

3

Comprehensive School Reform & Student Achievement in Kentucky Middle Schools

Melissa Evans-Andris

*University of Louisville
Louisville, Kentucky*

Wayne M. Usui

*University of Louisville
Louisville, Kentucky*

ABSTRACT

This project examines the effects of Comprehensive School Reform (CSR) models on the achievement of students in Kentucky middle schools. Previous studies exploring the effects of CSR on schools and student achievement have rendered mixed results (Berends, 2000; May & Supovitz, 2006; May, Supovitz, & Perda, 2004; RAND, 2002; Zhang, Shkolnik, & Fashola, 2005). The project examines student achievement in 74 schools identified as “low performing” in 1998-99 by the Kentucky Department of Education. It employs least squares regression to analyze data from the 2002-03 academic year, four years after 18 of the schools had implemented CSR models. Findings suggest that the impact of CSR was minor. Students in schools that had applied but did not receive CSR funding consistently scored higher than those in other comparison schools, whereas students in the CSR-funded schools performed the same or lower than those in comparison groups. These findings hold important implications for policy makers, model providers, and educators as they attempt to clarify the effects of CSR models on achievement among middle school students.

INTRODUCTION

The comprehensive school reform movement has been surrounded by debate regarding the effects of design models on student achievement. Since the late 1990s, thousands of schools nationwide have implemented schoolwide reforms through the federally funded Comprehensive School Reform Demonstration program (CSR). The program typically has targeted high risk, high poverty schools. Underlying the program is the notion that schools are more effective when they adopt a unified approach to improve student and school performance rather than adding fragmented programs or investing in personnel dedicated to small groups of students in pull-out programs (Berends, 2000). Although some researchers argue the positive effects of comprehensive school reform (Berends, 2000; May & Supovtiz, 2006; May et al., 2004; RAND, 2002; Slavin, 2002), other studies have found that, in fact, few of the existing reform models demonstrate strong evidence of effectiveness through rigorous research (Pogrow, 2002; Whitmore, 2000; Zhang et al., 2005).

This project examines the effects of CSR models on the performance among middle grades populations in low performing schools located in Kentucky. It specifically considers the relationship of CSR status to achievement of seventh and eighth grade students across four instructional content areas four years after the implementation of CSR models. CSR and student achievement among middle school students is understudied, and consequently this project offers insight into this important topic.

BACKGROUND

Over the last decade, research focusing on the effectiveness of whole school reform has produced mixed results. This pattern is no different at the middle school level.

Comprehensive School Reform

Despite the promise of Comprehensive School Reform (CSR), studies conducted during the early “scale up” stages of the movement found little evidence of effectiveness of these models on student performance (Herman et al., 1999; Wong & Meyer, 1998). For example, based on a three year study of 163 schools participating in the “scale up” phase of CSR models sponsored by the New American Schools (NAS), a private nonprofit organization, about half the schools posted gains in math relative to their districts and less than half posted gains in reading (Berends, Bodilly, & Kirby, 2002; Berends, Heilbrunn, McKelvey, & Sullivan, 1999; Kirby, Berends, & Naftel, 2001). No student scores were reported. Consistent with these findings were

studies by Ross, Sanders, and Stringfield (1998) and Ross, Wang, Weiping, et al. (2001) that examined the effects of the restructuring effort of 25 elementary schools in Memphis, Tennessee. Using the Tennessee Value-Added Assessment System, these researchers found that student achievement increased significantly, particularly in math, after two years of implementation but failed to show significant gains in reading. These findings taken together suggest that the adoption of CSR models did not promote significant overall improvement in school performance in a large portion of schools.

Another body of research suggests that positive achievement outcomes are associated with intensity and number of years of CSR implementation. For example, various research examining the effects of Success For All (SFA) has found that the SFA schools tended to improve at greater rates than the state average during the study period and the performance of SFA schools tended to continue to improve with each additional year of implementation (Borman & Hewes, 2002; Hurley, Chamberlain, Slavin, & Madden, 2001). In addition, Slavin (2002) argues that the achievement gap between minority students and others narrows more significantly in SFA schools. Most recently, however, when examined using a national randomized field trial, claims regarding the effects of SFA were more modest. After one year, the researchers found that SFA students posted greater gains compared to non-SFA students in only one of four posttests (Borman et al., 2005a). After two years, findings were only slightly more promising (Borman et al., 2005b).

Another study that found a relationship of CSR and improvement over time was that of May and colleagues (May & Supovitz, 2006; May et al., 2004) who, using eleven years of student performance data, tracked the effects of a CSR model, America's Choice (AC), for up to five years in one school district using Bayesian hierarchical growth curve analysis. The researchers found that students in schools using AC showed greater gains in math, and to a lesser extent in reading, than students attending non-AC schools. This pattern held for both elementary and middle grade students. No findings for science or social studies were reported. Importantly, the study also revealed that by the fifth year, the effects of AC on math gains had diminished and the effects for reading were negative. Consistent with other studies, the gains were most pronounced among low performing and minority students.

