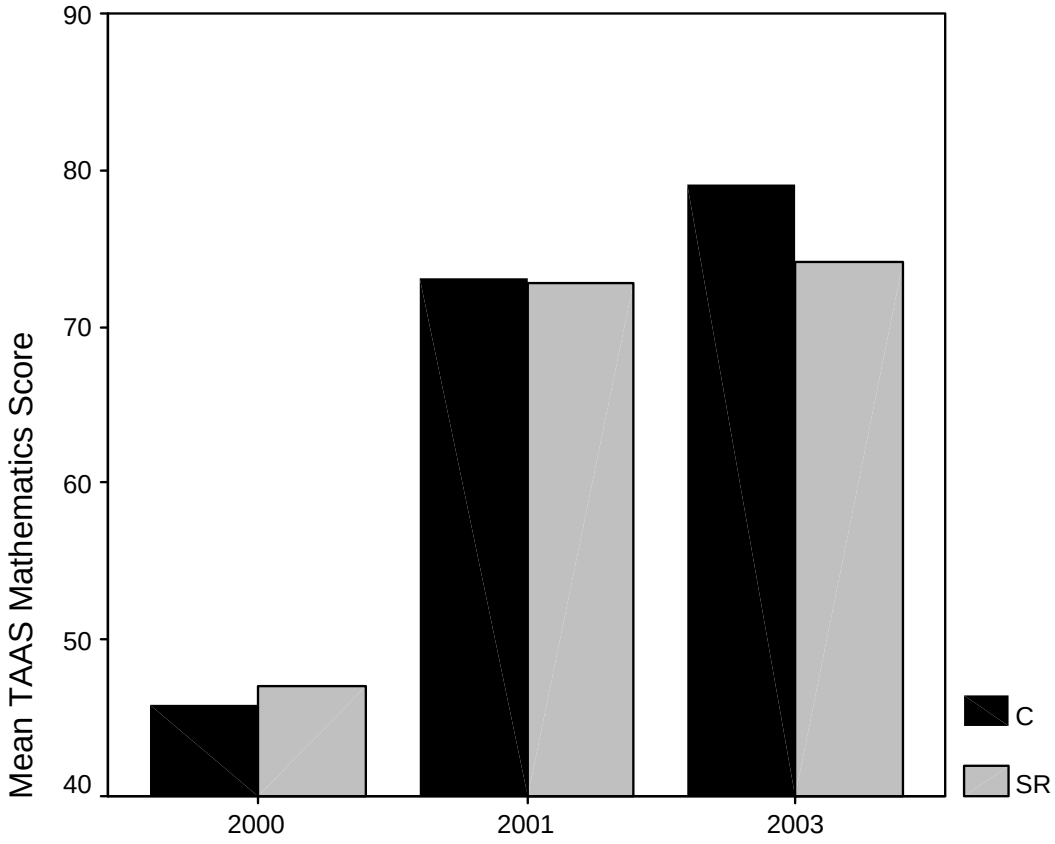

Figure 44. Mean TAAS Mathematics Scores by Year and Program Students who Scored At or Below Grade Level in 3rd Grade: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

