U.S. DEPARTMENT OF EDUCATION

NoChild LeftBehind

Evaluation of the Comprehensive School Reform Program Implementation and Outcomes

Third-Year Report

Evaluation of the Comprehensive School Reform Program Implementation and Outcomes

Third-Year Report

Martin Orland, WestEd
Brooke Connolly, WestEd
Tony Fong, WestEd
Lauren Davis Sosenko, WestEd
Naida C. Tushnet, WestEd

Robert K. Yin, COSMOS Corp.
Janeula M. Burt, COSMOS Corp.
Emily Warner, COSMOS Corp.

Prepared for:

U.S. Department of Education
Office of Planning, Evaluation and Policy Development
Policy and Program Studies Service

2008

This report was prepared for the U.S. Department of Education under Contract Number ED01CO0129. Menahem Herman served as the contracting officer's representative. The content of this report does not necessarily reflect the views or policies of the U.S. Department of Education, nor does the mention of commercial products or organizations imply endorsements by the U.S. government. The inclusion of such information is for the reader's convenience and is not intended to endorse any views expressed, or products, programs, models or services offered.

U.S. Department of Education
Margaret Spellings
Secretary

Office of Planning, Evaluation and Policy Development
Bill Evers
Assistant Secretary

Policy and Program Studies Service
Alan Ginsburg
Director

Program and Analytic Studies Division
David Goodwin
Director

December 2008

This report is available on the Department's Web site at:
www.ed.gov/about/offices/list/opepd/ppss/reports.html#title.

On request, this publication is available in alternate formats, such as Braille, large print, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-0852 or 202-260-0818.

CONTENTS

EXHIBITS

PREFACE

This *Third-Year Report* from the Evaluation of the Comprehensive School Reform Program Implementation and Outcomes (ECSRIO) presents findings about the relationship between participation in the U.S. Department of Education's Comprehensive School Reform (CSR) Program and student achievement. It follows the study's *First-Year Report* published in 2004.[1] The *Third-Year Report* examines the CSR program throughout the country and its relationship with gains in student achievement.

The CSR program was established as a demonstration program in 1998 and authorized as a full program in 2002 as part of the *No Child Left Behind Act (NCLB)*. It is one approach to help low-performing K–12 public schools meet state performance standards. CSR emphasizes two major concepts. First, the approach mandates that school reform should be *comprehensive* in nature, strengthening all aspects of school operations—curriculum, instruction, professional development, parental involvement, and school organization. Second, CSR should involve the use of *scientifically based research models*—that is, models with evidence of effectiveness in multiple settings.

In 2002, the U.S. Department of Education contracted with WestEd and COSMOS Corp. to conduct a five-year study of the CSR program, ECSRIO, which involved a survey of 500 CSR schools and 500 comparison schools, case studies of 30 sites, and analyses of student achievement in all schools receiving CSR funding in 2002. This report presents the third-year findings of this study.

[1] The *Second-Year Report* was an internal document to the U.S. Department of Education.

ACKNOWLEDGMENTS

We wish to thank many individuals who contributed to this report. We are especially grateful to the state, district, and school survey respondents, as well as to all of the school or district staff who planned or participated in site visits. We express our deep gratitude for your role in the study and the data you provided, without which this study would not have been possible.

The information in this report was provided through the congressionally mandated Evaluation of the Comprehensive School Reform Program Implementation and Outcomes (ECSRIO), which was conducted by WestEd and COSMOS Corp. under contract to the U.S. Department of Education. Naida Tushnet of WestEd and Robert Yin of COSMOS Corp. led ECSRIO, and Susan Cragle from DUERR Associates coordinated the survey data collection.

Several individuals at the U.S. Department of Education provided report guidance and direction. We would like to acknowledge the assistance of Alan Ginsburg, director, Policy and Program Studies Service (PPSS), and David Goodwin, director, Program and Analytic Studies Division, PPSS. In addition, Carol Chelemer, Menahem Herman, Aaron Neumann, and Dante Randazzo in PPSS contributed to the direction of the report.

We would like to acknowledge the contributions and guidance of the members of our Technical Working Group, including Carolyn Temple Adger, Geoffrey Borman, H. J. Green, Bryan Hassel, Elsie Leak, Valerie Lee, Paul Ruiz, Jean Rutherford, Malik Stewart, Sam Stringfield, and Ken Wong.

We also would like to thank Noel White, WestEd senior editor, and Colleen Montoya, WestEd senior communicator, for their editorial assistance.

While we appreciate the assistance and support of all of these individuals, any errors in judgment or fact are the responsibility of the authors.

EXECUTIVE SUMMARY

The Comprehensive School Reform (CSR) Program was established as a demonstration program in 1998 and authorized as a full program in 2002 as part of the *No Child Left Behind Act (NCLB)*. It is one approach to help low-performing K–12 public schools meet state performance standards. CSR emphasizes two major concepts. First, the approach mandates that school reform should be *comprehensive* in nature, strengthening all aspects of school operations—curriculum, instruction, professional development, parental involvement, and school organization. Second, CSR should involve the use of *scientifically based research models*—that is, models with evidence of effectiveness in multiple settings.

In 2002, the U.S. Department of Education contracted with WestEd and COSMOS Corp. to conduct a five-year study of the CSR program, the Evaluation of the Comprehensive School Reform Program Implementation and Outcomes (ECSRIO), which involved a survey of 500 CSR schools and 500 comparison schools, case studies of 30 sites, and analyses of student achievement in all schools receiving CSR funding in 2002.

This report presents the third-year findings of this study and focuses on the relationship between CSR award receipt and growth in student achievement and whether aspects of program implementation are associated with higher CSR achievement outcomes. Overall, the findings indicate that schools receiving CSR awards did no better in mathematics and reading achievement than comparable schools not receiving CSR funding.

The report's major findings, organized by research question, are:

Was receipt of a CSR award associated with improvements in school-level mathematics and reading achievement?

- Receipt of a CSR award was not associated with achievement gains in mathematics or reading achievement through the first three years of award.[2]

Were schools that received CSR awards more likely to implement the legislatively specified components of CSR than other schools?

- No, both CSR and non-CSR schools implemented an average of fewer than four components in 2003 and fewer than five in 2005 at both the elementary and middle school levels.

[2] Note that, in this report, all references to achievement gains refer to gains in school-level achievement measures, such as the percent proficient or higher or the school's mean scaled score. Readers are cautioned not to infer that any gains on school-level metrics refer to absolute improvements in students' achievement. See Robinson (1950) for a discussion of *ecological fallacies*, in which inferences about individuals are mistakenly made based on aggregate statistics.

Was fidelity of CSR implementation associated with gains in school-level mathematics and reading achievement?

Two analyses address this question:

Comprehensiveness of implementation

- The comprehensiveness of implementation, as measured by the number of CSR components implemented, was not related to mathematics and reading achievement gains in CSR schools.

Adoption of models with a recognized scientific research base

- Only one-third of 2002 CSR awardees chose reform approaches with recognized scientific research bases.
- Low-performing elementary schools that adopted models with stronger evidence of effectiveness had gains in mathematics achievement that were not found in higher-performing schools.
- Adoption of a CSR model independently determined to have had limited scientific evidence of effectiveness was associated with higher gains in middle school mathematics achievement in all CSR schools, whether they were low-performing or not. There is also weaker evidence that CSR middle schools that adopted models with limited scientific evidence may have experienced gains in middle school reading achievement relative to schools that adopted other models.
- There was weaker evidence that low-performing middle schools that adopted models with moderate or higher bases of evidence showed improvement in mathematics compared with schools using other models.
- In no other instances was adoption of models with a scientific research base related to achievement gains.

BACKGROUND

Originally funded as the Comprehensive School Reform Demonstration (CSRD) Program in 1998 with $145 million, the CSR program became part of *NCLB* in 2002 with funding of $310 million. In FY 2003 and FY 2004, Congress allocated $308 million for CSR, and in FY 2005, $205 million. In FY 2006, funding was appropriated only for a clearinghouse to support comprehensive school reform and not for school-based activities. Between 1998 and 2006, nearly 7,000 schools nationwide received three-year awards to implement CSR models.

NCLB defines CSR as containing 11 components, which are assumed to work together as schools undergo reform (Exhibit E.1).

Exhibit E.1
Eleven Components of Comprehensive School Reform
Described in the *No Child Left Behind Act*

1. **Proven methods** and strategies for student learning, teaching, and school management that are based on scientifically based research and effective practices, and have been replicated successfully in schools with diverse characteristics.
2. **Comprehensive design** for effective school functioning, integrating instruction, assessment, classroom management, and professional development and aligning these functions into a schoolwide reform plan designed to enable all students to meet challenging state content and performance standards and address needs identified through a school needs assessment.
3. **Professional development.** High-quality and continuous teacher and staff professional development and training.
4. **Measurable goals** for student performance and benchmarks for meeting those goals.
5. **Support from staff.** Support from school faculty, administrators, and staff.
6. **Support for staff.** Support for school faculty, administrators, and staff (added in 2001).
7. **Parent and community involvement.** Meaningful involvement of parents and the local community in planning and implementing school improvement activities.
8. **External assistance.** High-quality external support and assistance from a comprehensive school reform entity (which may be a university) with experience in schoolwide reform and improvement.
9. **Evaluation.** Plan to evaluate the implementation of school reforms and the student results achieved.
10. **Coordination of resources.** Identification of how other available resources (federal, state, local, and private) help the school coordinate services to support and sustain the school reform.
11. **Scientifically based research.** Scientifically based research to significantly improve the academic achievement of students participating in such programs as compared with students in schools who have not participated in such programs; or strong evidence that such programs will significantly improve the academic achievement of participating children (added in 2001).

Source*: No Child Left Behind Act*, Title I, Part F, Section 1606.

This evaluation, mandated by Section 1606 of *NCLB*, addresses four broad questions related to the CSR program.

- How were CSR funds targeted?
- How was reform implemented in CSR and non-CSR schools?
- How did state and district conditions influence reform implementation?
- How was reform related to student achievement outcomes?

ECSRIO examines implementation and outcomes of a cohort of CSR awardees that received their initial awards in 2002. The *First-Year Report* (2004) addressed the first three questions. It concluded that CSR funds primarily targeted low-performing, high-poverty elementary and middle schools in rural and urban areas. Although both CSR and non-CSR schools were engaged in reform, reform in CSR schools was more likely to include adoption of models and other activities closely associated with externally developed models. Furthermore, states that integrated CSR with their standards and testing programs provided more CSR implementation support than those states that did not integrate CSR. Building from these findings, this *Third-Year Report* answers three questions directly related to two of the mandated questions

above—how reform relates to student achievement outcomes and how reform is implemented in CSR and non-CSR schools:

- Was receipt of a CSR award associated with improvements in school-level mathematics and reading achievement?
- Were schools that received CSR awards more likely to implement the legislatively specified components of CSR than other schools?
- Was fidelity of CSR implementation associated with gains in school-level mathematics and reading achievement?

EVALUATION DESIGN

This study focuses on schools nationwide that received their initial CSR awards in 2002, the year most of them began implementing CSR. Awards spanned three years, although some awardees carried over funds for an additional year.

The evaluation includes four methodological approaches.

- Multivariate statistical analyses comparing all Title I CSR with Title I non-CSR matched comparison schools to determine whether receiving a CSR award was related to school-level achievement increases over time.[3]
- Qualitative case study analyses of CSR reform implementation in 15 pairs of schools (15 CSR schools and 15 matched comparison schools) to illustrate the ways CSR components were implemented in both CSR and non-CSR schools.
- Quantitative descriptive analyses of CSR reform implementation from a survey of principals and teachers in a random sample of 500 CSR and 500 matched non-CSR comparison schools to relate the comprehensiveness of implementation to achievement.[4]
- Multivariate statistical analyses of the universe of 2002 CSR awardees to examine the relationship between scientific research-based model adoption and achievement.

[3] The achievement analyses rely on standardized school-level achievement scores from the National Longitudinal School-Level State Assessment Score Database (NLSLSASD) maintained by the American Institutes for Research (AIR). The NLSLSASD includes school-level measures of student achievement for nearly all public schools. (For more information, see http://www.schooldata.org/.)

[4] Most CSR awards were given to elementary and middle schools. Consequently, the sample included too few high schools for analysis.

Only a small number of high schools received CSR awards in 2002. Consequently, the sample included too few high schools for analysis, so this report does not focus on them.[5] This study frequently found differences in CSR implementation and outcomes between elementary and middle schools. These include differences in the relationships between achievement and award receipt and component implementation. Such differences are possibly related to the different structures in the two types of schools, with middle schools having subject-area departments and elementary schools mainly teaching students in self-contained classrooms. These structural differences may affect the nature of the reform approach each level of school adopts, as well as how well reforms are implemented across all grades, subjects, and teachers. Consequently, the report includes separate findings for elementary and middle schools.

Frequently, this study examined the differences between low-performing schools and the remaining schools in the study. Low-performing schools, defined as those in bottom 25 percent of achievement in 2001–02 (the baseline year in this study), are of particular interest to policymakers given the recent focus on schools identified for restructuring under *No Child Left Behind*.

FINDINGS

Was Receipt of a CSR Award Associated With Improvements in School-level Mathematics and Reading Achievement?

Receipt of a CSR award was not associated with achievement gains in mathematics or reading achievement through the first three years of award.

While CSR schools made small but statistically significant gains in elementary mathematics and reading achievement during the time they were implementing their award, these changes were not statistically different from achievement changes in comparison schools (Exhibit E.2). Over the course of their three-year awards, CSR schools were no better at improving achievement than Title I schools that were similar in demographics and achievement in the baseline year.

[5] For more information on the number of high schools versus elementary and middle schools that received CSR awards, see Exhibit B-2 on page 57 of *Longitudinal Assessment of Comprehensive School Reform Program Implementation and Outcomes: First-Year Report* (Tushnet, Flaherty, and Smith, 2004).

Exhibit E.2
Average Gains in Percent Proficient or Higher from 2003 to 2005 for CSR and Non-CSR Schools

	N	Average Gain in Percent Proficient		
		CSR Schools	non-CSR Schools	Difference
Elementary Mathematics	634	0.96+	1.44**	n.s.
Elementary Reading	638	1.12*	1.12*	n.s.
Middle School Mathematics	318	n.s.	1.44**	n.s.
Middle School Reading	320	n.s.	n.s.	n.s.

Exhibit highlights: While CSR schools made small but statistically significant gains in elementary mathematics and reading achievement during the time they were implementing their award, these changes were not statistically different from achievement changes in non-CSR schools.

Note: "n.s." indicates not statistically significant; + p<.10; * p<.05; ** p<.01. Tests for the statistical significance of achievement gains for CSR and non-CSR schools are t-tests to assess whether the value is different from zero. The differences between CSR and non-CSR schools are assessed through paired t-tests. Values for percent proficient or higher are computed by multiplying the average standard deviation of the achievement measures by the statistically significant estimates for the relationships between scientifically based research model adoption and school-level achievement.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Were Schools That Received CSR Awards More Likely to Implement the Legislatively Specified Components of CSR Than Other Schools?

No, both CSR and non-CSR schools implemented an average of fewer than four components in 2003 and fewer than five in 2005 at both the elementary and middle school levels.

CSR awards were intended to stimulate the implementation of the 11 components identified in *NCLB*. However, as shown in Exhibit E.3, by 2005, CSR schools, on average, implemented fewer than five components after three years of their CSR awards. Furthermore, implementation of these components at the elementary school level was nearly identical in both CSR and non-CSR schools. In 2003, CSR elementary schools reported implementing slightly more components than non-CSR schools; this difference was statistically significant. However, this difference was no longer statistically significant by 2005. No significant differences existed in either 2003 or 2005 in the number of reform components implemented by CSR and non-CSR middle schools. Both CSR and non-CSR schools increased the number of reform components implemented from 2003 to 2005.

Exhibit E.3
Average Number of CSR Components Implemented by CSR and Non-CSR Schools in 2003 and 2005

	Average Number of Components Implemented					
	Elementary School			Middle School		
School Type	2003	2005	Change from 2003 to 2005	2003	2005	Change from 2003 to 2005
CSR Schools	3.8	4.7	0.9**	2.7	4.0	1.3**
Non-CSR Schools	3.3	4.4	1.1**	3.1	4.4	1.3**
Difference Between CSR Schools and Non-CSR Schools	0.5*	0.3	-0.2	0.4	0.4	0.0

Exhibit highlights: In 2003, CSR elementary schools reported implementing a somewhat higher average number of components than non-CSR schools. No other significant differences between CSR and non-CSR schools exist. Both CSR and non-CSR schools reported similar increases between 2003 and 2005 in the number of components implemented.

Note: + p<.10; * p<.05; ** p<.01. N = 292 for CSR elementary schools; N = 304 for non-CSR elementary schools; N = 128 for CSR and non-CSR middle schools.

Source: ECSRIO surveys.

Was Fidelity of CSR Implementation Associated With Gains in School-level Mathematics and Reading Achievement?

The comprehensiveness of implementation was not related to mathematics and reading achievement gains between 2002 and 2005.

While only one-third of schools implemented reform models with limited scientific research backing of effectiveness, those doing so were likely to experience achievement gains in middle school mathematics. There was also some evidence, albeit weaker, that schools adopting models with limited scientific evidence may have experienced achievement gains in middle school reading.

Additionally, low-performing schools that adopted CSR models with scientific evidence of effectiveness did improve in elementary mathematics.

The CSR program is based on two key implementation concepts: Schools implementing reforms that encompass the 11 components indicated in *NCLB* will have higher achievement than schools that do not; and schools implementing reform models with a scientific research base will have higher achievement than schools implementing models that lack a scientific research base.

Two different analyses, included in the next two sections, assess the relationship between implementation and achievement. The first analyzes the relationship between the comprehensiveness of CSR implementation, as measured by the number of components a school

implemented, and achievement outcomes. The second analyzes the relationship between the adoption of a scientifically based research model and achievement.

Comprehensiveness of CSR Reform Implementation and Achievement Outcomes

The comprehensiveness of CSR reform implementation, as measured by the number of CSR components implemented, was not related to achievement gains in mathematics and reading at the elementary or middle school levels. There were no statistically significant relationships between the number of components implemented and achievement gains from 2002 to 2005. This finding could be due to the overall low levels of implementation of the components; that is, since most schools implemented relatively few of the components, it is difficult to detect statistically a consistent relationship between component implementation and achievement gains.

Adoption of Scientifically Based Research Models and Achievement Outcomes

This evaluation used information from the Comprehensive School Reform Quality Center (CSRQ) ratings[6] of the scientific research base in commonly adopted reform models to assess the extent to which the use of CSR approaches with such a base was associated with higher achievement. The CSRQ reports were not available at the time that 2002 CSR awardees made their model selections; however, schools may have used data from two earlier reviews of CSR models: Herman (1999), and Borman, Hewes, Overman, and Brown (2003). The CSRQ reports draw heavily from the data used in previous reviews of CSR models. Of the 40 ratings conducted, 25 models were rated as having some evidence of effectiveness. CSRQ relies on the number and quality of research studies of the models to develop its ratings. However, only one-third of CSR schools adopted models that later received ratings from CSRQ indicating that they had a scientific research base. Most schools chose models that were not rated by CSRQ.

In general, the results demonstrate stronger relationships between the scientific research bases of comprehensive school reform models and mathematics achievement gains compared to reading achievement gains. Mathematics achievement improved in low-performing elementary schools,[7] as well as in all middle schools (including low-performing schools), that adopted models with scientific research bases. Mathematics achievement improved in low-performing CSR elementary schools that adopted models with scientific research bases, while it did not improve in non-low-performing CSR elementary schools (Exhibit E.4).[8] Low-performing CSR elementary schools gained about 2 percentage points in the percent of students proficient or

[6] The CSRQ ratings provide a scale for the breadth and quality of the research base for the 31 most widely adopted CSR models in elementary, middle, and high schools. The CSRQ Center conducted a total of 40 ratings because some models were rated at both the elementary and secondary levels. The categories are Very Strong, Moderately Strong, Moderate, Limited, Zero, and Negative. The CSRQ Center was funded under the U.S. Department of Education's Comprehensive School Reform Quality Initiatives program.

[7] Low-performing schools refer to those CSR schools that were in the lowest 25 percent of achievement of all CSR schools in 2001-02.

[8] The CSRQ rating scale includes Very Strong, Moderately Strong, Moderate, Limited, Zero, and Negative categories. No model received a "Negative" or "Very Strong" rating. The analyses that examine the relationship between strength of the research base and achievement treat scientific research base as a series of dummy coded variables for Limited and Moderate/Moderately Strong CSRQ ratings.

higher in mathematics if they adopted a model with a limited scientific research base and 4 percentage points if they adopted a model with a moderate or moderately strong scientific research base. There was no relationship between changes in elementary reading achievement and the selection of a model with a scientific research base.

At the middle school level, all CSR schools (including low-performing schools) that adopted models with limited scientific research bases realized gains in the percent proficient or higher of about 6 percentage points in mathematics and 3 percentage points in reading, although the relationship with reading achievement is only weakly significant.[9] Furthermore, low-performing CSR middle schools that adopted a model with a moderate or higher research base had gains of about 4 percentage points in the percent proficient or higher in mathematics, although this relationship is also only weakly significant.

[9] Throughout this report, findings that are reported as weakly significant are statistically significant at $p<.10$.

Exhibit E.4
Average Gains in Percent Proficient or Higher from 2002 to 2005, by Subject Area and Strength of Scientific Research Base

	Average Gains in Percent Proficient or Higher			
	Elementary School		Middle School	
	Mathematics	Reading	Mathematics	Reading
All CSR Schools				
Limited Scientific Research Base	n.s.	n.s.	5.92**	3.36+
Moderate/Moderately Strong Research Base	n.s.	n.s.	n.s.	n.s.
Low-performing Schools				
Limited Scientific Research Base	1.92**	n.s.	5.92**	3.36+
Moderate/Moderately Strong Research Base	4.00*	n.s.	4.16+	n.s.

Exhibit highlights: At the elementary level, low-performing CSR schools that adopted models with limited or moderate to moderately strong scientific research bases had higher mathematics achievement; there was no relationship between reading achievement and models with a scientific research base. At the middle school level, CSR schools that adopted models with limited scientific research bases had higher mathematics and reading achievement gains in the percent proficient or higher than those that did not; low-performing CSR schools that adopted models with moderate to moderately strong research bases had higher mathematics achievement than those that chose other models.

Note: n.s. indicates not statistically significant; + p<.10; * p<.05; ** p<.01. Values for percent proficient or higher are computed by multiplying the average standard deviation of the achievement measures by the statistically significant estimates for the relationships between scientifically based research model adoption and school-level achievement.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

One possible explanation for these inconsistent findings lies with the way CSRQ determined ratings. CSRQ determined "strength of the research base" by the nature of the evaluation design, with more weight given to more rigorous designs and the number of models evaluated that use such designs. It is possible that some highly potent strategies had not been evaluated rigorously, so the ratings underplay their potential for effectiveness. It is also possible that implementation in schools in this study may have been less comprehensive than it was in schools included in the studies that CSRQ used.

SUMMARY

These third-year evaluation findings demonstrate both the complexity of implementing CSR and a weak and uneven relationship between implementation and achievement. The relationship

varies by content area (mathematics or reading) and by school level (elementary or middle school). The findings indicate that as a reform strategy, comprehensive school reform grants were not systematically associated with gains in either student reading or mathematics achievement. It also indicates that relatively few schools implemented CSR as intended by the legislation and that by the third year of implementation (2005), CSR reform components were about equally likely to be present in non-CSR as in CSR schools.

CSR provided funds for schools to implement comprehensive reform using scientifically based research strategies. However, relatively few of the schools adopted reforms backed by scientifically based research. When they did, the findings provide preliminary evidence of the advantage of adopting scientifically research-based models, although the pattern of findings was not consistent across all school levels and subject areas.

I. INTRODUCTION

The Comprehensive School Reform (CSR) program was established as a demonstration program in 1998 and authorized as a full program in 2002 as part of the *No Child Left Behind Act (NCLB)*. It is one approach to help low-performing K–12 public schools meet state performance standards. CSR emphasizes two major concepts. First, the approach mandates that school reform should be *comprehensive* in nature, strengthening all aspects of school operations—curriculum, instruction, professional development, parental involvement, and school organization. Second, CSR involves the use of *scientifically based research models*—that is, models with evidence of effectiveness in multiple settings.

Originally funded as the Comprehensive School Reform Demonstration (CSRD) Program in 1998 with $145 million, the program became part of *NCLB* in 2002 with funding of $310 million. In FY 2003 and FY 2004, Congress allocated $308 million for CSR, and in FY 2005, $205 million. Between 1998 and 2006, nearly 7,000 schools nationwide received three-year awards to implement CSR models. In FY 2006, funding was appropriated only for a clearinghouse to support comprehensive school reform and not for school-based activities.

NCLB defines CSR as containing 11 components, which are assumed to work together as schools undergo reform (Exhibit 1). For example, the "support from staff" component of CSR dictates that the entire school should adopt the reform model, while the "support for staff" component focuses on helping teachers and other staff members learn to use the reform strategy. This report studies these components, their implementation, and their relationship to achievement.

In addition to implementing the 11 components, CSR schools are expected to use reform models with a strong scientific research base. According to Hale (2000), the unique aspect of CSR is its expectation that schools will collaborate with expert partners to implement research-based, whole-school reform methods with a successful replication record. This report examines the relationships between schools adopting models with strong scientific research bases and achievement.

Exhibit 1
Eleven Components of Comprehensive School Reform
Described in the *No Child Left Behind Act*

1. **Proven methods** and strategies for student learning, teaching, and school management that are based on scientifically based research and effective practices, and have been replicated successfully in schools with diverse characteristics.
2. **Comprehensive design** for effective school functioning, integrating instruction, assessment, classroom management, and professional development and aligning these functions into a schoolwide reform plan designed to enable all students to meet challenging state content and performance standards and address needs identified through a school needs assessment.
3. **Professional development.** High-quality and continuous teacher and staff professional development and training.
4. **Measurable goals** for student performance and benchmarks for meeting those goals.
5. **Support from staff.** Support from school faculty, administrators, and staff.
6. **Support for staff.** Support for school faculty, administrators, and staff (added in 2001).
7. **Parent and community involvement.** Meaningful involvement of parents and the local community in planning and implementing school improvement activities.
8. **External assistance.** High-quality external support and assistance from a comprehensive school reform entity (which may be a university) with experience in schoolwide reform and improvement.
9. **Evaluation.** Plan to evaluate the implementation of school reforms and the student results achieved.
10. **Coordination of resources.** Identification of how other available resources (federal, state, local, and private) help the school coordinate services to support and sustain the school reform.
11. **Scientifically based research.** Scientifically based research to significantly improve the academic achievement of students participating in such programs as compared with students in schools who have not participated in such programs; or strong evidence that such programs will significantly improve the academic achievement of participating children (added in 2001).

Source: *No Child Left Behind Act*, Title I, Part F, Section 1606.

In 2002, the U.S. Department of Education contracted with WestEd and COSMOS Corp. to conduct a five-year study of the CSR program. The Evaluation of the Comprehensive School Reform Program Implementation and Outcomes (ECSRIO), mandated by Section 1606 of *NCLB*, addresses four broad questions related to the CSR program.

- How were CSR funds targeted?
- How was reform implemented in CSR and non-CSR schools?
- How did state and district conditions influence reform implementation?
- How was reform related to achievement outcomes?

ECSRIO examines the implementation and outcomes of a cohort of CSR awardees that received their initial awards in 2002. The First-Year Report[10] (2004) addressed the first three questions. It concluded CSR funds primarily targeted low-performing, high-poverty elementary and middle schools in rural and urban areas. Although both CSR and non-CSR schools were engaged in reform, reform in CSR schools included adoption of models and other activities closely

[10] The *Second-Year Report* was an internal document to the U.S. Department of Education.

associated with externally developed models. Furthermore, states that integrated CSR with their standards and testing programs provided more support for CSR implementation than those without this integration. Building from these findings, this *Third-Year Report* answers questions directly related to two of the mandated questions—how reform related to achievement outcomes and how reform was implemented in CSR and non-CSR schools:

- Was receipt of a CSR award associated with improvements in school-level mathematics and reading achievement?
- Were schools that receive CSR awards more likely to implement the legislatively specified components of CSR than other schools?
- Was fidelity of CSR implementation associated with gains in school-level mathematics and reading achievement?

RELEVANT PAST RESEARCH

The remainder of this introduction reviews the findings from the first two years of the evaluation and other related research efforts.

Review of Findings From the First and Second Years of the Study

The first- and second-year reports addressed the first three study questions listed on page 2. The major findings, organized by study question, are:

How were CSR funds targeted?

States targeted CSR funds largely to high-poverty and low-performing urban and rural elementary and middle schools.[11] Most states helped these schools identify appropriate research-based methods and with applications for grant awards.

How was reform implemented in CSR and non-CSR schools?

Both CSR and non-CSR schools exhibited many aspects of comprehensive reform. However, CSR schools were more likely to adopt externally developed models than non-CSR schools.

How did state and district conditions influence reform implementation?

In nearly three-fourths of the states (72 percent), state CSR coordinators saw the program as a way to help schools meet state standards and succeed in state testing programs.

[11] Because only a small number of high schools received CSR awards, these reports focus only on elementary and middle schools.

States that aligned CSR with state standards and testing programs provided more support for implementation than states without such alignment. When states used CSR to advance statewide reform priorities, they more often targeted CSR funds to specific sites, provided support for schools, and maintained higher accountability standards. In 26 percent of states, coordinators indicated the absence of a state reform agenda, thereby relying on federal standards or guidelines. Such states provided less support and were less likely to monitor or evaluate CSR efforts than other states.

Other Research Efforts and Relationship to ECSRIO

There are numerous studies of individual CSR models, such as Success for All (Borman et al., 2005a, 2005b) and America's Choice (May and Supovitz, 2006). In addition, Borman, Hewes, and Overman (2003) conducted a meta-analysis of studies of 29 individual CSR models. Most recently, the American Institutes for Research's (AIR's) National Longitudinal Evaluation of Comprehensive School Reform (NLECSR) examined implementation and outcomes of eight models over a six-year period. The following summarizes the NLECSR findings.

Aladjem et al. (2006) reported on a six-year mixed-method study of eight CSR models in 650 elementary and middle schools. NLECSR addressed questions about the relationship between the characteristics of specific CSR models and individual student and school-level achievement.[12] The 650 schools were a sample of schools that received a CSR award between 1999 and 2001 and implemented one of the eight most popular models. The NLECSR included survey data from district administrators, principals, and teachers; achievement analyses from both CSR schools and matched non-CSR comparison schools that did not receive a CSR award; and case studies of 34 schools.

Aladjem et al. (2006) measured implementation as the difference between what the model developers consider to be full implementation and what the school actually does. To do so, they surveyed each of the model developers and asked them to respond as if they were a school implementing their model. The researchers compared each school's responses to this measure of implementation to develop a measure of fidelity. In order to measure implementation in the comparison schools, Aladjem et al. predicted what CSR model the comparison school would have chosen based on its school characteristics and those of the CSR schools. The researchers then compared the school's responses against those of the model developer of the predicted CSR model. This measure of implementation is consistent with the focus on models as compared with this study's focus on the 11 components identified in *NCLB*.

CSR schools and their matched comparisons in NLECSR differed little in the level of implementation. Furthermore, implementation was infrequently comprehensive in CSR schools, with only about 20 percent of the schools in 2002 (and 10 percent in 2004) implementing a CSR model comprehensively. However, CSR schools were more likely to implement CSR comprehensively than their comparison schools. One predictor of the level of implementation was the CSR model chosen, and this relationship did not change over time. That is,

[12] Student-level achievement data were available for a subset of five school districts during the study period.

implementation in schools with highly prescriptive models (e.g., Success for All) was higher than in schools with other models.

In the Aladjem et al. (2006) study, CSR schools with higher levels of implementation had higher levels of achievement gains than the matched comparison schools. This is particularly apparent when implementation was high three to five years after initially receiving a CSR award, and when high implementation was comprehensive (as opposed to focused on a few specific aspects of the model). Furthermore, models matter in that schools that selected Success for All typically had higher levels of implementation and achievement than schools that selected other models. However, achievement results of several other models (Accelerated Schools Project, ATLAS Communities, and Co-Nect) indicate that comprehensive implementation of these models consistent with the model developers' visions may result in higher achievement in the long term.

Zhang, Fashola, Shkolnik, and Boyle (2006) reanalyzed the NLECSR implementation and achievement data and largely came to the same conclusions in the NLECSR final report. Because some schools in the NLECSR study switched CSR models, and a number of comparison schools adopted CSR models, Zhang et al. used a smaller sample of 115 schools. They analyzed the relationship between implementation and achievement, using school-level mathematics and reading scores. Achievement gains (relative to comparison schools) were largest in CSR schools between years 3 and 5 in implementation. Zhang et al. also found the level of implementation is marginally associated with gains in achievement and confirmed the NLECSR finding that models matter, with similar patterns of results as those reported by AIR.

In contrast to these studies, ECSRIO focuses on how funds allocated through the CSR program were used. Consequently, the schools in this report did not necessarily adopt a single model, and the measure of implementation is different, focusing on fidelity to the 11 legislatively identified components of reform rather than to a particular model. Similarly, the focus on how awards were used by a representative sample of recipients allows policymakers to view the program as it was actually implemented. As such, this study provides insight into the extent to which comprehensive school reform prompts changes in school operations and student outcomes. This focus on the program in operation limits any causal claims. However, it provides meaningful and policy-relevant insights regarding the value and limitations of a broad program design in stimulating school reform nationally.

ORGANIZATION OF THE THIRD-YEAR REPORT

The next chapter describes the data sources used in the ECSRIO. The subsequent three chapters report on: the overall relationship between CSR award and achievement (Chapter III), the comprehensiveness of CSR implementation and its relationship to achievement (Chapter IV), and CSR schools' adoption of scientifically based research models and its relationship to achievement (Chapter V). These three chapters that report findings include discussions of methodological approaches because each question relies on different methodologies and on samples constructed in different manners. The final chapter presents a brief summary of key findings. The report appendixes include the CSRQ Center Quality Review Tool (Appendix A), standard error tables (Appendix B), and data collection instruments (Appendix C).

