CREP
Center for Research in Educational Policy

Center for Research in Educational Policy

The University of Memphis
325 Browning Hall
Memphis, Tennessee 38152
Toll Free: 1-866-670-6147

The Effect of School Renaissance on Student Achievement in Two Mississippi School Districts

Center for Research in Educational Policy

The University of Memphis
325 Browning Hall
Memphis, Tennessee 38152
Toll Free: 1-866-670-6147

THE EFFECT OF SCHOOL RENAISSANCE ON STUDENT ACHIEVEMENT IN TWO MISSISSIPPI SCHOOL DISTRICTS

January 2005

Steven M. Ross
John A. Nunnery

EXECUTIVE SUMMARY

Introduction

The purpose of the present study was to compare the achievement of students in 14 Pascagoula schools and 9 Biloxi schools implementing School Renaissance (SR) to those of students in matched Control schools. The achievement measures to be examined were from the 2003 administration of the Mississippi Curriculum Test (MCT) for Reading, Language Arts, and Mathematics in grades 2-8 and the Mississippi Writing Assessment (WA) in grade 4 and 7. The Pascagoula schools began implementing SR in January, 2002, and the Biloxi schools began implementation in August, 2000. Thus, the present analyses of 2003 achievement data relates to the first and third full years of SR in Pascagoula and Biloxi, respectively. Given that positive school climate is often a precursor to enacting successful school reform (Ross, 2003; Rowan, Camburn, & Barners, 2004), a supplementary part of the study examined school climate in SR relative to Control schools.

As background for the present, quantitative analyses, a qualitative study of the same schools was conducted by Sullivan and Harper (2004) from the North Mississippi Educational Consortium. Results indicated that program selection and implementation in both districts were based on (a) initial interest by several staff members at individual schools, creating broader commitment in the school and district; and (b) the desire by schools to improve test scores. Perceptions regarding implementation were that:

- Implementation of SR was generally easy.
- Positive aspects included the diagnostic component (STAR testing) and the ease with which students were assigned to their appropriate reading level.
- Student involvement and motivation were generally high.
- Administrator and teacher buy-in was strong.
- Challenges occurred in purchasing the required number of books, finding time to train teachers, and becoming accustomed to the management system used by the program.
- Test scores improved.

Methodology

Participants and Design

A matched program-control school quasi-experimental design was employed for the current study. Participants included 10,160 students in grades 3 through 8 attending one of 40 schools in 15 school districts in southern Mississippi. The majority (60.1%) of the students were eligible for free or reduced-price lunch, nearly equal proportions were male (50.9%) or female (49.9%), and 43.7% were minority students (40.0% of the total were African American). A matched Control school was selected for each SR school on the basis of student ethnicity, poverty, mobility, school location, grades served, size, prior achievement, and no or very limited usage of Accelerated Reader or Accelerated Math.

Measures

The School Climate Inventory (SCI) (Butler & Alberg, 1991) was administered at each SR and matched Control school. The Mississippi Curriculum Test (MCT) subtests in Reading, Language Arts, and Mathematics, and the Mississippi Writing Assessment (WA), were used to measure student achievement.

Analyses

Student records were matched by student identification number. Pretest-posttest match rates were computed for each grade, test, and treatment group. Match rates ranged from a low of 51.3% for sixth grade comparison students on the Mathematics test, to a high of 83.0% for eighth-grade Control students on the Mathematics test. Analyses of SR vs. Control school differences for students without pretest scores (unmatched) showed significant SR advantages, and thus a possible negative bias in analyzing posttest differences, on Reading, Language, and Mathematics in grades 7 and 8, and on Writing in grade 7.

Program effects were tested across grade levels by performing mixed-model analyses of covariance, with treatment (SR vs. Control) and grade level (3 to 8) serving as fixed effects, and matched school set serving as a random effect. Outcome variables were 2003 MCT scale scores on the Reading, Language Arts, and Mathematics subtests, with respective 2002 scores serving as covariates. Mississippi WA scores were used as outcome variables for 4^{th} and 7^{th} grades, with 2002 MCT Reading scores serving as a covariate.

Results

School Climate

School climate data were analyzed via a one-way MANOVA comparing SR and Control teachers on the seven climate dimensions. Results significantly favored SR schools on all dimensions, with effect sizes ranging from +0.20 (Involvement) to +0.54 (Collaboration).

Student Achievement

A summary of the major results is as follows:

- For both Reading and Language, SR students scored significantly higher than comparison students in 2003 after controlling for 2002 MCT scores. Program effects,

while small, were generally consistent and positive across grades 3 to 8. Median effect size estimates across grades were +0.11 for Reading and +0.12 for Language.

- In Writing, fourth grade SR students scored significantly and substantially higher than comparison students, with an effect size estimate of +0.45. Nearly double the percentage of fourth grade SR students (44.5%) scored 3 or 4 (on a 0 to 4 rubric) on the Writing Assessment relative to the comparison group (25.4%). No significant effects were observed for seventh-grade Writing.
- Mathematics outcomes were significantly and substantially higher for SR students in grades 6 and 7, with effect size estimates of +0.22 and +0.27, respectively. No differences were observed in other grades.
- Achievement effects were stronger for SR schools in Biloxi (three-year implementation) than in Pascagoula (only one year).

Conclusions

Taken as a whole, the present results suggest that SR is having positive effects on the participating schools. Drawing from Sullivan and Harper's (2004) qualitative study performed with the same schools, it appears that SR has clear teacher, administrator, and district support. An explicit theoretical assumption and goal of Comprehensive School Reform (CSR) is that improvements in student learning result from establishing a positive school culture in which teachers and school leaders work together to promote high expectations, pride in school environment, effective professional development, and implementation of research-based teaching practices (Desimone, 2002; Rowan et al., 2004; Ross & Gil, 2004). Although in the absence of a randomized experimental design, advantages found for SR schools over Control schools in school climate cannot be directly attributed to program impacts, they are nevertheless suggestive

that adoption of SR at least supports, if not directly engenders, the establishment of school climates that promote collegiality, improved teaching, and ultimately, higher student achievement.

It is becoming increasingly recognized in CSR evaluations that effect sizes in the +0.10 – +0.20 range are educationally meaningful and impactful in raising student achievement (Borman et al., 2003). Thus, it is encouraging to SR that in the present analyses, effect sizes were positive in 16 out of the 20 (80%) grade level analyses performed and exceeded +0.10 in 13 out of the 20 (65%) analyses. The evidence thus clearly supports continued implementation of SR at these schools as well as continued study of sustainability and cumulative impacts of program usage across grades.

The Effect of School Renaissance on Student Achievement in Two Mississippi School Districts

Overview

The purpose of the present study was to compare the achievement of students in 14 Pascagoula schools and 9 Biloxi schools implementing School Renaissance (SR) to those of students in matched Control schools. The achievement measures to be examined were from the 2003 administration of the Mississippi Curriculum Test (MCT) for Reading, Language Arts, and Mathematics in grades 2-8 and the Mississippi Writing Assessment (WA) in grade 4 and 7. The Pascagoula schools began implementing SR in January, 2002. The Biloxi schools began implementation in August, 2000. Thus, the present analyses of 2003 achievement data relates to the first and third full years of SR in Pascagoula and Biloxi, respectively.

Given that positive school climate is often a precursor to enacting successful school reform (Ross, 2003; Rowan, Camburn, & Barners, 2004), a supplementary qualitative or descriptive study examined school climate in SR relative to the same control schools. A reasonable hypothesis is that an effective Comprehensive School Reform (CSR) model should create effective climate conditions so that faculty, administrators, and parents are working collaboratively to raise expectations, improve teaching methods, and improve student achievement.

Program Description

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 (RR) 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 1992. 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.

Descriptive Study

In conjunction with the present, quantitative study, Sullivan and Harper (2004) conducted a descriptive or qualitative study, using interviews and site visits, to determine (a) most of the same SR and Control schools' program selection and implementation procedures (e.g., two schools declined participation), and (b) school administrators' perceptions of the strengths and weaknesses of their schools' reading and math programs, and (c) the administrators' perceptions of the impact of these programs on students' motivation and learning. As background for the present study, we summarize below key findings from the Sullivan and Harper (2004) technical report.

Findings indicated that six Biloxi SR schools and five Pascagoula SR schools were implementing both the reading and the math components of the School Renaissance (SR) program. The remaining ten SR schools were implementing only the reading component. Seven

Biloxi schools and seven Pascagoula schools reported full implementation of the SR reading component. Math implementations were generally in beginning or intermediate stages of implementation.