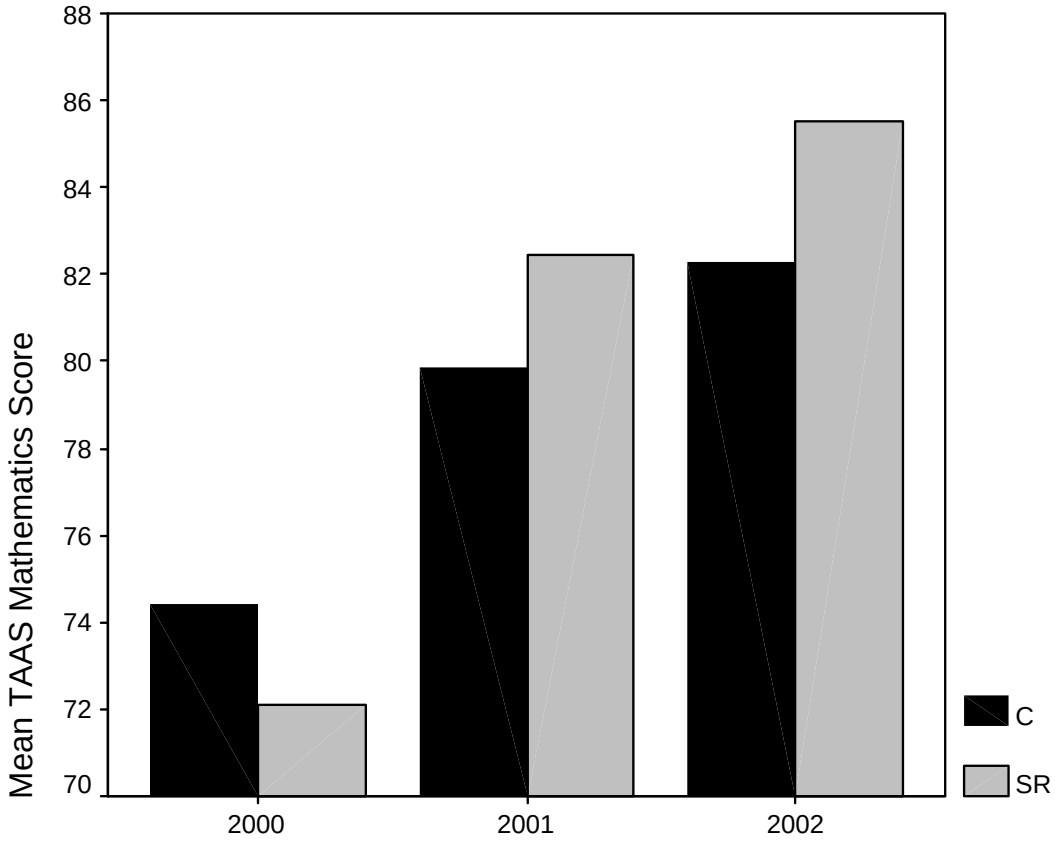

Figure 45. Mean TAAS Mathematics Scores by Year and Program for Students Eligible for Free or Reduced-price Lunch: 2002 Fifth Grade Longitudinal Cohort.
*Year represents spring testing (2001 = 2001-2002 school year).

Implementation Analyses

2002 5th Grade Cohort: Reading

Three consecutive years of matched data were available for a total of 448 participants in the C schools, 218 participants in low-intensity SR schools, and 224 participants in high-intensity SR schools. In the omnibus multivariate test, the treatment by reading achievement (repeated measures) interaction effect was not significant (Wilk's lambda = 0.997, $F_{2,870} = 1.35$, $p = 0.26$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (4th and 5th grade) transformed scores. Program type was found to have a significant main effect ($F_{2,870} = 3.02$, $p = .050$), as were 2000 transformed reading score ($F_{1,870} = 366.68$, $p < .001$) and free or reduced-price lunch status ($F_{1,870} = 4.74$, $p = .03$).

Program type accounted for about 0.7% of the variance in 2001 and 2002 reading scores (partial $\eta^2 = 0.070$), after controlling for free or reduced-price lunch status and 2000 reading scores (see Figure 46). Covariate-adjusted means on the averaged transformed variables were $M = 1.10$ for C group students, $M = 1.14$ for low-intensity SR students, and $M = 1.15$ for high-intensity SR students. Follow-up pairwise comparisons indicated that the averaged transformed mean for the high-intensity SR schools was significantly higher than the C school mean.