II. DATA SOURCES

This study compares implementation and school-level achievement between schools that received CSR awards and similar schools that did not receive such awards. Comparisons were done by employing multiple research methods for collecting and organizing data, and incorporating information from teachers and principals about reform in their schools.

- Multivariate statistical analyses comparing all Title I CSR with Title I non-CSR matched comparison schools to determine whether receiving a CSR award was related to school-level achievement increases over time.[13]

- Qualitative case study analyses of CSR reform implementation in 15 pairs of schools (15 CSR schools and 15 matched non-CSR comparison schools) to illustrate the ways CSR components were implemented in both CSR and non-CSR schools.

- Quantitative descriptive analyses of CSR reform implementation from a survey of principals and teachers in a random sample of 500 CSR and 500 matched non-CSR comparison schools to relate the comprehensiveness of implementation to achievement.[14]

- Multivariate statistical analyses of the universe of 2002 CSR awardees to examine the relationship between scientific research-based model adoption and achievement.

Using multiple methods increases the validity of key measures and results. In addition, this is an evaluation of a program as it operates in the field. Such an evaluation cannot control the conditions under which activities take place. Consequently, the study cannot use an experimental design, and no causal claims can be made. The evaluation uses a quasi-experimental design, matching schools that received CSR funds with schools that did not receive funding, in both the quantitative and qualitative aspects of data collection and analysis.

This section describes the samples and the three following data sources for ECSRIO.

- School-level achievement measures
- Surveys of school reform activities
- Case studies of school reform activities

[13] The achievement analyses rely on standardized school-level achievement scores from the National Longitudinal School-Level State Assessment Score Database (NLSLSASD) maintained by the American Institutes for Research (AIR). The NLSLSASD includes school-level measures of student achievement for nearly all public schools. (For more information, see http://www.schooldata.org/.)

[14] Most CSR awards were given to elementary and middle schools. Consequently, the sample included too few high schools for analysis.

Because each evaluation question relies on different methods and samples constructed in different ways, discussions of the specific methods used are included with the findings for each evaluation question. The last part of this section presents study limitations.

SCHOOL-LEVEL ACHIEVEMENT MEASURES

The achievement analyses rely on standardized school-level achievement scores from the National Longitudinal School-Level State Assessment Score Database (NLSLSASD) maintained by the American Institutes for Research (AIR). The NLSLSASD includes school-level measures of achievement for nearly all public schools. (For more information, see http://www.schooldata.org/.) This database includes seven years of data, spanning 1998–99 through 2004–05; however, many states did not have or provide assessment data for the first year or two. Also, prior to 2003–04, nearly all states only tested a few grades. Therefore, for this study, one grade (typically fourth) was chosen to represent elementary achievement, and one grade (typically eighth) was chosen to represent middle school achievement in the school. Exhibit 2 details each state's assessment measures used to compute the standardized achievement scores by year.

In most instances, the average scale score or percent proficient or higher was used to calculate a standardized z-score[15] within each state for each year. In several instances, however, other measures, such as percentile ranks, were the only available outcome measures and, as such, standardization was done in the following manner:

1. In each year, appropriate assessment measures for mathematics and reading or language arts in elementary and middle grades were selected.

2. In many cases, fourth-grade scores were used for elementary schools and eighth-grade scores for middle schools; however, when those scores were not available, other proximate grades were used.

3. The schools in each state in each year were ranked according to their achievement measure.

4. Percentile ranks were computed from these rankings.

5. Percentile ranks were converted to normal curve equivalent (NCE) scores.

6. NCE scores were converted to z-scores, so each state was standardized with 0 mean and a standard deviation of 1.

Where no assessment scores were available for a state in an academic year, the standardized score was estimated by averaging the standardized score from the year prior and the year after. For instance, there were no assessment measures available for Idaho in 2004. The elementary

[15] The z-score is calculated by subtracting the population mean from an individual score and dividing it by the population standard deviation.

mathematics score for each school in Idaho in 2004 was computed by averaging the 2003 and 2005 standardized scores in elementary school mathematics for each school.

Exhibit 2
Grade-level Tested and Achievement Measures Used to Compute Standardized Achievement Scores for Elementary and Middle Schools in Mathematics and Reading, by State and Year[16]

State	Year	Mathematics		Reading		Measure
		Elem	Middle	Elem	Middle	
Alaska	2005	4	8	4	8	Percent proficient or above
	2004	3	8	3	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
Alabama	2005	4	6	4	8	Percent meets academic standards
	2004	4	6	4	8	Percent meets academic standards
	2003	4	8	4	8	Percentile rank
	2002	4	8	4	8	Percentile rank
	2001	4	8	4	8	Percentile rank
	2000	4	8	4	8	Percentile rank
	1999	4	8	4	8	Percentile rank
Arkansas	2005	4	8	4	8	Percent proficient or above
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
Arizona	2005	3	8	3	8	Mean Scale Score
	2004	3	8	3	8	Mean Scale Score
	2003	3	8	3	8	Mean Scale Score
	2002	3	8	3	8	Mean Scale Score
	2001	3	8	3	8	Mean Scale Score
	2000	3	8	3	8	Mean Scale Score
California	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Mean Scale Score
	2001	4	8	4	8	Mean Scale Score
	2000	4	8	4	8	Mean Scale Score
	1999	4	8	4	8	Mean Scale Score
	1998	4	8	4	8	Mean Scale Score

[continued on next page]

[16] Note that many states did not have or provide school-level assessment data prior to 2000.

[continued from previous page]

Colorado	2005	4	8	4	8	Percent proficient or above
	2004	5	8	4	8	Percent proficient or above
	2003	5	8	4	8	Percent proficient or above
	2002	5	8	4	8	Percent proficient or above
	2001	5	8	4	8	Percent proficient or above
	2000	NA	8	4	7	Percent proficient or above
Connecticut	2005	4	8	4	8	Percent at goal (level 4)
	2004	4	8	4	8	Percent at goal (level 4)
	2003	4	8	4	8	Percent at goal (level 4)
	2002	4	8	4	8	Percent at goal (level 4)
	2001	4	8	4	8	Percent at goal (level 4)
	2000	4	8	4	8	Percent at goal (level 4)
District of Columbia	2005	3	8	3	8	Percent proficient or higher
	2004	4	8	4	8	Percent proficient or higher
	2003	4	8	4	8	Mean NCE (SAT-9)
	2002	4	8	4	8	Mean NCE (SAT-9)
	2001	4	8	4	8	Mean NCE (SAT-9)
	2000	4	8	4	8	Mean NCE (SAT-9)
Delaware	2005	3	8	3	8	Mean Scale Score
	2004	3	8	3	8	Mean Scale Score
	2003	3	8	3	8	Mean Scale Score
	2002	3	8	3	8	Mean Scale Score
	2001	3	8	3	8	Mean Scale Score
	2000	3	8	3	8	Mean Scale Score
	1999	3	8	3	8	Mean Scale Score
Florida	2005	4	8	4	8	Percent at or above level 3
	2004	4	8	4	8	Percent at or above level 3
	2003	4	8	4	8	Percent at or above level 3
	2002	4	8	4	8	Percent at or above level 3
	2001	5	8	4	8	Percent at or above level 3
	2000	5	8	4	8	Percent at or above level 3
	1999	5	8	4	8	Percent at or above level 3
Georgia	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Percent meeting/exceeding the standard
	2002	4	8	4	8	Percent meeting the standard
	2001	4	8	4	8	Percent meeting the standard
	2000	4	8	4	8	Percent meeting the standard
Hawaii	2005	3	8	3	8	Percent at or above proficiency
	2004	3	8	3	8	Percent at or above proficiency
	2003	3	8	3	8	Percent at or above proficiency
	2002	3	8	3	8	Percent at or above proficiency
	2001	3	8	3	8	Average of 2000 and 2002
	2000	3	8	3	8	Percent at or above proficiency

[continued on next page]

[continued from previous page]

Iowa	2005	4	8	4	8	Percent at or above proficiency
	2004	4	8	4	8	Percent at or above proficiency
	2003	4	8	4	8	Percent at or above proficiency
Illinois	2005	3	8	3	8	Percent at or above proficiency
	2004	3	8	3	8	Percent at or above proficiency
	2003	3	8	3	8	Percent at or above proficiency
	2002	3	8	3	8	Percent at or above proficiency
	2001	3	8	3	8	Percent at or above proficiency
	2000	3	8	3	8	Percent at or above proficiency
	1999	3	8	3	8	Percent at or above proficiency
Indiana	2005	4	8	4	8	Percent at or above proficiency
	2004	3	8	3	8	Percent at or above proficiency
	2003	3	8	3	8	Percent at or above proficiency
	2002	3	8	3	8	Percent at or above proficiency
	2001	3	8	3	8	Percent at or above proficiency
	2000	3	8	3	8	Percent at or above proficiency
Idaho	2005	4	8	4	8	Percent at or above proficiency
	2004	(1)	(1)	(1)	(1)	
	2003	4	8	4	8	Percent at or above proficiency
	2002	4	8	4	8	Percentile rank
	2001	4	8	4	8	Percentile rank
	2000	4	8	4	8	Percentile rank
Kansas	2005	4	7	5	8	Percent at or above proficiency
	2004	4	7	5	8	Percent at or above proficiency
	2003	4	7	5	8	Percent at or above proficiency
	2002	4	7	5	8	Percent at or above proficiency
	2001	4	7	5	8	Percent at or above proficiency
	2000	4	7	5	8	Percent at or above proficiency
Kentucky	2005	5	8	4	7	Mean Scale Score
	2004	5	8	4	7	Mean Scale Score
	2003	5	8	4	7	Mean Scale Score
	2002	5	8	4	7	Mean Scale Score
	2001	5	8	4	7	Mean Scale Score
	2000	5	8	4	7	Mean Scale Score
	1999	5	8	4	7	Mean Scale Score
Louisiana	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Mean Scale Score

[continued on next page]

[continued from previous page]

Massachusetts	2005	4	8	4	8	Percent at or above proficiency
	2004	4	8	4	8	Percent at or above proficiency
	2003	4	8	4	8	Percent at or above proficiency
	2002	4	8	4	8	Mean Scale Score
	2001	4	8	4	8	Mean Scale Score
	2000	4	8	4	8	Mean Scale Score
	1999	4	8	4	8	Mean Scale Score
	1998	4	8	4	8	Mean Scale Score
Maryland	2005	4	8	4	8	Percent at or above proficiency
	2004	4	8	4	8	Percent at or above proficiency
	2003	3	8	3	8	Percent at or above proficiency
	2002	3	8	3	8	Percent at or above proficiency
	2001	3	8	3	8	Percent Satisfactory
	2000	3	8	3	8	Percent Satisfactory
	1999	3	8	3	8	Percent Satisfactory
	1998	3	8	3	8	Percent Satisfactory
Maine	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Mean Scale Score
	2001	4	8	4	8	Mean Scale Score
	2000	4	8	4	8	Mean Scale Score
	1999	4	8	4	8	Mean Scale Score
Michigan	2005	4	8	4	7	Percent at or above proficiency
	2004	4	8	4	7	Percent at or above proficiency
	2003	4	8	4	7	Percent at or above proficiency
	2002	4	8	4	7	Percent at or above proficiency
	2001	4	(1)	4	7	Percent at or above proficiency
	2000	4	7	4	7	Percent at or above proficiency
Minnesota	2005	3	7	3	7	Mean Scale Score
	2004	3	7	3	7	Mean Scale Score
	2003	3	(1)	3	(1)	Mean Scale Score
	2002	3	8	3	8	Mean Scale Score
	2001	3	8	3	8	Mean Scale Score
	2000	3	8	3	8	Mean Scale Score
	1999	3	8	3	8	Mean Scale Score
Missouri	2005	4	8	3	7	Mean Scale Score
	2004	4	8	3	7	Mean Scale Score
	2003	4	8	3	7	Mean Scale Score
	2002	4	8	3	7	Achievement level
	2001	4	8	3	7	Mean Scale Score
	2000	4	8	3	7	Mean Scale Score
	1999	4	8	3	7	Mean Scale Score

[continued on next page]

[continued from previous page]

Mississippi	2005	4	8	4	8	Percent at or above proficiency
	2004	4	8	4	8	Percent at or above proficiency
	2003	4	8	4	8	Percent at or above proficiency
	2002	4	8	4	8	Percent at or above proficiency
	2001	4	8	4	8	Percent at or above proficiency
Montana	2005	4	8	4	8	Percent at or above proficiency
	2004	4	8	4	8	Percent at or above proficiency
	2003	4	8	4	8	Percent at or above proficiency
	2002	4	8	4	8	Percent at or above proficiency
	2001	4	8	4	8	Percent at or above proficiency
North Carolina	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Mean Scale Score
	2001	4	8	4	8	Mean Scale Score
	2000	4	8	4	8	Mean Scale Score
	1999	4	8	4	8	Mean Scale Score
North Dakota	2005	4	8	4	8	Percent meeting the standard
	2004	4	8	4	8	Percent meeting the standard
	2003	4	8	4	8	Percent meeting the standard
Nebraska	Omitted because Nebraska only tested in a few of the years and, in 2005, used a local assessment.					
New Hampshire	2005	NA	NA	NA	NA	
	2004	3	6	3	6	Mean Scale Score
	2003	3	6	3	6	Mean Scale Score
	2002	3	6	3	6	Mean Scale Score
	2001	3	6	3	6	Mean Scale Score
	2000	3	6	3	6	Mean Scale Score
	1999	3	6	3	6	Mean Scale Score
New Jersey	2005	4	8	4	8	Percent proficient or above (4^{th} grade mathematics is percent advance proficient)
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percent proficient or above
New Mexico	2005	4	8	4	8	Percent proficient or above
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Mean Scale Score

[continued on next page]

[continued from previous page]

Nevada	2005	3	8	3	8	Achievement level (percent at or above level 3)
	2004	3	8	3	8	Achievement level (percent at or above level 3)
	2003	3	8	3	8	Mean Scale Score
	2002	3	(1)	3	(1)	Elementary: Mean Scale Score
	2001	4	8	4	8	Mean Scale Score
	2000	4	8	4	8	Mean Scale Score
New York	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Mean Scale Score
	2001	4	8	4	8	Mean Scale Score
	2000	4	8	4	8	Mean Scale Score
	1999	4	8	4	8	Mean Scale Score
Ohio	2005	4	6	4	6	Percent proficient or above
	2004	4	6	4	6	Percent proficient or above
	2003	4	6	4	6	Percent proficient or above
	2002	4	6	4	6	Percent proficient or above
	2001	4	6	4	6	Percent proficient or above
	2000	4	6	4	6	Percent proficient or above
	1999	4	8	4	8	Percent proficient or above
Oklahoma	2005	4	8	4	8	Percent proficient or above
	2004	NA	NA	NA	NA	
	2003	5	8	5	8	Percent proficient or above
	2002	NA	NA	NA	NA	
	2001	5	8	5	8	Percent proficient or above
	2000	5	8	5	8	Percent proficient or above
Oregon	2005	3	8	3	8	Percent proficient or above (Elementary: Percent advanced)
	2004	3	8	3	8	Percent proficient or above
	2003	3	8	3	8	Percent proficient or above
	2002	3	8	3	8	Percent proficient or above
	2001	3	8	3	8	Percent proficient or above
	2000	3	8	3	8	Percent proficient or above
	1999	3	8	3	8	Percent proficient or above
Pennsylvania	2005	5	8	5	8	Percent proficient or above
	2004	5	8	5	8	Percent proficient or above
	2003	5	8	5	8	Percent proficient or above
	2002	5	8	5	8	Percent proficient or above
	2001	5	8	5	8	Percent proficient or above

[continued on next page]

[continued from previous page]

Rhode Island	2005	NA	NA	NA	NA	
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
	1999	4	8	4	8	Percent proficient or above
South Carolina	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
	1999	4	8	4	8	Percent proficient or above
South Dakota	2005	NA	NA	4	NA	
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percentile rank
	2001	4	8	4	8	Percentile rank
Tennessee	2005	4	8	4	8	Percent proficient or above
	2004	(1)	(1)	(1)	(1)	
	2003	4	8	4	8	Percentile rank
	2002	4	8	4	8	Percentile rank
	2001	4	8	4	8	Percentile rank
	2000	4	8	4	8	Percentile rank
	1999	4	8	4	8	Percentile rank
Texas	2005	4	8	4	8	Mean Scale Score
	2004	4	8	4	8	Mean Scale Score
	2003	4	8	4	8	Mean Scale Score
	2002	4	8	4	8	Percent mastering
	2001	4	8	4	8	Percent passing
	2000	4	8	4	8	Percent passing
	1999	4	8	4	8	Percent passing
Virginia	2005	3	8	3	8	No variable labels available
	2004	3	8	3	8	No variable labels available
	2003	3	8	3	8	Percent proficient or above
	2002	3	8	3	8	Percent proficient or above
	2001	3	8	3	8	Percent proficient or above
	2000	3	8	3	8	Percent proficient or above
	1999	3	8	3	8	Percent proficient or above

[continued on next page]

[continued from previous page]

Vermont	2005	NA	NA	NA	NA	
	2004	4	8	4	8	Percent achieving standard
	2003	4	8	4	8	Percent achieving standard
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
	1999	4	8	4	8	Percent proficient or above
Washington	2005	4	7	4	7	Achievement Level (Pct. at or above level 3 or 4)
	2004	4	7	4	7	Achievement level (percent at or above level 3)
	2003	4	7	4	7	Index score
	2002	4	7	4	7	Index score
	2001	4	7	4	7	Index score
	2000	4	7	4	7	Index score
	1999	4	7	4	7	Index score
Wisconsin	2005	4	8	4	8	Percent proficient or above
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
	1999	4	8	4	8	Percent proficient or above
West Virginia	2005	4	8	4	8	Percent proficient or above
	2004	4	8	4	8	Percent proficient or above
	2003	(1)	(1)	(1)	(1)	
	2002	Elem	Middle	Elem	Middle	Percent above 3rd quartile; WV only reported scores as elementary or middle grades
Wyoming	2005	4	8	4	8	Percent proficient or above
	2004	4	8	4	8	Percent proficient or above
	2003	4	8	4	8	Percent proficient or above
	2002	4	8	4	8	Percent proficient or above
	2001	4	8	4	8	Percent proficient or above
	2000	4	8	4	8	Percent proficient or above
	1999	4	8	4	8	Percent proficient or above

Exhibit highlights: The achievement measures used in the analyses varied across state and year. As a result, achievement measures were standardized.

Note: "NA" indicates scores not available that year and not imputed from other data. "(1)" indicates scores not available that year, imputed by averaging the standardized scores from the year prior and the year after.

Source: National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Because the standardized scores are standard deviations, coefficients are comparable to effect sizes; that is, coefficients are interpreted as changes in standard deviations of the outcome measure. Coefficients that result from analyses can also be converted to more interpretable

measures, such as percent proficient or higher (Gill et al., 2005). For example, if CSR accounts for an increase in achievement scores of 0.5 standard deviations, this value can be multiplied by the average standard deviation of the percentage of students who achieve at the proficient level or higher across states to develop a rough estimate of the increase in achievement levels associated with CSR. Exhibit 3 presents the standard deviations of elementary and middle school mathematics and reading achievement calculated from the 46 states in which achievement data were expressed as percent proficient or higher in 2005. The average standard deviation for each of the four achievement measures (elementary and middle school mathematics and reading) is approximately 16 percent. Therefore, a finding that CSR is associated with an increase in achievement scores of 0.5 standard deviations corresponds to an increase of about 8 percentage points.

Exhibit 3
Standard Deviations of Percent Proficient or Higher for 2005 Elementary and Middle School Achievement, by State

State	Elementary School		Middle School	
	Mathematics	Reading	Mathematics	Reading
AK	17.9	19.3	17.7	21.4
AL	15.5	10.8	16.9	15.0
AR	19.5	17.3	14.5	15.1
AZ	18.6	23.2	20.2	22.7
CA	19.7	19.9	21.2	20.3
CO	19.6	19.5	20.6	19.7
CT	20.2	21.4	22.0	20.0
DC	19.6	21.6	21.0	23.9
DE	14.6	26.8	12.1	18.0
FL	17.2	14.8	21.4	18.6
GA	13.9	17.2	9.2	13.2
HI	14.8	16.2	9.9	14.4
IA	11.1	11.1	10.8	11.1
ID	8.0	9.0	13.4	10.6
IL	18.0	19.7	22.7	15.9
IN	14.1	16.4	14.0	14.6
KS	13.6	13.6	16.6	12.8
KY	16.8	14.8	15.1	14.0
LA	19.2	27.5	18.1	24.1
MA	20.1	20.5	20.4	20.7
MD	17.0	13.9	22.8	20.1
ME	16.9	14.3	16.3	19.0
MI	19.2	14.5	20.9	17.6
MN	16.5	18.1	15.5	16.5

[continued on next page]

[continued from previous page]

MO	16.9	10.0	15.6	14.2
MS	14.4	8.2	16.8	15.1
MT	14.1	13.7	16.7	16.9
NC	7.3	15.8	10.7	12.3
ND	13.1	13.9	16.1	13.7
NJ	16.0	14.3	22.5	20.0
NM	17.7	18.0	13.4	15.2
NV	16.1	8.5	15.9	17.6
NY	19.7	8.9	14.3	8.5
OH	19.0	17.1	22.7	15.8
OK	15.0	16.1	14.8	13.3
OR	17.2	17.4	17.0	16.9
PA	19.1	19.8	19.6	18.5
SC	16.8	15.7	12.2	13.0
SD	NA	NA	15.2	NA
TN	17.5	15.9	17.3	15.7
TX	12.3	19.3	11.8	11.8
VA	8.8	11.6	22.9	20.2
WA	17.7	14.6	17.1	15.5
WI	17.1	13.3	18.5	13.9
WV	11.9	9.9	9.7	8.9
WY	17.8	16.3	15.9	14.3
Average	16.2	16.0	16.7	16.2

Exhibit highlights: The average standard deviation of all of the states was between 16.0 and 16.7 percent across school level and mathematics and reading subject areas.

Note: NA indicates not available.

Source: National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

SURVEY OF SCHOOL REFORM ACTIVITIES

The implementation analyses and the analyses of the relationship between reform implementation and outcomes rely on survey data from a random sample of CSR and matched comparison non-CSR schools developed at the outset of this study. The survey provides descriptive quantitative data on reform implementation—that is, the extent that schools implemented the 11 elements of CSR included in *NCLB* as well as the extent that schools engaged in other non-CSR reform-related activities. The survey also examines other factors that research associates with successful reform, such as school organization (e.g., instructors teach all subjects versus the same subject to most classes; common planning periods).

The evaluation team ensured the sample of schools in the ECSRIO survey represented the population of schools by randomly selecting a large enough number of CSR awardees for the survey sample. When the sample for the study was drawn, the initial CSR universe for calendar year 2002 comprised 38 states and 1,096 schools. The sample of 400 schools was about 36 percent of this universe, representing 35 states. The evaluation team then updated the sample based on new data, increasing the sample by 100 schools.

In order to select the sample, researchers obtained the most complete list of CSR schools from the U.S. Department of Education's CSR Awards Database. The Department of Education requires states to report their awardees to the CSR Awards Database. However, the database did not include all states from the year 2002 for two reasons. First, although data collection is ongoing, at any given time the data may be incomplete due to delinquent reporting by state education agencies. Second, many states award CSR funds to schools biennially or irregularly.

Of the initial 15 states in the 2002 CSR universe missing from the original sample, three states were left out as a by-product of the random selection process. Five additional states reported data after the initial sample selection. The remaining seven states did not make any awards in 2002 but reported 2003 data. During the first year of the study, the Department of Education allocated additional funds to gather data from these 15 states.

Researchers purposefully selected an additional 100 schools from a larger random sample of these missing states to ensure the sample represented all states and school levels. Also, to measure the value added by CSR over Title I schoolwide grants, a larger comparison group of schools with non-CSR Title I schoolwide grants was required. Thus, in the second sample, the choice of comparison schools was limited to those classified as having Title I schoolwide grants in 2002. The resulting combined sample of 500 CSR schools makes up 37 percent of the eventual (now relatively stable) universe of 1,340 awardees of the 2002 cohort. Survey data from the original sample have been combined with the follow-up sample for analyses in this report. Only a small number of high schools received CSR awards in 2002. Consequently, the high school survey sample was too small for statistical analyses, so this report focuses on elementary and middle schools, but not high schools.

Researchers used a two-step process to select potential matches for comparison with CSR schools. First, they created a school equivalency index for all schools in each state (where data were available). Second, they calculated a proximity score between each pair of schools within a state. Matching schools were selected that had the closest proximity on the index to CSR schools within the same district.

A regression-based approach to weighting and combining background characteristics was used to construct the index of school similarity for each state. This method is a simplified version of the California School Characteristics Index (Technical Design Group of the Advisory Committee for the Public Schools Accountability Act of 1999). Using the National Longitudinal School-Level State Assessment Score Database (NLSLSASD*)*, the team regressed measures of academic performance on measures of schoolwide participation in federally subsidized free or reduced-price lunch programs and on schoolwide counts of student ethnicity. The estimated coefficients led to a composite of background characteristics for each school. In short, each background characteristic was weighted by the amount it contributed to student performance.

The team selected potential matches for each first-year CSR school based upon a minimum distance criterion. This method was used in the majority of cases. However, alternative methods were used in two circumstances. First, in some districts (or states), either demographic or performance data were not available. Second, a suitable comparison non-CSR school was not available within the same district because the district was too small or all other comparable schools had previously participated in CSR. Each of these contingencies is discussed below.

In some cases, not enough data were available to construct a school equivalency index. Where states or districts did not report free or reduced-price lunch or ethnicity, schools were ranked within districts using only achievement scores. Examples of states where these data were not available are Tennessee and Washington. Achievement data were missing for some schools or districts. This was often the case in high schools where the SAT takes the place of district-administered standardized tests. The proximity scores in these cases were based on an unweighted composite of free or reduced-price lunch and ethnicity. Finally, in cases in which neither student performance nor demographic data were available, non-CSR comparison schools were matched by school grade span, size, and locale.

In districts where a non-CSR comparison school could not be selected, the team searched for a suitable comparison in an adjacent district or similar locale. Because the school equivalency index included all public schools in the state, the proximity of any school within the state could be calculated. The same criteria were used for selection across districts when data were available. In cases where data were not available, the team used the same procedures that applied to selecting comparison schools within districts.

Non-CSR schools in the sample were higher-performing than the CSR schools in the sample and among all newly awarded CSR schools in 2002. The baseline achievement in the ECSRIO sample of CSR schools was about 0.3 standard deviations lower than their matched non-CSR comparisons (Exhibit 4). This difference may be related to the procedures used to select CSR award recipients. The baseline achievement for the sample of CSR schools was also lower than the baseline achievement for the universe of CSR schools. This difference may be due to the purposeful selection of the additional 100 schools.

Exhibit 4
Average Baseline Standardized Achievement Measures of All 2002 CSR Schools and of the CSR and Non-CSR Schools in the ECSRIO Sample

	Standardized Achievement Measures (Z-scores)		
	CSR Universe	**ECSRIO CSR Sample Schools**	**ECSRIO non-CSR Sample Schools**
Elementary School Mathematics	-0.96	-0.82	-0.54
Elementary School Reading	-0.97	-0.86	-0.57
Middle School Mathematics	-0.90	-0.87	-0.55
Middle School Reading	-0.92	-0.87	-0.54

Exhibit highlights: The average baseline standardized achievement measures of the ECSRIO CSR schools was about 0.3 standard deviations lower than the average baseline standardized achievement measures of the ECSRIO non-CSR schools.

Sources: CSR Awards Database; National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

The sample of ECSRIO CSR and non-CSR schools were similar to each other and to the universe of schools that received their CSR awards in 2002 in their demographic compositions in 2005 (Exhibit 5). The non-CSR schools had a slightly lower percentage of free or reduced-price lunch eligible students than the CSR universe or the CSR schools. The universe of CSR schools had a somewhat higher percentage of black students than either the ECSRIO CSR schools or the ECSRIO non-CSR schools.

Exhibit 5
Average Demographic Measures in 2005 of All 2002 CSR Schools and of the CSR and Non-CSR Schools in the ECSRIO Sample

Demographic Measures	CSR Universe	ECSRIO CSR Sample Schools	ECSRIO non-CSR Sample Schools
Free or Reduced-Price Lunch (%)	71.24	69.23	64.91
Black (%)	43.05	35.75	31.48
Hispanic (%)	19.09	20.09	21.48
Type of School (%)			
Elementary School (includes K–8)	61.13	65.88	67.09
Middle School	19.68	16.20	14.35
High School	15.24	15.78	15.40
Other Configuration	3.95	2.13	3.16

Exhibit highlights: The average percentages of students receiving free or reduced-price lunch, average percentage of black students, and average percentage of Hispanic students in the ECSRIO CSR schools in 2005 were about the same as in the ECSRIO non-CSR schools.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

From the 1,000 schools originally selected for the survey sample, 961 agreed to participate in the study and were surveyed in 2003 and again in 2005. Principals and up to three teachers in each school were surveyed on various aspects of their respective school's reform activities (see Appendix C for the survey instruments). The response rate for the first-year data collection (2003) was 90 percent for the 479 CSR schools and 82 percent for the 482 non-CSR schools. The response rate for the third-year data collection (2005) was 88 percent for CSR schools and 78 percent for non-CSR schools. The response rate for surveyed staff was 77 and 64 percent for principals and teachers respectively in 2003 and 75 and 71 percent in 2005.

The survey instrument includes items that measure behaviors rather than attitudes and expectations. For example, rather than ask whether the reform strategy meets student needs, respondents were asked about procedures that monitor student achievement during implementation. Although this approach is likely to increase the objectivity of the responses, self-reported responses may still have limitations such as the accuracy of respondent recall and the tendency for respondents to give socially desirable responses (Flower, 1995; Tourangeau, Rips, and Rasinki, 2000). Survey findings were analyzed in conjunction with the results of the case studies to validate information on implementation from both sources. In general, both methods yield similar findings and, when differences exist, the case studies found greater reform implementation than did the surveys, indicating that survey respondents do not inflate their estimates of reform implementation.

CASE STUDIES OF SCHOOL REFORM ACTIVITIES

The case studies included in this evaluation provide in-depth insight into comprehensive school reform implementation processes and outcomes. They also provided clues for interpreting the survey data to determine if self-reported data reflect inflated or conservative responses.

Of the 1,000 schools selected for the survey sample, a subsample of 15 pairs of schools from 15 districts across 14 states—with each pair containing a CSR school and a non-CSR school—was selected to provide case studies that supplement the quantitative analyses with qualitative understanding of implementation dynamics. For example, although almost all schools reported having a comprehensive school plan, the case studies revealed differences in how the plan was used, with some schools developing a plan as a ritual activity and others making it into a living document. With such a small number of case study schools, this sample does not represent either the geographic distribution or the distribution of school levels in the universe of CSR schools. However, the selection ensured the case study sample included enough geographic and school-level diversity to reflect a variety of state policy environments, including schools in states with a long tradition of local control and others in states that exercise more centralized control over schools.

Researchers selected CSR and comparison pairs that resided in the same district for the field-based study because it was the only way to observe the differential effect of district policies. Also, visiting a single district for each pair minimized the data collection burden for both the evaluators and respondents.

The case studies included two visits to each site, with a "site" defined as a combination of four entities:

1. A CSR-funded school;
2. A demographically matched non-CSR school that did not receive any federal CSR funds prior to 2002 and is located in the same district as the CSR-funded school;
3. The district within which the two schools are located; and
4. The state education agency in the state within which the district is located.

The first phase of case study data collection occurred from spring 2003 to spring 2004. A second phase of site visits occurred during the 2004–05 school year. During visits to the 15 pairs of schools, the evaluation team observed classrooms using a formal observation instrument; reviewed school documents and materials; and interviewed state and district officials, school administrators, teachers, reform developers, and parents.

The case studies include information related to reform implementation as well as its relationship to district and state policies, including those that do not focus directly on CSR. The case studies follow a formal protocol that calls for integrating information from documents, interviews, focus groups, observations, and quantitative data (see Appendix C).

STUDY LIMITATIONS

The multi-method approach to this evaluation, like most studies of practice (Lipsey, 2003), has limitations.

All achievement analyses presented in this report rely on standardized school-level achievement scores from NLSLSASD. The use of school-level outcome measures presents methodological and psychometric challenges. School-level scores are aggregate measures of individual student performance within a school and, as such, do not account for variation in academic performance among students within a group. In addition, such analyses are not sensitive to the fact that performance may be contingent on both individual factors and factors related to group membership, such as membership in different classrooms and school environments. The multi-leveled structure of data is not adequately addressed in aggregate analyses, which limits the precision of the estimates produced. Three major areas have been widely acknowledged as contributing to imprecision in aggregate analyses: aggregation bias, misestimated standard errors, and heterogeneity of regression (Burstein, 1978; Burstein and Miller, 1981; Haney, 1980; Raudenbush and Bryk, 2002). Discussion of these issues follows:

1. Aggregation bias exists because variables at different organizational levels have different contextual meanings and may show varying relationships across organizational levels. Analyses based on aggregate outcomes do not account for differences in variables across organizational levels leading to biased estimates of effects. Raudenbush and Bryk (2002) use socioeconomic status to explicate aggregation bias. They note that both individual social class and socioeconomic status of the school influence student achievement. School-level socioeconomic measures may influence achievement beyond the effects estimated at the individual level.

2. Misestimation of standard errors occurs in single-level analyses because such analyses ignore the similarities based on common experiences among responses from individuals within a group. For example, all students in a classroom are likely to be exposed to the same curriculum. Multi-level analyses calculate standard errors in a way that accounts for the clustering of individuals and adjusts for the dependence of responses from a particular group.

3. Despite the common experiences, differences exist within all groups. Consequently, the relationships between individual characteristics and outcomes may vary among individuals. This heterogeneity is masked in school-level analyses that assume a single linear relationship between characteristics and outcomes for all individuals within a group. In contrast, multi-level models provide a mechanism to explore the heterogeneity of relationships and potential moderating factors that may account for differences in the relationships between outcomes and explanatory variables. For example, the amount of school resources or level of teacher professional development in schools may influence the relationship between CSR and student achievement.