Finally, using a quasi-experimental design and examining outcomes over five years, Zhang et al. (2005) considered student achievement in math and reading as it related to CSR model implementation. They found that student achievement in CSR schools was lower as compared to non-CSR schools in the first three years of implementation. During that time, students in CSR schools made greater gains, particularly in math, but not at a significant level.

However, once the researchers controlled for intensity and duration of implementation (three to five years) they found that achievement in CSR schools was greater relative to the comparison schools.

CSR & Middle School

Comprehensive school reform has been underrepresented and understudied at the middle school level. At the time when schools were selecting whole school reform models that they might wish to implement with first round CSR monies, only a few models targeted middle level populations exclusively. No model designed to meet the unique needs of middle schools was listed in the “Educators’ guide to schoolwide reform” published in 1998 by the American Institutes for Research (AIR) to assist schools in their selection of the appropriate CSR model (Herman et al., 1999). Moreover, information on the effectiveness of comprehensive school reform models at the middle school level continues to be scarce. One notable examination of five high-poverty middle schools in Philadelphia using the Talent Development Middle School model (Balafanz & MacIver, 2000) reveals that the three schools for which data were available all showed gains either in specific content areas including math, language arts, reading, or in terms of the school’s ability to meet its goal relative to the district’s performance index.

Even so, Lee and Smith (1993) argue that middle schools with larger numbers of students in each grade have a greater likelihood of being characterized as more bureaucratic, having a more impersonal climate, and being less apt to focus resources around a common goal such as whole school reform. Further, a report released by RAND (2004) cites additional problems surrounding middle schooling including inadequate implementation of the middle school concept and training of middle school educators, lack of parental support, and so forth. Following this, the research of Berends and colleagues (Berends et al., 1999; Berends, 2000; Berends et al., 2002), Desimone (2002), and others suggest that school reform may be more difficult in middle schools than elementary schools. Corcoran and Christman (2002) report this to be the case with most Philadelphia middle schools as compared to elementary schools undergoing systemic reform in the late 1990s. Similarly, Colgan (2002) reports that in the Memphis schools, no whole school reform models implemented in the middle and high schools positively affected student achievement.

Kentucky Context

Kentucky public schools participate in a state mandated accountability system whereby student-level results from the Kentucky Core Content Tests (KCCT) are aggregated to the school level, and in two-year cycles are fed – along with non-academic information – by formula into an accountability index. Overall school performance scores are calculated by a formula that combines reading, writing, science, math, social studies, and arts/humanities/practical living, as well as a non-academic index that takes into account the previous year's attendance, retention, and dropout rates. The index score is then used to indicate school improvement. This present Commonwealth Accountability Testing System (CATS) represents a system of high stakes "milepost testing" (Stecher & Barron, 2001) whereby certain subjects are tested at selected grades. For example, seventh graders are tested in reading, writing, and science whereas eighth graders are tested in math, social studies, and arts/humanities/practical living. Presently, approximately 350 schools serve middle school students in Kentucky. Despite efforts of school improvement initiated by the state's Department of Education, Kentucky's CATS results indicate that student performance in math, reading, and so forth is especially dismal at the middle school level (Kannapel & Clements, 2004). Roughly 25% of Kentucky's middle schools were identified as "low performing" in 1998 and thus, based on criteria set by the Kentucky Department of Education, were qualified to submit proposals for CSR funding to begin in 1999. Ultimately, 18 schools serving seventh and eighth graders were awarded CSR monies in the first round of funding. Little empirical research has documented the effects of the comprehensive school reform models on student achievement in these schools, particularly as compared to achievement in other schools identified as low performing at that time. Even so, performance at this level is of utmost concern to policy makers and educators. Our project is designed to address this concern.

RESEARCH OBJECTIVES

Certainly though the results of much of the research appears promising, the debate surrounding the issue of effectiveness of the CSR program remains unresolved and there is much yet to be determined. For instance, many of the whole school reform models continue to lack evidence of effectiveness, particularly across curricular content areas. Moreover, it appears that effectiveness is related to length of time of implementation. Focusing on seventh and eighth grade students in Kentucky schools identified as low performing in 1998, this paper examines questions regarding the effects of CSR models on student

achievement across various content areas four years after model implementation. Specifically, it asks the following:

Is there a relationship between CSR model and student achievement in the areas of reading, math, science, and social studies?

In order to address this question, least squares regression analysis was used.

METHODOLOGY & DATA SOURCES

This project is part of a study investigating the process of change and performance outcomes related to CSR models in Kentucky schools serving middle level populations. The overall study involves intense field observation, interviewing, and quantitative data analysis. Data for this article include school CSR funding status, student characteristics, and content-based student performance scores from Kentucky's Commonwealth Accountability Testing System (CATS).