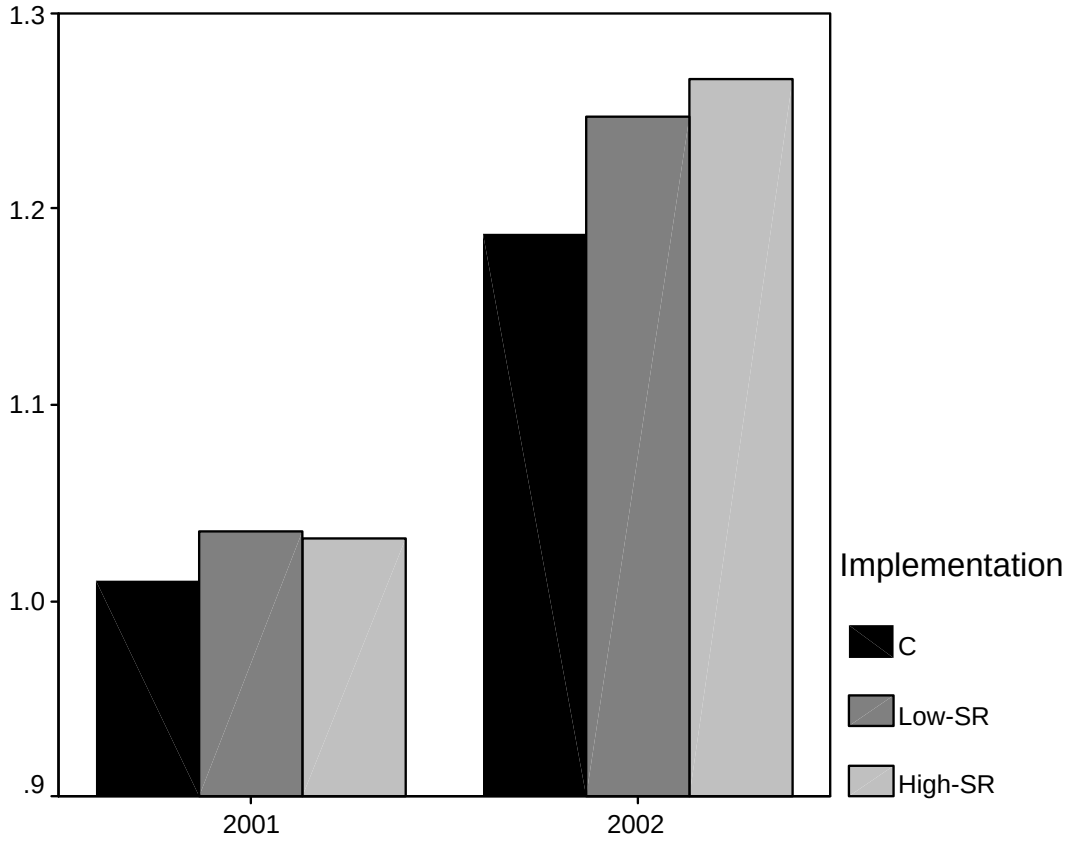

Figure 46. Covariate-adjusted means on transformed TLI Reading score by year and implementation level.
*Year represents spring testing (1999 = 1998-99 school year).

Note: Means for high-intensity School Renaissance implementation schools were significantly higher than comparison school means.

2002 5th Grade Cohort: Mathematics

Three consecutive years of matched data were available for a total of 448 participants in the C schools, 196 participants in low-intensity SR schools, and 221 participants in high-intensity SR schools. In the omnibus multivariate test, the treatment by mathematics achievement (repeated measures) interaction effect was not significant (Wilk's lambda = 1.00, $F_{1,845} = 0.57$, $p = 0.95$), indicating that the effect of treatment was constant across years. Thus, program comparisons were conducted on averaged (4th and 5th grade) transformed scores.

Program type was found to have a significant main effect ($F_{2,845} = 5.47$, $p = .004$), as was 2000 transformed mathematics score ($F_{1,845} = 387.90$, $p < .001$). Free or reduced-price lunch status was not significantly related to math achievement after controlling for prior math achievement ($F_{1,845} = 0.84$, $p = .77$).

Program type accounted for about 1.3% of the variance in 2001 and 2002 math scores (partial $\eta^2 = 0.013$), after controlling for free or reduced-price lunch status and 2000 math scores. Covariate-adjusted means on the averaged transformed variables were $M = 1.24$ for C group students, $M = 1.27$ for low-intensity SR students, and $M = 1.30$ for high-intensity SR students (see Figure 47). Follow-up procedures indicated that the averaged transformed mean for the high-intensity SR schools was significantly higher than the C school mean.