Despite these weaknesses, designating the school as the unit of analysis and using aggregate outcome measures is deemed appropriate given the aims of the evaluation and the nature of CSR. The analyses rely on aggregate outcome measures that are both meaningful to schools and

relevant to the current policy environment. School-level proficiency is the performance measure upon which all schools are held accountable. Although the use of aggregate measures may compromise precision, its relevance for school officials and policymakers offsets its limitations. Moreover, the CSR program was designed to stimulate comprehensive change through a set of coordinated reform actions that influence school operations across the whole school. As such, it is appropriate to use the school as the unit of analysis and school-level achievement measures to examine how CSR influences achievement.

Furthermore, although a multi-level model, in which students are nested within schools, would provide more precise estimates of the relationships between CSR and achievement outcomes than the models employed in this evaluation, collecting individual student records over an extended time period from a large nationally representative sample of CSR and comparison schools places a heavy burden on schools.

A second limitation of this study is shared with other efforts to assess program effectiveness across states. Such evaluations must overcome the difficulty of using the existing state assessments, which are designed to provide information about students' progress toward mastering the content established in each state's standards. Consequently, assessments differ. In fact, neither the content nor the criteria for determining proficiency are the same from state to state. Also, standards, assessments, and proficiency criteria often change, making scores within states difficult to compare over time. To analyze outcomes across states and assessment instruments, school-level assessment scores were standardized within states for each year. The standardization provides a common metric for all achievement outcomes; however, it does not account for all sources of heterogeneity among states resulting from differences in assessments and the stringency of state proficiency standards. A more precise analytic approach would employ a multi-level model of schools nested within states to account for the variation across states. Insufficient within-state samples, stemming from both missing data and the manner in which 2002 CSR grants were awarded across states, precluded the use of a multi-level model to explore the relationships between CSR and student academic achievement.

The study is further limited by the loss of a large number of schools from some analyses due to survey nonresponse in 2005 and the fact that several states did not report all data to the Common Core of Data. For example, nearly one-half of the elementary and middle schools in the original ECSRIO sample are not included in the analyses of the relationships between comprehensiveness of CSR implementation and achievement growth. If the survey had captured information about implementation from the excluded schools and found they were implementing fewer CSR components while still experiencing comparable achievement gains as the included schools (see Exhibits 24, 25, and 26 in Chapter IV for details), then a weaker relationship between CSR implementation and achievement may exist than reported from this study. This possibility seems plausible if one assumes that schools implementing more CSR components are more likely to complete the survey than those implementing fewer components.

EVALUATION QUESTIONS

The next three chapters provide findings related to the year-three evaluation questions:

- Was receipt of a CSR award associated with improvements in school-level mathematics and reading achievement?
- Were schools that received CSR awards more likely to implement the legislatively specified components of CSR than other schools?
- Was fidelity of CSR implementation associated with gains in school-level mathematics and reading achievement?

III. OVERALL RELATIONSHIP BETWEEN CSR AWARD AND ACHIEVEMENT

Key Finding

Was receipt of a CSR award associated with improvements in school-level mathematics and reading achievement?

- Receipt of a CSR award was not associated with achievement gains in mathematics and reading achievement through the first three years of award.

METHODOLOGY USED TO ASSESS THE RELATIONSHIP BETWEEN RECEIPT OF A CSR AWARD AND ACHIEVEMENT

Analyses of the relationship between CSR award and growth in achievement were restricted to Title I schools that received CSR program funding. The evaluation team selected Title I schools using demographic information from the Common Core of Data (CCD), the U.S. Department of Education's national database of school demographics, and the standardized achievement scores developed for this evaluation. Selection of the comparison group was based on data from 2001–02, the year before the CSR schools began implementing their awards.

States targeted CSR awards to low-performing Title I schools, resulting in 96 percent of CSR awards going to such schools. Therefore, analyzing Title I schools allows for inference regarding the population of greatest interest. However, Title I encompasses a wide range of schools; in 2003–04, over 54,000 schools received Title I assistance as either a Schoolwide Title I or Targeted Assistance School (U.S. Department of Education, 2007). Given the emphasis most states placed on making CSR awards to the lowest-performing and highest-poverty schools, it comes as no surprise that the population of Title I non-CSR schools are higher-performing, smaller, and contain lower percentages of students from traditionally underserved minority groups and those eligible for free and reduced-price lunch than Title I CSR schools (Exhibit 6).

Exhibit 6
Differences Between the 2002 Cohort of Title I CSR Schools and Title I Non-CSR Schools

	Non-CSR		CSR		
	N	Average	N	Average	Difference
Standardized Assessment Scores					
Elementary Math	28,366	-0.22	649	-0.99	0.77**
Elementary Reading	28,512	-0.25	654	-0.97	0.73**
Middle School Math	10,135	-0.19	310	-0.95	0.76**
Middle School Reading	10,282	-0.19	318	-0.97	0.78**
Demographic and School Characteristics					
Membership (N)	46,667	475	973	579	-104**
Percent Minority	45,635	52.0	940	72.1	-20.1**
Percent FRL	46,632	36.2	973	59.9	-23.8**

Exhibit highlights: Achievement in Title I CSR schools in the 2001–02 school year was approximately three-fourths of a standard deviation lower than in Title I non-CSR schools. The percentage of minority students in Title I CSR schools was about 20 percent higher, and the percentage of students eligible for free and reduced-price lunch was about 24 percent higher, than in Title I non-CSR schools.

Note: + p<.10; * p<.05; ** p<.01. "FRL" refers to students eligible for free or reduced-price lunches. All data are for the 2001–02 school year to demonstrate baseline achievement and demographics. Standard errors are included in Exhibit B.1.

Source: Common Core of Data (CCD).

Because of this disparity, the evaluation team developed a comparison group of non-CSR schools for the analysis of the relationship between receiving a CSR award and achievement through Mahalanobis propensity scoring (Rosenbaum and Rubin, 1983; Rosenbaum and Rubin, 1985; Rubin, 1980) based on mathematics and reading achievement, percent free and reduced-price lunch, and percent minority.[17] In this instance, CSR schools are counted as treated schools and non-CSR schools are counted as untreated schools. For each CSR school, Mahalanobis propensity scoring chooses a comparison school from the pool of untreated schools that most closely matches the CSR school based on the achievement and demographic measures used.[18]

Because some states did not report percent free and reduced-price lunch, and because sometimes CSR schools were located in small districts, the propensity score matching was run several times, first starting with the most restrictive matching procedure and then gradually relaxing the matching conditions. All matches are based on, at the minimum, mathematics and reading achievement and percent minority. Matches were restricted to being located in the same state and

[17] Note that the comparison group developed for the analysis of the relationship between receiving an award and achievement gains is different from the comparison group developed at the outset of this study. The comparison group used in this section is based only on Title I schools; furthermore, it relies on updated CSR award, Common Core of Data, and NLSLSASD data files not available at the outset of this study.
[18] The Mahalanobis propensity score matching was completed using the -mahapick- module in Stata (Kantor, 2006).

having the same school configuration (elementary, K–8, or 6–8 middle school) as the CSR school. The conditions that were used are listed from most restrictive to least restrictive. The comparison school was chosen from the most restrictive set of conditions where a successful comparison school was found.

1. The first set of matches is based on percent free and reduced-price lunch and being located in the same school district. Comparison schools selected in this group most closely matched their CSR school on mathematics and reading achievement, percent minority and percent free and reduced-price lunch; were located in the same school district as the CSR school; and had the same school configuration as the CSR school.

2. The second set of matches removes free and reduced-price lunch from the restrictions used for the first set.

3. The third set of matches is similar to the first, but instead of restricting the matching to the same district, restricts the matching to the same type of locality (e.g., large city, small city, suburban, or rural) in the state. This expands the pool of potential matches for schools in small school districts that may not have many potential matches.

4. The fourth set of matches removes free and reduced-price lunch from the third set of matches.

5. The fifth set of matches includes free and reduced-price lunch, and removes any locality restriction (choosing any school in the state).

6. The last set of matches removes free and reduced-price lunch from the restrictions used for the fifth set of matches.

Exhibit 7 presents average school-level achievement scores and demographics for schools with elementary mathematics and reading scores, while Exhibit 8 presents these measures for schools with middle school mathematics and reading scores. The CSR and comparison schools are very closely matched on baseline achievement and demographics.

Exhibit 7
Differences Between the 2002 Cohort of Title I CSR Schools and Non-CSR Title I Schools Chosen by Propensity Scoring: Schools with Elementary Mathematics and Reading Achievement

	Non-CSR		CSR		
	N	Average	N	Average	Difference
Standardized Assessment Scores					
Elementary Math	550	-0.90	645	-0.99	0.09
Elementary Reading	550	-0.89	645	-0.97	0.08
Demographic and School Characteristics					
Membership (N)	550	514	645	509	5
Percent Minority	533	72.4	618	74.1	-1.7
Percent FRL	550	58.8	645	61.2	1.4

Exhibit highlights: There are no statistically significant differences between matched CSR and non-CSR Title I schools in the baseline (2001–02) school year.

Note: + p<.10; * p<.05; ** p<.01. "FRL" refers to students eligible for free or reduced-price lunches. Standard errors are included in Exhibit B.2.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit 8
Differences Between the 2002 Cohort of Title I CSR Schools and Non-CSR Title I Schools Chosen by Propensity Scoring: Schools with Middle School Mathematics and Reading Achievement

	Non-CSR		CSR		
	N	Average	N	Average	Difference
Standardized Assessment Scores					
Middle School Math	248	-0.86	309	-0.96	0.10
Middle School Reading	248	-0.83	309	-0.94	0.11
Demographic and School Characteristics					
Membership (N)	248	664	309	663	1
Percent Minority	245	71.3	301	73.5	-2.1
Percent FRL	248	63.7	309	65.2	-1.5

Exhibit highlights: There are no statistically significant differences between matched CSR and non-CSR Title I schools in the baseline (2001–02) school year.

Note: + p<.10; * p<.05; ** p<.01. "FRL" refers to students eligible for free or reduced-price lunches. Standard errors are included in Exhibit B.3.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

After the selection of the comparison schools, paired t-tests were used to assess whether the change in achievement from 2002–03 to 2004–05 is different between CSR schools and their comparisons (Zhao, Li, Gao, and Tierney, 2001; Fraas, Newman, and Pool, 2007).[19]

FINDINGS

Was receipt of a CSR award associated with improvements in school-level mathematics and reading achievement?

- Receipt of a CSR award was not associated with achievement gains in mathematics and reading achievement through the first three years of award.

CSR schools made small but statistically significant gains in elementary mathematics and reading achievement during the time they were implementing their award (effect sizes of 0.06 and 0.07, respectively) (Exhibit 9). The comparison schools, however, made similar gains, indicating that CSR was no better than comparable Title I schools at improving reading achievement. At the middle school level, changes in mathematics and reading achievement are not statistically significant, except for non-CSR schools in mathematics. Again, differences in achievement change between CSR and non-CSR schools are not statistically significant. Over the course of their three-year awards, CSR schools were no better at improving achievement than Title I schools that were similar in demographics and achievement in the baseline year.

[19] T-tests are used to test the null hypothesis that the means of two groups (in this instance, CSR and comparison schools) are the same. A t-statistic is calculated from the two groups' means and standard deviations, which is then used to determine the probability of correctly rejecting the null hypothesis by comparing it to a table of students' t-distribution.

Exhibit 9
Changes in Standardized Assessment Scores in CSR and Non-CSR Schools from 2002–03 to 2004–05

	N	Changes in Standardized Assessment Scores		
		CSR Schools	Non-CSR Schools	Difference
Elementary Mathematics	634	0.06+	0.09**	-0.03
Elementary Reading	638	0.07*	0.07*	0.00
Middle School Mathematics	318	0.03	0.09**	-0.06
Middle School Reading	320	0.02	0.03	-0.01

Exhibit highlights: CSR schools had statistically significant increases between 2002–03 and 2004–05 in elementary mathematics and reading achievement of 0.06 and 0.07 standard deviations, respectively. There were, however, no statistically significant differences between CSR schools and their matched non-CSR schools.

Note: + p<.10; * p<.05; ** p<.01. Tests for the statistical significance of achievement gains for CSR and non-CSR schools are t-tests to assess whether the value is different from zero. The differences between CSR and non-CSR schools are assessed through paired t-tests. Standard errors are included in Exhibit B.4.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

DISCUSSION

Unlike these results, earlier CSR research shows an increase in mathematics and reading achievement over multiple years of implementation (Borman et al., 2003). Two possible reasons for the lack of findings here are suggested by the case studies. First, a number of schools ended implementation of an original CSR strategy, replacing it with a portfolio of programs. Second, some schools added reform programs, creating competition for teachers' time for professional development.

School W, for example, chose the Comer School Development model following award receipt in 2002. A new principal arrived in 2003 and instead chose the Effective Schools model. Teachers indicated they were not involved in this decision and had already scheduled professional development for the original model for the summer. Staff then engaged in what they termed a "whirlwind" of professional development on Effective Schools. However, key consultants for Effective Schools became ill, compounding staff indifference to the program. Ultimately, neither Effective Schools nor Comer School Development was implemented.

Another school, School E, was encouraged by the district's then superintendent to adopt and implement Accelerated Schools. The external assistance provided by the model developer helped the school arrange common planning time for teachers and offered over 20 days of professional development in 2002–03. However, in 2003, district leadership changed, and the new superintendent focused efforts on supplementary programs. Teachers were required to attend professional development sessions for the district's priority programs, creating scheduling conflicts. By 2004, Accelerated Schools was no longer present in the school.

Thus, implementation (or lack thereof) may explain the overall lack of a relationship between receiving a CSR award and achievement gains. The CSR program is based on two key implementation concepts: Schools implementing reforms that encompass the 11 components indicated in *NCLB* will have higher achievement than schools that do not; and schools that implement reform models with a scientific research base will have higher achievement than those schools that implement models that lack a scientific research base. Two different analyses, included in the next two sections, assess the implementation of CSR and its relationship to achievement. The first analyzes the relationship between the comprehensiveness of CSR implementation, as measured by the number of components a school implements, and achievement outcomes. The second analyzes the relationship between the adoption of a scientifically based research model and achievement.

IV. THE COMPREHENSIVENESS OF CSR IMPLEMENTATION AND ITS RELATIONSHIP TO ACHIEVEMENT

Key Findings

Were schools that received CSR awards more likely to implement the legislatively specified components of CSR than other schools?

- No, both CSR and non-CSR schools implemented an average of fewer than four components in 2003 and fewer than five in 2005 at both the elementary and middle school levels.

Was fidelity of CSR implementation associated with gains in school-level mathematics and reading achievement?

- The comprehensiveness of implementation, as measured by the number of CSR components implemented, was not related to mathematics and reading achievement gains in CSR schools.

This report includes analyses of changes in the extent of component implementation for both CSR and non-CSR schools over the three-year award period. The first section of this chapter focuses on the implementation of CSR, comparing CSR and non-CSR schools; the second section of this chapter focuses on the relationship between the implementation of the CSR components and achievement gains.

MEASURING IMPLEMENTATION OF THE CSR COMPONENTS

The analyses of the relationships between the CSR components and school-level achievement rely on measures of the CSR components developed from the surveys administered in 2003 and 2005 to both CSR and non-CSR schools. The surveys ask principals and teachers about the implementation of activities consistent with the 11 components identified in *NCLB*. For most questions, respondents were presented with a forced choice of "0" for not having implemented

the activity and "1" for implementing the activity.[20] Within each component, respondents' answers were averaged to develop a component score for each school.

The component scores were converted into measures of whether a school is counted as implementing a component. That is, schools were assigned a 0 if their component score or their responses to survey questions fell below a certain threshold and a 1 if they exceeded the threshold. For most of the components, the school had to report that it was implementing all of the activities to be counted as implementing the component. In some instances, however, this would have resulted in no schools implementing the component, and so the threshold was relaxed somewhat. Also, for parental involvement, the measures used were not dichotomies, but instead were estimates of the percentage of parents engaging in various activities in the school; and another threshold was developed based on the distribution of the estimates. The metric used to measure whether a school implemented a component is included in the discussions of how each component is measured. This was done to measure whether a school implemented a component and was used to calculate the number of components implemented. The number of components implemented was also used in regression analyses to measure the relationship between number of components implemented and achievement.

The surveys were modified between the two administrations as a result of findings from the 2003 survey concerning the existence of multiple reforms in schools, the need to develop more refined measures of professional development, and the fact that some of the questions in the 2003 survey were no longer applicable in 2005. For instance, in the 2003 survey, schools often reported implementing more than one reform method. The 2005 survey was modified to capture information on up to four reforms the school was undertaking (see Appendix C). As a result of these changes, some of the components were not measured consistently across the two time periods; these modifications also required changes in assessing whether a component was implemented. All such changes are documented below in the discussion of each component.

Components 1 and 11—Research-Based Design and Evidence-Based Practice

Given the substantial similarity between Component 1: Research-Based Design and Component 11: Evidence-Based Practice, the decision was made to combine them into one measure. As defined in the *No Child Left Behind Act (NCLB)* of 2001, evidence-based practice focuses on proven methods and strategies for student learning, teaching, and school management that have been replicated successfully in schools with diverse characteristics. Added in *NCLB*, the research-based design component calls for the adoption of programs that use scientifically based

[20] This evaluation measures implementation differently from Aladjem et al. (2006), who measured implementation as the difference between what the model developers consider to be full implementation and what the school actually does. To do so, they surveyed each of the model developers and asked them to respond as if they were a school that implemented their model. The researchers compared each school's responses to this measure of implementation to develop a measure of fidelity. In order to measure implementation in the comparison schools, Aladjem et al. predicted what CSR model the comparison school would have chosen based on its school characteristics and those of the CSR schools. The researchers then compared the school's responses with those of the model developer of the predicted CSR model. This measure of implementation is consistent with the focus on models as compared with this study's focus on fidelity with the 11 components identified in *NCLB*.

research to document significant improvement in the academic achievement of students participating in such programs.

The 2003 measure consisted of a principal's responses to three questions, each of which was a yes or no question on the survey: whether there was evidence based on independent research supporting the model (33 percent indicating yes), whether there was evidence based on research conducted by the reform developer (also 33 percent), and whether there was evidence that relied on the use of control or comparison groups (21 percent of respondents). A school was counted as implementing this component if it reported that at least two of the three types of evidence were present in 2003 (Exhibit 10).

Two of the items for the 2005 component measure were different from items used for the 2003 measure: whether the reform model had evidence that it improved student achievement and had been shown to improve student achievement at other schools. One item, whether evidence was based on research that relied on comparison or control groups, was consistent across the two years. Each of these was also a yes or no question. Schools were asked to report on up to four reform methods in their school at the time of the survey; these responses were averaged to derive a school-level score, which ranged from 0 to 1, for each of the three items. The school was counted as implementing this component if the score for each of the non-missing items was one. Thus, if a school only reported on two of the items and reported that they were both present, then that school was counted as implementing a research-based design. This was done to make the measure for year three relatively consistent with the measure for year one, which required that the school indicate that two types of evidence were present.

Exhibit 10
Means and Standard Deviations of Survey Items Used to Construct Research-based Design and Evidence-based Practice Measure, 2003 and 2005

Year	Evidence Items	N	Mean	Standard Deviation
2003	Independent Research	759	0.33	0.47
	Research by Reform Designer	759	0.33	0.47
	Comparison/Control Groups	759	0.21	0.41
2005	Improves Achievement	695	0.89	0.28
	Comparison/Control Groups	674	0.60	0.45
	Student Achievement at Other Schools	670	0.91	0.26

Exhibit highlights: The type of evidence to construct the research-based design and evidence-based practice measure differed from 2003 to 2005. The only consistent measure was comparison or control groups. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the mean score of 0.21 for comparison or control groups in 2003 indicates that 21 percent of respondents in 2003 indicated the reform method they chose had evidence based on comparison or control groups; in 2005, 60 percent of respondents said this was true.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various evidence items (0=No; 1=Yes).

Source: ECSRIO surveys.

Component 2: Comprehensive Planning

Questions on the survey form measured two aspects of comprehensive planning—classroom-based planning and school-based planning. Both were asked of respondents consistently in 2003 and 2005. On all questions, respondents were asked if their school improvement plans contained nine components of planning; they could answer yes or no to each. The classroom-based measures included curriculum and instruction, student assessment, classroom management, and professional development (Exhibit 11). The school-based planning measure included measurable goals for reform, periodic evaluation, parental involvement, professional development, participation in school management, and integration of new technology (Exhibit 12). Note that professional development occurred in both school- and classroom-level planning because professional development opportunities may be around subject-specific topics or school-reform related topics. In both years, a school was counted as implementing these components if it reported that the school improvement plan contained all of the items.

Exhibit 11
Means and Standard Deviations of Survey Items Used to Construct Comprehensive Planning-Classroom Measure, 2003 and 2005

	2003		2005	
Comprehensive Planning Items	Mean	Standard Deviation	Mean	Standard Deviation
Curriculum and Instruction	0.88	0.32	0.95	0.21
Student Assessment	0.50	0.50	0.65	0.48
Classroom Management	0.43	0.49	0.58	0.49
Professional Development	0.93	0.26	0.96	0.20

Exhibit highlights: Four survey items were used to measure comprehensive planning in the classroom. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, in 2003 the mean curriculum and instruction score of 0.88 indicates that 88 percent of respondents indicated that the school improvement plan covered curriculum and instruction.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various classroom planning items (0=No; 1=Yes). There are 764 observations in 2003 and 717 observations in 2005.

Source: ECSRIO surveys.

Exhibit 12
Means and Standard Deviations of Survey Items Used to Construct Comprehensive Planning-School Measure, 2003 and 2005

Comprehensive Planning Items	2003		2005	
	Mean	Standard Deviation	Mean	Standard Deviation
Measurable Goals	0.97	0.16	0.98	0.15
Periodic Evaluation	0.85	0.36	0.91	0.28
Parental Involvement	0.81	0.40	0.85	0.36
Professional Development	0.93	0.26	0.96	0.20
Participation in School Management	0.53	0.50	0.63	0.48
New Technology	0.72	0.45	0.78	0.42

Exhibit highlights: Six survey items were used to measure comprehensive planning in the classroom. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, in 2003 the mean parental involvement score of 0.81 indicates that 81 percent of respondents indicated that the school improvement plan covered parental involvement.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various school planning items (0=No; 1=Yes). There are 764 observations in 2003 and 717 observations in 2005.

Source: ECSRIO surveys.

Component 3: Professional Development

The professional development component relied on the reported number of days of professional development; however, it was measured differently in the two years, and any changes should be interpreted cautiously. On the 2003 survey, respondents were asked if their school provided at least 10 days of professional development. Principals could respond yes or no, and the school was counted as implementing the measure if they responded yes. On the 2005 survey, teachers were asked to report the number of hours of professional development they received in the last year. The research team divided the number of hours by six to convert their responses to days (because that is approximately the length of an average school day). Schools were counted as implementing professional development if the average number of days of professional development received was 10 or more (Exhibit 13).

Exhibit 13
Means and Standard Deviations of Professional Development Item, 2003 and 2005

	2003			2005		
Professional Development Item	**N**	**Mean**	**Standard Deviation**	**N**	**Mean**	**Standard Deviation**
Professional Development	736	0.48	0.50	718	0.61	0.49

Exhibit highlights: The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, on the 2003 survey, principals responded 48 percent of the time that at least 10 days of professional development were available to teachers. On the 2005 survey, 61 percent of teachers reported that they received at least 10 days of professional development during the last year.

Note: Means and standard deviations are computed from respondents' answers to questions pertaining to the professional development opportunities exceeding 10 days (0=No; 1=Yes). This question was asked differently in 2003 and 2005; any changes should be interpreted cautiously.

Source: ECSRIO surveys.

Component 4: Goals and Benchmarks

The 2003 and 2005 goals and benchmarks measures relied on a somewhat different set of items, and any change statistics should be interpreted cautiously. Four questions from the 2003 surveys were used to construct the measure for goals and benchmarks—whether the school improvement process (SIP) included student assessment rubrics and measurable goals and objectives as well as whether the school had end-of-year and interim student achievement goals (Exhibit 14). Non-missing responses were summed, and the school was counted as implementing this component if the average of the non-missing responses was one. Thus, if a school reported on two of the measures and indicated those were in place, the school was counted as implementing goals and benchmarks.

The 2005 goals and benchmarks measure relies on the same two SIP items as the 2003 measure. However, the two questions about student goals were not asked in the 2005 surveys. Instead, respondents were asked if the reform strategies were accompanied by implementation benchmarks. One of the earlier findings in this study is that schools were often implementing more than one reform strategy simultaneously. As a result, reform-specific questions were asked four times for each respondent. In some cases, the respondent indicated that his or her school was only undertaking one reform strategy; in other cases, respondents listed up to four reform strategies. The responses were averaged to derive a school-level measure for this item that varied between 0 and 1. The 2005 goals and benchmarks measure was created by averaging the non-missing responses. Schools with a score of one were counted as implementing goals and benchmarks.

Exhibit 14
Means and Standard Deviations of Survey Items Used to Construct the Goals and Benchmarks Measure, 2003 and 2005

Year	Goals and Benchmarks Items	N	Mean	Standard Deviation
2003	SIP Includes Student Assessment Rubrics	764	0.50	0.50
	SIP Includes Measurable Goals or Objectives	764	0.97	0.16
	School Has Student Goals at the End of the Year	561	0.88	0.33
	School Has Intermediate Student Goals	607	0.70	0.46
2005	SIP Includes Student Assessment Rubrics	717	0.65	0.48
	SIP Includes Measurable Goals or Objectives	717	0.98	0.15
	Strategies Accompanied by Implementation Benchmarks	694	0.75	0.38

Exhibit highlights: The survey items used to measure goals and benchmarks differed slightly from 2003 to 2005. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the mean score of 0.50 for whether the school improvement plan includes student assessment rubrics in 2003 indicates that one-half of respondents said that the school improvement plan includes student assessment rubrics.

Note: SIP refers to school improvement process. Means and standard deviations are computed from respondents' answers to questions about the existence of the various items measuring goals and benchmarks (0=No; 1=Yes).

Source: ECSRIO surveys.

Component 5: Faculty Participation

The 2003 and 2005 measures for faculty participation relied on almost the same survey items, but the 2003 measure included one additional item that was not asked on the 2005 survey (Exhibit 15). Both measures relied on yes or no questions about all teachers participating in reform and factors that limit participation. The 2003 measure included the estimate by the principal of the percentage of teachers who participate in reform. In both years, the faculty participation measure was created by averaging the non-missing responses. Schools with a score of one were counted as implementing the faculty participation component.

Exhibit 15
Means and Standard Deviations of Survey Items Used to Construct Faculty Participation Measure, 2003 and 2005

Year	Faculty Participation Items	N	Mean	Standard Deviation
2003	Percentage of teachers who participate in reform	694	0.89	0.21
	All teachers participate in reform (yes/no)	759	0.80	0.40
	Participation is not limited by subject-specific reform	759	0.79	0.41
	Participation is not limited because of funding	760	0.93	0.25
2005	All teachers participate in reform (yes/no)	718	0.78	0.42
	Participation is not limited by subject-specific reform	718	0.90	0.30
	Participation is not limited because of funding	718	0.88	0.33

Exhibit highlights: The survey items used to measure faculty participation differed slightly from 2003 to 2005. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the mean score of 0.80 for whether all teachers participate in reform in 2003 indicates that 80 percent of respondents indicated that all teachers participate in reform.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the faculty participation items (0=No; 1=Yes).

Source: ECSRIO surveys.

Component 6: District Support

In both years, the district support measure was a combination of six items pertaining to the types of district support (Exhibit 16): conducting a needs assessment, providing additional staff to support reform efforts, selecting a reform strategy, writing grant proposals, providing professional development, and providing release time for teachers to support the reform. Each of these items was a yes or no question. In each year, the non-missing scores were averaged, and a school with an average score of 0.8 or higher was counted as implementing this component.

Exhibit 16
Means and Standard Deviations of Survey Items Used to Construct the District Support Measure, 2003 and 2005

District Support Items	2003		2005	
	Mean	Standard Deviation	Mean	Standard Deviation
Needs Assessment	0.45	0.50	0.54	0.50
Additional Staff	0.40	0.49	0.45	0.50
Selecting a Reform Model	0.37	0.48	0.43	0.49
Writing Grant Proposals	0.58	0.49	0.57	0.49
Professional Development	0.77	0.42	0.81	0.39
Release Time for Teachers	0.61	0.49	0.63	0.48

Exhibit highlights: The survey items used to measure district support in 2003 and 2005 were the same. All district item means increased except for writing grant proposals from 2003 to 2005. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the 2003 needs assessment score of 0.45 indicates that 45 percent of respondents reported they received district assistance for their needs assessment.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various district support items (0=No; 1=Yes). There are 766 observations in 2003 and 715 observations in 2005.

Source: ECSRIO surveys.

Component 7: Parental Involvement

In both years, the parental involvement measures were calculated from the same five items. Principals were asked to estimate the percentage of parents who participated in their schools through parent-teacher conferences, by demanding frequent reports, actively volunteering, observing classrooms, and who were active in Parent-Teacher Associations or Parent-Teacher Organizations (Exhibit 17). For each year, non-missing responses were averaged, and a school was credited as implementing the parental involvement measure if the average was equal to or greater than 0.4.

Exhibit 17
Means and Standard Deviations of Survey Items Used to Construct the Parental Involvement Measure, 2003 and 2005

Parental Involvement Items	2003			2005		
	N	Mean	Standard Deviation	N	Mean	Standard Deviation
Parent-Teacher Conferences	750	0.64	0.26	781	0.57	0.28
Demanding Frequent Reports	699	0.36	0.32	782	0.30	0.22
Actively Volunteering	736	0.16	0.17	779	0.12	0.11
Observing Classrooms	704	0.11	0.16	775	0.09	0.12
Active in PTA/PTO	707	0.16	0.19	769	0.14	0.14

Exhibit highlights: The survey items used to measure parental involvement in 2003 and 2005 were the same. The mean of all parental involvement items decreased from 2003 to 2005. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the 2003 parent-teacher conferences mean of 0.64 indicates that principals reported, on average, that 64 percent of parents attended parent-teacher conferences.

Note: Means and standard deviations are computed from respondents' estimates of the percentage of parents who participated in school activities.

Source: ECSRIO surveys.

Component 8: External Assistance

The 2003 and 2005 measures of external assistance were constructed from different survey questions (Exhibit 18), and any year-to-year comparisons should be made with caution. The 2003 measure was constructed from six yes or no answers on the types of external assistance available. Non-missing responses were averaged, and a school was credited as implementing the external assistance measure if at least 80 percent of the non-missing items were reported as implemented.

The 2005 survey included an external assistance measure comprising four items on the number of reform training hours received, if strategies are included in the curriculum, if all teachers receive the training, and if support is ongoing. First, teachers were asked how many hours of training on reform they had received. The highest number reported was 959 hours. Some respondents may have interpreted this question as meaning over the life of the reform program, while others may have interpreted this as the number of hours in the last year. Also, because there is often high teacher turnover in lower-performing schools, some respondents may have only been in their school for a portion of the CSR award implementation period. However, to scale this number to vary between 0 and 1, it was divided by 1,000 and thus varies between 0 and 0.96. The next three questions were asked up to four times, once for each reform identified in the school. The responses for each question were averaged to derive a school-level measure for each item that varied between 0 and 1. The external assistance measure was constructed by taking the average of the non-missing four items, and schools with an average of at least 0.7 were counted as having implemented this item.

Exhibit 18
Means and Standard Deviations of Survey Items Used to Construct the External Assistance Measure, 2003 and 2005

Year	External Assistance Items	N	Mean	Standard Deviation
2003	On-site Consulting	769	0.71	0.46
	Professional Development	769	0.90	0.31
	Networking	769	0.56	0.50
	Written Materials for Students	769	0.44	0.50
	Written Materials for Teachers	769	0.67	0.47
	Software or Technology	769	0.46	0.50
2005	Hours (1,000) of Training on Reform Strategy	959	0.05	0.09
	Strategies Include Curriculum Materials	696	0.65	0.42
	All Teachers Received Training on all Strategies	695	0.72	0.39
	Ongoing Support is Available	694	0.77	0.37

Exhibit highlights: The survey items used to construct the external assistance measure were different in 2003 and 2005. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the mean score of 0.71 for whether onsite consulting was available in 2003 means that 71 percent of respondents indicated that on-site consulting was available.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various external assistance items (0=No; 1=Yes).

Source: ECSRIO surveys.

Component 9: Evaluation

The 2003 and 2005 measures for evaluation were nearly the same, but there was one item in the 2003 measure that was not asked of respondents in 2005 (Exhibit 19): whether there is a formal written plan to evaluate progress. All of the questions asked in both years were yes or no questions. In each year, non-missing responses were averaged and a school was credited as implementing the evaluation component if the average score was 1.0 in 2003 and at least 0.75 in 2005. This distinction was made due to the one item not present in the 2005 survey and the large drop in several of the items from 2003 to 2005.

Exhibit 19
Means and Standard Deviations of Survey Items Used to Construct the Evaluation Measure, 2003 and 2005

Evaluation Items	2003			2005		
	N	Mean	Standard Deviation	N	Mean	Standard Deviation
School Improvement Process has Mechanism for Periodic Evaluation	764	0.85	0.36	717	0.91	0.28
Evaluation Plan Topics						
Student Performance	752	0.96	0.19	722	0.86	0.35
Program Implementation	752	0.74	0.44	722	0.62	0.48
Parental Participation	752	0.63	0.48	722	0.92	0.28
Staff Development	752	0.86	0.35	722	0.46	0.50
Utility of External Assistance	752	0.34	0.48	722	0.35	0.48
Sources of Financial Support	752	0.44	0.50	722	0.06	0.23
Formal Written Plan to Evaluate Progress	756	0.85	0.36	NA	NA	NA

Exhibit highlights: The survey items used to construct the evaluation items were the same in 2003 and 2005, except in 2005, the survey did not ask about a formal written plan to evaluate progress. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the mean score of 0.85 for whether the school improvement process (SIP) has a mechanism for periodic evaluation in 2003 indicates that 85 percent of respondents said that the SIP did have such a mechanism.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various evaluation items (0=No; 1=Yes).

Source: ECSRIO surveys.