Program selection and implementation in both districts was based on (a) initial interest by several staff members at individual schools, creating broader commitment in the school and district; and (b) the desire by schools to improve test scores. Perceptions regarding implementation were that:

- Implementation of SR was generally easy.
- Positive aspects included the diagnostic component (STAR testing) and the ease with which students were assigned to their appropriate reading level.
- Student involvement and motivation were generally high.
- Administrator and teacher buy-in were strong.
- Challenges occurred in purchasing the required number of books, finding time to train teachers, and becoming accustomed to the management system used by the program.
- Test scores improved.

As described by Sullivan and Harper (2004), the most salient recurring theme in both SR and Control schools was the importance of the Mississippi Frameworks/Benchmarks and the Mississippi Curriculum Test (MCT). Respondents indicated that the main features desired in a curriculum were compatibility with the Frameworks and its ability to help students succeed on the MCT. SR respondents viewed the program as positively impacting teachers and students. However, they also expressed concerns about time or cost demands regarding teacher training, classroom time, and purchase of materials. Control teachers also expressed satisfaction with the academic effectiveness of their reading and math curricula, particularly the curricula's perceived

correlation with the State Frameworks/Benchmarks and effectiveness in promoting achievement on the MCT. On the other hand, compared to the SR group, Control school respondents expressed less satisfaction with the motivational effectiveness of their reading and math curricula and more criticism of the strong emphasis on the state Frameworks and state testing than did the intervention schools.

Methodology

Participants and Design

A matched program-control school quasi-experimental design was employed for the current study. Participants included 10,160 students in grades 3 through 8 attending one of 41 schools in 15 school districts in southern Mississippi. The majority (60.1%) of the students were eligible for free or reduced-price lunch, nearly equal proportions were male (50.9%) or female (49.9%), and 43.7% were minority students (40.0% of the total were African American). A listing of matched School Renaissance (SR) and Control school pairs is provided along with school demographic data for elementary and middle/high school samples in Table 1. In all, there were 23 SR schools and 18 comparison schools. Grade level configurations of treatment schools were seven K-6, eleven K-5, three 6-8, one 8-9, and one 7th grade only (see Table 1). Of the 18 comparison schools, three were PK-3, four were K-6, one was K-8, six were K-12, and four were middle schools (4-8).

Table 1.

School demographics for SR and matched Control schools.[1]

School	Grade Lvl (Per CCD)	Percentage Free or Reduced Lunch	Percentage Minority	Grade 3 %Prof&Adv Reading	Grade 5 %Prof&Adv Reading	Student/ Teacher Ratio
Rawls Springs Att. Cntr	K-6	*64.02*	*48.17*	*90.0*	*77.7*	*12.7*
Popp's Ferry Elem	K-6	63.39	50.30	94.9	90.7	16.1
Dixie Elem School	K-8	64.18	24.55	71.7	59.1	14.2
North Bay Elem	K-6	63.40	28.11	94.6	97.2	17.1
Wheeler School (K-12)	K-12	63.64	11.22	86.1	56.3	13.2
Lopez Elem	K-6	63.40	46.39	96.6	97.1	17.9
Dixie Elem School	K-6	64.18	24.55	71.7	59.1	14.2
Jeff Davis Elem	K-6	63.46	42.12	92.5	91.7	14.6
Columbia Primary School	PK-3	*67.31*	*49.84*	*91.3*	*NA*	*15.2*
Gorenflo Elem	K-6	63.31	66.91	89.4	95.6	12.3
Scott Central Attendance Center	K-12	72.60	50.55	55.0	35.6	18.4
DuKate Elem	K-6	63.24	81.61	77.4	96.0	11.3
Wheeler School (K-12)	K-12	63.64	11.22	86.1	56.3	13.2
Beauvoir Elem	K-6	63.40	34.53	93.3	96.4	13.6
Morton Elementary School	PK-3	*72.67*	*50.16*	*83.8*	*NA*	*16.5*
Singing River	K-5	60.79	56.85	81.7	91.2	19.4
Perkinston Elementary School	K-6	*56.05*	*20.97*	*95.2*	*81.4*	*16.5*
Martin Bluff Elem	K-5	60.77	23.60	70.7	95.0	18.5
North Pike Elementary School	K-6	*57.30*	*31.32*	*88.3*	*NA*	*19.6*
Lake Elem	K-5	60.74	43.70	93.8	84.6	14.8
North Pike Elementary School	K-6	*57.30*	*31.32*	*88.3*	*NA*	*19.6*
Jackson Elem	K-5	60.62	67.69	83.3	69.4	18.3
Kosciusko Upper Elementary Sch.	K-6	*59.88*	*55.02*	*NA*	*90.9*	*15.8*
Gautier Elem	K-5	60.80	56.81	78.4	87.2	16.9
Second Street Elementary Sch.	4-8/7-8	*64.99*	*24.44*	*NA*	*92.0*	*17.1*
Eastlawn	K-5	60.84	34.27	87.6	92.6	17.4
Thrasher High School	K-12	63.55	20.05	71.5	37.9	12.5
College Park	K-5	60.76	48.94	82.5	90.6	16.1
Lake Attendance Cntr	K-12	*72.77*	*50.00*	*90.5*	*NA*	*18.1*
Cherokee Elem.	K-5	60.75	50.81	88.0	86.1	15.2
Edinburg Attendance Center	K-12	72.84	13.17	57.1	70.6	15.0
Central Elem.	K-5	65.64	72.84	64.5	81.4	13.2
Walnut Attendance Center	K-12	65.21	17.72	71.4	68.2	14.9
Beach Elem.	K-5	61.21	30.17	100.0	100.0	14.5
North Bay Elem. School	PK-3	63.40	28.11	94.6	97.2	17.1
Arlington Heights	K-5	660.63	68.31	82.1	90.9	19.1

Table 1 (continued).

School demographics for SR and matched Control schools.[1]

School	Grade Lvl (Per CCD)	Percentage Free or Reduced Lunch	Percentage Minority	Grade 7 %Prof&Adv Reading	Grade 8 %Prof&Adv Reading	Student/ Teacher Ratio
Byram Middle School	4-8/7-8	54.02	36.85	83.4	62.1	18.7
Biloxi Jr. High	8,9	56.14	56.77	NA	71.0	NA
Middle School of Poplarville	4-8/7-8	60.12	11.90	80.2	64.9	17.2
Michel 7th Grade	7	61.94	44.64	86.4	NA	NA
Walnut Attendance Center	K-12	65.21	17.72	71.4	68.2	14.9
Trent Lott Middle	6,7,8	60.83	36.28	80.6	77.0	17.3
Solomon Jr. High School	4-8/7-8	89.27	97.13	40.2	35.4	14.0
Colmer Middle	6,7,8	60.70	49.81	72.2	60.3	15.9
Byram Middle School	4-8/7-8	54.02	36.85	83.4	62.1	18.7
Gautier Middle	6,7,8	60.80	44.84	68.3	62.8	14.8

[1]The Control school is listed first in each shaded or non-shaded pair.

A matched Control school was selected for each SR school using the following criteria:

- the percent of African American students enrolled for 2000-02
- the percent of White students enrolled for 2000-02
- the percent of economically-disadvantaged students enrolled for 2000-02
- the percent of mobile students as determined from 2000-02 cumulative attendance
- School location (rural, suburban, small city)
- Grades served
- School size
- School means on 2001-02 MCT in Reading and Mathematics
- No or very limited usage of Accelerated Reader or Accelerated Math

Candidate control schools were identified through an examination of state-wide school data on the above variables. Following compilation of an initial listing of the strongest matches, the district leaders (superintendent or designee) associated with each candidate schools were first contacted by the researchers to determine interest and gain support. The leaders were told that

the primary need for the study was to gain access to test scores on the state assessments; thus, Control schools would not be required to participate in any active way except to: (a) give permission for the researchers to obtain test scores from the MS DOE, (b) have teachers complete the School Climate Inventory (SCI) (Butler & Alberg, 1991) at a regularly scheduled faculty meeting in the spring, 2004 (see below), and (c) have the principal or literacy leader participate in a one-hour interview regarding literacy practices. The latter interview was conducted as part of the qualitative study by Sullivan and Harper (2004), as summarized above. In return for their cooperation, the Control schools received SCI data reports (valued at $350 each) and $500 in cash. Altogether, two out of the original 15 district superintendents contacted declined. Those Control schools were replaced by the next best available matches from other districts. All individual Control schools on the list agreed, through their principals, to participate.

Measures

School Climate Inventory (SCI©). The main purpose of the School Climate Inventory (SCI) is to assess teacher perceptions of impacts of reform initiatives in relation to seven dimensions logically and empirically linked with factors associated with effective school organizational climates (Butler & Alberg, 1991). The inventory contains 49 items, with 7 items comprising each scale. Responses are scored through the use of Likert-type ratings ranging from *strong disagreement* (1) to *strong agreement* (5). Each scale yields scores ranging from 1 to 5, with higher scores being more positive. Additional items solicit basic demographic information.