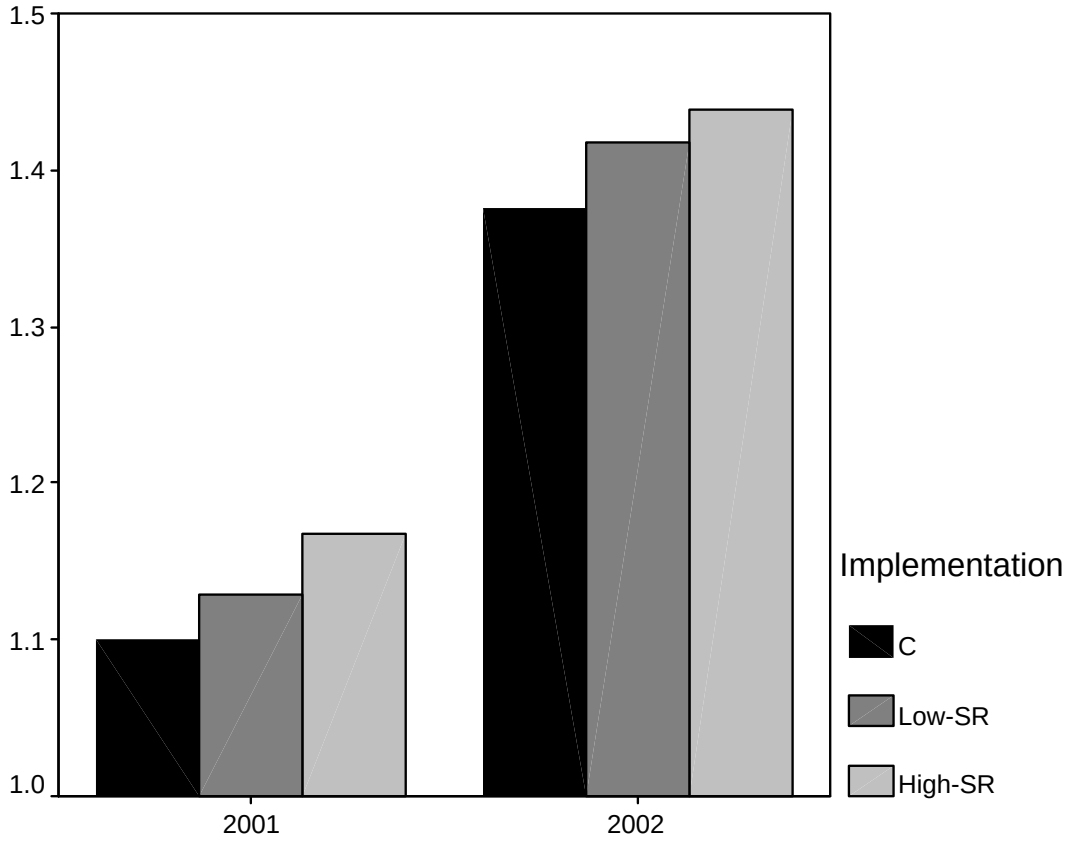

Figure 47. Covariate-adjusted means on transformed TLI Mathematics score by year and implementation level.
*Year represents spring testing (1999 = 1998-99 school year).

Note: Means for high-intensity School Renaissance implementation schools were significantly higher than comparison school means.

Summary and Conclusions

The major findings from the yearly cohort and repeated-measures analyses of reading and mathematics scores are summarized in the sections below.

Reading: Grades 3-5

- For separate cohorts, median effect size estimates indicating SR relative to C outcomes were generally small and stable from 1999 to 2002.
- The percentage of students reading at or above grade level improved substantially at all schools and was nearly equal at SR and C schools.
- Repeated-measures analyses of both 2001 and 2002 fifth-grade cohorts indicated statistically significant differences favoring SR schools after controlling for 3rd grade achievement and socioeconomic status.
- Participation in the SR program accounted for about the same amount of variance in student outcomes as socioeconomic status after controlling for prior achievement.
- Effect size estimates were $d = +0.22$ and $d = +0.17$ for 2001 and 2002 fifth grade cohorts, respectively, indicating that SR students performed about one-fifth of a standard deviation higher than comparison students in matched schools.

Reading: Grades 6-8

- Median effect size estimates indicating SR relative to C outcomes consistently improved from 1999 to 2002.
- The percentage of students reading at or above grade level improved at a greater rate in SR than in C schools. The C schools appeared to have a higher percent of students reading at or above grade level than the SR schools in the baseline year of 1999 (prior to SR implementation). By 2002, these percentages were nearly identical in both SR and C schools.

- Repeated-measures analyses revealed no significant differences between SR and C schools, although results directionally favored SR schools.

Math: Grades 3-5

- Median effect size estimates indicating SR relative to C outcomes for separate cohorts remained stable and near zero from 2000 to 2002.
- Similarly, the percentage of students performing at or above grade level rose at nearly equal rates for SR and C schools.
- Repeated-measures analysis showed a significant program effect favoring SR schools for the 2002 fifth-grade cohort. Program type explained substantially more variance in math outcomes than did socioeconomic status after controlling for prior math achievement.
- Repeated-measures analyses indicated that SR students performed about one-fifth of a standard deviation higher than comparison students (d = +0.20).