Component 10: Coordination of Resources

The 2003 and 2005 measures for the coordination of resources component were constructed from different survey responses (Exhibit 20), and any year-to-year comparisons should be made cautiously. All of the questions asked in both years were yes or no questions. In each year, non-missing responses were averaged, and a school was credited as implementing the evaluation component if the average score was 1.0 in 2003 and at least 0.7 in 2005. This distinction was made due to the increase in the types of funds that were asked about in 2005. That is, some schools may not have received foundation grants or local donations and may have reported on their survey that they did not have control over these funds.

Exhibit 20
Means and Standard Deviations of Survey Items Used to Construct Coordination of Resources Measure, 2003 and 2005

Year	Coordination of Resources Items	N	Mean	Standard Deviation
2003	School Has Control Over Budgetary Issues	754	0.77	0.42
	School Has Control Over Personnel Decisions	751	0.78	0.42
	How Existing Resources Have Been Coordinated			
	Align District Professional Development	754	0.86	0.35
	Align Title I Activities	754	0.69	0.46
	Align Other Funds	754	0.31	0.46
	Reallocate Staff	754	0.40	0.49
2005	**School Has Control Over the Following Resources**			
	Federal CSR Funds	704	0.52	0.50
	Title I Funds	704	0.68	0.47
	Other Federal Funds	704	0.33	0.47
	Special State Grants	704	0.40	0.49
	Discretionary District Funds	704	0.59	0.49
	Foundation Grants	704	0.25	0.43
	Local Community or Business Donations	704	0.62	0.49

Exhibit highlights: The survey items used to construct the coordination of resources measure were different in 2003 and 2005. The reader may convert any of the means into a percent by multiplying the mean value by 100. For example, the mean score of 0.77 for the item indicating that the school has control over budgetary issues in 2003 indicates that 77 percent of respondents said that the school did have such control.

Note: Means and standard deviations are computed from respondents' answers to questions about the existence of the various items used to measure the coordination of resources (0=No; 1=Yes).

Source: ECSRIO surveys.

FINDINGS

Were schools that received CSR awards more likely to implement the legislatively specified components of CSR than other schools?

- No, both CSR and non-CSR schools implemented an average of fewer than four components in 2003 and fewer than five in 2005 at both the elementary and middle school levels.

Both CSR and non-CSR schools implemented an average of fewer than four components in 2003 and fewer than five in 2005 at both the elementary and middle school levels (Exhibit 21). At the elementary school level, CSR schools reported slightly greater implementation (the difference is

statistically significant but substantively small) than did non-CSR schools in 2003. Three years after initial funding, both CSR and non-CSR schools reported implementing a more comprehensive reform (i.e., one that uses more components) in that the number of components implemented increased by 0.9 in CSR schools and 1.1 in non-CSR schools.

At the middle school level, CSR and non-CSR schools in both 2003 and 2005 reported implementing about the same number of components (Exhibit 21). Both reported statistically significant similar increases in the number of components implemented between these periods.

Exhibit 21
Average Number of CSR Components Implemented by CSR and Non-CSR Schools in 2003 and 2005

	Elementary Schools			Middle Schools		
School Type	2003	2005	Change from 2003 to 2005	2003	2005	Change from 2003 to 2005
CSR Schools	3.8	4.7	0.9**	2.7	4.0	1.3**
Non-CSR Schools	3.3	4.4	1.1**	3.1	4.4	1.3**
Difference Between CSR and Non-CSR Schools	0.5*	0.3	0.2	0.4	0.4	0.0

Exhibit highlights: In 2003, CSR elementary schools reported implementing a somewhat higher average number of components than non-CSR schools. No other significant differences between CSR and non-CSR schools were found. Both CSR and non-CSR schools reported similar increases between 2003 and 2005 in the number of components implemented.

Note: + p<.10; * p<.05; ** p<.01. N = 292 for CSR Elementary Schools; N = 304 for non-CSR Elementary Schools; N = 128 for CSR and non-CSR Middle Schools. Standard errors for elementary schools are included in Exhibit B.5; standard errors for middle schools are in Exhibit B.6.

Source: ECSRIO surveys.

While the average number of components reported implemented in CSR and non-CSR schools was similar, the distribution of components implemented indicates some modest difference in comprehensiveness between CSR and non-CSR schools. Somewhat more non-CSR than CSR elementary schools implemented between zero and three components in 2005 (Exhibit 22), while the proportion of CSR elementary schools that implemented seven components is higher than non-CSR elementary schools. At the middle school level, CSR schools outnumber non-CSR schools at all levels above four components implemented (Exhibit 23).

Exhibit 22
Number of Components Implemented in CSR and Non-CSR Elementary Schools in 2005

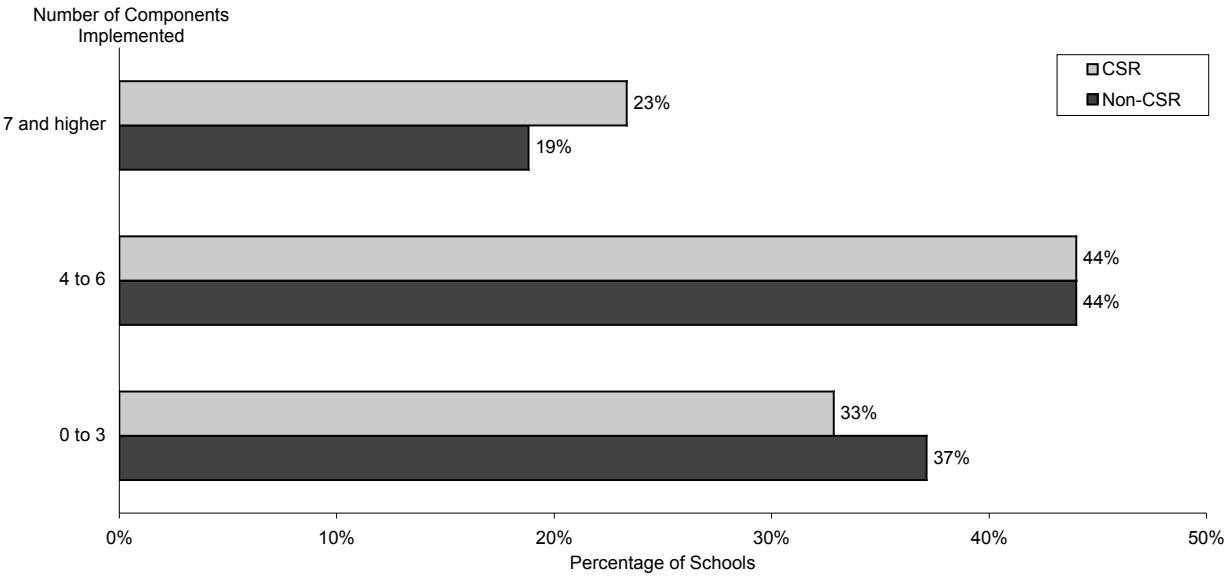

Exhibit highlights: A slightly higher percentage of CSR elementary schools reported implementing seven or more components in 2005 than did non-CSR schools. The same percentage of CSR and non-CSR schools implemented four to six components. A slightly higher percentage of non-CSR schools than CSR schools implemented zero to three components.

Source: ECSRIO surveys.

Exhibit 23
Number of Components Implemented in CSR and Non-CSR Middle Schools in 2005

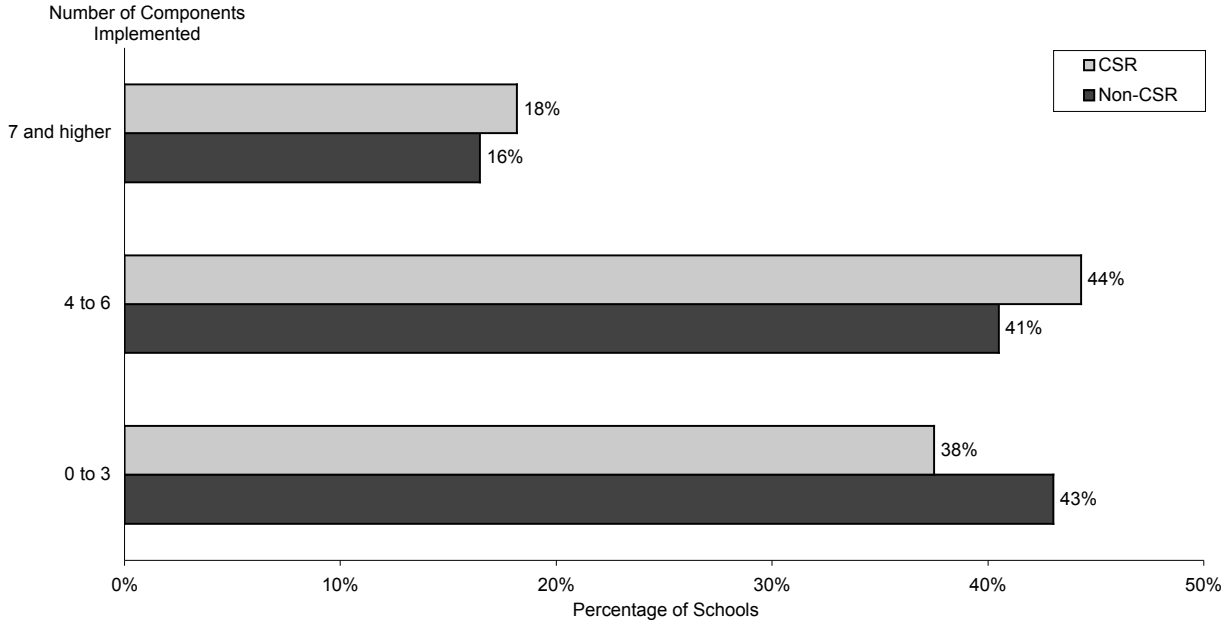

Exhibit highlights: A slightly higher percentage of CSR middle schools implemented seven or more components in 2005 than did non-CSR schools. A slightly higher percentage of CSR middle schools reported implementing at least four components. The percentage of non-CSR schools implementing zero to three components was 5 percentage points higher than that of CSR middle schools.

Source: ECSRIO surveys.

Three reasons may explain the similar implementation gains in both CSR and non-CSR school reform models. First, as shown in three of the case study sites, a school's receipt of a CSR award can allow its school district to shift resources away from that school and toward other non-CSR schools also in need of improvement. Second, *NCLB* was in its first full year in 2002–03; newly available assessment data may have prodded school districts to assist schools identified as low-performing. Third, as school districts gained experience with implementing CSR, they may have taken the lessons learned from those schools and applied them to non-CSR schools. This behavior was evident in two of the case study sites.

Was fidelity of CSR implementation associated with gains in school-level mathematics and reading achievement?

- The comprehensiveness of implementation, as measured by the number of CSR components implemented, was not related to mathematics and reading achievement gains in CSR schools.

The implementation analyses focus on assessing how a school's overall commitment to CSR is related to school-level achievement outcomes, rather than on the effects of individual CSR

components or particular component combinations on achievement. Regression analyses were used to examine the relationship between the number of components survey respondents reported as being implemented in their school as a measure of comprehensiveness and achievement outcomes and how the receipt of a CSR award may have moderated this relationship. These analyses relied on the ECSRIO sample of schools from the 2002 cohort of CSR recipients and the matched non-CSR comparison schools. Ordinary Least Squares (OLS) regression analyses predicted achievement in 2005 from the number of components implemented as reported on the 2005 surveys, controlling for achievement in 2002 and demographics in 2005.[21] These regression analyses also used the interaction between the receipt of a CSR award and the number of components implemented to test whether implementation comprehensiveness had different effects on achievement for CSR and non-CSR schools. These analyses, then, captured whether differences existed between CSR and non-CSR schools in the relationship between component implementation and achievement, controlling for school demographics. Equation 1 presents the model specifications for the analyses of CSR reform implementation comprehensiveness.

Equation 1

$$Y_{2005} = \alpha + \beta_1 y_{2002} + \beta_2 FRL + \beta_3 Black + \beta_4 HISP + \beta_5 CSR + \beta_6 COMP + \beta_7 (CSR * COMP)$$

where

Y_{2005} is the school's standardized achievement score in 2005;

y_{2002} is the school's standardized achievement score in 2002;

FRL is the percent free or reduced-price lunch eligible in 2005;

Black is the percent black in 2005;

HISP is the percent Hispanic in 2005;

CSR is a 0/1 variable that indicates whether the school received a CSR award in 2002;

COMP is the number of components a school reported implementing in 2005;

CSR*COMP is the interaction between a school receiving a CSR award in 2002 and the number of components the school implemented in 2005; and

$\beta_1, \beta_2, \ldots, \beta_7$ are the unstandardized regression coefficients for each of the variables.

Differences Between Included and Excluded Schools

Before introducing the regression results, it is important to note that there were several significant differences between schools included and those excluded from these analyses (Exhibit 24). These differences warrant additional caution when interpreting the results.

[21] The regression analyses relied on robust standard errors clustered at the state level.

Specifically, elementary schools included in the analyses had significantly higher achievement in both 2002 and 2005 and somewhat lower percentages of students eligible for free or reduced-price lunches (FRL) than those schools that were excluded from the analyses because of missing data. In addition, CSR elementary schools excluded from the analyses had significantly lower mathematics and reading achievement in 2005 than the CSR schools included in the analyses (Exhibit 25). CSR middle schools excluded from the analyses had slightly lower mathematics achievement than CSR schools included in the analyses. No other significant differences existed at the middle school level (Exhibit 26).

Exhibit 24
Differences in Achievement and Demographic Characteristics Between Schools Included and Excluded in Analyses of the Relationships Between Implementation and Achievement, 2002 and 2005

	Elementary School			Middle School		
	Excluded	Included	Difference	Excluded	Included	Difference
Standardized Achievement Score						
Mathematics Achievement, 2002	-0.84	-0.56	-0.28**	-0.77	-0.67	-0.10
Mathematics Achievement, 2005	-0.76	-0.45	-0.31**	-0.75	-0.62	-0.13
Reading Achievement, 2002	-0.90	-0.58	-0.32**	-0.82	-0.61	-0.21
Reading Achievement, 2005	-0.77	-0.46	-0.31**	-0.77	-0.63	-0.14
Demographic Characteristics						
Percent FRL, 2005	0.61	0.69	-0.08**	0.64	0.66	0.02
Percent Black, 2005	0.35	0.33	0.02	0.34	0.32	0.02
Percent Hispanic, 2005	0.20	0.20	0.00	0.19	0.23	-0.04

Exhibit highlights: Elementary schools included in the analyses had higher 2002 and 2005 mathematics and reading achievement than those schools excluded from the analyses. Furthermore, schools excluded from the analyses had somewhat lower percent free or reduced-price lunch (FRL) than the included schools. No other differences existed between excluded and included schools in either school level.

Note: + p<.10; * p<.05; ** p<.01. Standard errors are included in Exhibit B.7.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit 25
Differences in Standardized Mathematics and Reading Achievement Between CSR and Non-CSR Elementary Schools Included and Excluded in Analyses of the Relationships Between Implementation and Achievement, 2005

Subject and School Type	Excluded	Included	Difference
Mathematics			
CSR Schools	-0.98	-0.49	-0.44**
Non-CSR Schools	-0.57	-0.41	-0.15
Difference	-0.41**	-0.08	
Reading			
CSR Schools	-0.96	-0.54	-0.42**
Non-CSR Schools	-0.61	-0.38	-0.23*
Difference	-0.34*	-0.16+	

Exhibit highlights: CSR elementary schools excluded from the analyses had substantially lower mathematics and reading achievement than CSR schools included in the analyses as well as non-CSR schools excluded from the analyses.

Note: + p<.10; * p<.05; ** p<.01. Measures in table are standardized achievement scores. Standard errors are included in Exhibit B.8.

Sources: ECSRIO surveys; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit 26
Differences in Standardized Mathematics and Reading Achievement Between CSR and Non-CSR Middle Schools Included and Excluded in Analyses of the Relationships Between Implementation and Achievement, 2005

Subject and School Type	Excluded	Included	Difference
Mathematics			
CSR Schools	-0.91	-0.63	-0.28+
Non-CSR Schools	-0.61	-0.63	0.02
Difference	-0.30	0.00	
Reading			
CSR Schools	-0.80	-0.72	-0.08
Non-CSR Schools	-0.75	-0.54	-0.21
Difference	-0.05	-0.18	

Exhibit highlights: CSR middle schools excluded from the analyses had slightly lower mathematics achievement than CSR schools included in the analyses. All other differences were not statistically significant.

Note: + p<.10; * p<.05; ** p<.01. Measures in table are standardized achievement scores. Standard errors are included in Exhibit B.9.

Sources: ECSRIO surveys; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Comprehensiveness of CSR Reform Implementation and Achievement Outcomes

Comprehensiveness, as measured by the number of CSR components a school implemented, was not related to achievement gains in elementary or middle schools (Exhibit 27). CSR posits that schools that more comprehensively implement CSR should have higher achievement growth than schools that implement a less comprehensive approach. More comprehensive CSR implementation implies that the CSR award affects school operations in multiple areas, such as professional development, parental involvement, school planning, and external support. The synergies that arise from this comprehensive implementation should result in greater opportunities for learning compared with a piecemeal implementation across a few areas of school operations. However, when prior achievement and demographics were controlled for, the number of components implemented in 2005 was not associated with achievement gains in CSR schools. Our case studies also showed varied patterns in the relationship between component implementation and achievement. Of the 15 case study CSR schools, five that had implemented over four components improved in both mathematics and reading, two improved in either reading or mathematics, while eight failed to improve in either subject.

Exhibit 27
Relationships Between Number of CSR Components Implemented and Achievement Gains from 2002 to 2005

Effect	Elementary School		Middle School	
	Mathematics	Reading	Mathematics	Reading
Constant	0.42*	0.59**	0.23	0.12
Achievement (2002)	0.45**	0.47**	0.79**	0.80**
Percent FRL (2005)	-0.76**	-0.90**	-0.44	-0.39
Percent Black (2005)	-0.07	-0.08	-0.31	-0.01
Percent Hispanic (2005)	-0.02	-0.28+	-0.05	0.25
CSR Award (0=No, 1=Yes)	-0.04	-0.15	0.07	-0.14
Number of Components, CSR and Non-CSR Schools	-0.02	-0.02	0.01	-0.02
Number of Components, CSR Schools Only	0.03	0.04	0.01	0.04
R-Squared	0.34	0.39	0.62	0.61
N	349	353	129	131

Exhibit highlights: There is no relationship between the number of CSR components implemented and mathematics and reading achievement gains at the elementary or middle school levels, controlling for school demographics and baseline achievement.

Note: + p<.10; * p<.05; ** p<.01. Percent FRL, percent black, and percent Hispanic have been rescaled to a ratio varying between 0 and 10. The "Number of Components, CSR and Non-CSR Schools" row presents the change in achievement for implementing one more component among both CSR and non-CSR schools. The "Number of Components, CSR Schools Only" row presents the change in achievement among CSR schools when implementing one more component over and above the effects seen among all schools in previous first row. Standard errors are included in Exhibit B.10.

Sources: ECSRIO surveys; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

DISCUSSION

What might explain the overall weak relationship between implementation and achievement in this study, especially given stronger relationships found in previous work? One likely explanation is that prior research (e.g., Borman et al., 2003; Aladjem and Borman, 2006; Aladjem et al., 2006) focused on the adoption and implementation of clearly defined research-based reform models. The difference in the fairly strong findings from these works and the weaker conclusions in this report may lie in the large number and representativeness of schools included in this study and the relatively small percentage of these schools that implemented all, or even a majority of, the CSR components. That is, those studies focused on models with somewhat high fidelity of implementation. In this study, the average number of components

implemented after three years did not exceed five, and less than one in five schools implemented more than seven components. The relative lack of schools implementing a majority of the components may result in the analysis not being sensitive enough to identify any relationship.

One high-implementing site observed in our case studies illustrates how the components can work together to improve achievement. School I is a rural, pre-K school located in the Southeast. The state, as do other states, mandated a school improvement process (SIP) as a comprehensive strategy for increasing achievement (Component 2). School I faculty and staff were active participants in the SIP process (Component 5). As a result, the SIP was considered to be a "living document," as indicated by the handwritten comments included in monthly revisions. Furthermore, staff members used student assessment outcomes (Component 4), input from parents (Component 7), and student surveys and interviews to measure the needs, accomplishments, and perceptions of stakeholders. From 2002–05, School I's teachers participated in formal professional development associated with the reform strategies adopted by the school (Component 3). Teachers also participated in grade-level teacher meetings, cross-grade planning meetings, and school improvement committee meetings to adjust the instructional methods used. In 2003–04, the teachers began participating in a district-designed professional development strategy, which required them to observe instructional strategies implemented at other schools (Component 6). By 2003–04, the school moved out of corrective action status by making Adequate Yearly Progress (AYP) for two years. As illustrated by the survey results, however, this case is the exception rather than the rule.

Another explanation may be in the length of time needed for the achievement impacts of comprehensive school reform to become manifest. Other studies (as documented by Borman et al., 2003) have typically found the strongest effects for CSR in the fifth year and beyond, while the time period covered in this study is only three years.[22]

[22] The lack of a relationship between implementation of CSR after three years and achievement may also partly be explained by a reduction in the sample size due to nonresponse and missing survey data.

V. CSR SCHOOLS' ADOPTION OF A SCIENTIFICALLY BASED RESEARCH MODEL AND ITS RELATIONSHIP TO ACHIEVEMENT

Key Findings

What was the distribution of CSR schools across CSRQ scientific research base ratings?

- Only one-third of 2002 CSR awardees chose reform approaches with recognized scientific research bases.

Was the use of a scientifically based research model related to positive achievement outcomes? Did low-performing schools that adopted models with stronger scientific research bases have different achievement gains than other schools?

- Low-performing elementary schools that adopted models with stronger evidence of effectiveness had gains in mathematics achievement that were not found in higher-performing schools.
- Adoption of a CSR model independently determined to have had limited scientific evidence of effectiveness was associated with higher gains in middle school mathematics achievement in all CSR schools, whether they were low-performing or not. There also was weaker evidence that CSR middle schools that adopted models with limited scientific evidence may have experienced gains in middle school reading achievement relative to schools that adopted other models.
- There was weaker evidence that low-performing middle schools that adopted models with moderate or higher bases of evidence showed improvement in mathematics compared with schools using other models.
- In no other instances was adoption of models with a scientific research base related to achievement gains.

NCLB emphasizes the adoption of reform methods with strong research-based evidence of improving achievement. Specifically, two components described in *NCLB* emphasize the use of methods and strategies that are supported by scientifically based research:

- **Proven methods** and strategies for student learning, teaching, and school management that are based on scientifically based research and effective practices and have been replicated successfully in schools with diverse characteristics.

- **Scientifically based research** to significantly improve the academic achievement of students participating in such programs as compared with students in schools who have not participated in such programs; or strong evidence that such programs will significantly improve the academic achievement of participating children.

Within the 11 CSR components, two focus on the use of programs with a scientific research base. Focusing on the universe of 2002 CSR awardees, this evaluation examines the extent to which CSR schools adopted scientifically based research models and whether such models affected achievement outcomes.

The CSR Awards Database identifies 1,340 schools that received a CSR award in 2002. As with the implementation analyses that examine the relationship between comprehensiveness and achievement, the outcome measures for these analyses rely on standardized school-level achievement scores from NLSLSASD. Schools that implemented CSR models for fewer than three years were eliminated from the analyses (N = 60). The analyses are restricted to schools with full data, the number of which is provided in the results tables by school level and subject area for each analysis.

MEASURES OF THE SCIENTIFIC RESEARCH BASE OF CSR MODELS

The measure of any given reform method's scientific strength was derived by using rating scales from two Comprehensive School Reform Quality Center (CSRQ) reports on CSR reform models (CSRQ, 2006a; 2006b).[23] The CSRQ ratings provided a scale for the breadth and quality of the research base for the 31 most widely adopted CSR models.[24] The final CSRQ sample of reviewed models was based on three factors: 1) total number of schools implementing the model nationally; 2) replicability of the CSR model; and 3) comprehensiveness of model design. CSRQ ratings for the *evidence of positive effects* on student achievement were based on a rigorous, systematic review of the current literature on CSR models. The ratings were based on the number of published research studies meeting standards for rigorous research; the number of studies that provide conclusive versus suggestive findings for achievement; and the percentage of findings in either category that demonstrate a positive effect on achievement. (The CSRQ Center Quality Review Tool is presented in Appendix A.)

CSRQ recently reviewed the evidence of positive effects on achievement for 31 of the most widely adopted CSR models for elementary and secondary levels (Exhibit 28). The Comprehensive School Reform Quality Center (CSRQ) reports were not available at the time that 2002 CSR awardees made their model selections; however, schools may have used data from two earlier reviews of CSR models: Herman (1999), and Borman et al. (2003). The CSRQ reports draw heavily from the data used in previous reviews of CSR models.

These models addressed multiple aspects of school practices. Note, though, that most of the 2002 cohort of CSR schools chose models that were not subsequently rated in the CSRQ analyses. Among elementary schools, one-fourth chose models that received a moderate or moderately strong rating; most chose School Renaissance, Success for All, or America's Choice. Eleven percent of elementary schools chose models that later received limited ratings; Pearson Achievement Solutions (earlier known as Co-Nect) was the most frequently adopted of these. Among middle schools, just over 10 percent chose models that received a moderate or higher

[23] The CSRQ Center is funded under the U.S. Department of Education's Comprehensive School Reform Quality Initiatives program.

[24] The CSRQ Center conducted a total of 40 ratings because some models were rated at both the elementary and secondary levels.

rating (primarily America's Choice, Success for All, and School Development Program), while less than 5 percent chose models that received a limited rating (with most choosing Middle Start).

Exhibit 28
Number and Percentage of Schools Using CSR Models Included in CSRQ Reports

		Elementary School		Middle School	
Model Rating	**School Level**	**N**	**%**	**N**	**%**
Moderately Strong					
Direct Instruction (Full Immersion Model)	Elementary	30	3.5%		
Success for All	Elementary	54	6.2%		
Moderate					
Accelerated Schools PLUS	Elementary	12	1.4%		
America's Choice	Both	46	5.3%	20	4.1%
Core Knowledge	Elementary	13	1.5%		
First Things First	Secondary			0	0.0%
Literacy Collaborative	Elementary	13	1.5%		
National Writing Project	Elementary	4	0.5%		
School Development Program	Both	14	1.6%	12	2.5%
School Renaissance	Elementary	54	6.2%		
Success for All	Secondary			18	3.7%
Talent Development High Schools	Secondary			2	0.4%
Total Moderate and Moderately Strong		*240*	*27.6%*	*52*	*10.7%*
Limited					
ATLAS Learning Communities	Elementary	5	0.6%		
Different Ways of Knowing	Elementary	7	0.8%		
Expeditionary Learning	Secondary			5	1.0%
Integrated Thematic Instruction	Elementary	13	1.5%		
Knowledge Is Power Program (KIPP)	Secondary			0	0.0%
Middle Start	Secondary			16	3.3%
Modern Red School House	Elementary	17	2.0%		
More Effective Schools	Secondary			0	0.0%
Pearson Achievement Solutions (formerly Co-Nect)	Elementary	37	4.3%		
Project GRAD	Secondary			0	0.0%
Ventures Initiative and Focus System	Elementary	17	2.0%		
Total Limited		*96*	*11.0%*	*21*	*4.3%*
Zero					
Accelerated Schools PLUS	Secondary			3	0.6%
ATLAS Learning Communities	Secondary			7	1.4%
Breakthrough to Literacy	Elementary	11	1.3%		
Coalition of Essential Schools	Both	13	1.5%	5	1.0%
Community for Learning	Elementary	1	0.1%		
Comprehensive Early Literacy	Elementary	1	0.1%		
Expeditionary Learning	Elementary	9	1.0%		
First Steps	Elementary	2	0.2%		
High Schools That Work	Secondary			3	0.6%
Making Middle Grades Work	Secondary			4	0.8%
Modern Red School House	Secondary			3	0.6%
Onward to Excellence II	Both	6	0.7%	8	1.7%
Turning Points	Secondary			30	6.2%
Total Zero		*43*	*4.9%*	*63*	*13.0%*
Total Rated Models		**379**	**43.6%**	**136**	**28.1%**
Other (Unrated) Models		**490**	**56.4%**	**348**	**71.9%**

Exhibit highlights: More than one-fourth of elementary schools, and just over 10 percent of middle schools, chose CSR models with a moderate or moderately strong research base. Well over half of elementary schools, and nearly three-quarters of middle schools, chose CSR models that were not subsequently rated in CSRQ Center reports.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models; CSR Awards Database.

The CSRQ builds from work by Borman, Hewes, Overman, and Brown (2002) and Herman (1999) "to quantitatively evaluate CSR models as well as to provide qualitatively a narrative description of each reviewed model" (CSRQ, 2006a, p. 23). Furthermore, Kidron and Darwin (2007) state the CSRQ systematic review approach is aligned with standards articulated in a 2003 Institute of Education Sciences' technical report (Coalition for Evidence-Based Policy, 2003).

CSRQ authors conducted a systematic review (CSRQ, 2006a) of the research base on comprehensive school reform models consisting of five steps. The first step was a literature search. In order to be considered for the review, the full text publication needed to be released between 1980 and April 2005, must have examined a CSR model, and must have used quantitative methods. The study team then completed a Study Description Outcome Form, which coded studies based on the research design, outcome variables, and demographic information. At this stage of the analysis, researchers considered the use of quantitative data and the strength of the study design in determining the inclusion of studies for further review. "Research designs that passed this stage included experimental designs and quasi-experimental research designs with both pre- and posttests that evaluated the CSR model with a control group…and longitudinal and cohort designs with multiple testing periods" (CSRQ, 2006a, p. 27). Acceptable quasi-experimental designs needed to appropriately account for non-equivalence of comparison groups through statistical matching or employing statistical controls during data analysis.

The next step included the use of a Quality Indicators Form (QLIF — see Appendix A.1). Two independent researchers used this form to assess the quality of the research based on factors such as internal and external validity, and face and psychometric validity of the outcome measures. The researchers also used this form to gather statistical information, including effect sizes and other raw statistical information. The absence of valid equivalent comparison groups, failure to assess the degree and quality of implementation, or invalid assessment measures were all considered to be critical threats to validity and precluded a study's inclusion in the later stages of the review. The research team then reconciled any disagreement between the QLIF coders, and rated the study's overall causal validity as inconclusive, suggestive, or conclusive. Suggestive and conclusive studies had zero critical threats to validity and fewer than two noncritical threats to validity.[25] Conclusive studies included experimental and quasi-experimental designs, while the suggestive category included less rigorous studies (e.g., longitudinal and cohort research designs). Only suggestive and conclusive studies remained in the analysis.

Individual study ratings were then synthesized for each comprehensive reform model to determine the overall strength and nature of the scientific evidence regarding model effects on student achievement. The following categories were used in rating the evidence behind the models: Very Strong, Moderately Strong, Moderate, Limited, Zero, Negative, and No Rating (see Exhibit 29). No model reviewed by CSRQ met the criteria for very strong evidence of positive effects on student achievement, although several models had a large enough body of rigorous research to be categorized as having moderate or moderately strong effects.

[25] Examples of noncritical threats to validity included differential attrition, changes in instrumentation, disruptive or novelty events, maturation, and mismatches between level of assignment and unit of analysis (Crowley and Hauser, 2007).

Exhibit 29
Description of CSRQ Rating Scale for Strength of Research Base

Very Strong: Minimum of 10 studies meeting CSRQ standard for rigor of research design, with at least 50 percent presenting conclusive results, and 75 percent showing positive effect sizes of +.25 or higher.
Moderately Strong: Five to nine studies meeting CSRQ standard for rigor of research design, with a minimum of 50 percent presenting conclusive results, and 51–75 percent showing positive effect sizes of +.20 to +.24 or higher.
Moderate: Two to four studies meeting CSRQ standard for rigor of research design, with minimum of 50 percent presenting conclusive results, and 26–50 percent showing positive effect sizes of +.15 to +.19 or higher.
Limited: One study meeting CSRQ standard for rigor of research design, and 1–25 percent showing positive effect sizes at p< .05.
Zero: No studies meeting CSRQ standard for rigor of research design, or 0 percent showing statistically significant effects for achievement outcomes.
Negative: Minimum of 10 studies meeting CSRQ standard for rigor of research design, with at least 50 percent presenting conclusive results, and 75 percent showing negative effect at p<.05.

Exhibit highlights: CSRQ rating scale for the research base demonstrating evidence of positive student achievement effects.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models.

As with all analyses that rely on secondary data sources, the analyses examining the relationship between scientifically based research model adoption and achievement are contingent on the quality of data collection efforts and analyses conducted by external entities. As a widely distributed CSR model guidance tool for schools, the CSRQ reports aim to provide evidence from the research field in a consumer-friendly format, but the reports are not without their limitations, as acknowledged by the researchers and the broader education research community. For example, models that have shown promising effects over time are likely to be studied more often than other models in an effort to examine replicability of positive interventions across different populations. Given that the CSRQ ratings are determined in part by the number of high-quality studies conducted, the ratings may be biased toward those models that have demonstrated a more positive pattern of findings.

FINDINGS

What was the distribution of CSR schools across CSRQ scientific research base ratings?

- Only one-third of 2002 CSR awardees chose reform approaches with recognized scientific research bases.[26]

This finding is in line with prior research on the use of scientifically based research models that indicates a relatively low proportion of schools choose these evidence-based methods (Center on Education Policy, 2004). Among the CSR schools that use models with at least some evidence of effectiveness (i.e., limited to moderately strong evidence), about one-half selected approaches with moderate evidence of effectiveness in improving mathematics achievement, and two-thirds selected approaches with moderate to moderately strong evidence of effectiveness in improving reading achievement. The CSRQ reports were not available at the time that 2002 CSR awardees made their model selections; however, schools may have used data from two earlier reviews of CSR models: Herman (1999), and Borman, Hewes, Overman, and Brown (2003). The CSRQ reports draw heavily from the data used in previous reviews of CSR models.

Exhibit 30
Distribution of 2002 CSR Awardees Across CSRQ Evidence Ratings

Category	Mathematics		Reading	
	N	Percent	N	Percent
Zero	777	68	841	67
Limited Scientific Research Base	169	15	123	10
Moderate to Moderately Strong Scientific Research Base	200	17	296	23
Total	1,146	100	1260	100

Exhibit highlights: Two-thirds of schools did not adopt a scientifically based research reform model that meets CSRQ criteria.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models; CSR Awards Database.