Three groups of Kentucky public schools serving middle level populations were identified for this project. Group A consists of all 18 schools that received and completed three years of CSR funding beginning 1999-00 through 2001-02. Group B consists of the 13 schools that applied for but did not receive CSR funding in 1998. Group C consists of the remaining 43 schools identified as "low performing" by KDE in 1998 that, despite eligibility, did not apply for CSR funding for 1999-00 or receive CSR funding in subsequent funding cycles.

The 74 schools involved in this study were dispersed geographically across the state and about half of them were in rural districts. They ranged in size and configuration with 27 schools having grades PS/K-8, 7 schools having grades 7-8, 36 traditional middle schools having grades 6-8, and four having either 6-12 or 7-12. Among the 18 Group A schools with CSR funding, five design models¹ were represented including America's Choice (2), Community for Learning (1), Core Knowledge (7), Success For All (6), or National Writing Project (2). At the time of model selection, none of these models targeted middle level populations specifically but one school was identified by Success For All as a pilot site for a middle school SFA model. Three of the models (America's Choice, Success For All, and the National Writing Project) focused primarily on reading, literacy and writing skills whereas the other two (Community for Learning and Core Knowledge) had a more general focus across the curriculum.

¹ For further information regarding these design models see Herman et al., 1999.

Achievement scores for 2002-03 and corresponding student characteristic data for all seventh and eighth graders from the 74 schools used in this project were acquired from the Kentucky Department of Education (KDE). These data have been shown to have high validity and reliability (Kentucky Department of Education, 2005). Consistent with literature stating that optimal gains in performance scores should appear in the third to fifth implementation year (Borman, Hewes, Overman, & Brown, 2002; May & Supovitz, 2006; Zhang et al., 2005), 2002-03 represents the fourth academic year after which the 18 Group A schools had begun model implementation with CSR funding. Schools themselves were classified within CSR groups (i.e., Groups A, B, and C). Relevant student-level variables included the following:

- school identification
- student age
- student gender
- student race/ethnicity
- student socioeconomic status (free or reduced lunch)
- 2003 reading scores (7th grade only)
- 2003 science scores (7th grade only)
- 2003 math scores (8th grade only)
- 2003 social studies scores (8th grade only)

Hierarchical least squares regression analysis was applied in order to address our research questions and was conducted for each of the core content areas, reading, science, mathematics, and social studies. Student-level regressors included AGE, MALE, MINORITY, FREE, and REDUCED. AGE was student's chronological age. The remaining four student-level regressors were dummy variables. In addition to the student-level variables two dummy variables, GROUPA and GROUPB, were used to indicate which of the three CSR school groups students attended. Using these two dummy variables, interaction terms were computed in which each of the five student-level variables was multiplied by each of the two group dummy variables. The resulting interaction variables were used to assess whether student-level variables had different effects depending on CSR-group.

Regressors were added hierarchically in three blocks: first, the five student-level characteristics; second, the two CSR-group indicators; third, the interaction terms. In this way it was possible to test the significance of the improvement in R^2 resulting from adding (1) CSR-group while controlling for student characteristics and (2) the student-level-by-CSR group interaction variables.

There are several limitations to this project. First, embedded in any correlational technique is the possibility of bias when trying to tease out causal implications in the event that a causally antecedent variable has not been included in the analysis. Further, the regression

model assumes perfect measurement in the independent variables, an impossibility given our interests. For example, the free/reduced lunch variables are proxy, but commonly accepted, measures for socioeconomic status. Second, schools represented in this study were not selected randomly, but rather selected because of their eligibility to apply for CSR funds in the first round of funding. The schools were then placed in non-equivalent groups based on their subsequent funding status thus enabling us to make comparisons among students attending schools in each group. Third, though the KDE is working to remedy this situation presently, at the time the data for this project were obtained, data storage techniques used by KDE prohibited the ability to link scores by student in order to track changes in individual scores over time or to measure student mobility rates. Thus the analyses used in this project were based on cross-sectional data. Finally, the state's CATS testing system makes it difficult to track performance gains longitudinally as the subject areas tested vary by year in school. Seventh-graders are tested in reading, writing, and science. Eighth graders are tested in mathematics, social studies, arts and humanities, and practical living. In our sample of middle schools only the test scores from core subject areas for seventh and eighth grades (two separate cohorts) were relevant. Thus, scores for writing, arts and humanities, and practical living were not analyzed for this project. This milepost testing strategy (Stecher & Barron, 2001) coupled with the data storage techniques described above, led us to examine student achievement scores for one point in time.

Despite these limitations, the findings presented in the section that follows capture performance patterns in core subject areas among Kentucky middle school students at a time when some schools had implemented whole school reform models as a result of federal funding for CSR for as many as four years. The findings are illuminating and critical to further the understanding of CSR effects on student performance at the middle school level.