Math: Grades 6-8

- Median effect size estimates indicating SR relative to C outcomes improved at all grade levels from 2000 to 2002.
- The percentage of students performing at or above grade level rose at a greater rate in SR than in C schools. The C schools appeared to have a higher percentage of students performing at or above grade level than the SR schools in the baseline year of 2000 (prior to SR implementation). By 2002, these percentages were nearly identical in both SR and C schools.

- Repeated-measures analysis revealed a significant program effect favoring SR schools for the 2002 eighth-grade cohort after controlling for prior math achievement and socioeconomic status.
- Repeated-measures analyses indicated that SR students performed about one-fifth of a standard deviation higher than comparison students (d = +0.17).

Subgroup Performances

- Descriptive analyses showed trends indicating that SR students who were eligible for free or reduced-price lunch demonstrated larger gains than C students in both Reading and Mathematics in the 2001 and 2002 5^{th} grade cohorts.
- Similar trends, showing larger SR than C gains, were indicated for the LEP subgroup.
- Trends for the subgroup that performed at- or below grade level in third grade were comparable for SR and C students in the 2001 cohort. C students in the 2002 cohort, however, demonstrated a slightly higher gain than SR students in both subjects.

Implementation Effects

- Repeated-measures analyses indicated that students in high-implementation SR schools achieved at significantly higher levels in both reading and mathematics than students in similar comparison schools, after controlling for students' prior achievement and socioeconomic status.
- Results in low-implementation SR schools were directionally higher than comparison schools in both reading and mathematics, but the differences were not large enough to attain statistical significance.

Conclusions

The major conclusions from the study are highlighted below. In view of the overall positive results obtained in this study, it is recommended that future research examine school environment variables such as school climate and teacher buy-in to better explain varied program effectiveness across schools and identify the optimum conditions for schools to realize benefits from SR adoption. Conclusions from this study are:

- In the elementary grades, SR schools showed a clear trend of improved performance. Effects of the program were small, but statistically significant, for both reading (2001 and 2002 5th grade longitudinal cohorts) and mathematics (2002 longitudinal cohort). However, program effects were generally comparable to or greater than the effects of socioeconomic status after controlling for prior achievement.
- In the middle school grades, SR schools had noticeably improved performance relative to C schools when viewing separate cohorts in both reading and mathematics. Prior to SR implementation, the SR schools appeared to have a smaller percent of students performing at or above grade level, as compared to the C schools. By 2002, in reading and in math, however, the percentage of students performing at or above grade level was nearly identical for both SR and C schools. However, performance was significantly better only in mathematics in the 2002 8th grade longitudinal cohort.

In a recent meta-analytic study of 29 models, Borman et al. (2002) found an overall effect size of from +0.10 to +0.14, with the range for the most successful category, labeled by the authors as "proven models," being +0.17 to +0.21. Only 3 out of the 29 models achieved this high status (Direct Instruction, School Development Program, and Success For All). Turning to the present research, it is noteworthy that four out of six significant inferential comparisons were

associated with effect sizes for SR ranging from +0.17 to +0.22. Including the two nonsignificant effects (the two eight-grade cohorts in reading), both of which directionally favored SR, would still yield a relatively high overall median ES of +0.17 for the entire study. Compared to the above three "proven models," SR has been made available to schools for only a short time and has had many fewer years to be researched.

Taken as a whole, the present results are clearly suggestive of its benefits for student achievement, and if consistently replicated in future studies would strongly imply proven effectiveness as CSR model**.** Not surprisingly, the present results also imply that the program impacts are significantly more positive when implementation intensity is high. It is noteworthy that the program schools participating in this study were relatively high-performing, with some having 80-90% of their students performing at or above grade level. Thus, the achievement effects occurred in a context where potential gains might have been limited by ceiling effects. Also, descriptive trends indicated higher SR than C gains for longitudinal cohorts who were academically at risk due to being socioeconomically disadvantaged or second language learners, but not for the low-achieving subgroup. Further study is needed with larger samples and more experienced program schools to obtain a clearer impression of aptitude x treatment effects over time.