The average baseline scores for schools adopting moderate to moderately strong research models were lower than schools adopting models with more limited evidence of effectiveness (Exhibit 31).

[26] The CSRQ Center conducted a total of 40 ratings because some models were rated at both the elementary and secondary levels. Of the 40 ratings conducted, 25 models were rated as having some evidence of effectiveness. CSRQ relies on the number and quality of research studies of the models to develop its ratings. Further detail on CSRQ methodology is provided in Chapter II of this report.

Exhibit 31
Average Baseline Standardized Achievement Scores for CSR Awardees Used in Scientifically Based Research Analyses by Evidence Category, 2002

	Average Baseline Standardized Achievement Scores							
	Elementary School				Middle School			
	Mathematics		Reading		Mathematics		Reading	
Evidence Category	N	Mean	N	Mean	N	Mean	N	Mean
Zero	312	-0.94	377	-0.95	264	-0.81	274	-0.84
Limited Scientific Research Base	129	-0.75	81	-0.88	24	-0.89	30	-0.69
Moderate to Moderately Strong Scientific Research Base	132	-1.13	203	-0.91	88	-1.27	51	-1.30

Exhibit highlights: The standardized achievement scores of CSR schools that adopted models with moderate or moderately strong research bases were over one-half a standard deviation lower in middle school reading and one-third of a standard deviation lower in middle and elementary school mathematics than schools that adopted models with limited evidence of effectiveness. There were no differences in baseline reading achievement by evidence category at the elementary school level.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Was the use of a scientifically based research model related to positive achievement outcomes? Did low-performing schools that adopted models with stronger scientific research bases have different achievement gains than other schools?

- Low-performing elementary schools that adopted models with moderate or moderately strong evidence of effectiveness had gains in mathematics achievement that were not found in higher-performing schools.
- Adoption of a CSR model with limited scientific evidence of effectiveness was associated with higher gains in middle school mathematics achievement in all CSR schools, whether they were low-performing or not. There also was weaker evidence that CSR middle schools that adopted models with limited scientific evidence may have experienced gains in middle school reading achievement relative to schools that adopted other models.
- There was weaker evidence that low-performing middle schools that adopted models with moderate or moderately strong bases of evidence showed improvement in mathematics compared with schools using other models.
- In no other instances was adoption of models with a scientific research base related to achievement gains.

To examine the relationship between adoption of scientifically based research models and growth in achievement, the CSRQ ratings for each of the models included in the elementary and middle school reports were converted into a four-point scale, with 0 indicating the model did not

meet the standards for the initial CSRQ review or received a CSRQ rating of 0 for having an insufficient number of quality research studies demonstrating effectiveness. A 3 indicated the model received a "Moderate" or "Moderately Strong" rating.[27] Of the 40 ratings conducted, 25 models were rated as having some evidence of effectiveness. No model received a "Negative" or "Very Strong" rating. Separate codes were created for each subject area because several CSRQ rated models, such as Literacy Collaborative and Comprehensive Early Literacy Learning, were designed to address learning in a specific subject area.

Additional analyses were conducted to test the appropriateness of including both nonrated models and models that received 0 ratings in the referent group. The same pattern of results emerged. Therefore, the larger sample size was retained in the analyses to promote the generalizability of the findings.[28] Eleven percent of schools reported implementing multiple CSR models simultaneously (N = 143). In those cases where schools adopted multiple models with different CSRQ ratings (N = 4), schools were assigned the higher evidence rating.

The regression analyses used to examine relationships between scientifically based research strategy adoption and school-level achievement included the rating of the scientific research base for models as a series of dummy coded variables, with models with no research base serving as the comparison group (Equation 2).[29]

Because of the differences in baseline achievement in schools that adopted models with zero, limited, and moderate to moderately strong research bases, the analyses include a measure for low-performing schools and interactions between the low-performing schools measure and the measure for the scientific research base of the model chosen. This allowed for the determination of differential achievement gains for low-performing schools that adopted models with higher levels of scientific research bases. Low-performing schools were those in the lowest quartile of mathematics or reading achievement in the baseline year (2001–02).[30]

---

[27] Due to the small number of middle schools in the moderately strong category with non-missing data, the moderate and moderately strong CSRQ ratings were collapsed into a single category in the analyses.

[28] The coefficients for "scientifically based research strategy" adoption were slightly smaller in the restricted sample analyses for both middle school mathematics and reading (0.22 versus 0.26 and 0.27 versus 0.29, respectively). In the analyses that examine the strength of the research base, the coefficients for the limited research base category are slightly smaller for both middle school mathematics and reading (0.43 versus 0.47 and 0.40 versus 0.42, respectively) in the restricted sample analyses.

[29] The regression analyses in this section relied on robust standard errors clustered at the state level.

[30] The incorporation of the interaction term results in some shared variance in the estimates of the relationships between scientifically based research and achievement. For instance, there are two estimates for limited scientifically based research—one for all schools and one for low-performing schools that represents the additional change in achievement over and above the estimate for all schools. Because the group of all schools overlaps with the group of low-performing schools, the two groups share some variance in their relationship with achievement. Substantively, this makes sense, in that there are characteristics of all CSR schools that are shared with or similar to characteristics found in low-performing schools. This overlap is not included in any of the regression coefficients because there is no way to tell how much of the variance in the overlap to attribute to all schools and how much to attribute to low-performing schools.

Equation 2

$$Y_{2005} = \alpha + \beta_1 y_{2002} + \beta_2 FRL + \beta_3 Black + \beta_4 HISP + \beta_5 Lim + \beta_6 Mod + \beta_7 LP + \beta_8 (LP * Lim) + \beta_9 (LP * Mod)$$

where

Y_{2005} is the school's standardized achievement score in 2005;

y_{2002} is the school's standardized achievement score in 2002;

FRL is the percent free or reduced-price lunch eligible in 2005;

Black is the percent black in 2005;

HISP is the percent Hispanic in 2005;

Lim and *Mod* are a series of dummy coded variables indicating strength of the scientifically based research model (limited and moderate to moderately strong, respectively);

LP is a dummy coded variable for low-performing schools (those in the lowest quartile of achievement); and,

(LP*Lim) and (LP*Mod) are the interaction terms between low-performing schools and the scientific base rating.

The results when examining the relationship between the strength of the scientific research base of the adopted model and achievement are mixed (Exhibit 32). In general, the results demonstrate stronger relationships between the scientific research bases of comprehensive school reform models and mathematics achievement gains compared to reading achievement gains. Note that Wald tests indicate that in elementary mathematics and in both middle school mathematics and reading, the models with the interaction terms (denoted as (2) in the exhibit) should be used. In elementary reading, the Wald test was not significant, so the model without the interaction terms (model (1)) should be used.

Low-performing schools that adopted models with stronger evidence of effectiveness had gains in elementary mathematics achievement that were not found in higher-performing schools. In elementary mathematics, low-performing schools choosing models with moderate or moderately strong evidence had a net gain of 0.25 standard deviations, or a 4.0 percentage point gain in percent proficient or higher.[31] There was less of a gain for schools that chose a model with

[31] The net gain of 0.25 standard deviations was calculated by adding the decrease of 0.44 standard deviations that low-performing elementary schools experience in mathematics relative to higher-performing elementary schools to the increase of 0.69 standard deviations that low-performing schools that adopt a model with moderate or moderately strong scientifically based research experience relative to low-performing schools that do not adopt such models.

limited evidence, which had net increase of 0.12 standard deviations, or about a 1.9 percentage point increase in the proportion of students scoring proficient or higher.[32]

In elementary reading, the Wald test indicates that model (1) should be used. In it, there were no statistically significant relationships between the scientific research base of the model chosen and achievement.

In middle school mathematics, CSR schools overall that chose a model with limited evidence had a gain of 0.37 standard deviations, or about a 5.9 percentage point gain in proficiency over schools choosing other models. Low-performing schools that chose models with moderate or higher evidence had an average gain of 0.26 standard deviations, or about 4.2 percentage points more than low-performing schools choosing other models. Note that this relationship is only statistically significant at $p<.10$ and should be interpreted with caution.

In middle school reading, CSR schools that chose a model with limited evidence experienced a gain of 0.21 standard deviations, or about a 3.4 percentage point increase in proficiency over schools choosing other models. This, too, was statistically significant only at $p<.10$, warranting caution in placing too much emphasis on the results.

[32] The net gain of 0.12 standard deviations was calculated by adding the decrease of 0.44 standard deviations that low-performing elementary schools experience in mathematics relative to higher-performing elementary schools to the increase of 0.56 standard deviations that low-performing schools that adopt a model with limited scientifically based research experience relative to low-performing schools that do not adopt such models.

Exhibit 32
Regression Coefficients Demonstrating the Relationships Between Scientific Research Base and Achievement Gains from 2002 to 2005

	Elementary Math		Elementary Reading		Middle School Math		Middle School Reading	
Model	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.20	0.24+	0.20	0.22	0.15	0.16	0.13	0.16
Achievement (2002)	0.52**	0.47**	0.50**	0.47**	0.71**	0.74**	0.60**	0.66**
Percent FRL (2005)	-0.03+	-0.03	-0.04*	-0.04+	-0.02	-0.02	-0.02	-0.02
Percent Black (2005)	-0.05**	-0.05**	-0.04*	-0.04*	-0.04**	-0.04**	-0.06**	-0.06**
Percent Hispanic (2005)	-0.04*	-0.04*	-0.05**	-0.05**	-0.03+	-0.03+	-0.05*	-0.05*
Limited Scientific Research Base	0.16	0.03	0.11	-0.01	0.38**	0.37**	0.29+	0.21+
Moderate/Moderately Strong Research Base	0.11	-0.09	0.02	-0.04	0.10	0.00	0.20+	0.06
Low-performing Schools		-0.44**		-0.21+		0.06		0.05
Limited Scientific Research Base in Low-performing Schools		0.56**		0.45+		0.00		0.32
Moderate/Moderately Strong Research Base in Low-performing Schools		0.69*		0.23		0.26+		0.38
r2	0.40	0.42	0.39	0.40	0.63	0.63	0.56	0.56
N	593	593	684	684	332	332	360	360
Wald Test (3 degrees of freedom)	3.43*		1.56		2.40+		2.89+	

Exhibit highlights: At the elementary level, low-performing CSR schools that adopted models with limited or moderate to moderately strong scientific research bases had higher mathematics achievement; there was no relationship between reading achievement and models with a scientific research base. At the middle school level, CSR schools that adopted models with limited scientific research bases had higher mathematics and reading achievement than those that did not; low-performing CSR schools that adopted models with moderate to moderately strong research bases had higher mathematics achievement than those that chose other models.

Note: + p<.10; * p<.05; ** p<.01. Percent FRL, percent black, and percent Hispanic have been rescaled to a ratio varying between 0 and 10. Regression coefficients presented in the table represent the change in standard deviations resulting from a one-unit change in the explanatory variable and may thus be thought of as effect sizes. Model 1 refers the basic model, which does not include the term for low-performing schools and the interaction terms between low-performing schools and the strength of the research base. Model 2 refers to the full model, which includes those terms. The incorporation of interaction terms results in shared variance not reflected in the regression estimates but included in the r2. See footnote 30 above. The Wald Test tests the hypothesis that the regression estimates that do not include the term for low-performing schools and the interaction terms between being low-performing and adopting a model with a scientific research base (labeled 1) are different from the regression estimates that do include them (labeled 2). Each Wald Test is similar to an F-test with three degrees of freedom. A statistically significant result implies that the analysis should include the terms for low-performing and the interaction terms; a result that is not statistically significant implies that the analysis should not include these terms. In the four models above, the one for elementary math and the two middle school models should include the terms for low-performing schools and the interaction terms, while the model for elementary reading should not. Robust standard errors were computed by clustering by state. Standard errors are included in Exhibit B.11.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

DISCUSSION

These findings overall do demonstrate some statistical association between the adoption by CSR grantees of "evidence-based" models and achievement improvements. However, they are also inconsistent, being more strongly manifest in mathematics than reading, and in the lowest-performing schools. Furthermore, they do not offer compelling evidence that the strength of the evidence base is strongly and consistently associated with achievement improvements.

Why does it appear that models with a scientific research base are consistently more strongly related to achievement gains in mathematics than reading? One possible explanation lies with the way CSRQ determined ratings. CSRQ determined "strength of the research base" by the nature of the evaluation design, with more weight given to more rigorous designs and the number of models evaluated that use such designs. It is possible that some highly potent strategies had not been evaluated rigorously, so the ratings underplay their potential for effectiveness. It is also possible that implementation in schools in this study may have been less comprehensive than it was in schools included in the studies that CSRQ used. Implementation problems could well have outweighed the research base for the strategy.

In addition, the measure of rigor itself may be an issue, because the CSRQ scales were originally developed to guide school decision-making in selecting a program model with strong scientific evidence of effectiveness. The most rigorous research, according to the CSRQ rating system, was that with randomized treatment (randomized controlled trials). Such research may have shown highly positive outcomes for the groups with which the model was tested. However, the adopting sites may have had very different populations, and the replicability in such sites could have been limited.

As Lipsey (2003) noted, evaluations of research and demonstration programs, such as the individual models, and what he calls "practice" programs (programs "administered by a service agency in an ongoing fashion," p. 75) yield consistently different results. The research and demonstration programs show larger effects, within each type of evaluation design, than practice programs. The results reported in this study, then, provide additional evidence underscoring the complex association between evidence of effectiveness from evaluations of educational interventions in demonstration programs and their adoption on a broader scale.

VI. CONCLUSIONS AT THE END OF THREE YEARS

SUMMARY OF FINDINGS

A summary of study findings, categorized by key evaluation questions, follows:

Was Receipt of a CSR Award Associated With Improvements in School-level Mathematics and Reading Achievement?

- Receipt of a CSR award was not associated with achievement gains in mathematics or reading achievement through the first three years of award.

Were Schools That Received CSR Awards More Likely to Implement the Legislatively Specified Components of CSR Than Other Schools?

- No, both CSR and non-CSR schools implemented an average of fewer than four components in 2003 and fewer than five in 2005 at both the elementary and middle school levels.

Was Fidelity of CSR Implementation Associated With Gains in School-level Mathematics and Reading Achievement?

Two analyses address this question:

Comprehensiveness of implementation

- The comprehensiveness of implementation, as measured by the number of CSR components implemented, was not related to mathematics and reading achievement gains in CSR schools.

Adoption of models with a recognized scientific research base

- Only one-third of 2002 CSR awardees chose reform approaches with recognized scientific research bases.

- Low-performing elementary schools that adopted models with stronger evidence of effectiveness had gains in mathematics achievement that were not found in higher-performing schools.

- Adoption of a CSR model with limited scientific evidence of effectiveness was associated with higher gains in middle school mathematics achievement in all CSR schools, whether they were low-performing or not. There also was weaker evidence that CSR middle

schools that adopted models with limited scientific evidence may have experienced gains in middle school reading achievement relative to schools that adopted other models.

- There was weaker evidence that low-performing middle schools that adopted models with moderate or higher bases of evidence showed improvement in mathematics compared with schools using other models.

- In no other instances was adoption of models with a scientific research base related to achievement gains.

What do these results suggest about the theory underlying comprehensive school reform? The CSR program awards schools relatively small increases in funds to supplement their existing resources. The funds are intended to stimulate holistic, systemic approaches to reform. However, few schools implemented even a majority of the components identified in *NCLB* or adopted models based on scientific research. Furthermore, the list of components is based on an image that they will work in a synergistic, coherent manner to bring about change. The data from this evaluation provide little evidence of such synergy and coherence in actual practice.

Consequently, the findings really do not address whether the existence of comprehensive and coherent approaches to school reform and improvement would help students perform at dramatically higher levels, only that achieving this objective is much more difficult than the designers of this program probably had envisioned.

REFERENCES

Aladjem, D., and Borman, K. M. (eds.). (2006). *Examining Comprehensive School Reform*. Washington, D.C.: Urban Institute Press.

Aladjem, D. K., LeFloch, K. C., Zhang, Y., Kurki, A., Boyle, A., Taylor, J. E., Herrman, S., Uekawa, K., Thomsen, K., and Fashola, O. (2006). *Models Matter—The Final Report of the National Longitudinal Evaluation of Comprehensive School Reform*. Washington, D.C.: American Institutes for Research. Retrieved May 31, 2007, from http://www.air.org/publications/documents/NLECSR_2006.pdf.

Borman, G. D., Hewes, G. M., Overman, L. T., and Brown, S. A. (2002). *Comprehensive School Reform and Student Achievement: A Meta-Analysis*. CRESPAR Technical Report, No. 59. Baltimore, Md.: Center for Research on the Education of Students Placed at Risk, Johns Hopkins University.

Borman, G. D., Hewes, G. M., and Overman, L. T. (2003). "Comprehensive School Reform and Achievement: A Meta-Analysis." *Review of Educational Research*, *73*(2), 125–230.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., and 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. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., and Chambers, B. (2005b). "The National Randomized Field Trial of Success for All: Second-Year Outcomes." *American Education Research Journal*, *42*(4), 673–696.

Burstein, L. (1978). "Assessing Differences Between Grouped and Individual-Level Regression Coefficients: Alternative Approaches." *Sociological Methods and Research*, *7*, 5–28.

Burstein, L., and Miller, M. D. (1981). "Regression-Based Analyses of Multi-Level Educational Data." In R. F. Boruch, P. M. Wortman, D. S. Cordray, and Associates (eds.), *Reanalyzing Program Evaluations*. San Francisco, Calif.: Jossey-Bass.

Center on Education Policy (2004). *From the Capital to the Classroom: Year 2 of the* No Child Left Behind Act. Washington, D.C.

Coalition for Evidence-Based Policy (2003, December). *Identifying and Implementing Education Practices Supported by Rigorous Evidence: A User Friendly Guide.* Retrieved October 15, 2006, from http://www.ed.gov/rschstat/research/pubs/rigorousevid/index.html.

Comprehensive School Reform Quality Center (2006a). *CSRQ Center Report on Elementary School Comprehensive School Reform Models.* Washington D.C.: American Institutes for Research.

Comprehensive School Reform Quality Center (2006b). *CSRQ Center Report on Middle and High School Comprehensive School Reform Models.* Washington D.C.: American Institutes for Research.

Crowley, J., and Hauser, A. (2007). "Evaluating Whole School Improvement Models: Creating Meaningful and Reasonable Standards of Review." *Journal of Education for Students Placed at Risk*, *12*(1), 37–58.

Flower, F. (1995). *Improving Survey Questions: Design and Evaluation.* Thousand Oaks, Calif.: Sage Publications.

Fraas, J. W., Newman, I., and Pool, S. (2007). "The Use of Propensity Score Analysis to Address Issues Associated with the Use of Adjusted Means Produced by Analysis of Covariance." *Multiple Linear Regression Viewpoints*, *33*(1).

Gill, B. P., Hamilton, L. S., Lockwood, J. R., Marsh, J. A., Zimmer, R. W., Hill, D., and Pribesh, S. (2005). *Inspiration, Perspiration, and Time: Operations and Achievement in Edison Schools*. Santa Monica, Calif.: RAND Corporation. Retrieved March 10, 2007, from http://www.rand.org/pubs/monographs/MG351/.

Hale, S. (2000). *Comprehensive School Reform: Research-Based Strategies to Achieve High Standards.* San Francisco, Calif.: Comprehensive Assistance Center, Region XI, WestEd.

Haney, W. (1980). "Units and Level of Analysis in Large-Scale Evaluation." *New Directions for Methodology and Social and Behavioral Science*, *6*, 1–15.

Hausman, J. A. (1978). "Specification Tests in Econometrics." *Econometrica*, *46*(6), 1251–1271.

Herman, R. (1999). *An Educators' Guide to Schoolwide Reform*. Arlington, Va.: Educational Research Service.

Kantor, D. (2006). "MAHAPICK: Module to Select Matching Observations Based on a Mahalanobis Scoring." Retrieved from http://ideas.repec.org/c/boc/bocode/s456703.html.

Kidron, Y., and Darwin, M. (2007). "A Systematic Review of Whole School Improvement Models." *Journal of Education for Students Placed at Risk*, *12*(1), 9–35.

Lipsey, M. (2003). "Assessing Systematic Evidence in Crime and Methodological Concerns and Empirical Outcomes." *The Annals of the American Academy of Political and Social Sciences*, *587*, 69–81.

May, H., and Supovitz, J. A. (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.

No Child Left Behind Act of 2001, Public Law 110, 107th Congress, 1st Session (Jan. 8, 2002).

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed.* Thousand Oaks, Calif.: Sage Publications.

Robinson, W. S. (1950). "Ecological Correlations and the Behavior of Individuals." *American Sociological Review, 15*, 351–357.

Rosenbaum, P. R., and Rubin, D. B., (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika, 70*, 41–55.

Rosenbaum, P. R., and Rubin, D. B. (1985). "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician, 39*(1), 33–38.

Rubin, D. B. (1980). "Bias Reduction Using Mahalanobis-Metric Matching." *Biometrics, 36*, 293–298.

Tushnet, N., Flaherty, J., and Smith, A. (2004). *Longitudinal Assessment of Comprehensive School Reform Program Implementation and Outcomes: First-Year Report*. Los Alamitos, Calif.: WestEd.

Tourangeau, R., Rips, L. J., and Rasinki, K. (2000). *The Psychology of Survey Responses.* New York, N.Y.: Cambridge University Press.

U.S. Department of Education. (2002). *Elementary and Secondary Education: Part F — Comprehensive School Reform*. Washington, D.C. Retrieved July 26, 2004, from http://www.ed.gov/policy/elsec/leg/esea02/pg13.html.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, and Office of Elementary and Secondary Education. (2007). *State* ESEA *Title I Participation Information for 2003–04*. Washington, D.C.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Common Core of Data: School Years 1998–99 through 2004–05, [http://nces.ed.gov/ccd/pubschuniv.asp], cited August 2008.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Mass.: MIT Press.

Zhang, Y., Fashola, O., Shkolnik, J., and Boyle, A. (2006). "Implementation of Comprehensive School Reform and Its Impact on Increases in Student Achievement." *Journal of Education for Students Placed at Risk*, *11*(3–4), 309–329.

Zhao, X. H., Li, C. M., Gao, S. J., and Tierney, W. M. (2001). "Methods for Testing Equality of Means of Health Care Costs in a Paired Design Study." *Statistics in Medicine*, *11*(20), 1703–20.

APPENDIX A.
CSRQ CENTER QUALITY REVIEW TOOL

The following form, from the American Institutes for Research, is reprinted here with permission.

Exhibit A.1
CSRQ Center Quality Review Tool

CSRQ Center Quality Review Tool
Part 2: Evaluating the Quality of Research Studies
STEP 3: Evaluation of Study Quality Indicators

Using the QLIF form, STEP 3 documents dimensions of validity for those quantitative studies that employ research designs deemed sufficient for drawing causal conclusions. Those studies that meet the specific causal validity requirements in the QLIF will move on to QRT 3.

Two QLIFs will be independently completed for each study by both a Lead and Secondary Coder. Because an error could disqualify an otherwise qualified study or, alternatively, qualify an ineligible study for advancement to QRT 3, the two QLIFs will be reconciled. If the two coders disagree on whether or not the study should qualify for QRT 3, a ruling will be requested by a Review Coordinator, whose judgment will be binding.

QLIF Electronic Form

During coding, if the coder would like to flag an item for reconciliation, the coder checks the ☐ next to the item.

NOTE: This form is best viewed in "NORMAL" (in "View" menu on top toolbar).

			NOTE: Complete One QLIF per CSR program per Outcome
			DATE:
			Check if you are: ☐ **Primary Coder** ☐ **Secondary Coder**
			PROGRAM NAME (Choose only one from menu):
			1. Reviewer ID code:
			2. Study citation (Author, Date, Title):
			3. Content area: (e.g. Reading, Writing, Math, etc.)
Applies to all programs/ outcomes in study?	*Pg #*	*Reconcile ?*	**A. Characterizations of the Outcome Measure**
☐		☐	**4.1** Name of outcome measure (include publication date and author if available):
			Recommendations about the item itself:
☐		☐	**5.1** List any subtests/components of the outcome measure described in the Methods section of the study.
			5.2 Notes:

☐		☐	**6.1** What type of outcome measure does this QLIF address? (Choose only one.) ☐ a. National standardized achievement test ☐ b. State standardized achievement test ☐ c. District-developed achievement test ☐ d. Standardized assessment measure (e.g., Woodcock Reading Mastery test, Peacock Vocabulary Picture Test, etc.) ☐ e. CSR-developed test ☐ f. Other test measuring achievement (e.g., teacher- or researcher-developed worksheets) ☐ g. Administrative data collected at the class, school or district level (e.g., grades, attendance, graduation rates, office referrals, detention and suspension rates, etc.) ☐ h. Survey data from students, teachers, principals, parents, etc. ☐ i. Observational data ☐ j. Other (please specify): **6.2** **6.3** Notes:
			Recommendations about the item itself:
Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**7.1-7.7** What was evidence pertaining to reliability and face validity of the instruments measuring outcomes? (Check all that apply.) ☐ 7.1 Temporal stability/test-retest reliability of the measure is at least 0.40 ☐ 7.2 The internal consistency of the measure is at least 0.60 ☐ 7.3 Inter-rater reliability is at least 0.50 ☐ 7.4 The outcome is measured using a standardized test ☐ 7.5 Face validity can be established based on sample items or author's description ☐ 7.6 Neither reliability nor face validity can be established ☐ 7.7 Not applicable (if response to 6.1 is "f") 7.8 Notes:
			Recommendations about the item itself:
☐		☐	**B. Internal Validity** **8.1** How many posttests were administered? (Choose from menu) 8.2 If more than 5, please specify: 8.3 Notes:

Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**9.1-9.5** In years, please specify when any posttests of this outcome were administered relative to the beginning of the CSR program, if not reported choose "NR" (for instance, if the first posttest occurred one year after the CSR program was implemented, fill in "1"):

	9.1 *Posttest 1*	9.2 *Posttest 2*	9.3 *Posttest 3*	9.4 *Posttest 4*	9.5 *Posttest 5*
No. of years relative to beginning of CSR program					

9.6 Notes:

Recommendations about the item itself:

☐ ☐ **10.1-10.6** Was the CSR program discontinued prior to any of the administered posttests? If so, please mark the first posttest in which the CSR program was no longer being implemented. (Choose only one from menu.) →

10.7 Notes:

Recommendations about the item itself:

☐ ☐ **10.1.1-10.1.8**. At the time of this study, which components of the CSR program were being implemented by the sample school(s)? (Check all that apply.)

- ☐ 10.1.1 Reading/English Language Arts
- ☐ 10.1.2 Writing
- ☐ 10.1.3 Mathematics
- ☐ 10.1.4 Science
- ☐ 10.1.5 Social Studies
- ☐ 10.1.6 Professional Development
- ☐ 10.1.7 Other (please specify):

  10.1.8 _____

- ☐ 10.1.9 Not Reported

10.1.10 Notes:

Recommendations about the item itself:

Applies to all?	Pg #	Reconcile ?	
☐		☐	**11.1-11.5** Which best describes the information provided by the study authors concerning the fidelity with which the CSR program was implemented? (Choose only one from each column.) [Note: both moderate and high levels of fidelity will be considered as sufficient fidelity].

	11.1 *Posttest 1*	11.2 *Posttest 2*	11.3 *Posttest 3*	11.4 *Posttest 4*	11.5 *Posttest 5*
a. Sufficient fidelity (according to the authors, implementation description, or reports on fidelity checks)	☐	☐	☐	☐	☐
b. Insufficient fidelity (according to description of implementation or fidelity checks)	☐	☐	☐	☐	☐
c. Insufficient fidelity (according to the authors)	☐	☐	☐	☐	☐
d. Not enough information	☐	☐	☐	☐	☐
e. Not reported	☐	☐	☐	☐	☐

11.6 Notes:

Recommendations about the item itself:

☐ ☐ **12.1** At what level was the assignment of students performed? (Choose only one.)

☐ a. School

☐ b. District

☐ c. Other (please specify):

12.2

12.3 Notes:

Recommendations about the item itself:

Applies to all?	Pg #	Reconcile ?	
☐		☐	**13.1-13.24** Please enter sample size of the **CSR** school(s) included in this measurement at the start of the study (pre-test) and at each posttest (if there was more than one pretest please describe the different pretests, and the reason for selecting the pretest entered in the table below):

	Pretest	*Posttest 1*	*Posttest 2*	*Posttest 3*	*Posttest 4*	*Posttest 5*
a. Total number of CSR schools	13.1	13.5	13.9	13.13	13.17	13.21
b. Total number of classes across CSR schools	13.2	13.6	13.10	13.14	13.18	13.22
c. Total number of students across all classes in all CSR schools	13.3	13.7	13.11	13.15	13.19	13.23
d. Not reported	13.4 ☐	13.8 ☐	13.12 ☐	13.16 ☐	13.20 ☐	13.24 ☐

13.25 If more than one pre-test was conducted, please specify why this pretest was selected for coding purposes:

13.26 Notes:

Recommendations about the item itself:

Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**14.1-14.6** How were the participants identified? What eligibility requirements were used to include students in the experimental sample? (Choose all that apply.) ☐ 14.1 SES ☐ 14.2 Minority status ☐ 14.3 Achievement ☐ 14.4 Urbanicity ☐ 14.5 Grade Level ☐ 14.6 Other (please specify): 14.7 14.8 Notes: Recommendations about the item itself:
☐		☐	**15.1-15.24** Please enter sample size of the **comparison** school(s) included in this measurement at the start of the study (pre-test) and at each posttest (if there was more than one pretest please describe the different pretests, and the reason for selecting the pretest entered in the table below):

	Pretest	*Posttest 1*	*Posttest 2*	*Posttest 3*	*Posttest 4*	*Posttest 5*
a. Total number of comparison schools	15.1	13.5	15.9	15.13	15.17	15.21
b. Total number of classes across comparison schools	15.2	15.6	15.10	15.14	15.18	15.22
c. Total number of students across all classes in all comparison schools	15.3	15.7	15.11	15.15	15.19	15.23
d. Not reported	15.4 ☐	13.8 ☐	15.12 ☐	15.16 ☐	15.20 ☐	15.24 ☐

15.25 If more than one pre-test was conducted, please specify why this pretest was selected for coding purposes:

15.32 Notes:

Recommendations about the item itself:

16.1 What was the unit of analysis? (Choose only one.)

- ☐ a. Student
- ☐ b. Class/Teacher
- ☐ c. School
- ☐ d. Other (please specify):

16.3 Notes:

Recommendations about the item itself:

Applies to all?	*Pg #*	*Reconcile ?*

17.1-17.5 Did the CSR or comparison groups experience severe attrition? (Choose only one from each column.) [Note: severe attrition is regarded as attrition of more than 20 percent of the students].

	17.1 *Posttest 1*	17.2 *Posttest 2*	17.3 *Posttest 3*	17.4 *Posttest 4*	17.5 *Posttest 5*
a. Yes	☐	☐	☐	☐	☐
b. No	☐	☐	☐	☐	☐
c. Not reported	☐	☐	☐	☐	☐
d. Not applicable (no comparison group)	☐	☐	☐	☐	☐

17.6 Notes:

Recommendations about the item itself:

Applies to all?	*Pg #*	*Reconcile ?*

18.1-18.5 Were there different reasons for the attrition observed from the samples of CSR and comparison students? (Choose only one from each column.) [Note: differential attrition is coded when at least 7 percent of the total students in a group discontinued their participation due to reasons different from the reasons for drop out in the other group].

	18.1 *Posttest 1*	18.2 *Posttest 2*	18.3 *Posttest 3*	18.4 *Posttest 4*	18.5 *Posttest 5*
a. Yes	☐	☐	☐	☐	☐
b. No	☐	☐	☐	☐	☐
c. Not reported	☐	☐	☐	☐	☐
d. Not applicable (answered "no" or "Not applicable" to question #17)	☐	☐	☐	☐	☐

18.6 Notes:

Recommendations about the item itself:

☐ ☐ **19.1** For designs with comparison groups, at the start of the study (at or before pre-test) did the study show there to be an initial equivalence with respect to the outcome measure evaluated between the CSR and comparison students? (Choose only one.)

- ☐ a. Initial equivalence with respect the outcome measure was not addressed
- ☐ b. The CSR and comparison schools were shown to be equivalent at the start of the study
- ☐ c. The CSR and the comparison groups were shown to be non-equivalent at the start of the study
- ☐ d. Not applicable (no comparison group)

19.2 Notes:

Recommendations about the item itself:

Applies to all?	*Pg #*	*Reconcile ?*

☐ ☐ **20.1-20.5** For designs with comparison groups, was a lack of initial group equivalence with respect to the outcome being measured taken into account at the time of analysis? (Choose only one from each column.)

	20.1 *Posttest 1*	20.2 *Posttest 2*	20.3 *Posttest 3*	20.4 *Posttest 4*	20.5 *Posttest 5*
a. Yes	☐	☐	☐	☐	☐
b. No	☐	☐	☐	☐	☐
c. Not reported	☐	☐	☐	☐	☐
d. Not applicable (no comparison group)	☐	☐	☐	☐	☐

20.6 Notes:

Recommendations about the item itself:

☐ ☐ **21.1** For time-series designs: Did the authors establish a **baseline** for the outcome measure? [Note: An appropriate baseline is a pretest of the outcome administered prior to the start of the treatment, measured with the ***same*** instrument used for the posttest(s).]