Face validity of the school climate items and logical ordering of the items by dimensions were established during the development of the inventory (Butler & Alberg, 1991). Subsequent analysis of responses collected through administration of the inventory in a variety of school

sites, including a concurrent validity study by Sterbinsky (2001), substantiated validity of the items and scales. Dimension descriptions and current internal reliability coefficients on the seven dimensions of the inventory, obtained using Cronbach's alpha, are as follows:

Scale	Internal Reliability	Description
Order	α=.8394	The extent to which the environment is ordered and appropriate student behaviors are present
Leadership	α=.8345	The extent to which the administration provides instructional leadership
Environment	α=.8094	The extent to which positive learning environments exist
Involvement	α=.7582	The extent to which parents and the community are involved in the school
Instruction	α=.7453	The extent to which the instructional program is well developed and implemented
Expectations	α=.7275	The extent to which students are expected to learn and be responsible
Collaboration	α=.7417	The extent to which the administration, faculty, and students cooperate and participate in problem solving

Student achievement. Mississippi Curriculum Test (MCT) subtests in Reading, Language Arts, and Mathematics, and the Mississippi Writing Assessment (WA), were used to measure student achievement. Three forms of the MCT are used in each subject area in grades 2 through 8. The Mississippi State Department of Education reported Cronbach's alpha reliability coefficients across forms and grade levels ranging between 0.88 and 0.90 for Reading, 0.87 and 0.91 for Language, and 0.85 to 0.90 for Mathematics in 2001 (MSDOE, 2002). The MCT provides vertically-equated scale scores that use the same metric for student performance across grade levels, allowing for comparison of gains across grade levels and tracking of individual growth patterns (Tomkowicz & Schaeffer, 2002). Content validity of the MCT's was addressed by statewide teacher committees, who formed consensus about what specific skills and objectives were taught in particular subjects and grade levels (MSDOE, 2002). The WA is

administered in grades 4 and 7, and yields an integer score between 0 and 4. Rubrics for scoring each writing assessment are provided in the appendix.

Procedure

Student records from participating schools for 2002 MCT, 2003 MCT, and 2003 WA were provided by the Mississippi State Department of Education. Records were matched by student identification number. Pretest-posttest match rates were computed for each grade, test, and treatment group. Match rates ranged from a low of 51.3% for sixth grade comparison students on the Mathematics test, to a high of 83.0% for eighth-grade comparison students on the Mathematics test (see Table 2). Given low match rates and differential attrition of subjects by treatment for some grade levels, analyses were performed to determine whether the differential match rates might bias results.

A two-way (treatment X grade) analysis of variance was performed on each 2003 achievement outcome for those subjects who did not have matching pretest (2002) scores. For reading, a significant treatment X grade interaction effect was observed ($F_{5,2581}$=85.1, $p < .001$). Follow-up tests showed that SR students in grades 7 and 8 who did not have matching pretests had significantly higher 2003 Reading scores than comparison students without pretests (see Figure A1 in the Appendix). Likewise, a significant treatment X grade interaction effect was observed for Language ($F_{5,2572}$=4.10, $p = .001$) and Mathematics ($F_{5,2630}$=6.15, $p < .001$), with non-matched SR students in grades 7 and 8 having significantly higher 2003 Language (see Figure A2) and Mathematics (see Figure A3) scores than their comparison counterparts. Finally, a significant interaction was observed for Writing scores ($F_{1,922}$=16.96, $p < .001$), which showed that non-matched SR students in 7th grade had significantly higher mean Writing scores than non-matched Control students (see Figure A4).

While no bias was observed for other grades, the final comparisons are likely to underestimate the effect of SR on Reading, Language, and Mathematics in grades 7 and 8, and on Writing in grade 7. No relationship was observed between treatment and matching status in other grades.

Table 2

Percentage of Cases with Matching Pretest-posttest data by Grade, Treatment Group, and Test

Grade and Treatment Group	Reading	Language	Writing	Mathematics
Third				
SR	80.8	80.7		80.7
Comparison	81.5	81.5		80.9
Fourth				
SR	80.4	79.9	81.7	79.1
Comparison	66.3	67.0	66.6	65.9
Fifth				
SR	81.9	81.8		80.9
Comparison	78.7	78.4		77.7
Sixth				
SR	79.2	79.5		79.2
Comparison	52.2	51.5		51.3
Seventh				
SR	80.4	80.5	81.6	79.9
Comparison	71.8	57.9	58.1	57.1
Eighth				
SR	57.1	57.5		57.5
Comparison	82.4	82.8		83.0

Analyses

Program effects were tested across grade levels by performing mixed-model analyses of covariance, with treatment (SR vs. Control) and grade level (3 to 8) serving as fixed effects, and matched school set serving as a random effect. In four instances, matched sets were comprised of two treatment schools and one comparison school. In all other instances, matched sets were comprised of one comparison school and one treatment school. By modeling matched school set

as a random effect, treatment effects are estimated independent of the effects of school-level demographic differences. Outcome variables were 2003 MCT scale scores on the Reading, Language Arts, and Mathematics subtests, with respective 2002 scores serving as covariates. Mississippi WA scores were used as outcome variables for 4th and 7th grades, with 2002 MCT Reading scores serving as a covariate. Where ANCOVA results indicated a significant effect for treatment or a treatment X grade interaction, follow-up tests were conducted within each grade level by performing independent *t*-tests on predicted scores from a regression of the outcome variable on the covariate. Effect size estimates were computed for each outcome within each grade level by subtracting the comparison group adjusted mean from the treatment group adjusted mean, and dividing by the pooled within-groups standard deviation for the outcome variable.

Results

School Climate

School climate data were analyzed via a one-way MANOVA comparing SR and Control teachers on the seven climate dimensions. Because the multivariate outcome was highly significant, F (7, 648) = 8.01, p <.0001, we conducted follow-up ANOVAs on the seven dimensions. Table 3 presents descriptive statistics, ANOVA outcomes, and effect sizes. As indicated, all comparisons significantly favored the SR over the Control group, with effect sizes ranging from +0.20 (Involvement) to +0.54 (Collaboration).

Table 3

Descriptive statistics, and univariate results for the SR and the Control groups on the seven School Climate Inventory (SCI) dimensions

	SR (*n*=479)		Control (*n*=319)			
SCI-R Dimension	*M*	*SD*	*M*	*SD*	*F*(1,654)	*ES*
Collaboration	4.13	0.54	3.81	0.67	44.02***	+0.54
Environment	4.19	0.56	3.85	0.78	41.64***	+0.52
Expectations	4.36	0.51	4.10	0.67	30.27***	+0.45
Instruction	4.35	0.48	4.21	0.53	13.67***	+0.29
Involvement	4.10	0.55	3.99	0.56	5.55*	+0.20
Leadership	4.39	0.59	4.08	0.80	34.53***	+0.46
Order	3.78	0.73	3.48	0.89	22.52***	+0.38

p≤ .05, p≤.01**, p≤ .001****

Student Achievement

Reading, 2003

The ANCOVA indicated a significant treatment effect on Reading scale scores favoring SR, $F_{1, 23.7} = 8.40$, $p = .008$. The treatment X grade interaction ($F_{5,23.9} = 0.62$, $p = 0.69$) was not significant. Results directionally favored SR in all but 3rd grade (see Figure 1). Effect size estimates ranged from -0.06 in third grade, to a high of +0.16 in seventh grade (see Table 4). The median effect size across grade levels was *ES* = +0.11. The largest differences in mean scale scores were 7.2 points in seventh grade ($M'_{SR} = 541.5$ versus $M'_{C} = 534.3$), 5.5 points in eighth grade ($M'_{SR} = 540.0$ versus $M'_{C} = 534.5$), and 5.3 points in sixth grade ($M'_{SR} = 532.7$ versus $M'_{C} = 527.4$).

Language, 2003

The ANCOVA, controlling for 2002 Language scale scores, yielded a significant main effect for treatment on 2003 Language scale scores, $F_{1,24.4}=7.81$, $p = .01$. Again, no treatment X grade interaction was evident ($F_{5,23.3} = 0.19$, $p = 0.97$). Adjusted mean scores were higher for SR students at all grade levels except grade 4 (see Figure 2, Table 5). The median effect size

estimate across grade levels was +0.12. As with Reading, the largest difference in adjusted mean scale scores was 9.9 points in seventh grade (M'_{SR} = 547.4 versus M'_{C} = 537.5), followed by 7.5 points in sixth grade (M'_{SR} = 542.8 versus M'_{C} = 535.3) and 6.0 points in fifth grade (M'_{SR} = 542.3 versus M'_{C} = 536.3).