References

Berenson, M.L., Levine, D.M., & Goldstein, M. (1983). *Intermediate statistical methods and applications*. Englewood Cliffs, NJ: Prentice-Hall.

Borman, G. D. (2002). *Experiments for educational evaluation and improvement*. Unpublished manuscript. University of Wisconsin-Madison.

Borman, G. D., Hewes, G., Rachuba, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis*. Baltimore, MD: Johns Hopkins University, Center for Research on Students Placed at Risk.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, 31(8), 4-14.

Romesburg, H. C. (1990). *Cluster analysis for researchers*. Melbourne, FL: Krieger Publishing.

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

Shavelson, R. J., & Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15-21.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn & Bacon.

U.S. Congress (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.

U.S. Department of Education (2002). *Strategic plan, 2002-2007*. Washington, DC: Author.

APPENDIX A: Summary of Student Sample Sizes

Table A.1

Number of students by year and matched pair: Third grade.

	1998		1999		2000		2001		2002	
Pair	**C**	**SR**	**C**	**SR**	**C**	**SR**	**C**	**SR**	**C**	**SR**
1	68	64	63	46	76	83	59	56	54	40
2	41	78	34	67	34	90	44	48	38	47
3	83		84	67	62	73	53	86	55	129
4	106	65	111	65	91	86	79	47	79	37
5	31	114	37	120	30	133	31	104	30	87
6	92	87	104	121	103	136	99	131	85	90
7	110	91	129	117	97	101	79	102	93	107
8	93	101	104	134	102	124	81	116	80	95
9	78	57	83	64	79	64	54	54	52	46

Note: C = Comparison school. SR = School Renaissance school.

Note: The Pair 3 SR school did not open until the 1998-99 school year.

Table A.2

Number of students by year and matched pair: Fourth grade.

	1998		1999		2000		2001		2002	
Pair	**C**	**SR**	**C**	**SR**	**C**	**SR**	**C**	**SR**	**C**	**SR**
1	81	65	55	65	59	44	56	61	42	50
2	43	86	38	69	37	68	44	63	44	42
3	75		80	27	83	131	55	83	55	123
4	124	68	94	66	103	67	87	44	79	34
5	38	135	31	112	39	132	31	109	36	107
6	94	91	92	113	94	132	104	117	104	103
7	131	89	113	98	103	122	85	82	90	106
8	120	111	94	114	106	134	103	109	77	88
9	78	68	78	60	70	66	60	56	59	58

Note: C = Comparison school. SR = School Renaissance school.

Note: The Pair 3 SR school did not open until the 1998-99 school year.

Table A.3

Number of students by year and matched pair: Fifth grade.

	1998		1999		2000		2001		2002	
Pair	**C**	**SR**	**C**	**SR**	**C**	**SR**	**C**	**SR**	**C**	**SR**
1	59	62	68	71	52	62	46	57	58	51
2	30	70	42	83	41	61	38	64	44	55
3	87		70	55	74	84	64	71	54	109
4	100	70	117	73	81	74	94	46	89	33
5	46	120	37	106	28	117	38	106	30	105
6	94	73	104	117	92	109	103	110	109	86
7	126	70	115	90	107	110	91	89	94	88
8	101	100	125	120	97	120	107	122	99	91
9	80	77	90	63	60	58	54	68	52	52

Note: C = Comparison school. SR = School Renaissance school.

Note: The Pair 3 SR school did not open until the 1998-99 school year.

Table A.4

Number of students by year and matched pair: Sixth, seventh, and eighth grades.

6th Grade										
	1998		1999		2000		2001		2002	
Pair	C	SR	C	SR	C	SR	C	SR	C	SR
10	393	373	415	370	435	463	443	321	419	375
11	257	307	279	367	299	363	291	272	249	302
7th Grade										
	1998		1999		2000		2001		2002	
Pair	C	R	C	R	C	R	C	R	C	R
10	383	331	406	413	437	410	428	348	443	315
11	255	308	234	343	295	358	307	293	187	268
8th Grade										
	1998		1999		2000		2001		2002	
Pair	C	R	C	R	C	R	C	R	C	R
10	416	335	393	353	423	425	467	323	476	352
11	281	271	277	336	249	356	144	265	235	332

Note: C = Comparison school. SR = School Renaissance school.