(Choose only one from menu.) →

21.2 Notes:

Recommendations about the item itself:

☐		☐	**22.1** For time-series designs: Did the authors measure the outcome in approximately equaltime intervals? (Choose only one from menu.) →
			22.2 Notes:
			Recommendations about the item itself:
Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**23.1-23.10** Was there a recent local history event (for definition, please refer to the glossary)? (choose only one from menu for each test)
			Posttest 1: 23.1 — 23.2 Specify:
			Posttest 2: 23.3 — 23.4 Specify:
			Posttest 3: 23.5 — 23.6 Specify:
			Posttest 4: 23.7 — 23.8 Specify:
			Posttest 5: 23.9 — 23.10 Specify:
			23.11 Notes:
			Recommendations about the item itself:
☐		☐	**24.1-24.10** Did students in the CSR sample experience a changed expectancy, novelty or disruption effect that was not also experienced by the comparison students that could account for differences in any of the posttest outcome measures between these groups? (choose only one from menu for each test)
			Posttest 1: 24.1 — 24.2 Specify:
			Posttest 2: 24.3 — 24.4 Specify:
			Posttest 3: 24.5 — 24.6 Specify:
			Posttest 4: 24.7 — 24.8 Specify:
			Posttest 5: 24.9 — 24.10 Specify:
			24.11 Notes:
			Recommendations about the item itself:
Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**25.1-25.11** Were there changes in the instruments or procedures used to measure outcomes during the collection of these data? (choose only one from menu for each test)
			Posttest 1: 25.1 — 25.2 Specify:
			Posttest 2: 25.3 — 25.4 Specify:
			Posttest 3: 25.5 — 25.6 Specify:
			Posttest 4: 25.7 — 25.8 Specify:
			Posttest 5: 25.9 — 25.10 Specify:

			25.11 Notes:
			Recommendations about the item itself:
☐		☐	**26.1-26.10** Did maturation occur that could account for differences in any of the posttest outcome measures between the CSR and comparison students? (choose only one from menu for each test)
			Posttest 1: 26.1 — 26.2 Specify: *Posttest 2*: 26.3 — 26.4 Specify: *Posttest 3*: 26.5 — 26.6 Specify: *Posttest 4*: 26.7 — 26.8 Specify: *Posttest 5*: 26.9 — 26.10 Specify:
			26.11 Notes:
			Recommendations about the item itself:
Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**27.1-27.10** For research designs with control groups only, were there disparities in the assignment or selection of participants in comparison groups? (choose only one from menu for each test)
			Posttest 1: 27.1 — 27.2 Specify: *Posttest 2*: 27.3 — 27.4 Specify: *Posttest 3*: 27.5 — 27.6 Specify: *Posttest 4*: 27.7 — 27.8 Specify: *Posttest 5*: 27.9 — 27.10 Specify:
			27.11 Notes:
			Recommendations about the item itself:
☐		☐	**28.1-28.10** Was there a selection by maturation interaction? (choose only one from menu for each test)
			Posttest 1: 28.1 — 28.2 Specify: *Posttest 2*: 28.3 — 28.4 Specify: *Posttest 3*: 28.5 — 28.6 Specify: *Posttest 4*: 28.7 — 28.8 Specify: *Posttest 5*: 28.9 — 28.10 Specify:

			28.11 Notes:
			Recommendations about the item itself:
Applies to all?	*Pg #*	*Reconcile ?*	
☐		☐	**29.1-29.10** Was there evidence that initial scores at the start of the study (i.e., pre-test scores) were "extreme" such that posttest scores seemed to move towards the mean, also known as statistical regression. (choose only one from menu for each test)
			Posttest 1: 29.1 — 29.2 Specify:
			Posttest 2: 29.3 — 29.4 Specify:
			Posttest 3: 29.5 — 29.6 Specify:
			Posttest 4: 29.7 — 29.8 Specify:
			Posttest 5: 29.9 — 29.10 Specify:
			29.11 Notes:
			Recommendations about the item itself:
			C. Statistical Reporting
		☐	**30.1** Questions 34 and 35 contain tables to record elements of the authors' statistical analysis. To aid in this exercise, you are required to describe the statistical analysis in narrative format addressing the following bulleted points in order: • Please indicate a detailed account of **all** the types of statistical analyses that were conducted (e.g., t-test of means, ANCOVA). • Please identify the authors' *preferred statistical method and model specification*. If the author did not explicitly state their preferred specification, please pick the most inclusive specification using the following list as a guide (list ranges from most to least inclusive): (1) HLM/SEM; (2) ANCOVA/Regression (that controls for relevant covariate); (3) ANOVA/Regression (no covariate controlled); (4) t-test of means (paired or independent samples); (5) simple correlations; (6) descriptive, quantitative stats (e.g., crosstabs, simple descriptives, change scores not analyzed with parametric analyses); (7) nonparametric statistics (e.g., Kruskall-Wallis, Wilcoxon Mann-Whitney). • Please provide a thorough description of how the outcome measures were defined [e.g. gain scores, normal curve equivalent (NCE) scores, etc.]. • If the authors provided an effect size, please report the name of the effect size (e.g., Cohen's d, Hedge's g, etc.) and/or the formula used to calculate it if these have been made available.

			30.2 Notes:
		☐	**31.** Please provide a summary of the author's results for each group

Sec1

32.1-32.55 What were the results of the statistical analysis? (Check all that apply)?

Applies to all?	*Pg #*	*Reconcile?*		*Posttest 1*	*Posttest 2*	*Posttest 3*	*Posttest 4*	*Posttest 5*
		☐	a. Positive main effects for at least part of the subtests/components of the measure, p<.05	32.1 ☐	32.12 ☐	32.23 ☐	32.34 ☐	32.45 ☐
			b. If positive main effects were reported in the statistical results for subtests/components, please indicate their names	32.2 Subtests/ Components	32.13 Subtests/ Components	32.24 Subtests/ Components	32.35 Subtests/ Components	32.46 Subtests/ Components
			c. Negative main for at least part of the subtests/components of the measure, p<.05	32.3 ☐	32.14 ☐	32.25 ☐	32.36 ☐	32.47 ☐
			d. If negative main effects were reported in the statistical results for subtests/components, please indicate their names	32.4 Subtests/ Components	32.15 Subtests/ Components	32.26 Subtests/ Components	32.37 Subtests/ Components	32.48 Subtests/ Components
			e. Non-significant	32.5 ☐	32.16 ☐	32.27 ☐	32.38 ☐	32.49 ☐
			f. If non-significant main effects were reported in the statistical results for subtests/components, please indicate their names	32.6 Subtests/ Components	32.17 Subtests/ Components	32.28 Subtests/ Components	32.39 Subtests/ Components	32.50 Subtests/ Components
			g. Positive impact due to CSR interaction effect(s) for at least part of the subtests/components of the outcome measure, p<.05	32.7 ☐	32.18 ☐	32.29 ☐	32.40 ☐	32.51 ☐
			h. If positive effects due to CSR interaction effect(s) were reported in the statistical results for subtests/components, please indicate their names	32.8 Subtests/ Components	32.19 Subtests/ Components	32.30 Subtests/ Components	32.41 Subtests/ Components	32.52 Subtests/ Components
			i. Negative impact due to CSR interaction effect(s) for at least part of the subtests/components of the outcome measure, p<.05	32.9 ☐	32.20 ☐	32.31 ☐	32.42 ☐	32.53 ☐
			j. If negative effects due to CSR interaction effect(s) were reported in the statistical results for subtests/components, please indicate their names	32.10 Subtests/ Components	32.21 Subtests/ Components	32.32 Subtests/ Components	32.43 Subtests/ Components	32.54 Subtests/ Components
			k. Non-significant interaction effect(s)	32.11 ☐	32.22 ☐	32.33 ☐	32.44 ☐	32.55 ☐

32.56 Notes:

Recommendations about the item itself:

Applies to all?	*Pg #*	*Reconcile?*	
☐		☐	**33.1-33.5** Can effect size be determined using available information? (Choose only one from each column.)

	33.1 *Posttest 1*	33.2 *Posttest 2*	33.3 *Posttest 3*	33.4 *Posttest 4*	33.5 *Posttest 5*
a. Effect size can not be determined given the available information	☐	☐	☐	☐	☐
b. The authors report sufficient statistical information to calculate effect size	☐	☐	☐	☐	☐
c. Effect size was reported by the authors	☐	☐	☐	☐	☐

33.6 Notes:

Recommendations about the item itself:

Mark HERE if completing a section	Page	Recon?
☐		☐
☐		☐

◇ **34.1.1-34.6.273** Can separate effect sizes be calculated for various subgroups of students defined by social background, need and/or other student- or school-level characteristics? (Check all that apply.)

Overall and Subgroup Characteristics	*Pretest*	*Posttest 1*	*Posttest 2*	*Posttest 3*	*Posttest 4*	*Posttest 5*
a. Overall	34.1.1 ☐	34.2.1 ☐	34.3.1 ☐	34.4.1 ☐	34.5.1 ☐	34.6.1 ☐
1. Mean Effect Treatment	34.1.2	34.2.2	34.3.2	34.4.2	34.5.2	34.6.2
2. Mean Effect Control	34.1.3	34.2.3	34.3.3	34.4.3	34.5.3	34.6.3
3. Mean Standard Deviation Treatment	34.1.4	34.2.4	34.3.4	34.4.4	34.5.4	34.6.4
4. Mean Standard Deviation Control	34.1.5	34.2.5	34.3.5	34.4.5	34.5.5	34.6.5
5. Number of Observations Treatment	34.1.6	34.2.6	34.3.6	34.4.6	34.5.6	34.6.6
6. Number of Observations Control	34.1.7	34.2.7	34.3.7	34.4.7	34.5.7	34.6.7
7. t-statistic		34.2.8	34.3.8	34.4.8	34.5.8	34.6.8
8. F-statistic		34.2.9	34.3.9	34.4.9	34.5.9	34.6.9
9. _[1] (chi-square)		34.2.10	34.3.10	34.4.10	34.5.10	34.6.10
10. P-value		34.2.11	34.3.11	34.4.11	34.5.11	34.6.11
11. Effect Size		34.2.12	34.3.12	34.4.12	34.5.12	34.6.12
12. What formula was used to calculate effect size?		34.2.13	34.3.13	34.4.13	34.5.13	34.6.13
b. Socio-economic status (SES): **High**	34.1.8 ☐	34.2.14 ☐	34.3.14 ☐	34.4.14 ☐	34.5.14 ☐	34.6.14 ☐
1. Mean Effect Treatment	34.1.9	34.2.15	34.3.15	34.4.15	34.5.15	34.6.15
2. Mean Effect Control	34.1.10	34.2.16	34.3.16	34.4.16	34.5.16	34.6.16
3. Mean Standard Deviation Treatment	34.1.11	34.2.17	34.3.17	34.4.17	34.5.17	34.6.17
4. Mean Standard Deviation Control	34.1.12	34.2.18	34.3.18	34.4.18	34.5.18	34.6.18
5. Number of Observations Treatment	34.1.13	34.2.19	34.3.19	34.4.19	34.5.19	34.6.19
6. Number of Observations Control	34.1.14	34.2.20	34.3.20	34.4.20	34.5.20	34.6.20
7. t-statistic		34.2.21	34.3.21	34.4.21	34.5.21	34.6.21
8. F-statistic		34.2.22	34.3.22	34.4.22	34.5.22	34.6.22
9. _[1] (chi-square)		34.2.23	34.3.23	34.4.23	34.5.23	34.6.23
10. P-value		34.2.24	34.3.24	34.4.24	34.5.24	34.6.24
11. Effect Size		34.2.25	34.3.25	34.4.25	34.5.25	34.6.25
12. What formula was used to calculate effect size?		34.2.26	34.3.26	34.4.26	34.5.26	34.6.26

Note: Item replicated for the following additional subgroups: Socioeconomic status (Low); Special needs: Learning disabled; Special Needs: other; Limited English proficient; Gender: female; Gender: male; Achievement: high; Achievement: low; Ethnicity: African American; Ethnicity: Hispanic; Ethnicity: Asian; Ethnicity: Native American; Grade: Grade 1; Grade: Grade 2; Grade: Grade 3; Grade: Grade 4; Grade: Grade 5; Grade: Grade 6; Other subgroup.

Mark HERE if completing a section	Page	Recon?
☐		
☐		

◇ **35.1.1-35.6.156** Can separate effect sizes be calculated for variations in posttest scores across various settings or setting characteristics? (Check all that apply.)

Setting Characteristics	*Pretest*	*Posttest 1*	*Posttest 2*	*Posttest 3*	*Posttest 4*	*Posttest 5*
a. Geographical region: Northeast (e.g. Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont)	35.1.1 ☐	35.2.1 ☐	35.3.1 ☐	35.4.1 ☐	35.5.1 ☐	35.6.1 ☐
1. Mean Effect Treatment	35.1.2	35.2.2	35.3.2	35.4.2	35.5.2	35.6.2
2. Mean Effect Control	35.1.3	35.2.3	35.3.3	35.4.3	35.5.3	35.6.3
3. Mean Standard Deviation Treatment	35.1.4	35.2.4	35.3.4	35.4.4	35.5.4	35.6.4
4. Mean Standard Deviation Control	35.1.5	35.2.5	35.3.5	35.4.5	35.5.5	35.6.5
5. Number of Observations Treatment	35.1.6	35.2.6	35.3.6	35.4.6	35.5.6	35.6.6
6. Number of Observations Control	35.1.7	35.2.7	35.3.7	35.4.7	35.5.7	35.6.7
7. t-statistic		35.2.8	35.3.8	35.4.8	35.5.8	35.6.8
8. F-statistic		35.2.9	35.3.9	35.4.9	35.5.9	35.6.9
9. χ^2 (chi-square)		35.2.10	35.3.10	35.4.10	35.5.10	35.6.10
10. P-value		35.2.11	35.3.11	35.4.11	35.5.11	35.6.11
11. Effect Size		35.2.12	35.3.12	35.4.12	35.5.12	35.6.12
12. What formula was used to calculate effect size?		35.2.13	35.3.13	35.4.13	35.5.13	35.6.13
b. Geographical region: South Atlantic (e.g. Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia)	35.1.8 ☐	35.2.14 ☐	35.3.14 ☐	35.4.14 ☐	35.5.14 ☐	35.6.14 ☐

Note: Item replicated for the following additional settings/setting characteristics: Geographical region: South Central; Geographical region: Midwest; Geographical region: Mountain; Geographical region: Pacific; Geographical region: Outside of U.S.; Urbanicity: urban; Urbanicity: rural; School- or District-level SES: low; School- or District-level SES: high; Other settings/setting characteristics.

Applies to all?	*Pg #*	*Reconcile?*	
☐		☐	**36.1-36.5** Were posttest scores approximately normally distributed within the CSR and/or comparison groups? (Choose only one from the menu for each.)
			36.1 *Posttest 1*:
			36.2 *Posttest 2*:
			36.3 *Posttest 3*:
			36.4 *Posttest 4*:
			36.5 *Posttest 5*:
			36.6 Notes:
			Recommendations about the item itself:
Applies to all?	*Pg #*	*Reconcile?*	
☐		☐	**37.1-37.5** Were the within-group variances of the posttest(s) approximately equivalent between the CSR and comparison groups (where approximate equivalence is defined as a ratio of within-group variances no greater than 3-to-1)? (Choose only one from the menu for each.)
			37.1 *Posttest 1*:
			37.2 *Posttest 2*:
			37.3 *Posttest 3*:
			37.4 *Posttest 4*:
			37.5 *Posttest 5*:
			37.6 Notes:
			Recommendations about the item itself:

AFTER COMPLETION, both coders reconcile their QLIFS. The Lead coder creates an MQLIF and sends it to his/her assigned Tracker.

Exhibit A.2
CSRQ Overall Causal Validity Mapping Rubrics

	1 (Inconclusive)	2 (Suggestive)	3 (Conclusive)
Category	**Study is limited to inconclusive with respect to its causal validity with this outcome measure if one of the conditions below *has been met*.**	**Study is limited to suggestive with respect to this outcome measure if, for the most recent posttest occurring at least one year after implementation of the CSR model (QLIF Questions 9.x is greater than or equal to 1), one of the conditions below *has been met*.**	**Study is conclusive with respect to this outcome measure if:** ▪ ***none*** **of the conditions in Column 1 hold** **and** ▪ **for the most recent posttest occurring at least one year after implementation of the CSR model (QLIF Questions 9.x is greater than or equal to 1), *none* of the conditions in Columns 2 hold.**
Threats to Face Validity/Instrument Reliability			
Face Validity of Instrument Measuring Outcome	The testing instrument used is not reliable or does not have face validity. This is the case if: • QLIF Question 7.6 is checked.		Adequately meets face validity and/or instrument reliability requirement.
Threats to Internal Validity			
Fidelity of CSR Model Implementation	The authors report that the implementation was conducted without sufficient fidelity to the adopted model. This is the case if all of the post-tests have: • QLIF Questions 11.1 to 11.5 are checked **"b"** or "c".	The authors offer few or no details as to the implementation of the reform model and do not address fidelity issues. This is the case if, for the given posttest: • QLIF Question 11.x is checked **"d"**.	Adequately meets model fidelity requirement.
Non-Equivalence of Control and Treatment Groups	For non-equivalent control group or cohort research designs only (responses **"b"** and **"d"** on SDOF Question 4.1), the control and CSR groups were not shown to be equivalent before the treatment occurred or controlled for in the analysis. This is the case if: • QLIF Questions 19.1 is checked **"a"** or **"c"** and • QLIF Questions 20.1 to 20.5 are all checked **"b"**.	For randomized control trial or regression discontinuity research designs only (responses **"a"** and **"c"** on SDOF Question 4.1), the control and CSR groups were not shown to be equivalent before the treatment occurred or controlled for in the analysis. This is the case if, for the given posttest: • QLIF Questions 19.1 is checked **"a"** or **"c"** and ▪ QLIF Question 20.x is checked **"b"**.	Adequately meets equivalency requirement.
Timing of Outcome Measure	This is the case if all of post-tests occurred less than one year after implementation of the CSR model or: ▪ QLIF Questions 9.1 to 9.5 are less than 1.		Adequately meets timing requirement.
Non-Random Attrition		The CSR and/or comparison groups experienced severe attrition for differing reasons. This is the case if: • QLIF Question 17.x is checked **"a"** and • QLIF Question 18.x is checked **"a"**.	Adequately meets attrition requirement.
Baseline		Authors did not report baseline for the outcome measure. This is the case if: • QLIF Question 21.1 is checked **"b"**.	Adequately meets baseline requirement.
History		A local history event may have influenced results. This is the case if: ▪ QLIF Question 23.x is checked **"a"**.	Adequately meets history requirement.
Disruption		The CSR students experienced a changed expectancy, novelty, or disruption effect. This is the case if: ▪ QLIF Question 24.x is checked **"a"**.	Adequately meets disruption requirement.
Instrumentation		The authors reported changes in the instruments or procedures during data collection. This is the case if: ▪ QLIF Question 25.x is checked **"a"**.	Adequately meets instrumentation requirement.
Maturation		The authors reported a maturation effect. This is the case if: ▪ QLIF Question 26.x is checked **"a"**.	Adequately meets maturation requirement.
Selection bias		There were disparities in selection or assignment of participants and/or there was selection by maturation interaction. This is the case if: ▪ QLIF Question 27.x is checked **"a"**. or ▪ QLIF Question 28.x is checked **"a"**.	Adequately meets selection bias requirement.
Regression to the mean		There was evidence of regression to the mean. This is the case if: ▪ QLIF Question 29.x is checked **"a"**.	Adequately meets regression requirement.

Note: A study is promising with respect to a specific outcome measure if none of the answer patterns listed above is rated as "YES"

Exhibit A.3
CSRQ Evidence of Positive Effects on Student Achievement

Sub-Categories	NEG rating	NO rating	ZERO rating	1=Limited	2=Moderate	3 =Moderately strong	4 = Very strong
1A) Overall Effects	**At least 10** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **5** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least 50% of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant negative model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of **ES<0**	A model has **NO** articles to evaluate (i.e. no evidence is available.)	Although **1** or more studies exist for a particular model, **zero** quality study exist. **OR** The evidence from **zero%** of the APOs does not indicate "statistically significant positive effects" as required for a LIMITED rating.	**1** quality study (suggestive or conclusive from QRT 2) exists. AND **1-50%** of the APOs across the quality studies show positive model effects that are statistically significant (p≤.05, Confidence Intervals = 95%).	**2-4** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **1** study is rated CONCLUSIVE (and/or conclusive study/ies constitute at least **50%** of the total studies coded). AND **50%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of **ES = +0.15 to +0.19**.	**5-9** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **3** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least **50%** of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of **ES = +0.20 to +0.24**.	**At least 10** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **5** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least **50%** of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of at least **ES = +0.25**.

Sub-Categories	*NEG rating*	NO rating	ZERO rating	1=Limited	2=Moderate	3=Moderately strong	4=Very strong
1B) Evidence of positive effects for diverse student populations[1,2]						**5-9** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **3** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least 50% of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of **ES = +0.20 to +0.24**.	**At least 10** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **5** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least 50% of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of at least ES = **+0.25**.
1C) Evidence of positive effects in subject areas[3,4]						**5-9** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **3** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least 50% of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of **ES = +0.20 to +0.24**.	**At least 10** quality studies (suggestive or conclusive from QRT 2) exist. AND Among the quality studies, at least **5** studies are rated CONCLUSIVE (and/or conclusive studies constitute at least 50% of the total studies coded). AND **75%** of the APOs across the quality studies show statistically significant positive model effects (p≤.05, Confidence Intervals = 95%). AND These statistically significant APOs have an overall mean model achievement effect of at least ES = **+0.25**.

[1] This table repeats for the following sub-groups: 1) High-poverty students; 2) Minority students (African American, Hispanic, Asian, Native American, Other Race/Ethnicity); 3) ELL students; 4) Learning disabled students; 5) Other special needs students; 6) Urban; and 7) Rural.

[2] We will only report information from the MODERATELY STRONG and VERY STRONG categories. For efficiency in conducting QRT 3, we will therefore conduct Quality Control on these two categories only.

[3] This table repeats for the following subject areas: 1) Reading; 2) Writing; 3) Mathematics; 4) Science; 5) Social Studies; and 6) Other subjects.

[4] We will only report information from the MODERATELY STRONG and VERY STRONG categories. For efficiency in conducting QRT 3, we will therefore conduct Quality Control on these two categories only.

APPENDIX B.
STANDARD ERROR TABLES FOR ANALYSES

This appendix includes the standard error tables for the analyses presented throughout the report.

STANDARD ERRORS FOR REPORT EXHIBITS

Exhibit B.1
Standard Errors for Differences Between the 2002 Cohort of Title I CSR Schools and Title I Non-CSR Schools

	Non-CSR		CSR	
	N	Standard Error	N	Standard Error
Standardized Assessment Scores				
Elementary Math	28,366	0.006	649	0.038
Elementary Reading	28,512	0.006	654	0.036
Middle School Math	10,135	0.010	310	0.047
Middle School Reading	10,282	0.010	318	0.051
Membership	46,667	1.631	973	14.972
Percent Minority	45,635	0.125	940	0.645
Percent Free or Reduced-price Lunch	46,632	0.166	973	1.131

Exhibit highlights: Standard errors for estimates presented in Exhibit 6 in the main body of the report.

Source: Common Core of Data (CCD).

Exhibit B.2
Standard Errors for Differences Between the 2002 Cohort of Title I CSR Schools and Non-CSR Title I Schools Chosen by Propensity Scoring: Schools with Elementary Mathematics and Reading Achievement

	Non-CSR		CSR	
	N	Standard Error	N	Standard Error
Standardized Assessment Scores				
Elementary Math	550	0.038	645	0.038
Elementary Reading	550	0.037	645	0.037
Membership	550	11.396	645	10.028
Percent Minority	533	0.868	618	0.789
Percent Free or Reduced-price Lunch	550	1.514	645	1.392

Exhibit highlights: Standard errors for estimates presented in Exhibit 7 in the main body of the report.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.3
Standard Errors for Differences Between the 2002 Cohort of Title I CSR Schools and Non-CSR Title I Schools Chosen by Propensity Scoring: Schools with Middle School Mathematics and Reading Achievement

	Non-CSR		CSR	
	N	Standard Error	N	Standard Error
Standardized Assessment Scores				
Middle School Math	248	0.051	309	0.047
Middle School Reading	248	0.050	309	0.051
Membership	248	26.536	309	23.495
Percent Minority	245	1.339	301	1.127
Percent Free or Reduced-price Lunch	248	2.346	309	2.038

Exhibit highlights: Standard errors for estimates presented in Exhibit 8 in the main body of the report.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.4
Standard Errors for Changes in Standardized Assessment Scores in CSR and Non-CSR Schools from 2002–03 to 2004–05

	N	CSR Schools	Non-CSR Schools
Elementary Mathematics	634	0.032	0.031
Elementary Reading	638	0.070	0.068
Middle School Mathematics	318	0.032	0.029
Middle School Reading	320	0.033	0.031

Exhibit highlights: Standard errors for estimates presented in Exhibit 9 in the main body of the report.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.5
Standard Errors for Average Number of CSR Components Implemented by CSR and Non-CSR Elementary Schools in 2003 and 2005

School Type	2003	2005
CSR Schools (N=292)	0.137	0.155
Non-CSR Schools (N=304)	0.137	0.146

Exhibit highlights: Standard errors for estimates presented in Exhibit 21 in the main body of the report.

Source: ECSRIO surveys.

Exhibit B.6
Standard Errors for Average Number of CSR Components Implemented by CSR and Non-CSR Middle Schools in 2003 and 2005

School Type	2003	2005
CSR Schools (N=128)	0.187	0.217
Non-CSR Schools (N=128)	0.208	0.220

Exhibit highlights: Standard errors for estimates presented in Exhibit 21 in the main body of the report.

Source: ECSRIO surveys.

Exhibit B.7
Standard Errors for Differences in Achievement and Demographic Characteristics Between Schools Included and Excluded in Analyses of the Relationships Between Implementation and Achievement, 2002 and 2005

Achievement or Demographic Characteristics	Elementary School		Middle School	
	Excluded	Included	Excluded	Included
Mathematics Achievement, 2002	0.062	0.050	0.068	0.091
Mathematics Achievement, 2005	0.062	0.047	0.091	0.077
Reading Achievement, 2002	0.060	0.047	0.095	0.076
Reading Achievement, 2005	0.063	0.047	0.096	0.082
Percent Free or Reduced-price Lunch, 2005	0.012	0.012	0.010	0.020
Percent Black, 2005	0.015	0.017	0.012	0.028
Percent Hispanic, 2005	0.012	0.014	0.009	0.026

Exhibit highlights: Standard errors for estimates presented in Exhibit 24 in the main body of the report.

Sources: CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.8
Standard Errors for Differences in Mathematics and Reading Achievement Between CSR and Non-CSR Elementary Schools Included and Excluded in Analyses of the Relationships Between Implementation and Achievement, 2005

Subject and School Type	Excluded	Included
Mathematics		
CSR Schools	0.091	0.064
Non-CSR Schools	0.083	0.069
Reading		
CSR Schools	0.092	0.065
Non-CSR Schools	0.084	0.067

Exhibit highlights: Standard errors for estimates presented in Exhibit 25 in the main body of the report.

Sources: ECSRIO surveys; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.9
Standard Errors for Differences in Mathematics and Reading Achievement Between CSR and Non-CSR Middle Schools Included and Excluded in Analyses of the Relationships Between Implementation and Achievement, 2005

Subject and School Type	Excluded	Included
Mathematics		
CSR Schools	0.120	0.114
Non-CSR Schools	0.135	0.104
Reading		
CSR Schools	0.134	0.128
Non-CSR Schools	0.140	0.099

Exhibit highlights: Standard errors for estimates presented in Exhibit 26 in the main body of the report.

Sources: ECSRIO surveys; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.10
Standard Errors for Relationship Between Number of Components Implemented and Achievement Gains from 2002 to 2005

Effect	Elementary School		Middle School	
	Mathematics	Reading	Mathematics	Reading
Constant	0.169	0.161	0.237	0.178
Achievement (2002)	0.055	0.054	0.087	0.087
Percent Free or Reduced-price Lunch (2005)	0.195	0.191	0.368	0.302
Percent Black (2005)	0.154	0.165	0.188	0.244
Percent Hispanic (2005)	0.171	0.160	0.204	0.236
CSR Award (0=No, 1=Yes)	0.181	0.167	0.233	0.254
Number of Components	0.025	0.023	0.030	0.027
Number of Components by CSR Status	0.038	0.034	0.054	0.053
N	349	353	129	131

Exhibit highlights: Standard errors for the estimates presented in Exhibit 27.

Sources: ECSRIO Surveys; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Exhibit B.11
Standard Errors for the Relationships Between Scientific Research Base and Achievement Gains from 2002 to 2005

	Elementary Math		Elementary Reading		Middle School Math		Middle School Reading	
Model	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.124	0.131	0.125	0.131	0.153	0.154	0.118	0.114
Achievement (2002)	0.058	0.057	0.041	0.044	0.056	0.056	0.097	0.075
Percent Free or Reduced-price Lunch (2005)	0.018	0.019	0.019	0.020	0.023	0.023	0.023	0.022
Percent Black (2005)	0.019	0.017	0.016	0.015	0.010	0.009	0.017	0.017
Percent Hispanic (2005)	0.018	0.019	0.019	0.019	0.016	0.016	0.021	0.019
Limited Scientific Research Base	0.112	0.087	0.104	0.110	0.119	0.111	0.151	0.118
Moderate to Moderately Strong Research Base	0.147	0.083	0.100	0.081	0.100	0.106	0.102	0.141
Low-performing Schools		0.155		0.111		0.082		0.171
Limited Scientific Research Base in Low-performing Schools		0.185		0.238		0.339		0.386
Moderate to Moderately Strong Research Base in Low-performing Schools		0.257		0.230		0.149		0.228

Exhibit highlights: Standard errors for Exhibit 32 in the main body of the report.

Sources: CSRQ Center Report on Elementary CSR Models; CSRQ Center Report on Middle and High School CSR Models; CSR Awards Database; Common Core of Data (CCD); National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

APPENDIX C.
DATA COLLECTION INSTRUMENTS

Longitudinal Assessment of Comprehensive School Reform

PRINCIPAL INVENTORY

INVENTORY OF SCHOOL REFORM EFFORTS

Dear Principal,

WHO IS CONDUCTING THIS INVENTORY?
WestEd and COSMOS Corporation, educational research organizations under contract to the U.S. Department of Education, request your participation in this inventory. WestEd and the Department of Education are conducting this inventory by the authority of Section 1607 of the *No Child Left Behind Act of 2001* (P.L. 107-110).

WHAT IS THE PURPOSE OF THIS INVENTORY?
This inventory is designed to understand how reform at your school changes the ways that you and your staff approach teaching and learning. This inventory focuses on changes that can affect every facet of school operations including classrooms, professional development, schoolwide operations (*e.g.* scheduling) and management, and administration of the school building.

HOW WILL THE RESULTS BE USED?
The data from this inventory will be used by the U.S. Department of Education and Congress to evaluate the implementation and effectiveness of federal school reform legislation. Data will be reported only in statistical summaries; your individual responses will be kept confidential.

WHY SHOULD YOU PARTICIPATE IN THIS INVENTORY?
We are conducting this inventory with only a sample of schools. Therefore, the value of your individual contribution is greatly increased because it represents many other schools. We encourage you to participate in this voluntary inventory.

WHERE SHOULD YOU MAIL YOUR COMPLETED QUESTIONNAIRE?
Please return your completed questionnaire in the enclosed envelope. If you do not have the return envelope, or have other inquiries, call toll free 1–866-880-2773 or email csr@duerrevaluation.com. You may also mail your questionnaire to:

WestEd
Attn: John Flaherty
4665 Lampson Avenue
Los Alamitos, CA 90720

THANK YOU FOR YOUR COOPERATION IN THIS IMPORTANT EFFORT! As soon as we receive your questionnaire, remuneration for you or your school will be mailed immediately.

OMB No. 1875-0222 Approval Expires 01/31/2006

Instructions:

The questions on this inventory apply to reform being implemented at your school. All questions refer to reform occurring at your school *during the last school year* (2002-2003).

Please refer to the state of reform at your school at the end of last year when answering each question. Some questions ask that you mark the *one* best response, while others ask you to mark *all that apply*.