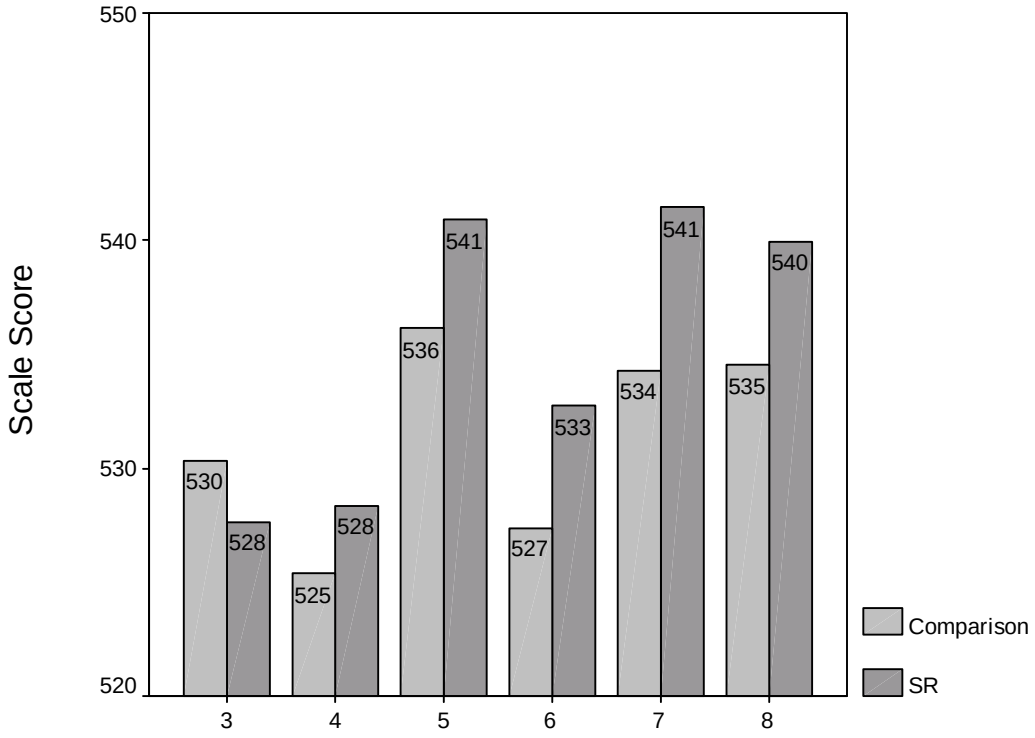

Figure 1. 2003 Adjusted MCT Reading Scale Score Means by Grade Level and Treatment Group.

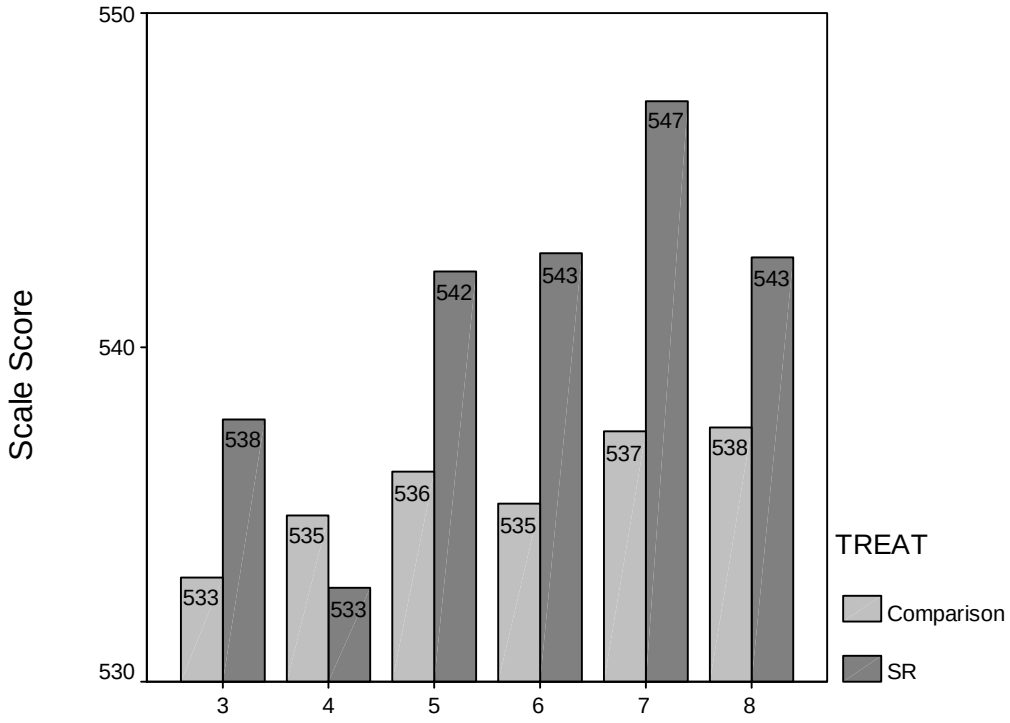

Figure 2. 2003 Adjusted MCT Language Scale Score Means by Grade Level and Treatment Group.

Table 4

2003 Reading Scale Scores Means, Adjusted Means, and Effect Size Estimates by Grade and Treatment Group (SR = School Renaissance)

Grade/Group	*n*	Mean	Adjusted Mean	*SD*[1]	*ES*
Third					
Control	698	500.5	530.4	48.9	-0.06
SR	788	500.7	527.7		
Fourth					
Control	597	511.8	525.4	43.7	+0.07
SR	754	519.3	528.4		
Fifth					
Control	547	533.7	536.1	42.8	+0.11
SR	812	538.7	540.9		
Sixth					
Control	237	534.8	527.4	45.6	+0.11
SR	720	546.6	532.7		
Seventh					
Control	413	550.6	534.3	44.4	+0.16
SR	818	567.6	541.5		
Eighth					
Control	474	549.6	534.5	47.2	+0.12
SR	568	568.3	540.0		

[1] Pooled within-groups standard deviation.

Writing, 2003

A significant treatment effect was observed on adjusted WA scores for fourth grade ($F_{1,10.7}$=20.3, $p < .001$), but not for seventh grade ($F_{1,2.5}$ =1.08, $p = 0.39$). For fourth grade, adjusted mean scores were significantly higher for SR students ($M'_{SR} = 2.55$) than for comparison students ($M'_{C} = 2.28$; see Figure 3 and Table 5). The effect size estimate of +0.45 for fourth grade was substantial. In seventh grade, the adjusted means were nearly identical (M'_{SR}= 2.39 versus M'_{C}=2.39), yielding an effect size estimate of +0.00 (see Figure 4 and Table 6). In fourth grade, 44.5% of SR students received WA scores of 3 or 4 (see rubric in appendix) versus 25.4% of comparison students. In seventh grade, the percentage of students scoring 3 or 4 was 43.3% in SR schools and 29.6% in comparison schools.

Table 5

2003 Language Scale Scores Means, Adjusted Means, and Effect Size Estimates by Grade and Treatment Group (SR = School Renaissance)

Grade/Group	*n*	Mean	Adjusted Mean	*SD*[1]	*ES*
Third					
Control	700	505.1	533.1	49.1	+0.10
SR	789	506.9	537.9		
Fourth					
Control	601	515.9	535.0	46.8	-0.05
SR	752	523.8	532.8		
Fifth					
Control	545	535.3	536.3	41.1	+0.12
SR	812	540.4	542.3		
Sixth					
Control	236	544.8	535.3	48.1	+0.16
SR	721	556.9	542.8		
Seventh					
Control	413	553.0	537.5	52.2	+0.19
SR	814	573.5	547.4		
Eighth					
Control	477	533.0	537.6	44.0	+0.12
SR	565	570.8	542.7		

[1] Pooled within-groups standard deviation.

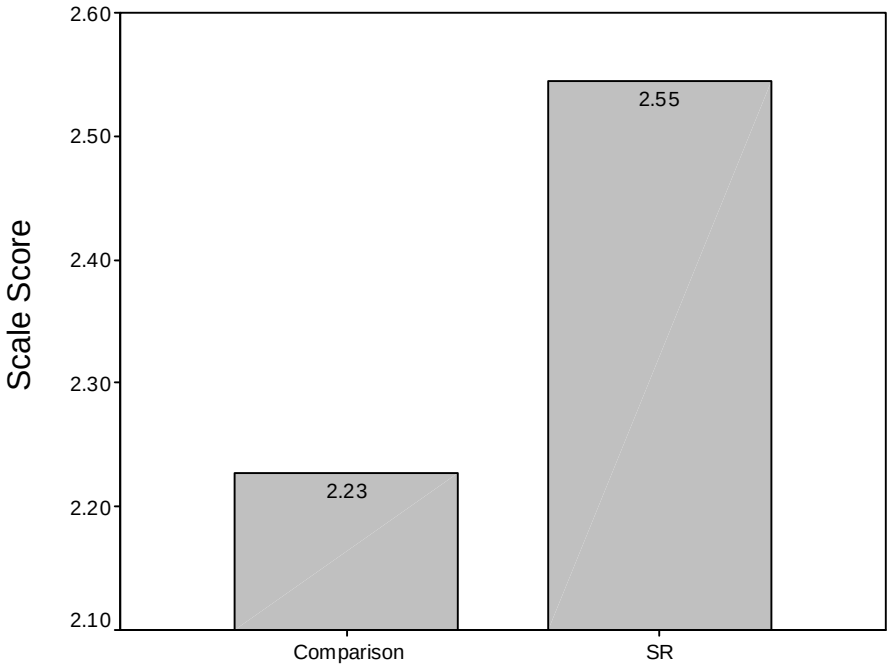

Figure 3. 2003 Adjusted MCT Writing Assessment Score Means by Treatment Group, 4th Grade.