FINDINGS & RESULTS

Seventh Grade

A total of 7682 seventh graders had sufficient data to score reading for the 2003 CATS. (See *Table 1*.) This was an average of 108 seventh graders per school. The KDE accountability data included (1) student "equated scale scores" with possible range of 325 to 800, (2) state percentile ranking, and (3) performance level. The latter variable was an eight-point scale ranging from non-performance novice to distinguished. Only the results from the equated scale scores are presented here; there was no meaningful difference from the results of the other scores.

Table 1

Student-level Descriptive Statistics by CSR Group, Seventh Grade

CSR Group (A, B, or C)		Age	Male	Minority	Free	Reduced	Reading	Science
Group A	Mean	13.333	0.52	0.26	0.69	0.10	501.18	487.54
	N	1884.00	1884.00	1884.00	1884.00	1884.00	1884.00	1884.00
	Std. Deviation	0.53	0.50	0.44	0.46	0.30	33.51	36.34
Group B	Mean	13.42	0.52	0.17	0.63	0.13	504.85	494.57
	N	1468.00	1468.00	1468.00	1468.00	1468.00	1468.00	1468.00
	Std. Deviation	0.63	0.50	0.37	0.48	0.33	33.74	34.41
Group C	Mean	13.34	0.53	0.16	0.64	0.13	503.26	491.88
	N	4330.00	4330.00	4330.00	4330.00	4330.00	4330.00	4330.00
	Std. Deviation	0.57	0.50	0.37	0.48	0.33	33.41	33.934
Total	Mean	13.36	0.53	0.18	0.65	0.12	503.05	491.33
	N	7682.00	7682.00	7682.00	7682.00	7682.00	7682.00	7682.00
	Std. Deviation	0.57	0.50	0.39	0.48	0.32	33.52	34.71

Reading (Seventh Grade Only)

Approximately one-sixth (16.8%) of the variance in student reading scores was accounted for by the student-level variables (Block 1). All student-level regressors were significant and negative. Older students most likely were in this grade level because of prior poor performance. Boys had lower reading scores than girls, and minority students (typically African-American students in our sample) on average tended to perform more poorly on reading tests than did white students. Students who received free lunches, and to a lesser extent those receiving reduced-cost lunches, also scored lower than others on the reading test (Table 2).

Table 2

*Reading Regressed on Student-Level Variables & CSR Group
(N= 7682^a)*

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	662.656	8.234		80.474	0.00
	Age	-10.352	0.624	-0.176	-16.599	0.00
	Male	-13.548	0.707	-0.202	-19.173	0.00
	Minority	-13.096	0.910	-0.152	-14.397	0.00
	Free	-16.599	0.867	-0.236	-19.152	0.00
	Reduced	-7.846	1.248	-0.076	-6.287	0.00
2	(Constant)	663.677	8.245		80.492	0.00
	Age	-10.457	0.625	-0.178	-16.737	0.00
	Male	-13.523	0.707	-0.201	-19.140	0.00
	Minority	-13.008	0.914	-0.151	-14.234	0.00
	Free	-16.529	0.867	-0.235	-19.067	0.00
	Reduced	-7.852	1.248	-0.076	-6.292	0.00
	Group A	-0.439	0.849	-0.006	-0.517	0.605
	Group B	2.139	0.925	-0.025	2.313	0.021

^a Dependent Variable: Reading Equated Scale Score, with possible range of 325 to 800.

The addition of the two CSR group indicators (Block 2) increased R^2 to 0.169, which was not a meaningful amount (albeit statistically significant due to the large sample size). We did note that students in Group A schools were not significantly different from those in comparison schools (i.e., Group C); whereas those in Group B schools averaged significantly *higher* reading scores than those in Group C schools. The addition of the interaction terms (Block 3) also made little difference and was not statistically significant. Thus, the regression analyses suggested that student-level variables of age, gender, race/ethnicity, and socioeconomic status were significant predictors of performance in reading, but CSR group was of only minimal consequence.

Science (Seventh Grade Only)

Only 12.9% of the variation in science scale scores was explained by the student-level variables. Slightly more of the variation was accounted for by the addition of CSR group (R^2 change = .003) and the interaction terms (R^2 change = .004). These increases were statistically significant, and although not substantial in magnitude will be addressed here. As in the case of reading scale scores, statistically significant negative regression coefficients were found for age, minority status, and the two socioeconomic status indicators (*Table 3*). Unlike reading scores, however, the coefficient for MALE was positive.

The two indicators of CSR group were significant. Students in Group A averaged significantly lower science scores than those of Group C schools. On the other hand, students in Group B schools were significantly *higher* than students in Group C. The addition of the interactions (Block 3) suggested that to an extent the student-level variables operated differently in the three groups. For example, the negative effect of being a minority student was even greater in both Group A and B schools than in C schools. Interestingly, the positive impact on science for boys was primarily located among Group A schools. Finally, the negative effect of age on science scores was even more pronounced in Group B schools.