I. School Planning

1. How formal is the school improvement plan at your school?	Last year (2002-03) (mark one response)
a. Comprehensive written plan	❒
b. Outline of a plan	❒
c. Written mission statement only	❒
d. No formal plan but teachers generally share the same ideas	❒
e. Multiple plans	❒
f. No plan (**Skip to question 4**)	❒

2. What aspects of reform are covered by the school improvement plan?	Last year (2002-03) (mark all that apply)
a. Measurable goals or objectives	❒
b. Mechanism for periodic evaluation of goals	❒
c. Curriculum and instruction content	❒
d. Student assessment rubrics	❒
e. Classroom management guidelines	❒
f. Professional development activities	❒
g. Parental involvement plan	❒
h. Framework for participation in school management	❒
i. Integration of new technology	❒
j. Other (specify) _____	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

3. What factors influenced the content of your formal school improvement plan?	Last year (2002-03) (mark all that apply)
a. State or district content standards	❒
b. State or district performance standards	❒
c. Needs identified through a school needs assessment	❒
d. School performance standards	❒
e. Specifications of adopted/adapted reform design	❒
f. Assigned by district/state	❒
g. Other (specify) _____	❒

4. Has your school been *identified* as a low performing school?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

5. Has your school been *sanctioned* because of low performance?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

II. Reform Characteristics

6. How is reform at your school designed?	Last year (2002-03) (mark one response)
a. Totally designed at this school	❒
b. Adapted with modifications from external source	❒
c. Adapted selected parts from multiple external sources	❒
d. Adopted unmodified from external source	❒

7. Indicate the primary designer for the reform efforts at your school:	Last year (2002-03) (mark one response)
a. Locally developed	❒
b. School district	❒
c. University	❒
d. Private developer or publisher	❒
e. Other (specify) _____	❒

8. Does the *primary* reform effort at your school have a name?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. Don't know	❒

9. If so, write the name here: _____

10. Is your school currently implementing more than one reform?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. Don't know	❒

11. If so, list the names here: _____

12. At which grade levels in your school are reforms mainly focused?

K 1 2 3 4 5 6 7 8 9 10 11 12 **(mark all that apply)**

13. Which categories best describe the focus of reform efforts at your school?

	Last year (2002-03) (mark all that apply)
a. Whole-school	❒
b. Reading/language arts	❒
c. Mathematics	❒
d. Science	❒
e. Social studies	❒
f. Arts	❒
g. Other (specify) _____	❒

14. Who is primarily responsible for selecting the school reform in which your school participates?

	Last year (2002-03) (mark all that apply)
a. School board	❒
b. District central office	❒
c. School administrators	❒
d. School improvement team	❒
e. Teachers	❒
f. Parents	❒
g. Other (specify) _____	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

15. Which of the following factors contribute to the selection of reform at your school?

	Last year (2002-03) (mark all that apply)
a. Cost of reform	❒
b. Theoretical or research foundation	❒
c. Successful implementation at other schools	❒
d. "Fit" with school needs	❒
e. State or district mandate	❒
f. Compatibility with assessment tools	❒
g. Best published results	❒
h. Don't know	❒
i. Other (specify) _____	❒

16. Does your school have control over budgetary issues at the school site?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

17. Does your school have control over personnel decisions at the school site?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

III. Faculty Role in Reform

18. Did the faculty formally vote to adopt the current reform in your school?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. N/A No adopted reform	❒

19. Estimate the percentage of teachers who participate in reform at your school (write percentage in space at right): _____

20. What factors determine the extent of teacher participation in reform at your school?

	Last year (2002-03) (mark all that apply)
a. All teachers participate at this school	❒
b. Our reform is subject-specific	❒
c. Our funding limits the number of teachers who can participate	❒
d. Our school reform is being phased in over time	❒
e. Other reason (specify) _____	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

21. What evidence shows that the reform chosen at your school improves student achievement?

	Last year (2002-03) (mark all that apply)
a. Independent research	❒
b. Research conducted by reform designer	❒
c. Research that included comparison or control groups	❒
d. Results from students at this school	❒
e. Other (specify) _____	❒
f. Don't know	❒

22. Does your school have performance goals for students *at the end* of each school year?

	Last year (2002-03) (mark one response)
a. Yes, by grade level	❒
b. Yes, by content area	❒
c. Other (specify) _____	❒
d. No	❒

23. Does your school have *intermediate* student performance goals within each school year to gauge student progress?

	Last year (2002-03) (mark one response)
a. Yes, by grade level	❒
b. Yes, by content area	❒
c. Other (specify) _____	❒
d. No	❒

24. What factors influenced the creation of performance goals at your school?

	Last year (2002-03) (mark all that apply)
a. State content standards	❒
b. State testing requirements	❒
c. Local standards or testing requirements	❒
d. Postsecondary eligibility requirements	❒
e. Reform efforts	❒
f. School priorities	❒
g. Parent concerns	❒
h. Other (specify) _____	❒
i. None, our school does not have performance goals	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

25. Which academic subjects are covered by goals or benchmarks at your school for student achievement?

	Last year (2002-03) (mark all that apply)
a. Reading/language arts	❒
b. Mathematics	❒
c. Science	❒
d. Social Studies	❒
e. Arts	❒
f. Other (specify) _____	❒

26. Does your school have a formal test preparation program to prepare students for mandated state tests?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No **(Skip to question 29)**	❒

27. If so, which grade levels participate in the test preparation program? (mark all that apply)

K 1 2 3 4 5 6 7 8 9 10 11 12 **(mark all that apply)**

28. If so, how much time do students spend in the test preparation program each year?

	Last year (2002-03) (mark one response)
a. 2 days or less	❒
b. Between 2 and 5 days	❒
c. Between 5 and 10 days	❒
d. More than 10 days	❒

V. Evaluation

29. Does your school have a formal written plan to evaluate progress toward improvement goals?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

30. Which of the following factors influence the direction of evaluation?

	Last year (2002-03) (mark all that apply)
a. School-developed requirements	❒
b. Requirements of reform	❒
c. District requirements	❒
d. State requirements	❒
e. Federal Title I requirements	❒
f. Other (specify) _____	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

31. What topics are addressed in the formal evaluation plan?

	Last year (2002-03) (mark all that apply)
a. Student performance	❒
b. Program implementation	❒
c. Parental participation	❒
d. Staff development	❒
e. Utility of external assistance	❒
f. Sources of financial support	❒
g. Other (specify) _____	❒

VI. Support for School Reform

32. Does the *state* reward districts or schools for meeting student achievement goals?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

33. Does the *state* sanction districts or schools for failing to meet student achievement goals?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

34. Which sources of funding contribute to implementation and operation of reform efforts at your school?

	Last year (2002-03) (mark all that apply)
a. Federal CSR funds	❒
b. Title I	❒
c. Other federal funds	❒
d. Special state grants	❒
e. Discretionary district funds	❒
f. Foundation grants	❒
g. Local community or business donations	❒
h. Other (specify) _____	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

35. How have existing resources been coordinated to support reform efforts at your school?

		Last year (2002-03) (mark all that apply)
a.	Alignment of district professional development	❒
b.	Alignment of Title I activities	❒
c.	Alignment of other special funds (*e.g.* bilingual education, or magnet schools)	❒
d.	Reallocation of staff positions	❒
e.	Other (specify) _____	❒

36. Which one of these external entities is *primarily* responsible for supporting reform efforts at your school? (choose only one)

		Last year (2002-03) (mark one response)
a.	University	❒
b.	Regional Education Laboratory	❒
c.	State agency	❒
d.	School district	❒
e.	Reform program developer	❒
f.	Teachers from other schools	❒
g.	Non-affiliated consultants	❒
h.	Other (specify) _____	❒
i.	None, our school does not use external support	❒

37. What kind of assistance does this *primary* external entity provide to support reform efforts at your school?

		Last year (2002-03) (mark all that apply)
a.	Onsite consulting	❒
b.	Professional development	❒
c.	Networking opportunities	❒
d.	Written materials for students	❒
e.	Written materials for teachers	❒
f.	Software or technological assistance	❒
g.	Other (specify) _____	❒
h.	None, our school does not use external support	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

38. What type of additional support has been available through your *district* for school reform efforts?

	Last year (2002-03) (mark all that apply)
a. Administering a needs assessment	❒
b. Providing additional school staff to support school reform	❒
c. Selecting a school reform model	❒
d. Writing grants to support school reform	❒
e. Providing professional development for school reform	❒
f. Release time for teachers	❒
g. Other (specify) _____	❒
h. None, the district does not supply additional support	❒

39. What type of additional support has been available through your *state* for school reform efforts?

	Last year (2002-03) (mark all that apply)
a. Administering a needs assessment	❒
b. Selecting a school reform model	❒
c. Writing grants to support school reform	❒
d. Providing professional development for school reform	❒
e. Release time for teachers	❒
f. Other (specify) _____	❒
g. None, the state does not supply additional support	❒

VII. Instructional Practice and Professional Development

40. What teacher enhancement opportunities are available to teachers in your school?

	Last year (2002-03) (mark all that apply)
a. Mentoring new teachers	❒
b. Coaching other teachers	❒
c. Making management decisions	❒
d. Making decisions related to curriculum development	❒
e. Participating on grade-level teams	❒
f. Participating on content area teams across grades	❒
g. Other (specify) _____	❒

41. Are teachers scheduled to have a common planning period during the day?	Last year (2002-03) (mark all that apply)
a. Yes, by grade level	❒
b. Yes, by content area	❒
c. Other (specify) _____	❒
d. No, no common planning time is scheduled	❒

42. Who organizes the scope and sequence of the curriculum at your school?	Last year (2002-03) (mark all that apply)
a. State	❒
b. District	❒
c. School	❒
d. Individual teachers	❒

43. How is professional development organized at your school?	Last year (2002-03) (mark all that apply)
a. Teachers select activities sponsored by the district or other organizations	❒
b. The district identifies common professional development themes for all teachers	❒
c. The school reform plan includes professional development activities for all teachers	❒
d. All teachers participate in the same professional development activities	❒
e. Each teacher has an individual professional development plan	❒

44. Which kinds of professional development opportunities are available to teachers at your school?	Last year (2002-03) (mark all that apply)
a. Reading/language arts instruction	❒
b. Mathematics instruction	❒
c. Instructional strategies for low-achieving, limited-English-proficient, special education, and/or migrant students	❒
d. Ensuring that curriculum and instruction are consistent with state and/or district content standards	❒
e. Ensuring that curriculum and instruction are consistent with state and/or district assessments	❒
f. Implementation of a school reform model	❒
g. Monitoring individual students' progress toward learning goals	❒
h. Interpreting reports of student achievement data	❒
i. Other (specify) _____	❒

45. In what format or setting is professional development offered to teachers at your school?

	Last year (2002-03) (mark all that apply)
a. Seminars	❒
b. Workshops	❒
c. Study groups	❒
d. Make and take	❒
e. Regional and national conferences	❒
f. Courses for college credit	❒
g. Other (specify) _____	❒
h. None, our school does not provide formal professional development	❒

46. How is time allocated to enable teachers to participate in formal professional development opportunities?

	Last year (2002-03) (mark all that apply)
a. Pupil-free days	❒
b. Release time	❒
c. Faculty meeting time	❒
d. Personal time (evenings and weekends)	❒
e. Holidays (including summer vacation)	❒
f. Our school does not provide formal professional development	❒

47. Does your school provide the equivalent of over 10 days of professional development per year?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

VIII. Role of Parents

48. By what means does your school formally communicate with parents?

	Last year (2002-03) (mark all that apply)
a. Designated parent coordinator	❒
b. Regular newsletter	❒
c. Telephone calls	❒
d. Email	❒
e. Website	❒
f. In the language spoken at home (other than English)	❒
g. Other (specify) _____	❒
h. No formal communication with parents	❒

49. In what ways does your school encourage parents to be involved in governance?

	Last year (2002-03) (mark all that apply)
a. Defining school mission and goals	❒
b. Choosing instructional materials	❒
c. Hiring teachers and staff	❒
d. Scheduling and school calendar	❒
e. Evaluating school improvement	❒
f. Fundraising	❒
g. Other (specify) _____	❒
h. School governance does not involve parents	❒

50. Estimate the percentage of parents who <u>are engaged</u> with the school in the following ways (write percentage in space at right):

	Last year (2002-03)
a. Attend parent-teacher conferences	_____
b. Demand frequent reports on their children's progress	_____
c. Actively volunteer	_____
d. Observe classroom activities	_____
e. Actively participate in a formal parent organization (PTA/PTO)	_____
f. Other (specify) _____	_____

IX. Principal Characteristics

51. How many years have you been a principal? _____

52. How many years have you been a principal at this school? _____

OMB No. 1875-0222 Approval Expires 01/31/2006

Are any of the following design(s) currently in use at your school? (mark all that apply)

Entire-School

- ❒ Accelerated Schools
- ❒ America's Choice
- ❒ ATLAS Communities
- ❒ Audrey Cohen College: Purpose-Centered Education
- ❒ Center for Effective Schools
- ❒ Child Development Project
- ❒ Coalition of Essential Schools
- ❒ Community for Learning
- ❒ Co-nect
- ❒ Core Knowledge
- ❒ Different Ways of Knowing
- ❒ Direct Instruction
- ❒ Edison Schools
- ❒ Expeditionary Learning Outward Bound
- ❒ High Schools That Work
- ❒ High/Scope Primary Grades Approach to Education
- ❒ Integrated Thematic Instruction
- ❒ MicroSociety®
- ❒ Modern Red Schoolhouse
- ❒ Montessori
- ❒ Onward to Excellence
- ❒ Paideia
- ❒ QuEST
- ❒ Roots & Wings
- ❒ School Development Program (Comer)
- ❒ Success for All
- ❒ Talent Development HS with Career Academies
- ❒ Talent Development Middle School
- ❒ The Learning Network
- ❒ Turning Points
- ❒ Urban Learning Centers
- ❒ Ventures Initiative and Focus® System
- ❒ Other whole-school reform

Reading/Language Arts

- ❒ Breakthrough to Literacy
- ❒ Carbo Reading Styles Program
- ❒ CELL/ExLL
- ❒ CORE
- ❒ Early Intervention in Reading
- ❒ Exemplary Center for Reading Instruction
- ❒ First Steps
- ❒ Junior Great Books
- ❒ Literacy Collaborative
- ❒ National Writing Project
- ❒ Reading Recovery
- ❒ Other reading/language arts design

Mathematics

- ❒ Connected Mathematics Project
- ❒ Core Plus Mathematics Project
- ❒ Growing with Mathematics
- ❒ Interactive Mathematics Program
- ❒ MATH Connections®
- ❒ U. of Chicago School Mathematics Project
- ❒ Other mathematics design

Science

- ❒ Developmental Approaches to Science, Health and Technology
- ❒ Foundational Approaches in Science Teaching
- ❒ GALAXY Classroom Science
- ❒ Iowa Chautauqua Program
- ❒ Other science design

Other

- ❒ ACCESS
- ❒ Basic Skill Builders
- ❒ COMP: Creating Conditions for Learning
- ❒ Feuerstein's Instrumental Enrichment
- ❒ HOSTS
- ❒ HOTS
- ❒ Lightspan Achieve Now
- ❒ Positive Action
- ❒ Responsive Classroom
- ❒ Success-in-the-Making
- ❒ Other - not listed on this page

Thanks for your participation! **As soon as we receive your questionnaire, remuneration for you or your school will be mailed immediately.**

OMB No. 1875-0222 Approval Expires 01/31/2006

Longitudinal Assessment of

Comprehensive School Reform

PRINCIPAL SURVEY

2004-2005 School Year

Dear Principal,

WHO IS CONDUCTING THIS SURVEY?
WestEd and Cosmos Corporation, educational research organizations under contract to the U.S. Department of Education, request your participation in this survey. WestEd and the Department of Education are conducting this survey by the authority of Section 1607 of the *No Child Left Behind Act of 2001* (P.L. 107-110).

WHAT IS THE PURPOSE OF THIS SURVEY?
This survey is designed to understand how comprehensive reform at your school changes the ways that you and your staff approach teaching and learning. This survey focuses on changes that can affect every facet of school operations including classrooms, professional development, school-wide operations (*e.g.* scheduling) and management, and administration of the school building.

HOW WILL THE RESULTS BE USED?
The data from this survey will be used by the U.S. Department of Education and Congress to evaluate the implementation and effectiveness of federal school reform legislation. Data will be reported only in statistical summaries; your individual responses will be kept confidential.

WHY SHOULD YOU PARTICIPATE IN THIS SURVEY?
We are conducting this survey with only a sample of schools. Therefore, the value of your individual contribution is greatly increased because it represents many other schools. We encourage you to participate in this voluntary survey.

WHERE SHOULD YOU MAIL YOUR COMPLETED QUESTIONNAIRE?
Please return your completed questionnaire in the enclosed envelope. If you do not have the return envelope, you can mail the survey to the address below, or contact the study team via email or using our toll-free number provided below.

Longitudinal Assessment of Comprehensive School Reform
Attn: Susan Cragle
55 Hanover Lane
Chico, CA 95973
1-866-880-2773
csr@duerrevaluation.com

THANK YOU FOR YOUR COOPERATION IN THIS IMPORTANT EFFORT!

OMB No. 1875-0222 Approval Expires 01/31/2006

Adequate Yearly Progress

1. Has your school been identified as having failed to make adequate yearly progress (AYP) as defined by the accountability provisions of NCLB during the 2003-04 school year? *(mark one response)*

a. Yes ☐
b. No (skip to question 3) ☐

2. How many consecutive years has your school failed to meet AYP? _____ years

School Improvement Plan

3. Does your school improvement plan cover the following? *(mark all that apply)*

a. Measurable goals or objectives ☐
b. Mechanism for periodic evaluation of goals ☐
c. Curriculum and instruction content ☐
d. Student assessment rubrics ☐
e. Classroom management guidelines ☐
f. Professional development activities ☐
g. Parental involvement plan ☐
h. Framework for participation in school management ☐
i. Integration of new technology ☐
j. Other (specify) _____ ☐
k. No school improvement plan ☐

4. What factors influenced the content of your formal school improvement plan? *(mark all that apply)*

a. State or district content standards ☐
b. State or district performance standards ☐
c. Needs identified through a school needs assessment ☐
d. School performance standards ☐
e. Specifications of adopted/adapted reform design ☐
f. Assigned by district/state ☐
g. Other (specify) _____ ☐
h. No school improvement plan ☐

OMB No. 1875-0222 Approval Expires 01/31/2006

School Reform

The following questions refer to the most important reform strategy or strategies in your school. Such reform strategies are those intended to improve school operations and student outcomes.

5. Please name the most important reform strategy or strategies in use at your school. (The table at the end of the survey will ask for more descriptive information about each strategy.)

_____

_____

_____

_____

6. At which grade levels in your school is reform mainly focused?

☐ K ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12 *(mark all that apply)*

Evaluation

7. In what ways does your school evaluate progress toward school reform goals? *(mark all that apply)*

a. Student performance ☐
b. Program implementation ☐
c. Parental participation ☐
d. Staff development ☐
e. Utility of external assistance ☐
f. Sources of financial support ☐
g. Other (specify) _____ ☐
h. No evaluation plan at our school ☐

OMB No. 1875-0222 Approval Expires 01/31/2006

Teacher Participation in Reform

8. To what extent do you believe reform at your school: *(Circle one number in each row.)*

	Not at all	To a small extent	To some extent	To a fairly large extent	To a great extent
a. helps improve student learning	1	2	3	4	5
b. helps teachers to teach more effectively	1	2	3	4	5
c. detracts from more important efforts	1	2	3	4	5
d. improves communication among teachers	1	2	3	4	5
e. helps to improve student behavior	1	2	3	4	5

9. What factors limit teacher participation in reform at your school? *(mark all that apply)*

a. No limitations - all teachers participate at this school	☐
b. Our reform is subject-specific – only teachers in a specific subject participate	☐
c. Our funding limits the number of teachers who can participate	☐
d. Our school reform is being phased in over time	☐
e. Other reason (specify) _____	☐

OMB No. 1875-0222 Approval Expires 01/31/2006

Professional Development

This question refers to all activities intended to help teachers develop and improve their content knowledge and classroom instruction. Examples of professional development activities include mentoring programs and coaching in addition to more traditional activities such as internships, workshops, conferences, institutes, and college courses. These may be conducted within the school or outside the school setting.

10. How is professional development organized at your school? Include only workshops, coursework, and conferences sponsored by your school, district, or state during the 2003-04 school year (including summer 2004). *(mark all that apply)*

a. Teachers select activities sponsored by the district or other organizations	☐
b. The district identifies common professional development themes for all teachers	☐
c. The school reform plan includes professional development activities for all teachers	☐
d. All teachers participate in the same professional development activities	☐
e. Each teacher has an individual professional development plan	☐

Parent Involvement

11. In what ways does your school encourage parents to be involved in governance-related activities? *(mark all that apply)*

a. Defining school mission and goals	☐
b. Choosing instructional materials	☐
c. Hiring teachers and staff	☐
d. Scheduling and school calendar	☐
e. Evaluating school improvement	☐
f. Fundraising	☐
g. Other (specify) _____	☐
h. School governance does not involve parents	☐

OMB No. 1875-0222 Approval Expires 01/31/2006

Support for Reform

12. Over which of the following resources do you have control at the school site for implementation and operation of reform efforts at your school? *(mark all that apply)*

a. Federal Comprehensive School Reform funds	☐
b. Title I	☐
c. Other federal funds	☐
d. Special state grants	☐
e. Discretionary district funds	☐
f. Foundation grants	☐
g. Local community or business donations	☐
h. Other (specify) _____	☐

13. What type of support has been available through your *district* for school reform efforts? *(mark all that apply)*

a. Administering a needs assessment	☐
b. Providing additional school staff to support school reform	☐
c. Selecting a school reform model	☐
d. Writing grants to support school reform	☐
e. Providing professional development for school reform	☐
f. Release time for teachers	☐
g. Other (specify) _____	☐
h. None, the district does not supply support	☐

14. Which of the following external entities supports reform efforts at your school? *(mark all that apply)*

a. University	☐
b. Regional Education Laboratory	☐
c. State agency	☐
d. Reform program developer	☐
e. Teachers from other schools	☐
f. Non-affiliated consultants	☐
g. Other (specify) _____	☐
h. None, our school does not use external support	☐

OMB No. 1875-0222 Approval Expires 01/31/2006

Reform Strategies

Schools often use single or multiple reform strategies designed locally or by external program developers. Some strategies are externally designed, for example Success for All or Co-nect. Other strategies are local and may not have a formal name. On this page we are collecting specific information about each strategy in use at your school. Begin by writing the name of the most important reform strategy at the top of the first column and then answer each question considering that strategy only. Then, using the next column, repeat the process again for each additional strategy at your school. If your school is using a single strategy, complete the first column only. If your school has no reform strategy, write "none" in the first column.

Reform Strategy Name ➔ ➔ ➔ ➔ ➔ ➔ ➔ ➔ ➔ ➔				
	(circle one)	**(circle one)**	**(circle one)**	**(circle one)**
Description of Reform Strategy				
Does the strategy target curricular areas for improvement?	Yes No	Yes No	Yes No	Yes No
If yes, indicate curricular area(s) targeted: *(Circle all that apply.)*	Language Arts Mathematics Social Studies Science Arts Other	Language Arts Mathematics Social Studies Science Arts Other	Language Arts Mathematics Social Studies Science Arts Other	Language Arts Mathematics Social Studies Science Arts Other
Does the strategy target school organization and management for improvement?	Yes No	Yes No	Yes No	Yes No
Does the strategy target community and parent involvement?	Yes No	Yes No	Yes No	Yes No
Do you have evidence that this strategy improves student achievement?	Yes No	Yes No	Yes No	Yes No
Does the evidence include comparison groups?	Yes No	Yes No	Yes No	Yes No
Has the strategy been demonstrated to improve student achievement at other schools?	Yes No	Yes No	Yes No	Yes No
Status of Reform Strategy				
Are all grades included in the strategy?	Yes No	Yes No	Yes No	Yes No
Is this strategy being phased-in?	Yes No	Yes No	Yes No	Yes No
Is the strategy accompanied by benchmarks for implementation?	Yes No	Yes No	Yes No	Yes No
Did your district mandate this strategy?	Yes No	Yes No	Yes No	Yes No
Will sanctions result from failure to implement this strategy?	Yes No	Yes No	Yes No	Yes No
Support Provided through Reform Strategy				
Was this strategy developed externally?	Yes No	Yes No	Yes No	Yes No
Does the strategy include specific curricular materials?	Yes No	Yes No	Yes No	Yes No
Do guidelines for classroom practice accompany the strategy?	Yes No	Yes No	Yes No	Yes No
How many hours of training on the strategy did teachers receive from the program developer?	____ hours	____ hours	____ hours	____ hours
How many hours of training on the strategy did teachers receive from the district?	____ hours	____ hours	____ hours	____ hours
How many hours of training on the strategy did teachers receive from the state?	____ hours	____ hours	____ hours	____ hours
How many of these total training hours (from the developer, district and state) occurred at your school?	____ total hrs.	____ total hrs.	____ total hrs.	____ total hrs.
Did all teachers at your school receive training to use the strategy?	Yes No	Yes No	Yes No	Yes No
Is ongoing support for the strategy available through an onsite facilitator?	Yes No	Yes No	Yes No	Yes No

OMB No. 1875-0222 Approval Expires 01/31/2006

Background

15. How many years have you been a principal? _____ years

16. How many years have you been a principal at this school? _____ years

THANK YOU FOR PARTICIPATING!

Please return your completed questionnaire in the enclosed envelope. If you do not have the return envelope, you can mail the survey to the address below, or contact the study team via email or using our toll-free number provided below.

Longitudinal Assessment of Comprehensive School Reform
Attn: Susan Cragle
55 Hanover Lane
Chico, CA 95973
1-866-880-2773
csr@duerrevaluation.com

OMB No. 1875-0222 Approval Expires 01/31/2006

Longitudinal Assessment of Comprehensive School Reform

TEACHER INVENTORY

INVENTORY OF SCHOOL REFORM EFFORTS

Dear Teacher,

WHO IS CONDUCTING THIS INVENTORY?
WestEd and COSMOS Corporation, educational research organizations under contract to the U.S. Department of Education, request your participation in this inventory. WestEd and the Department of Education are conducting this inventory by the authority of Section 1607 of the *No Child Left Behind Act of 2001* (P.L. 107-110).

WHAT IS THE PURPOSE OF THIS INVENTORY?
This inventory is designed to understand how reform at your school changes the ways that you and your staff approach teaching and learning. This inventory focuses on changes that can affect every facet of school operations including classrooms, professional development, schoolwide operations (*e.g.* scheduling) and management, and administration of the school building.

HOW WILL THE RESULTS BE USED?
The data from this inventory will be used by the U.S. Department of Education and Congress to evaluate the implementation and effectiveness of federal school reform legislation. Data will be reported only in statistical summaries; your individual responses will be kept confidential.

WHY SHOULD YOU PARTICIPATE IN THIS INVENTORY?
We are conducting this inventory with only a sample of schools. Therefore, the value of your individual contribution is greatly increased because it represents many other schools. We encourage you to participate in this voluntary inventory.

WHERE SHOULD YOU MAIL YOUR COMPLETED QUESTIONNAIRE?
Please return your completed questionnaire in the enclosed envelope. If you do not have the return envelope, or have other inquiries, call toll free 1–866-880-2773 or email csr@duerrevaluation.com. You may also mail your questionnaire to:

WestEd
Attn: John Flaherty
4665 Lampson Avenue
Los Alamitos, CA 90720

THANK YOU FOR YOUR COOPERATION IN THIS IMPORTANT EFFORT! As soon as we receive your questionnaire, remuneration for you or your school will be mailed immediately.

OMB No. 1875-0222 Approval Expires 01/31/2006

Instructions:

The questions on this inventory apply to reform being implemented at your school. All questions refer to reform occurring at your school *during the last school year* (2002-2003).

Please refer to the state of reform at your school at the end of last year when answering each question. Some questions ask that you mark the *one* best response, while others ask you to mark *all that apply*.

I. Teacher Characteristics

1. **How many years have you been a teacher?** _____

2. **How many years have you taught at this school?** _____

3. **How many years have you taught the same grade/subject area?** _____

4. **What type of credential(s) do you hold?**

	Last year (2002-03) (mark all that apply)
a. Professional	❒
b. Preliminary	❒
c. Emergency	❒
d. Substitute	❒
e. Other _____	❒

II. School Planning

5. **How formal is the school improvement plan at your school?**

	Last year (2002-03) (mark one response)
a. Comprehensive written plan	❒
b. Outline of a plan	❒
c. Written mission statement only	❒
d. No formal plan but teachers generally share the same ideas	❒
e. Multiple plans	❒
f. No plan **(Skip to question 8)**	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

6. What aspects of reform are covered by the school improvement plan?	Last year (2002-03) (mark all that apply)
a. Measurable goals or objectives	❒
b. Mechanism for periodic evaluation of goals	❒
c. Curriculum and instruction content	❒
d. Student assessment rubrics	❒
e. Classroom management guidelines	❒
f. Professional development activities	❒
g. Parental involvement plan	❒
h. Framework for participation in school management	❒
i. Integration of new technology	❒
j. Other (specify) _____	❒

7. What factors influenced the content of your formal school improvement plan?	Last year (2002-03) (mark all that apply)
a. State or district content standards	❒
b. State or district performance standards	❒
c. Needs identified through a school needs assessment	❒
d. School performance standards	❒
e. Specifications of adopted/adapted reform design	❒
f. Assigned by district/state	❒
g. Other (specify) _____	❒

8. Has your school been *identified* as a low performing school?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

9. Has your school been *sanctioned* because of low performance?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

III. Reform Characteristics

10. Does the *primary* reform effort at your school have a name?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. Don't know	❒

11. If so, write the name here: _____

12. Is your school currently implementing more than one reform?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. Don't know	❒

13. If so, list the names here: _____

14. At which grade levels in your school are reforms mainly focused?

K 1 2 3 4 5 6 7 8 9 10 11 12 **(mark all that apply)**

15. Which categories best describe the focus of reform efforts at your school?

	Last year (2002-03) (mark all that apply)
a. Whole-school	❒
b. Reading/language arts	❒
c. Mathematics	❒
d. Science	❒
e. Social studies	❒
f. Arts	❒
g. Other (specify) _____	❒

16. Which of the following factors contribute to selection of a reform at your school?

	Last year (2002-03) (mark all that apply)
a. Cost of reform	❒
b. Theoretical or research foundation	❒
c. Successful implementation at other schools	❒
d. "Fit" with school needs	❒
e. State or district mandate	❒
f. Compatibility with assessment tools	❒
g. Best published results	❒
h. Don't know	❒
i. Other (specify) _____	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

IV. Faculty Role in Reform

17. Did the faculty formally vote to adopt the current reform in your school?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. N/A No adopted reform	❒

18. Were you employed at this school when faculty voted formally to adopt the current reform?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒
c. N/A No formal vote or adopted reform	❒

19. Do you willingly participate in reform-related activities at your school?	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

V. School Performance Goals

20. Does your school have performance goals for students *at the end* of each school year?	Last year (2002-03) (mark all that apply)
a. Yes, by grade level	❒
b. Yes, by content area	❒
c. Other (specify) _____	❒
d. No	❒

21. Does your school have *intermediate* student performance goals within each school year to gauge student progress?	Last year (2002-03) (mark all that apply)
a. Yes, by grade level	❒
b. Yes, by content area	❒
c. Other (specify) _____	❒
d. No	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

22. Which academic subjects are covered by goals or benchmarks at your school for student achievement?

	Last year (2002-03) (mark all that apply)
a. Reading/language arts	❒
b. Mathematics	❒
c. Science	❒
d. Social Studies	❒
e. Arts	❒
f. Other (specify) _____	❒

23. Does your school have a formal test preparation program to prepare students for mandated state tests?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No **(Skip to question 26)**	❒

24. If so, which grade levels participate in the test preparation program? (mark all that apply)

K 1 2 3 4 5 6 7 8 9 10 11 12 **(mark all that apply)**

25. If so, how much time do students spend in the test preparation program each year?

	Last year (2002-03) (mark one response)
a. 2 days or less	❒
b. Between 2 and 5 days	❒
c. Between 5 and 10 days	❒
d. More than 10 days	❒

26. Which one of these external entities is *primarily* responsible for supporting reform efforts at your school? (choose only one)

	Last year (2002-03) (mark one response)
a. University	❒
b. Regional Education Laboratory	❒
c. State agency	❒
d. School district	❒
e. Reform program developer	❒
f. Teachers from other schools	❒
g. Non-affiliated consultants	❒
h. Other (specify) _____	❒
i. None, our school does not use external support	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

27. What kind of assistance does this *primary* external entity provide to support reform efforts at your school?

	Last year (2002-03) (mark all that apply)
a. Onsite consulting	❒
b. Professional development	❒
c. Networking opportunities	❒
d. Written materials for students	❒
e. Written materials for teachers	❒
f. Software or technological assistance	❒
g. Other (specify) _____	❒
h. None, our school does not use external support	❒

28. What type of additional support has been available through your *district* for school reform efforts?

	Last year (2002-03) (mark all that apply)
a. Administering a needs assessment	❒
b. Providing additional school staff to support school reform	❒
c. Selecting a school reform model	❒
d. Writing grants to support school reform	❒
e. Providing professional development for school reform	❒
f. Release time for teachers	❒
g. Other (specify) _____	❒
h. None, the district does not supply additional support	❒

29. What type of additional support has been available through your *state* for school reform efforts?

	Last year (2002-03) (mark all that apply)
a. Administering a needs assessment	❒
b. Selecting a school reform model	❒
c. Writing grants to support school reform	❒
d. Providing professional development for school reform	❒
e. Release time for teachers	❒
f. Other (specify) _____	❒
g. None, the state does not supply additional support	❒

VI. Instructional Practice and Professional Development

30. What teacher enhancement opportunities are available to teachers in your school?	Last year (2002-03) (mark all that apply)
a. Mentoring new teachers	❒
b. Coaching other teachers	❒
c. Making management decisions	❒
d. Making decisions related to curriculum development	❒
e. Participating on grade-level teams	❒
f. Participating on content area teams across grades	❒
g. Other (specify) _____	❒

31. Are teachers scheduled to have a common planning period during the day?	Last year (2002-03) (mark all that apply)
a. Yes, by grade level	❒
b. Yes, by content area	❒
c. Other (specify) _____	❒
d. No, no common planning time is scheduled	❒

32. Do any of the following grouping patterns occur in your school (for some or all of the students in grades participating in reform)?	Last year (2002-03) (mark all that apply)
a. Students remain with one teacher for most subjects	❒
b. Students are divided into groups such as "houses" or "families"	❒
c. Student groups remain two or more years with the same teacher	❒
d. Interdisciplinary teaching (e.g., two or more teachers with different academic specializations collaborate to teach an interdisciplinary program to the same group of students)	❒
e. Paired or team teaching (e.g., two teachers are jointly responsible for teaching a single group of students)	❒
f. None of the above	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

33. ***If you are a secondary school teacher*, what kind of student groups exist to achieve particular instructional goals?**

		Last year (2002-03) (mark all that apply)
a.	Ability grouping	❒
b.	Academic or technical academies	❒
c.	School-to-career paths	❒
d.	No such grouping exists	❒
e.	Other (specify) _____	❒
f.	Not applicable	❒

34. Who organizes the scope and sequence of the curriculum at your school?

		Last year (2002-03) (mark all that apply)
a.	State	❒
b.	District	❒
c.	School	❒
d.	Individual teachers	❒

35. How is professional development organized at your school?

		Last year (2002-03) (mark all that apply)
a.	Teachers select activities sponsored by the district or other organizations	❒
b.	The district identifies common professional development themes for all teachers	❒
c.	The school reform plan includes professional development activities for all teachers	❒
d.	All teachers participate in the same professional development activities	❒
e.	Each teacher has an individual professional development plan	❒

36. Which kinds of professional development opportunities are available to teachers at your school?

		Last year (2002-03) (mark all that apply)
a.	Reading/language arts instruction	❒
b.	Mathematics instruction	❒
c.	Instructional strategies for low-achieving, limited-English-proficient, special education, and/or migrant students	❒
d.	Ensuring that curriculum and instruction are consistent with state and/or district content standards	❒
e.	Ensuring that curriculum and instruction are consistent with state and/or district assessments	❒
f.	Implementation of a school reform model	❒
g.	Monitoring individual students' progress toward learning goals	❒
h.	Interpreting reports of student achievement data	❒
i.	Other (specify) _____	❒

37. In what format or setting is professional development offered to teachers at your school?

	Last year (2002-03) (mark all that apply)
a. Seminars	❒
b. Workshops	❒
c. Study groups	❒
d. Make and take	❒
e. Regional and national conferences	❒
f. Courses for college credit	❒
g. Other (specify) _____	❒
h. None, our school does not provide formal professional development	❒

38. How is time allocated to enable teachers to participate in formal professional development opportunities?

	Last year (2002-03) (mark all that apply)
a. Pupil-free days	❒
b. Release time	❒
c. Faculty meeting time	❒
d. Personal time (evenings and weekends)	❒
e. Holidays (including summer vacation)	❒
f. Our school does not provide formal professional development	❒

39. Does your school provide the equivalent of over 10 days of professional development per year?

	Last year (2002-03) (mark one response)
a. Yes	❒
b. No	❒

VII. Role of Parents

40. By what means does your school formally communicate with parents?

	Last year (2002-03) (mark all that apply)
a. Designated parent coordinator	❒
b. Regular newsletter	❒
c. Telephone calls	❒
d. Email	❒
e. Website	❒
f. In the language spoken at home (other than English)	❒
g. Other (specify) _____	❒
h. No formal communication with parents	❒

OMB No. 1875-0222 Approval Expires 01/31/2006

41. Estimate the percentage of parents who support their children's learning in the following ways (write percentage in space at right): **Last year (2002-03)**

i.	Monitor attendance from home	_____
j.	Help students with homework	_____
k.	Volunteer to assist in classrooms	_____
l.	Other (specify) _____	_____

42. Estimate the percentage of parents who are engaged with the school in the following ways (write percentage in space at right): **Last year (2002-03)**

m.	Attend parent-teacher conferences	_____
n.	Demand frequent reports on their children's progress	_____
o.	Actively volunteer in the school	_____
p.	Observe classroom activities	_____
q.	Actively participate in a formal parent organization (PTA/PTO)	_____
r.	Other (specify) _____	_____

VIII. Reform Designs in Use

Are any of the following design(s) currently in use at your school? (mark all that apply)

Entire-School

- ❒ Accelerated Schools
- ❒ America's Choice
- ❒ ATLAS Communities
- ❒ Audrey Cohen College: Purpose-Centered Education
- ❒ Center for Effective Schools
- ❒ Child Development Project
- ❒ Coalition of Essential Schools
- ❒ Community for Learning
- ❒ Co-nect
- ❒ Core Knowledge
- ❒ Different Ways of Knowing
- ❒ Direct Instruction
- ❒ Edison Schools
- ❒ Expeditionary Learning Outward Bound
- ❒ High Schools That Work
- ❒ High/Scope Primary Grades Approach to Educ.
- ❒ Integrated Thematic Instruction
- ❒ MicroSociety®
- ❒ Modern Red Schoolhouse
- ❒ Montessori
- ❒ Onward to Excellence
- ❒ Paideia
- ❒ QuEST
- ❒ Roots & Wings
- ❒ School Development Program (Comer)
- ❒ Success for All
- ❒ Talent Development HS with Career Academies
- ❒ Talent Development Middle School
- ❒ The Learning Network
- ❒ Turning Points
- ❒ Urban Learning Centers
- ❒ Ventures Initiative and Focus® System
- ❒ Other whole-school reform

Reading/Language Arts

- ❒ Breakthrough to Literacy
- ❒ Carbo Reading Styles Program
- ❒ CELL/ExLL
- ❒ CORE
- ❒ Early Intervention in Reading
- ❒ Exemplary Center for Reading Instruction
- ❒ First Steps
- ❒ Junior Great Books
- ❒ Literacy Collaborative
- ❒ National Writing Project
- ❒ Reading Recovery
- ❒ Other reading/language arts design

Mathematics

- ❒ Connected Mathematics Project
- ❒ Core Plus Mathematics Project
- ❒ Growing with Mathematics
- ❒ Interactive Mathematics Program
- ❒ MATH Connections®
- ❒ U. of Chicago School Mathematics Project
- ❒ Other mathematics design

Science

- ❒ Developmental Approaches to Science, Health and Technology
- ❒ Foundational Approaches in Science Teaching
- ❒ GALAXY Classroom Science
- ❒ Iowa Chautauqua Program
- ❒ Other science design

Other

- ❒ ACCESS
- ❒ Basic Skill Builders
- ❒ COMP: Creating Conditions for Learning
- ❒ Feuerstein's Instrumental Enrichment
- ❒ HOSTS
- ❒ HOTS
- ❒ Lightspan Achieve Now
- ❒ Positive Action
- ❒ Responsive Classroom
- ❒ Success-in-the-Making
- ❒ Other - not listed on this page

Thanks for your participation! **As soon as we receive your questionnaire, remuneration for you or your school will be mailed immediately.**

OMB No. 1875-0222 Approval Expires 01/31/2006

Longitudinal Assessment of

Comprehensive School Reform

TEACHER SURVEY

2004-2005 School Year

Dear Teacher,

WHO IS CONDUCTING THIS SURVEY?
WestEd and Cosmos Corporation, educational research organizations under contract to the U.S. Department of Education, request your participation in this survey. WestEd and the Department of Education are conducting this survey by the authority of Section 1607 of the *No Child Left Behind Act of 2001* (P.L. 107-110).