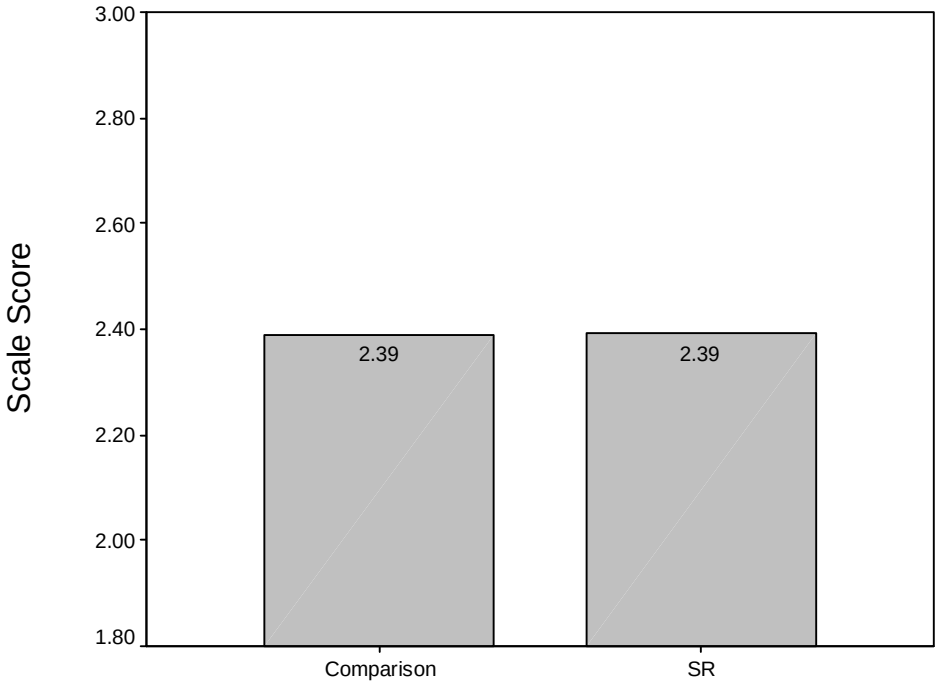

Figure 4. 2003 Adjusted MCT Writing Assessment Score Means by Treatment Status, 7th Grade.

Table 6

2003 Writing Assessment Score Means, Adjusted Means, and Effect Size Estimates by Grade and Treatment Group (SR = School Renaissance)

Grade/Group	*n*	Mean	Adjusted Mean	*SD*[1]	*ES*
Fourth					
Control	585	2.16	2.28	0.60	+0.45
SR	730	2.54	2.55		
Seventh					
Control	403	2.38	2.39	0.65	+0.00
SR	792	2.47	2.39		

[1] Pooled within-groups standard deviation.

Mathematics, 2003

The ANCOVA failed to yield a significant main effect for program on Mathematics scale scores, $F_{1,22} = 3.39$, $p = 0.08$, although there was a significant treatment X grade interaction, ($F_{5,7377} = 6.85$, $p < 0.001$). Follow-up tests indicated that SR students performed significantly higher than comparison students in grades 6 and 7 only, where the effect size estimates were +0.22 and +0.27, respectively (see Table 7). However, as shown in Figure 5, the means were nearly equal in grades 4, 5 and 8. The effect size estimate for grade three was +0.11, and for other grades was near zero.

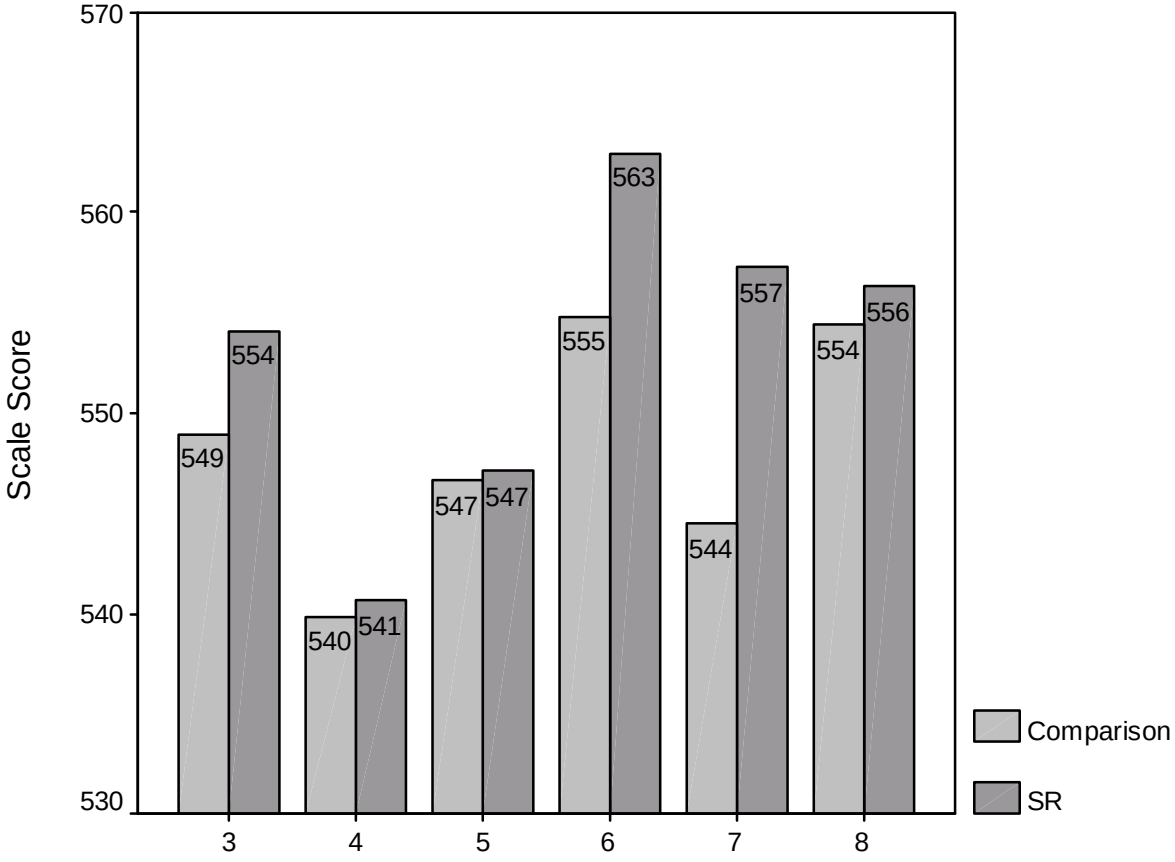

Figure 5. 2003 Adjusted MCT Mathematics Scale Score Means by Grade Level and Treatment Group.

Table 7

2003 Mathematics Scale Scores Means, Adjusted Means, and Effect Size Estimates by Grade and Treatment Status (SR = School Renaissance)

Grade/Group	*n*	Mean	Adjusted Mean	*SD*[1]	*ES*
Third					
Control	701	502.4	549.0	45.8	+0.11
SR	798	508.0	554.1		
Fourth					
Control	600	524.0	540.0	39.7	+0.00
SR	746	529.3	540.1		
Fifth					
Control	540	547.8	546.6	42.3	+0.01
SR	806	547.3	547.1		
Sixth					
Control	235	561.7	554.8	41.6	+0.22
SR	714	575.2	563.9		
Seventh					
Control	407	568.3	544.5	47.9	+0.27
SR	806	589.8	557.3		
Eighth					
Control	474	579.4	554.4	41.8	+0.05
SR	563	594.2	556.3		

[1] Pooled within-groups standard deviation.

Exploratory Analyses of District Effects

To determine whether length of experience with SR implementation was associated with program effects on student achievement, exploratory analyses of covariance were conducted on 2003 MCT Reading, Language, Mathematics scores, and WA scores. District (comparison, Pascagoula, or Biloxi) served as the independent variable, and corresponding 2002 MCT scale scores served as covariates. Analyses were performed for each grade level. Pascagoula had been implementing SR for one year, and Biloxi for three years, at the time of 2003 MCT testing. Descriptive statistics relating to the comparisons between district schools and control schools are reported in Table 8.