Eighth Grade

A total of 7789 eighth graders had sufficient data to score mathematics and social studies for the 2003 CATS. (See *Table 4*.) This was an average of 110 eighth graders per school. As with the seventh grade data, KDE provided equated scale scores with possible range of 325 to 800, state percentile ranking, and performance level. The results below deal only with the equated scale scores.

Table 3
Science Regressed on Student-Level Variables, CSR Group, & Interactions
 (N= 7682^a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	644.771	8.725		73.899	0.000
	Age	-10.456	0.661	-0.172	-15.822	0.000
	Male	1.623	0.749	0.023	2.168	0.000
	Minority	-17.667	0.964	-0.198	-18.330	0.000
	Free	-16.196	0.918	-0.222	-17.637	0.000
	Reduced	-6.876	1.322	-0.064	-5.200	0.000
2	(Constant)	647.137	8.726		74.159	0.000
	Age	-10.656	0.661	-0.175	-16.115	0.000
	Male	1.663	0.748	0.024	2.224	0.026
	Minority	-17.367	0.967	-0.194	-17.957	0.000
	Free	-16.046	0.917	-0.220	-17.489	0.000
	Reduced	-6.929	1.321	-0.065	-5.247	0.000
	Group A	-2.136	0.898	-0.026	-2.378	0.017
	Group B	3.457	0.979	0.039	3.532	0.000
3	(Constant)	628.857	11.692		53.785	0.000
	Age	-9.338	0.886	-0.154	-10.542	0.000
	Male	0.159	0.993	0.002	0.160	0.873
	Minority	-13.166	1.357	-0.147	-9.702	0.000
	Free	-15.158	1.218	-0.208	-12.442	0.000
	Reduced	-4.922	1.721	-0.046	-2.860	0.004
	Group A	32.096	22.181	0.398	1.447	0.148
	Group B	50.325	21.647	0.570	2.325	0.020
	A-Age	-2.355	1.679	-0.390	-1.402	0.161
	A-Male	3.949	1.807	0.038	2.185	0.029
	A-Minority	-8.718	2.205	-0.061	-3.953	0.000
	A-Free	-3.667	2.260	-0.040	-1.622	0.105
	A-Reduced	-5.241	3.370	-0.023	-1.555	0.120
	B-Age	-3.445	1.640	-0.525	-2.101	0.036
	B-Male	2.893	1.980	0.025	1.461	0.144
	B-Minority	-7.859	2.641	-0.040	-2.976	0.003
	B-Free	-0.555	2.382	-0.005	-0.233	0.816
	B-Reduced	-4.718	3.387	-0.021	-1.393	0.164

^a Dependent Variable: Science Scale Score

Table 4
Student-level Descriptive Statistics by CSR Group, Eighth Grade

CSR Group (A, B, or C)	Age	Male	Minority	Free	Reduced	Math	Social St.
Group A	Mean	14.36	0.50	0.27	0.66	0.11	515.37
	N	1908.00	1908.00	1908.00	1908.00	1908.00	1908.00
	Std. Deviation	0.57	0.50	0.45	0.47	0.31	46.06
Group B	Mean	14.35	0.54	0.16	0.59	0.12	526.65
	N	1466.00	1466.00	1466.00	1466.00	1466.00	1466.00
	Std. Deviation	0.59	0.50	0.37	0.49	0.33	41.50
Group C	Mean	14.35	0.52	0.18	0.62	0.13	518.04
	N	4415.00	4415.00	4415.00	4415.00	4415.00	4415.00
	Std. Deviation	0.57	0.50	0.38	0.48	0.34	44.52
Total	Mean	14.35	0.52	0.20	0.63	0.12	519.01
	N	7789.00	7789.00	7789.00	7789.00	7789.00	7789.00
	Std. Deviation	0.57	0.50	0.40	0.48	0.33	44.51

Mathematics (Eighth Grade Only)

R^2 using the student-level variables AGE, MALE, MINORITY, FREE, REDUCED equaled 0.146. Addition of the indicators of CSR group increased R^2 to 0.150, a small but statistically significant amount. The addition of the interaction variables did not significantly increase R^2 . Table 5 gives the regression coefficients for the equation containing the student-level variables and the two CSR group indicators. Each of these variables was statistically significant except for GROUPA; that is, student math scores in schools receiving CSR funding did not differ from those in the comparison schools (i.e., schools that did not apply for funding). On the other hand, students in schools that applied for but did not receive funding (i.e., Group B) were significantly higher than those in the comparison schools, even after controlling student-level variables. In addition, student age, gender, minority status, and socioeconomic status were significant. Higher age, being male, being nonwhite, and receiving free or reduced lunch all were associated with poorer math scores.

Table 5
Mathematics Regressed on Student-level Variables & CSR Group (N= 7789^a)

Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	t
1	(Constant)	758.66	11.32		67.022
	Age	-15.125	0.795	-0.202	-19.022
	Male	-5.493	0.940	-0.062	-5.8430
	Minority	-17.573	1.185	-0.157	-14.835
	Free	-23.438	1.125	-0.255	-20.826
	Reduced	-12.828	1.626	-0.095	-7.8900
2	(Constant)	757.288	11.295		67.046
	Age	-15.133	0.793	-0.202	-19.082
	Male	-5.613	0.938	-0.063	-5.9850
	Minority	-17.193	1.188	-0.154	-14.469
	Free	-23.205	1.123	-0.252	-20.663
	Reduced	-12.666	1.623	-0.094	-0.7806
	Group A	-0.500	1.132	-0.005	-0.4420
	Group B	7.634	1.240	0.067	6.1570

^a Dependent Variable: Math Equated Scale Score

Social Studies (Eighth Grade Only)

Results of the analysis of social studies scores were similar to those of math scores: R^2 equaled 0.155 with student-level variables only and increased significantly to 0.161 when the two dummy variables for CSR group were added. (See Table 6.) Significant interactions were not found. One difference between social studies and math scores among these eighth-graders should be noted, however. In

the case of social studies scores, students in Group A schools were significantly lower than those in Group C schools; whereas, this difference was not significant in the case of math scores.

Table 6

Social Studies Regressed on Student-Level Variables & CSR Group (N= 7789^a)

Model		Unstandardized	Standardized	t	Sig.
		Coefficients	Coefficients		
		B	Beta		
1	(Constant)	724.988		62.499	0.00
	Age	-13.994	-0.182	-17.174	0.00
	Male	-9.711	-0.106	-10.081	0.00
	Minority	-17.445	-0.151	-14.37	0.00
	Free	-26.403	-0.278	-22.893	0.00
	Reduced	-14.977	-0.107	-8.988	0.00
2	(Constant)	724.416		62.652	0.00
	Age	-13.981	-0.181	-17.22	0.00
	Male	-9.908	-0.108	-10.319	0.00
	Minority	-16.639	-0.145	-13.678	0.00
	Free	-26.156	-0.276	-22.751	0.00
	Reduced	-14.958	-0.107	-9.005	0.00
	Group A	-4.514	-0.042	-3.894	0.00
	Group B	6.786	0.58	5.346	0.00

^a Dependent Variable: Social Studies Equated Scale Score

SUMMARY & CONCLUSIONS

This project examined the relationship between CSR models and student achievement in the areas of reading, math, science, and social studies in 74 Kentucky schools serving middle grade populations. All of the schools had been identified as “low performing” in 1998-99 and thus, as determined by the Kentucky Department of Education, eligible to apply for CSR funding. Cross-sectional data used in the project were from the 2002-03 academic year, four years after 18 of the schools had implemented CSR models.

The project first determined student and school effects and their interactions on student achievement. It is clear from these analyses that in both the seventh and the eighth grades older students, boys, nonwhites, and those of lower socioeconomic status tended to perform more poorly on the accountability tests in the four content areas of reading and science (seventh grade) and mathematics and social studies (eighth grade). The one exception to the relationships we found between scores and student-level data was the positive, albeit small, male-science relationship.

The more interesting research question of the impact of CSR group remains perplexing. From the least squares analyses it appeared that the impact of CSR group and the interactions with student-level variables

was minor. More importantly, in none of the four content areas did the students in CSR-funded schools (Group A) average higher scores than those students in other schools. Instead, students in schools that had applied but did not receive funding (Group B) consistently scored higher than those in the comparison Group C schools. Students in the CSR-funded schools (Group A) were significantly lower than those in the comparison schools in science and social studies, but not in reading or math.

A significant interaction between being male and CSR group on science scores suggested that boys in CSR-funded schools, more so than boys in other schools, performed better than girls. Why this should be so is not evident. There was nothing in the descriptions of CSR models to suggest that boys might be helped in science more than girls. Certainly, this is a finding that will be explored more extensively in future analyses.

The findings presented in this article raise issues for consideration regarding CSR and middle grades populations in terms of model selection, implementation, and sustainability, and hold important implications for policy makers, model providers, and educators at the middle school level.

Model Selection

Literature suggests that implementation efforts are more likely to be favorable when the model adoption process is inclusive, when the information regarding model provision is accessed critically, and when the buy-in of teachers and other stakeholders is legitimate (Aladjem & Borman, 2005; Datnow & Stringfield, 2000; Le Floch, Zhang, & Herrmann, 2005). We know from qualitative portions of our study that KDE provided strong guidance to schools regarding model selection. For example, early on the Department compiled and distributed an evaluation of models to assist schools in making effective choices about model selection and encouraged schools to conduct site visits to model-using schools. Even so, at the time the schools in our study were making those selections, only several of the models were designed or otherwise well suited to address the needs of middle schools. Thus, many schools with seventh and eighth grades chose models that were intended for grades K-6. It may be that the minor impact of CSR revealed in this project is, in part, due to the inappropriate selection of models to be implemented at the middle school level.