In 70% (14 of 20) of the comparisons, students attending Biloxi SR (three-year) schools scored directionally higher than those at both the Control schools and Pascagoula SR (one-year) schools. The inferential tests showed the Biloxi means to be significantly higher than both comparison and Pascagoula means in Reading in third grade ($F_{2,1482}$=3.77, p < .05), sixth grade ($F_{2,975}$=13.59, p <.001), and seventh grade ($F_{2,1411}$=9.03, p < .001; see Table 8). On the Language test, Biloxi students scored significantly higher than Control students in third grade ($F_{2,1485}$=4.81, p < .01), and significantly higher than both Control students and Pascagoula students in fourth grade ($F_{2,1349}$=5.53, p < .01), sixth grade ($F_{2,976}$=17.17, p <.001), and seventh grade ($F_{2,1410}$=9.73, p < .001). Biloxi students also scored significantly higher than Control students in Mathematics in seventh grade ($F_{2,1394}$=29.05, p <.001), and significantly higher than both Control and Pascagoula students in Mathematics in third grade ($F_{2,1495}$=11.17, p <.001) and sixth grade ($F_{2,967}$=33.28, p <.001). On the Writing Assessment, Biloxi students scored significantly higher than comparison students in fourth grade ($F_{2,1311}$=57.0, p < .001), and higher than both comparison and Pascagoula students in seventh grade ($F_{2,1377}$=16.28, p < .001).

Pascagoula SR student means were directionally higher than Control student means in 65% (13 of 20) of the comparisons. Pascagoula students had significantly higher means than Control students in 8th grade Reading ($F_{2,1273}$=3.95, p = .02); 5th grade ($F_{2,1353}$= 3.89, p =.02) and 8th grade ($F_{2,1274}$=6.00, p <.01) Language; and 7th grade ($F_{2,1394}$=29.05, p <.001) and 8th grade ($F_{2,1267}$=5.67, p < .01) Mathematics. On the 4th grade Writing Assessment, Pascagoula students scored significantly higher than both Control students and Biloxi students ($F_{2,1311}$=57.0, p < .001). Control students, however, had a significantly higher 7th grade Writing Assessment mean than Pascagoula students ($F_{2,1377}$=16.28, p < .001).

Table 8

2003 Reading, Language, Mathematics Scale Scores, and Writing Rubric Adjusted Means by Grade Level and District

	Reading			Language			Mathematics			Writing		
	C	P	B	C	P	B	C	P	B	C	P	B
3rd Grade	499.7	498.8	505.0[1,2]	503.3	506.6	511.0[1]	501.2	505.4	513.8[1,2]			
4th Grade	515.3	516.6	516.4	518.8	518.1	525.6[1,2]	525.8	526.8	529.2	2.18	2.59[1,3]	2.45[1]
5th Grade	535.1	537.4	538.2	536.1	541.0[1]	538.5	548.4	546.5	547.4			
6th Grade	540.4	539.8	550.6[1,2]	548.5	548.0	562.6[1,2]	567.2	564.8	582.6[1,2]			
7th Grade	560.0	560.8	568.8[1,2]	564.9	565.6	575.3[1,2]	575.8	588.2[1,3]	582.9[1]	2.48[2]	2.33	2.60[1,2]
8th Grade	559.0	563.9[1]	561.8	561.7	567.9[1]	563.4	585.7	591.2[1]	590.9			

[1]Significanty higher than comparison district mean.
[2]Significantly higher than Pascagoula mean.
[3]Significantly higher than Biloxi mean.

Note. Schools from multiple districts were included in the comparison group means. Pascagoula had implemented School Renaissance for one year, and Biloxi for three years, at the time of the study.

C = Comparison
P = Pascagoula
B = Biloxi

Discussion and Conclusions

The present findings were supportive of School Renaissance in the MS districts, showing positive school climate and achievement outcomes compared to the Control schools. A summary of the major results is as follows:

- SR teachers were significantly more favorable than Control teachers on all SCI dimensions, with a strong median effect size of +0.45. The dimensions reflecting the largest SR advantages were Collaboration, Environment, and Leadership.
- For both Reading and Language, SR students scored significantly higher than comparison students in 2003 after controlling for 2002 MCT scores. Program effects, while small, were generally consistent and positive across grades 3 to 8. Median effect size estimates across grades were +0.11 for Reading and +0.12 for Language.
- In Writing, fourth grade SR students scored significantly and substantially higher than comparison students, with an effect size estimate of +0.45. Nearly double the percentage of fourth grade SR students (44.5%) scored 3 or 4 on the Writing Assessment relative to the comparison group (25.4%). No significant effects were observed for seventh grade Writing.
- Mathematics outcomes were significantly and substantially higher for SR students in grades 6 and 7, with effect size estimates of +0.22 and +0.27, respectively. No differences were observed in other grades.
- Achievement effects were stronger for SR schools in Biloxi (three-year implementation) than in Pascagoula (only one year).

The results of this study must be interpreted in view of the differential rates with which pretest and posttest scores could be matched for SR and comparison students. In particular, SR

students who were excluded from the analyses had significantly higher scores on all 2003 tests than did comparison students who were excluded. This may have caused SR program effects to be underestimated, or may have affected the precision with which adjusted means were computed.

Taken as a whole, the present results suggest that SR is having positive impacts on the participating schools. Drawing from Sullivan and Harper's (2004) qualitative study performed with the same schools, it appears that SR has clear teacher, administrator, and district support. Teachers believe that SR, especially the reading program, Reading Renaissance, is easy to implement, and is increasing students' interest in and motivation to read. Most importantly, they perceive student reading skills to have improved as a result of using the program. Although Math Renaissance was not as well or fully implemented, for interpreting the positive mathematics outcomes in the middle grades 6 and 7, it stands to reason that improved reading ability should facilitate student performance at comprehending problems.

An explicit theoretical assumption and goal of Comprehensive School Reform (CSR) is that improvements in student learning result from establishing a positive school culture in which teachers and school leaders work together to promote high expectation, pride in school environment, effective professional development, and implementation of research-based teaching practices (Desimone, 2002; Rowan et al., 2004; Ross & Gil, 2004). Although in the absence of a randomized experimental design, advantages found for SR schools over Control schools in school climate cannot be directly attributed to program impacts, they are nevertheless suggestive that adoption of SR at least supports, if not directly engenders, the establishment of school climates that promotes collegiality, improved teaching, and ultimately, higher student achievement. It is also noteworthy that consistent with the extensive literature indicating that

CSR requires time to implement fully and effectively (e.g., see Desimone, 2002), our results showed clearly stronger SR benefits for achievement in Biloxi, the three-year SR district, than in Pascagoula, the one-year district. Further study of the interrelationship of SR implementation quality and duration, school climate, and achievement, using a pretest-posttest design and a longitudinal research time frame, is recommended to corroborate and extend the above results.

It is becoming increasingly recognized in CSR evaluations that effect sizes in the +0.10-+0.20 range are educationally meaningful and impactful in raising student achievement (Borman et al., 2003). Thus, it is encouraging to SR that in the present analyses, effect sizes were positive in 16 out of the 20 (80%) grade-level analyses performed and exceeded +0.10 in 13 out of the 20 (65%) analyses. The evidence thus clearly supports continued implementation of SR at these schools as well as continued study of sustainability and cumulative impacts of program usage across grades.

References

Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125-230.

Butler, E. D., & Alberg, M. J. (1991). *The Tennessee School Climate Inventory: Resource Manual*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*, 433-479.

Mississippi Department of Education (2002). *Mississippi Curriculum Test: Summary of Technical Information*. Jackson, MS: Office of Research and Statistics, Mississippi Department of Education.

Ross, S. M. (2003). How to get off of the reform roller coaster. *Principal Leadership*, *4*(1), 16-23.

Ross, S. M., & Gil, L. (2004). The past and future of comprehensive school reform: Perspectives from a researcher and practitioner. In C.T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 151-174). Washington, DC: The National Clearinghouse for Comprehensive School Reform.

Rowan, B., Camburn, E., & Barnes, C. (2004). Benefiting for comprehensive school reform: A review of research on CSR implementation. In C.T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 1-52). Washington, DC: The National Clearinghouse for Comprehensive School Reform.

Sullivan, K., & Harper, M. (2004). *Report on Phase I of the qualitative component: Biloxi/Pascagoula School Renaissance Research Project*. Oxford, MS: North Mississippi Education Consortium.