Though school administrators and educators need to continue to strive to become more informed, savvy consumers of whole school reform models, certainly the “mismatch” of models designed for younger children rather than targeting the unique needs of middle school students in the selection process should be monitored and addressed both by model providers and state agencies. Moreover,

although there has been improvement in this area, even now, model providers need to continue to develop, research, and refine models specifically tailored to address the unique needs of middle school populations.

Implementation

An issue undoubtedly related to model selection is that of implementation. This project considered implementation in terms of duration in that it analyzed performance data four years after schools had introduced CSR models. Though it did not measure quality of implementation or fidelity to model strategies, it is likely that the misguided selection of models geared toward elementary school students jeopardized the implementation process as educators realized that the models they had selected were intended for a younger population and therefore not meeting their needs as teachers or the needs of their middle level students.

Additionally, Berends et al. (2002), Zhang et al. (2005), and others identify barriers to implementation including school characteristics relating to level, poverty and achievement. The schools in our sample served high poverty students, and many were located in rural areas. All of our schools were identified by the state as low performing in 1998. Group A (CSR-funded) schools were the lowest performing compared to the other two groups (Evans-Andris & Usui, 2005). Thus, given these characteristics, we would expect that CSR models would be implemented less effectively and thus have less impact in these schools. Developing effective strategies of implementation that meet the needs of educators in high risk, low performing schools represents a challenge that model providers need to address head-on.

Effects of Time & Issues of Sustainability

This project has also enabled us to make some determinations regarding the sustainability of student performance in schools one year after the termination of CSR funding. Only recently has the issue of sustainability begun to be addressed (Evans-Andris & Usui, 2004; Datnow, 2005; Taylor, 2005). However, literature suggests that analyses conducted between three to five years of implementation may render the most promising results (Borman et al., 2002; May & Supovitz, 2006; May et al., 2004; Zhang et al., 2005). It is the case, based on qualitative data collected in another portion of the larger study, that approximately two-thirds of the CSR funded school in this project terminated a formal relationship with their respective model providers in 2002, when their CSR awards expired. Even though some of the remaining schools attempted to sustain their model informally by practicing reform strategies learned through their CSR experience, it

seems that this certainly would reduce the effect size of Group A schools for 2003. Certainly, model providers, state and district administrators along with school educators need to anticipate and develop strategies of sustainability to assist schools as their CSR awards expire so as to maximize the longer lasting effects of whole school reform efforts on student gains.

Findings presented in this article suggest areas for further study. We intend to continue examining linkages between model selection, patterns of implementation, and the sustainability of reform efforts. We are also interested in refining our understanding of the effects of the proactive context within which the Kentucky schools operate. Finally, future analyses will explore factors that may be contributing to the higher performance of students in non-CSR funded comparison Group B schools. It may be that at least some of these schools have implemented comprehensive school reform models or other reform programs without CSR funding, thus masking the effects of these models on performance. These findings regarding the effects of CSR on student performance add critical understanding to the literature on comprehensive school reform particularly at the middle school level.

REFERENCES

- Aladjem, D., & Borman, K. (2005). *The lifecycle of Comprehensive School Reform*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Balfanz, R., & MacIver, D. (2000). Transforming high-poverty urban middle schools into strong learning institutions: Lessons from the first five years of the Talent Development Middle School. *Journal of Education for Students Placed at Risk*, 5(1/2), 137-158.
- Berends, M. (2000). Teacher-reported effects of New American Schools designs: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis*, 22(1), 65-82.
- Berends, M., Bodilly, S., & Kirby, S. (2002). Looking back over a decade of whole-school reform: The experience of New American Schools. *Phi Delta Kappan*, 84(2), 168-177.
- Berends, M., Heilbrunn, J., McKelvey, C., & Sullivan, T. (1999). *Assessing the progress of New American Schools: A status report* (MR-1085-ED). Santa Monica, CA: RAND.
- Borman, G., & Hewes, G. (2002). The long term effects and cost effectiveness of Success For All. *Educational Evaluation and Policy Analysis*, 24(4), 243-266.
- Borman, G., Hewes, G., Overman, L., & Brown, S. (2002). *Comprehensive School Reform and student achievement: A meta-analysis* (Report No. 59). Baltimore, MD: The Johns Hopkins