Sterbinsky, A. (2001). Comprehensive School Reform Teacher Questionnaire: Criterion Related Validity study. Technical Report. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Tomkowicz, J., & Schaeffer, G.A. (2002, April). Vertical scaling for custom criterion-referenced tests. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Appendix

Fourth Grade Mississippi Writing Assessment Scoring Rubric

SCORE 4

The student's writing

a. is about the topic (fully develops the writing prompt)
b. includes several details that support the topic
c. is organized (maintains logical sequence)
d. frequently contains interesting words (grade-level vocabulary or above)
e. contains complete sentences
f. follows punctuation, capitalization, spelling (both correct and phonetic), and usage rules (Two types of errors may occur: 1) those that appear as a consequence of risk-taking and 2) those that do not detract from overall quality.)

SCORE 3

The student's writing

a. is about the topic (partially develops the writing prompt)
b. includes some details that support the topic
c. is organized (may not maintain logical sequence throughout)
d. contains some interesting words (grade-level vocabulary)
e. contains complete sentences (may have occasional fragments and/or run-on sentences)
f. follows punctuation, capitalization, spelling (both correct and phonetic), and usage rules (Occasional errors occur that may detract from overall quality.)

SCORE 2

The student's writing

a. is about the topic (minimally develops the writing prompt)
b. includes only a few details that support the topic
c. shows minimal organization
d. contains only a few interesting words (grade-level vocabulary)
e. contains complete sentences (may have numerous fragments and/or run-on sentences)
f. rarely follows correct punctuation, capitalization, spelling, and usage rules

SCORE 1

The student's writing

a. attempts to address the topic (may digress from the writing prompt)
b. includes vague or no details that support the topic
c. shows no organization
d. includes no interesting words (below grade-level vocabulary)
e. contains numerous fragments and/or run-on sentences (may contain a complete sentence)
f. does not follow correct punctuation, capitalization, spelling, or usage rules

SCORE 0

The student's writing

a. is incomprehensible
b. is insufficient to score

Seventh Grade Mississippi Writing Assessment Scoring Rubric

SCORE 4

The student's response

a. addresses the specific writing prompt (fully develops the topic)
b. contains a clearly stated main idea (thesis)
c. shows a sense of audience and purpose
d. contains a minimum of three indented (or clearly delineated) paragraphs
e. has a clear beginning, middle, and end
f. has a main idea developed by supporting details that are well elaborated
g. exhibits logical order and appropriate sequencing of steps or ideas with adequate transitions
h. contains precise and vivid language (grade-level vocabulary or above)
i. maintains a consistent point of view
j. contains no errors in grammar usage that detract from the overall delivery (Grammar/usage includes subject-verb agreement, verb tense, pronoun case and reference, and complete and varied sentences.)
k. may contain a few errors in the correct use of mechanics (i.e., underlining, quotation marks, commas, semicolons, apostrophes, capitalization, and spelling), but errors do not detract from overall delivery

SCORE 3

The student's response

a. addresses the specific writing prompt (partially develops the topic)
b. contains a stated or implied main idea (thesis)
c. shows a sense of audience and purpose
d. contains a minimum of three indented (or clearly delineated) paragraphs
e. has a clear beginning, middle, and end
f. has a main idea developed by supporting details, but these are not consistently well elaborated
g. exhibits some logical order; sequences most steps or ideas with transitions
h. contains appropriate language, but word choice may be repetitive (grade-level vocabulary)
i. maintains a consistent point of view
j. may contain occasional errors in grammar/usage that may detract somewhat from the delivery (Grammar/usage includes subject-verb agreement, verb tense, pronoun case and reference, and complete and varied sentences.)
k. may contain some errors in the correct use of mechanics (i.e., underlining, quotation marks, commas, semicolons, apostrophes, capitalization, and spelling) that may detract somewhat from delivery

SCORE 2

The student's response

a. addresses the specific writing prompt (minimally develops the topic)
b. contains a vaguely implied main idea (thesis)
c. shows little regard for audience and/or purpose
d. may not exhibit indented (or clearly delineated) paragraphing
e. has a beginning, middle, and end
f. addresses the main idea with minimal supporting details
g. exhibits some evidence of organization but does not sequence steps consistently and/or does not use transitions
h. contains some appropriate language; word choice is repetitive (grade-level vocabulary)
i. may not maintain a consistent point of view
j. may contain frequent errors in grammar/usage that may impede communication (Grammar/usage includes subject-verb agreement, verb tense, pronoun case and reference, and complete and varied sentences.)
k. may contain frequent errors in the correct use of mechanics (i.e., underlining, quotation marks, commas, semicolons, apostrophes, capitalization, and spelling) that may impede communication

SCORE 1

The student's response

a. attempts to address the writing prompt (may digress from the topic)
b. does not contain a main idea (thesis) or contains only an implied focus on the topic
c. shows no regard for audience and/or purpose
d. may not exhibit indented (or clearly delineated) paragraphing
e. may lack a beginning, middle, and/or end
f. contains vague or no details that support the topic
g. lacks organization; presentation is rambling and repetitive
h. contains vague and imprecise language (below grade-level vocabulary)
i. does not maintain a consistent point of view
j. exhibits serious errors in grammar/usage that may severely impede communication (Grammar/usage includes subject-verb agreement, verb tense, pronoun case and reference, and complete and varied sentences.)
k. contains repeated errors in the correct use of mechanics (i.e., underlining, quotation marks, commas, semicolons, apostrophes, capitalization, and spelling) that may severely impede communication

SCORE 0

The student's response

a. is incomprehensible
b. is insufficient to score

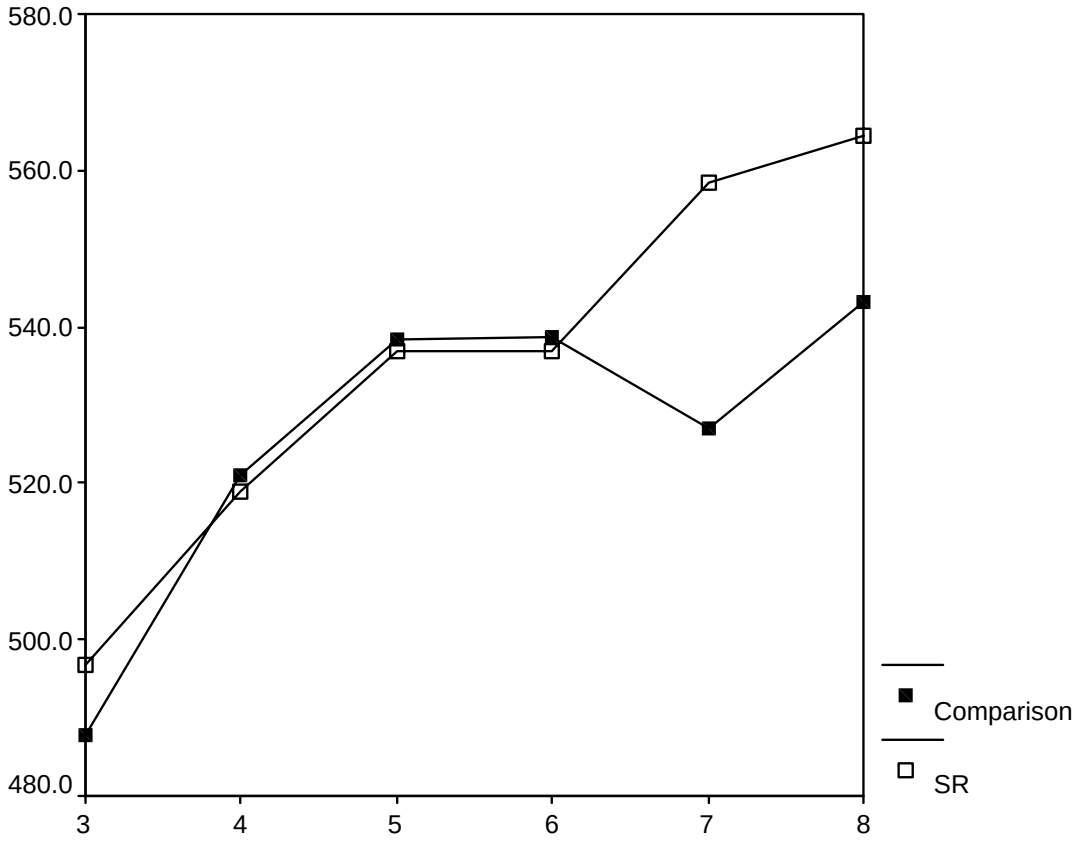

Figure A1
2003 MCT Reading Scores for Non-matched Students by Grade and Treatment Group

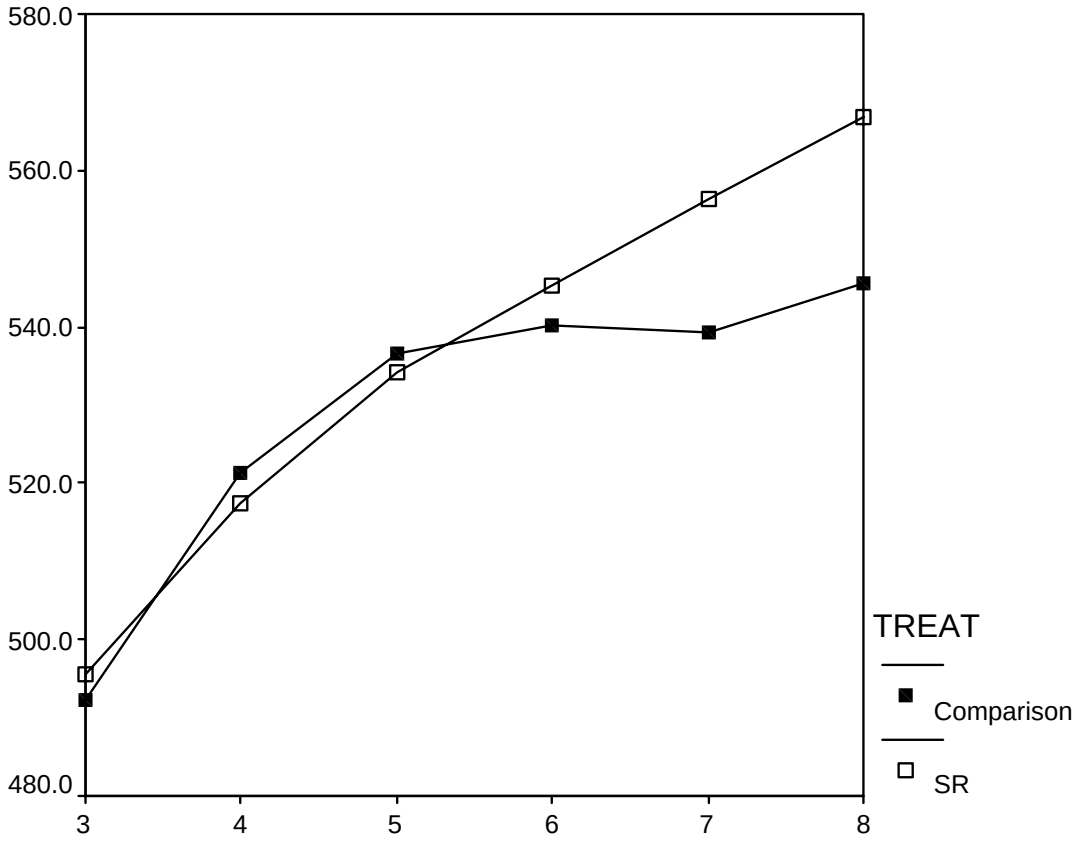

Figure A2

2003 MCT Language Scores for Non-matched Students by Grade and Treatment Group

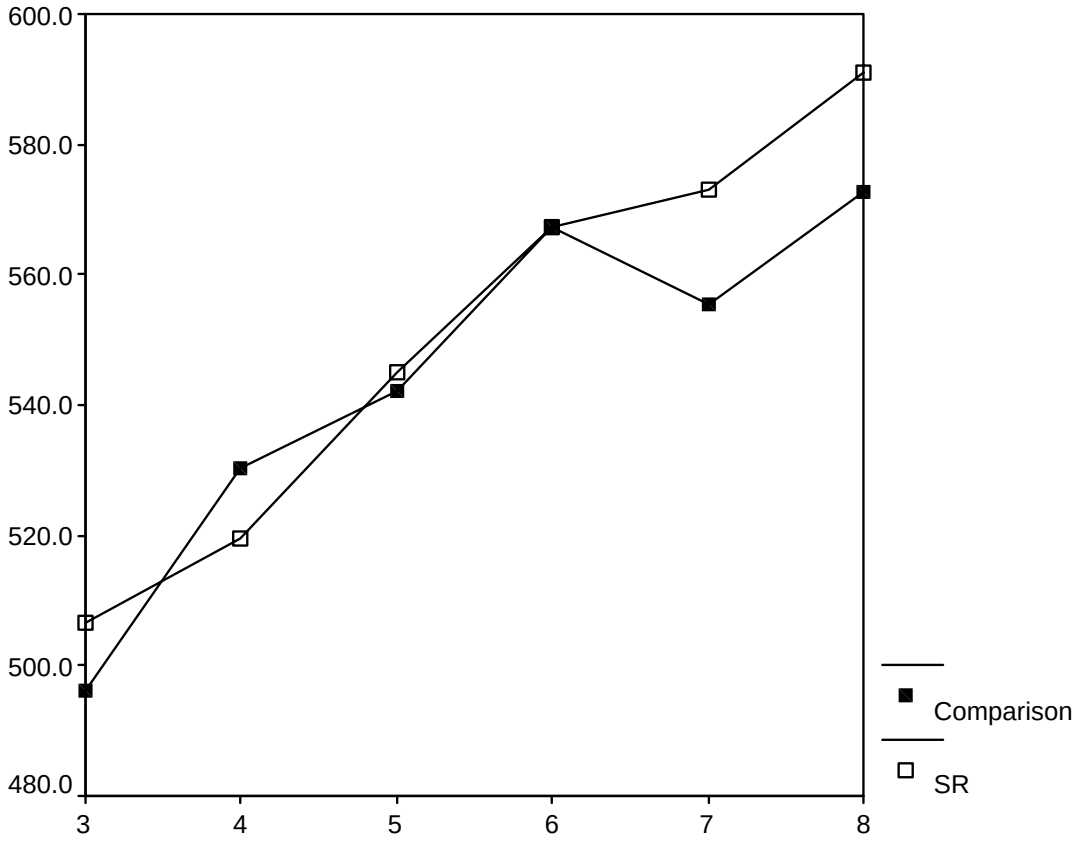

Figure A3

2003 MCT Mathematics Scores for Non-matched Students by Grade and Treatment Group

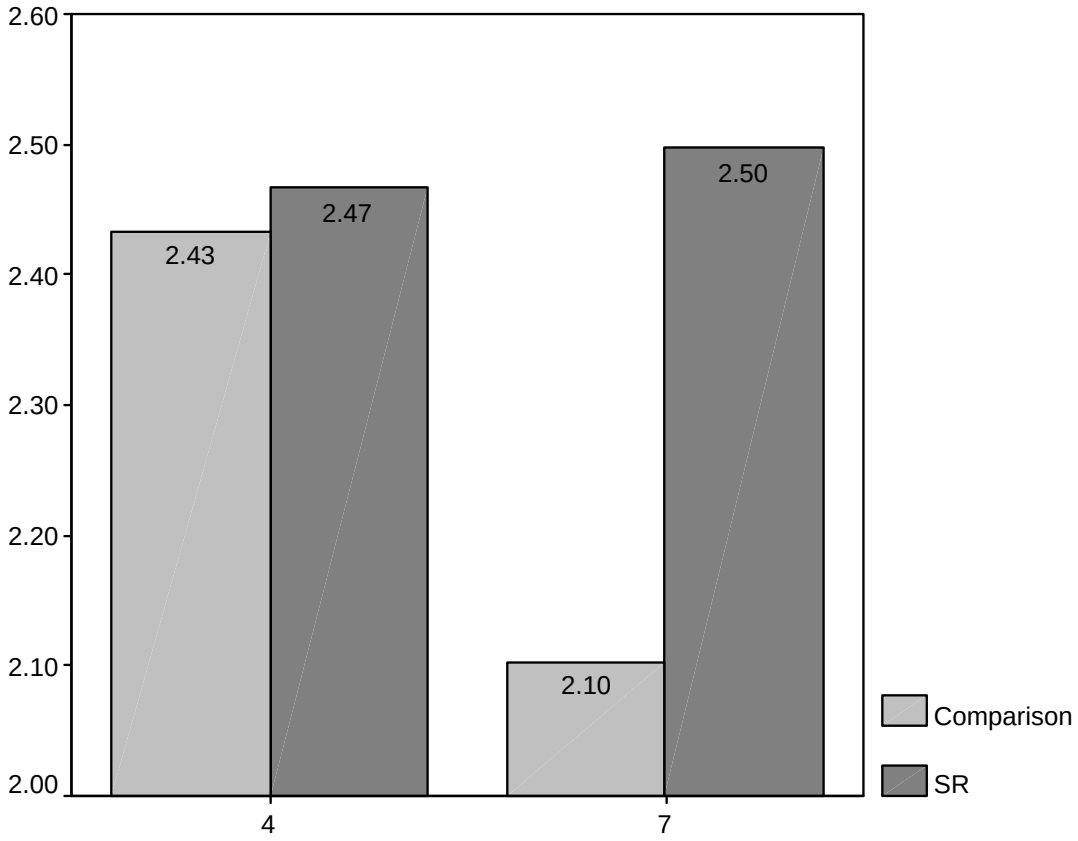

Figure A4

2003 Writing Scores for Non-matched Students by Grade and Treatment Group