- University, Center for Research on the Education of Students Placed at Risk.
- Borman, G., Slavin, R., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2005a). Success For All: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis*, 27(1), 1-22.
- Borman, G., Slavin, R., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2005b). The national randomized field trial of Success For All: Second-year outcomes. *American Educational Research Journal*, 42(4), 673-696.
- Colgan, C. (2002). Memphis blues. *Principal Leadership*, 2(5), 10-16.
- Corcoran, R., & Christman, J. (2002). *The limits and contradictions of systemic reform: The Philadelphia story*. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Datnow, A. (2005). The sustainability of Comprehensive School Reform models in changing district and state contexts. *Educational Administration Quarterly*, 41(1), 121-153.
- Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk*, 5(1), 183-204.
- Desimone, L. (2002). How can Comprehensive School Reform models be successfully implemented? *Review of Educational Research*, 72(3), 433-480.
- Evans-Andris, M., & Usui, W. (2004). *Comprehensive School Reform and sustainability in Kentucky middle schools*. Paper presented at a conference of the National Clearinghouse for Comprehensive School Reform, Washington, DC.
- Evans-Andris, M., & Usui, W. (2005). *Comprehensive School Reform and reading growth curves in Kentucky middle schools*. Paper presented at the annual meeting of the American Sociological Association, Philadelphia, PA.
- Herman R., Aladjem, D., McMahon, P., Masem, E., Muligan, I., Smith, O., O'Malley, A., Quinones, S., Reeve, A., & Woodruff, D. (1999). *An educator's guide to schoolwide reform*. Washington, DC: The American Institute for Research.
- Hurley, E., Chamberlain, A., Slavin, R., & Madden, N. (2001). Effects of SFA on TAAS reading scores: A Texas statewide evaluation. *Phi Delta Kappan*, 82(10), 750-757.
- Kannapel, P., & Clements, S. (2004). *Inside the black box of high performing, high poverty schools*. A report to the Ford Foundation on a funded project submitted by the Prichard Committee for Academic Excellence, Lexington, Kentucky.
- Kentucky Department of Education. (2005). *Kentucky Core Content Tests* (2004 Technical Report). Frankfort, Kentucky.

- Kirby, S., Berends, M., & Naftel, S. (2001). *Implementation in a longitudinal sample of New American Schools: Four years into scale-up* (MR-1413-EDU). Santa Monica, CA: RAND.
- Lee, V., & Smith, J. (1993). Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education*, 66(3), 164-187.
- Le Floch, K.C., Zhang, Y., & Herrmann, S. (2005). *Adoption: Exploring the initiation of Comprehensive School Reform models*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- May, H., & Supovitz, J. (2006). Capturing the cumulative effects of school reform: An 11-year study of the impacts of America's Choice on student achievement. *Educational Evaluation and Policy Analysis*, 28(3), 231-257.
- May, H., Supovitz, J., & Perda, D. (2004). *A Longitudinal Study of the Impact of America's Choice on Student Performance in Rochester, New York, 1998-2003*. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Pogrow, S. (2002). Success for all is a failure. *Phi Delta Kappan*, 83(6), 463-469.
- RAND Research Brief. (2002). *A decade of whole-school reform: The New American Schools experience* (RB-8019). Santa Monica, CA: RAND.
- RAND Research Brief. (2004). *Problems and promise of the American middle school* (RB-8025). Santa Monica, CA: RAND.
- Ross, S., Sanders, W., & Stringfield, S. (1998). *The Memphis restructuring initiative: Achievement results for years 1 and 2 on the Tennessee Value-Added Assessment System (TVASS)*. A special report prepared for Memphis City Schools.
- Ross, S., Wang, M., Weiping, L., Alberg, M., Sanders, W., Wright, S., & Stringfield, S. (2001). *Fourth-year achievement results on the Tennessee Value-Added Assessment system for restructuring schools in Memphis*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Slavin, R. (2002). Mounting evidence support the achievement effects of SFA. *Phi Delta Kappan*, 83(6), 469-473.
- Stecher, B., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment*, 7(4), 259-281.
- Taylor, J. (2005). *Sustainability: Examining the survival of schools' Comprehensive School Reform efforts*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Whitmore, R. (2000, February 14). Schools grant program alarming researchers. *Gannet News Service*.
[<http://www.ewa.org/infocenter/reporters/lightspan.htm>].

- Wong, K., & Meyer, S. (1998). Title I schoolwide programs: A synthesis of findings from recent evaluation. *Educational Evaluation and Policy Analysis*, 20(2), 115-136.
- Zhang, Y., Shkolnik, J., & Fashola, O. (2005). *Evaluating the implementation of Comprehensive School Reform and its impact on growth in student achievement*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

The study upon which this paper is based was funded by Grant # R305T010812 from the U.S. Department of Education, Office of Education Research and Improvement, Field Initiated Study. An earlier version of this paper was presented at the annual meetings of the American Education Research Association, Montreal, Canada. Please do not cite without permission of the authors.