WHAT IS THE PURPOSE OF THIS SURVEY?
This survey is designed to understand how comprehensive reform at your school changes the ways that you and your staff approach teaching and learning. This survey focuses on changes that can affect every facet of school operations including classrooms, professional development, school-wide operations (*e.g.* scheduling) and management, and administration of the school building.

HOW WILL THE RESULTS BE USED?
The data from this survey will be used by the U.S. Department of Education and Congress to evaluate the implementation and effectiveness of federal school reform legislation. Data will be reported only in statistical summaries; your individual responses will be kept confidential.

WHY SHOULD YOU PARTICIPATE IN THIS SURVEY?
We are conducting this survey with only a sample of schools. Therefore, the value of your individual contribution is greatly increased because it represents many other schools. We encourage you to participate in this voluntary survey.

WHERE SHOULD YOU MAIL YOUR COMPLETED QUESTIONNAIRE?
Please return your completed questionnaire in the enclosed envelope. If you do not have the return envelope, you can mail the survey to the address below, or contact the study team via email or using our toll-free number provided below.

Longitudinal Assessment of Comprehensive School Reform
Attn: Susan Cragle
55 Hanover Lane
Chico, CA 95973
1-866-880-2773
csr@duerrevaluation.com

THANK YOU FOR YOUR COOPERATION IN THIS IMPORTANT EFFORT!

Professional Development

The questions in this section refer to all activities intended to help teachers develop and improve their content knowledge and classroom instruction. Examples of professional development activities include mentoring programs and coaching in addition to more traditional activities such as internships, workshops, conferences, institutes, and college courses. These may be conducted within the school or outside the school setting.

1. **During the last school year (2003-04, including summer 2004), how many of the following types of professional development activities did you participate in?** *(Circle one number in each row.)*

Type of Professional Development Activity	None	1- 2 activities	3 - 5 activities	6 - 10 activities	11 or more activities
a. Conferences lasting two days or longer	1	2	3	4	5
b. Institutes (i.e., an intensive course of instruction on a particular topic or set of topics) lasting two days or longer	1	2	3	4	5
c. A series of connected workshops lasting two days or longer	1	2	3	4	5
d. Workshops lasting one day or less	1	2	3	4	5
e. Courses for college credit	1	2	3	4	5
f. Internships	1	2	3	4	5

2. **During the last school year (2003-04, including summer 2004), how frequently did you engage in each of the following professional development activities?** *(Circle one number in each row.)*

Type of Professional Development Activity	Never	Once or a few times a year	Once or twice a month	Once or twice a week	Daily or almost daily
a. Planned lessons or courses with other teachers	1	2	3	4	5
b. Consulted with other teachers about individual students (e.g., discussing specific students and arranging appropriate help)	1	2	3	4	5
c. Exchanged feedback with other teachers based on classroom observations (e.g., a teachers' observation of your class, your observation of another teachers' class, or observation of a class via video)	1	2	3	4	5
d. Exchanged feedback with other teachers based on student work	1	2	3	4	5
e. Acted as a formal or informal coach or mentor to other teachers or staff	1	2	3	4	5
f. Received formal or informal coaching or mentoring from other teachers or staff	1	2	3	4	5
g. Was observed/evaluated by the school principal or other staff (e.g., department chair, master teacher)	1	2	3	4	5
h. Participated in a learning community (e.g., teacher collaborative, network, or study group)	1	2	3	4	5
i. Participated in a district or school committee focused on curriculum, instruction, or student assessment	1	2	3	4	5
j. Visited other schools to observe classroom teaching and learning	1	2	3	4	5

3. **During the last school year (2003-04, including summer 2004), about how many hours of professional development did you receive in each of the following areas?** *(Circle one number in each row.)*

Area of Professional Development	None	1- 5 hours	6 - 24 hours	25 – 40 hours	41 - 80 hours	More than 80 hours
a. Instructional strategies for teaching reading/language arts/English	1	2	3	4	5	6
b. In-depth study of topics in reading/language arts/English	1	2	3	4	5	6
c. Instructional strategies for teaching mathematics	1	2	3	4	5	6
d. In-depth study of topics in mathematics	1	2	3	4	5	6
e. Instructional strategies or in-depth study of topics in other academic subject (e.g., science, social studies, foreign language, etc.)	1	2	3	4	5	6
f. Instructional strategies for students with limited English proficiency (LEP)	1	2	3	4	5	6
g. Instructional strategies for students with Individualized education programs (IEPs)	1	2	3	4	5	6
h. Preparing students to take the annual state assessment	1	2	3	4	5	6
i. Analyzing and interpreting student achievement data	1	2	3	4	5	6
j. Classroom and behavior management	1	2	3	4	5	6
k. Use of technology to improve classroom instruction	1	2	3	4	5	6
l. Use of appropriate assessment accommodations	1	2	3	4	5	6
m. Family/community involvement	1	2	3	4	5	6

4. **Thinking of all of the different professional development activities that you participated in during the last school year (2003-04, including summer 2004) and reported in this section, approximately how many total hours did you spend in professional development?** *(Write in the number of hours.)*

_____ hours

5. **Indicate the number of hours allocated for professional development from each of the following sources during the last school year (2003-04, including summer 2004). Include only workshops, coursework, and conferences sponsored by your school, district, or state.** *(Circle one number in each row.)*

Source of Professional Development Time	None	1- 5 hours	6 - 24 hours	25 – 40 hours	41 - 80 hours	More than 80 hours
a. Pupil-free days	1	2	3	4	5	6
b. Release time	1	2	3	4	5	6
c. Faculty meeting time	1	2	3	4	5	6
d. Personal time (evenings and weekends)	1	2	3	4	5	6
e. Holidays (including summer vacation)	1	2	3	4	5	6

Professional Community

6. **Are teachers at your school scheduled to have a common planning period during the day?** *(mark all that apply)*

a. Yes, by grade level	☐
b. Yes, by content area	☐
c. Other (specify) _____	☐
d. No, no common planning time is scheduled	☐

7. **What professional opportunities are available to teachers in your school?** *(mark all that apply)*

a. Mentoring new teachers	☐
b. Coaching other teachers	☐
c. Making management decisions	☐
d. Making decisions related to curriculum development	☐
e. Participating on grade-level teams	☐
f. Participating on content area teams across grades	☐
g. Other (specify) _____	☐

School Reform

The following questions refer to the most important reform strategy or strategies in your school. Such reform strategies are those intended to improve school operations and student outcomes.

8. To what extent do you believe reform at your school: *(Circle one number in each row.)*

	Not at all	To a small extent	To some extent	To a fairly large extent	To a great extent
a. helps to improve student learning	1	2	3	4	5
b. helps you to teach more effectively	1	2	3	4	5
c. detracts from more important efforts	1	2	3	4	5
d. improves communication among teachers	1	2	3	4	5
e. helps to improve student behavior	1	2	3	4	5

Parent Involvement

9. Thinking about the parents of the students in your classroom(s), what percentage of those parents support their children's learning in the following ways: *(write percentage in space at right)*

a. Monitor attendance from home	**%**
b. Help students with homework	**%**
c. Volunteer to assist in classrooms	**%**
d. Other (specify) _____	**%**

10. Thinking about the parents of the students in your classroom(s), what percentage of those parents are engaged with the school in the following ways: *(write percentage in space at right)*

a. Attend parent-teacher conferences	**%**
b. Demand frequent reports on their children's progress	**%**
c. Actively volunteer in the school	**%**
d. Observe classroom activities	**%**
e. Actively participate in a formal parent organization (PTA/PTO)	**%**
f. Other (specify) _____	**%**

Reform Strategies

Schools often use single or multiple reform strategies designed locally or by external program developers. Some strategies are externally designed, for example Success for All or Co-nect. Other strategies are local and may not have a formal name. On this page we are collecting specific information about each strategy in use at your school. Begin by writing the name of the most important reform strategy at the top of the first column and then answer each question considering that strategy only. Then, using the next column, repeat the process again for each additional strategy at your school. If your school is using a single strategy, complete the first column only. If your school has no reform strategy, write "none" in the first column.

Reform Strategy Name ➔ ➔ ➔ ➔ ➔ ➔ ➔ ➔ ➔ ➔				
	(circle one)	**(circle one)**	**(circle one)**	**(circle one)**
Description of Reform Strategy				
Does the strategy target any curricular areas for improvement?	Yes No	Yes No	Yes No	Yes No
If yes, indicate curricular areas targeted: *(Circle all that apply.)*	Language Arts Mathematics Social Studies Science Arts Other	Language Arts Mathematics Social Studies Science Arts Other	Language Arts Mathematics Social Studies Science Arts Other	Language Arts Mathematics Social Studies Science Arts Other
Does the strategy target school organization and management for improvement?	Yes No	Yes No	Yes No	Yes No
Does the strategy target community and parent involvement?	Yes No	Yes No	Yes No	Yes No
Do you have evidence that this strategy improves student achievement?	Yes No	Yes No	Yes No	Yes No
Does the evidence include comparison groups?	Yes No	Yes No	Yes No	Yes No
Has the strategy been demonstrated to improve student achievement at other schools?	Yes No	Yes No	Yes No	Yes No
Status of Reform Strategy				
Are all grades included in the strategy?	Yes No	Yes No	Yes No	Yes No
Is this strategy being phased-in?	Yes No	Yes No	Yes No	Yes No
Is the strategy accompanied by benchmarks for implementation?	Yes No	Yes No	Yes No	Yes No
Did your district mandate this strategy?	Yes No	Yes No	Yes No	Yes No
Will sanctions result from failure to implement this strategy?	Yes No	Yes No	Yes No	Yes No
Support Provided through Reform Strategy				
Was this strategy developed externally?	Yes No	Yes No	Yes No	Yes No
Does the strategy include specific curricular materials?	Yes No	Yes No	Yes No	Yes No
Do guidelines for classroom practice accompany the strategy?	Yes No	Yes No	Yes No	Yes No
How many hours of training on the strategy did you receive from the program developer?	____ hours	____ hours	____ hours	____ hours
How many hours of training on the strategy did you receive from the district?	____ hours	____ hours	____ hours	____ hours
How many hours of training on the strategy did you receive from the state?	____ hours	____ hours	____ hours	____ hours
How many of these total training hours (from the developer, district and state) occurred at your school?	____ total hrs.	____ total hrs.	____ total hrs.	____ total hrs.
Is ongoing support for the strategy available through an onsite facilitator?	Yes No	Yes No	Yes No	Yes No

Background

11. How many years have you been a teacher? _____ years

12. How many years have you been a teacher at this school? _____ years

13. What type of credential(s) do you hold? *(mark one response)*

a. Professional	☐
b. Preliminary	☐
c. Emergency	☐
d. Substitute	☐
e. Other _____	☐

14. Which best describes your MAIN teaching assignment? *(mark one response)*

a. Self-contained classroom teacher	☐
b. Specialist teacher *(mark below your primary subject area assignment for this year).*	☐

☐ English as a second language	☐ Science
☐ Fine Arts	☐ Special Education
☐ Language Arts	☐ Social Studies, History, Government
☐ Mathematics	☐ Speech, Communication
☐ Physical Education	☐ Writing Specialist
☐ Reading Specialist	☐ Other _____

THANK YOU FOR PARTICIPATING!

Please return your completed questionnaire in the enclosed envelope. If you do not have the return envelope, you can mail the survey to the address below, or contact the study team via email or using our toll-free number provided below.

Longitudinal Assessment of Comprehensive School Reform
Attn: Susan Cragle
55 Hanover Lane
Chico, CA 95973
1-866-880-2773
csr@duerrevaluation.com

CHECKLIST OF MATERIALS
TO BE OBTAINED AND REVIEWED
BEFORE THE 2ND SITE VISIT

Instruments to Bring:

1. A blank "Summary Measures of 11 CSR Components" - 2 copies (5 pages);
2. The original site visit protocol, entitled Field-based Component Site Visit Inventory for LACIO (4 pages);
3. The original Adjustment to Bodilly Scale (1 page);

4a. A blank Classroom Observation Form (6 copies per visitor) (5 pages);

*4b. A new addendum to the Site Visit Protocol (1 page);

*5. A list of School-Level Loose Ends for Field Study Teams for 2004-05 (1 page); and

*6. Data Collection Topics for District and State Officials for Field Study Teams, 2004-05 (1 page)

First Round Data Collection To be Reviewed:

8. DO NOT REVIEW THE EARLIER (2003-2004) "Summary Measures of 11 Components";
9. Long version of the case study report with Robert's hand-written comments, if any;
10. Short version of the case study report for each school (1 page each);
11. The filled out instrument, if any, from the Telephone Inventory with the specific district and state officials associated with your site.

Other:

12. Honorarium Form- WestEd - 2 copies (1 page); and
13. Background information about the 11 CSR components (9 pages); and
14. Review any materials received from the schools prior to the upcoming visit.

*Items to be discussed in phone prep (group) meeting prior to the site visit

Guidance for Completing the LACIO[1] Summary Measures of 11 Components

Please complete the attached *Summary of Measure of 11 Components Form* during the 2003 field-based site visits. This form should be completed for both the intervention and the comparison sites.

- **Many items on this form require teams to make judgment calls.** Keep in mind that this form is intended to quantify, as best possible, the level of implementation and reform at 30 different schools, using widely varying reform methods, and implementing these methods in widely varying ways. Use your best judgment and feel free to write clarifying comments as necessary. A good technique is for each team member to complete the form separately, share the results, and eventually arrive at a consensus.

- **It also should be noted that several items ask for judgments on activities or events that may be occurring over time.** For example, Item 8.1 asks for the level of developer support over time, and 11.2 asks for student achievement increases since CSR funding. Site visit teams should respond to these measures for the current period (i.e., is developer support "high" at the time of the first visit).

- **This is a summary measure of "Comprehensive School Reform," and is not limited to the research-based method (components 1 and 11).** When completing this form, site visitors should respond to components 2 through 10 with this broader vision in mind. For example, if the school participates in professional development not related to the adopted research-based method, that professional development should still be considered when completing component 3. The team should also make notations indicating whether this additional professional development is aligned to the comprehensive reform design or plan.

[1]Site visit team members will be notified if it becomes necessary to amend these instructions.

Summary Measures of 11 Components

Component	Measure	Score*
1. Research-Based Method or Strategy		
1.1 Implementation score (adjusted Bodily Scale):	5 4 3 2 1	1-5
1.2 Percentage of classrooms implementing (that should have been using the method):	_____%	0.0- 1.0
1.3 Fidelity rating by developer or consultant (high, medium, low, defined as follows):		
high: developer/consultant considers school to be among the best seen	high	3
medium: developer/consultant considers school to be using method in acceptable manner	medium	2
low: developer/consultant has major complaints about school's use of method	low	1
Total Possible Score for Component 1		**9**
2. Comprehensive Design		
2.1 Existence of written design or plan (name it and give its date):	yes no	1
Name:	_____	
Date:	_____	
2.2 Contents of plan (yes/no to each item):		
2.2.1 Inclusion of needs assessment or other performance *data*	yes no	1
2.2.2 Reference to specific financial resources	yes no	1
2.2.3 Indication of strategic use of financial resources	yes no	1
2.2.4 Statement of quantitative performance goals	yes no	1
2.2.5 Discussion of specific curricula	yes no	1
2.2.6 Discussion of assessment tools/data-driven instruction	yes no	1
2.2.7 Discussion of a professional development plan/strategy	yes no	1
2.2.8 Suggestions for organizational or structural changes within the school	yes no	1
2.2.9 Discussion of reform method sustainability.	yes no	1
2.3 Breadth of plan in covering all school operations (including, implicitly, all other CSR components) (high, medium, low, defined as follow):		
high: covers all CSR components (whether implicitly or explicitly)	high	3
medium: covers four or five components, but not all	medium	2
low: covers one to three components only (also name them)	low	1
Total Possible Score for Component 2		**13**

* yes=1 and no=0

(Continued)

Component	Measure	Score*
3. Professional Development		
3.1 Strong content focus:	yes no	1
3.2 Use of non-traditional teaching strategies:	yes no	1
3.3 Range of PD hours required or taken by average teacher per year:	54+ 36-53 8-35	54+ =3 36-53 =2 8-35 =1
3.4 Aligned with the student needs/data-driven instruction:	yes no	1
3.5 Evidence of collective participation of groups of teachers within the same school:	yes no	1
3.6 Evidence of some PD taking place in the teacher's classroom:	yes no	1
3.7 Explicit guidance to align PD with standards, curriculum, or assessment tools:	yes no	1
Total Possible Score for Component 3		**9**
4. Measurable Goals and Benchmarks		
4.1 Number of academic subjects covered: (Math, science, social science, English, language arts, social studies, vocational education, etc.)	No.: _____	4+ = 3 2-3 =2 0-1 =1
4.2 Number of grades covered and total no. of grades in the school:	No.: ___ No.:___	0.0- 1.0 (%)
Total Possible Score for Component 4		**4**
5. Support within the School		
5.1 Existence of formal faculty votes on reform or research-based method:	yes no	1
5.2 Formal faculty vote(s) on reform or research based method shows 75% support:	yes no	1
5.3 Interviewees/new teachers voice strong support or enthusiasm:	yes no	1
5.4 Fewer than two interviewees/teachers voice dissent or indicate lack of use:	yes no	1
5.5 Faculty maintains or increases support for method over time:	yes no	1
Total Possible Score for Component 5		**5**

* yes=1 and no=0

(Continued)

Component	Measure	Score*
6. Support for Teachers, Administrators, and Staff		
6.1 District supports model implementation:	yes no	1
6.2 District provides resources to support research-based method:	yes no	1
Total Possible Score for Component 6		**2**
7. Parent and Community Involvement		
7.1 Emergence of new forms of parent involvement during CSR years:	yes no	
7.1.1 Special parent events	yes no	3 - 4 =1
7.1.2 Programs or opportunities for parents in instructional roles	yes no	0 - 2 =0
7.1.3 Parent advisory or other committees	yes no	
7.2 Level of parental involvement (high, medium, or low, defined as follows):		
high: you've observed parents in the school and interviewees voice strong or satisfactory level of parental involvement in school activities	high	high=2
medium: school get traditional level of parental involvement (e.g., 10% attendance)	medium	medium=1
low: no evidence of parental involvement beyond a handful of parents and interviewees voice low levels of participation	low	low=0
7.3 Evidence of at least one community organization and one school/ community event or program:	yes no	1
Total Possible Score for Component 7		**4**
8. External Technical Support and Assistance		
8.1 Developer support and assistance (high, medium, or low, defined as follows):		
high: all 3 CSR years plus during year after CSR	high	high=3
medium: at least two of these four years	medium	medium=2
low: one or none of these four years	low	low=1
8.2 Other external (but non-district) support and assistance:		
yes: evidence for a specific source and function on two or more occasions	yes no	1
no: no such evidence (evidence can be documentation, interviewee mentions, or direct observation)		
Total Possible Score for Component 8		**4**

* yes=1 and no=0

(Continued)

Component	Measure		Score*
9. Evaluation Strategies			
9.1 Existence of a written evaluation plan:	yes	no	1
9.2 Evidence of written evaluation findings (could even be a memo):	yes	no	1
9.3 Data are being used to suggest school enhancement:	yes	no	1
9.4 Data are used annually to improve instruction:	yes	no	1
Total Possible Score for Component 9			**4**
10. Coordination of Resources			
10.1 Evidence of some coordination of funds from different external sources (e.g., federal):	yes	no	1
10.2 Evidence of some coordination of external and local funds (i.e., core building):	yes	no	1
Total Possible Score for Component 10			2
11. Evidence of Academic Improvement			
11.1 Annual student data indicates positive improvement in one or more areas:	yes	no	1
11.2 Student achievement data has increased incrementally since CSR funding was received:	yes	no	1
Total Possible Score for Component 11			2
Total Score			

* yes=1 and no=0

Field-based Component Site Visit Inventory for LACIO:
11 Provisions of Comprehensive School Reform (CSR)

A major purpose of the site visit is to observe a school's progress in implementing the 11 provisions of CSR as specified by Sec. 1606 of P.L. 107-110 (No Child Left Behind Act of 2001). Your observations should be based on direct field evidence, including observations of key documents, copies of which should be obtained for the project files. The needed observations are organized according to the 11 components and are as follows.

1. Employs Proven Strategies and Methods for Student Learning, Teaching, and School Management;
combined with
11. Has Support Based on: A) scientifically-based research, or B) strong evidence:

1.1 Identify the proven method or strategy being used by the school (observe the method in action or examine samples of students' work or other evidence showing how the method is being used).

1.2 From documentation about the method, cite the following:

1.1.1 The scientifically-based or other research evidence;
1.1.2 The evidence of positive findings relative to comparison schools;
1.1.3 The evidence of successful replication at other schools.

1.2 Obtain or create a roster of classrooms that should be using the method; from this roster, select a random sample for classroom observation. Using the Classroom Observation Instrument (see separate instructions), estimate the percentage of classrooms using the method that should have been using it: ___.

1.3 Ask the developer or consultant providing external assistance to describe the school's *fidelity* in implementing the method or strategy, and ask for the school's summary rating, based on the scale that the developer or consultant uses. Determine that the developer or consultant has had no knowledge of student outcomes in making this rating.

1.4 Score the extent of *implementation* of the method or strategy, using the adjusted Bodilly Scale, providing observational examples to support your score.

2. Integrates a Comprehensive Design:

2.1 Cite the plan, summary of faculty discussions, or other documentation, if any, that covers the full range of a school's operations, and which therefore should embrace the school's design for CSR (the documentation may be limited to CSR or go beyond it).

2.2 From the documentation, cite the following:

2.2.1 Needs assessment *data*;
2.2.2 Quantitative performance goals;
2.2.3 Discussion of specific curricula or instructional methods;
2.2.4 Identification of assessment tools;
2.2.5 Discussion of professional development;
2.2.6 Indication of how financial resources will be strategically used.

2.3 How many of the 11 CSR provisions can you find in the documentation? ___.

2.4 Observe the salience of the design at the school (e.g., does anybody talk about it or its provisions?).

3. Provides High Quality Professional Development:

3.1 Cite the *formal* requirements for PD at this school (also check the district for PD requirements).

3.2 Cite whether these requirements:

3.2.1 Limit the substantively relevant topics eligible for PD;
3.2.2 Discuss any preferred alignment between a) PD and b) specific standards, curriculum, or assessment tools;
3.2.3 Specify the required number of PD days, per year: ___. (Adjust the preceding estimate to *exclude* days for traditional teacher set-up in the fall and teacher clean-up in the spring).

OMB Clearance # 1875-0222
Expiration date: 01/31/2006

3.3 Observe or obtain evidence of *PD related to the CSR method or strategy*, and review records of the extent of this PD (total number of teacher-hours) during the past year or semester.

3.4 Observe or obtain evidence of *all other PD*, and review records of the extent of this PD (total number of teacher-hours) during the past year or semester.

3.5 Review PD records and determine the extent to which the PD:

- 3.5.1 Has been content-focused and content-based
- 3.5.2 Provides for coherent and continuous development
- 3.5.3 Is classroom-based (e.g., assistance occurring in the classrooms, not at workshops)

3.6 Collect evidence on the extent to which the school has used coaches, mentors, lead teachers, or other ongoing forms of assistance to teachers

4. Includes Measurable Goals and Benchmarks for Student Academic Achievement:

4.1 Cite the school's documentation for goals (endpoints) and benchmarks (intermediate points).

4.2 From the documentation, cite whether these goals and benchmarks cover:

- 4.2.1 Student achievement, not just intermediate or other outcomes (e.g., attendance);
- 4.2.2 Implementation variables, and not just outcome variables;
- 4.2.3 The full range of grades at the school;
- 4.2.4 All the core academic subjects at the school.

4.3 Observe the salience of the goals and benchmarks at the school (e.g., are there explicit reward systems for accomplishing the goals and benchmarks; does anybody talk about them; are there posters or wallcharts about them?).

5. Is Supported by School Staff:

5.1 Cite the dates and results of any formal faculty vote(s) on CSR or on the research-based method.

5.2 Cite whether at least 75% of interviewed staff members cite support for CSR or the research-based method; ditto dissent (at least 25%); ditto lack of use (at least 25%).

5.5 Cite any other observations illustrating support for reform or for the research-based method (e.g., are there wall posters, banners, or other readily displayed materials referring to CSR?).

6. Provides Support for School Staff:

6.1 Cite ways in which the school supports the staff in the reform effort:

- 6.1.1 Re-arrangement of school or teachers' schedules
- 6.1.2 Funds to teachers for supplemental classroom materials
- 6.1.3 Salary or other formal credit for professional development specifically related to reform activities
- 6.1.4 Salary for attending summer or weekend professional development
- 6.1.5 Establishment of new positions (e.g., reform or parent coordinator) related to reform
- 6.1.6 Other

7. Provides for Meaningful Parent and Community Involvement (and is consistent with Sec. 1118 of P.L. 107-110):

7.1 Cite any *written* parent involvement policy.

7.2 Observe or obtain evidence of parents' role in their children's learning (e.g., do you see parents in classrooms or otherwise assisting at the school? Is there any indication that parental support for children's homework has changed?).

OMB Clearance # 1875-0222
Expiration date: 01/31/2006

7.3 Observe or obtain evidence about parent-teacher communication (e.g., the frequency of parent-teacher conferences reflected by sign-in logs).

7.4 Observe or obtain evidence about parent involvement in planning, implementing, or evaluating reform

7.5 Corroborate items 7.2, 7.3, and 7.4 with events referenced in discussions with parents and teachers, also indicating how these parents and teachers were selected.

8. Uses High Quality External Technical Support and Assistance:

8.1 Identify the source(s) of external technical support and assistance, and review records indicating the frequency of such assistance over the past year or semester.

8.2 Ascertain the principal's satisfaction with the *quality* of the external support and assistance, as well as the reasons for this level of satisfaction.

8.3 Observe a technical assistance event (e.g., workgroup or in-class session), reviewing any materials (e.g., agendas, exercises, hands-on materials) actually used in the event. Comment on teachers' reactions and responses during the event:

8.3.1 Teachers' understanding of the substantive topics;

8.3.2 Teachers' motivation and support in participating in the event.

8.4. Determine the annual amount of funds, if any, paid for the external assistance: __

8.5 Obtain documentation about the expertise of the external provider, in relation to experience with comprehensive school reform or research-based methods

9. Includes Plan for Annual Evaluation of Implementation and Student Results:

9.1 Cite the written evaluation plan.

9.2 Review the plan for its coverage of implementation and student outcomes.

9.3 Cite any observations or evidence showing that evaluation findings have been used.

10. Identifies Other Resources to Coordinate and Sustain Reform:

10.1 Indicate the resources (cite specific amounts and fiscal years), other than CSR, being used to coordinate or sustain reform, as cited in any budget, comprehensive plan, or other similar document.

10.2 In what way and to what extent does the documentation indicate that coordination is occurring (e.g., a single plan, but different sources identified to support different subtasks)?

List of School-Level *Process* Influences that Help a School to Implement CSR

The list below deliberately ignores endemic conditions such as whether budgets are being cutback, staff turnover, student turnover, and other conditions that can impede change. The reason for ignoring these conditions is that the research question should be whether reform-oriented activities are better than non-reform-oriented activities, given the same endemic conditions.

The list also ignores state or district policies because of some of the same reasoning and because they will be covered elsewhere in the research.

A. An ongoing performance *gap*, either real or perceived:

- school wants to be the best it can be (e.g., embraces an *all students can learn* vision)
- school has not scored well in accountability systems
- other external pressure to perform (e.g., busing or effects of desegregation court order)

OMB Clearance # 1875-0222
Expiration date: 01/31/2006

Field Question: *Identify and cite the actual data demonstrating the existence or size of any performance gap, including the period of years covered.*

B. A principal with instructional and administrative expertise:

- can serve as instructional leader
- can create needed flexibility, adaptiveness, and resources to support change
- can motivate staff

Field Question: *Identify and document the three most important actions taken by the principal to support comprehensive school reform.*

C. A match between the research-based method and the school's needs (e.g., Stringfield and Datnow):

- availability of some type of needs assessment (can be informal)
- an informed process whereby alternative methods were considered, and the school understood well the grounds for making its final choice of methods

Field Question: *Document the process whereby such a match may or may not have taken place.*

D. Sufficient slack resources to support change (CSR award helps a lot if used properly; must then convert Title I and other funds later on, to maintain needed slack).

Field Question: *Identify three incidences whereby slack resources were needed and determine whether they were made available.*

E. Sufficient professional development days and other staff resources (e.g., teacher substitutes) to permit teachers to obtain needed training and knowledge.

Field Question: *What arrangements are made for professional development?*

F. Analytic ability to connect school operations with student outcomes (e.g., district or external assisters may help to define relevant portions of the curriculum to be covered when a student underperforms some subtest).

Field Question: Cite specific occasions when student performance is connected to needed curriculum work.

OMB Clearance # 1875-0222
Expiration date: 01/31/2006

ADJUSTMENT TO THE BODILLY SCALE

(see Appendix 4-1)

Although the Bodilly Scale is was specifically designed to describe the level of observed implementation in a CSR school, during the field-focused study of CSRD schools it was determined that the definition of scale points was limited. While anchor points 0-3 primarily focus on the early stages of implementation, the scale utilizes only two anchors (4-5) to describe the richness and variability of the implementation process. Therefore, in order to capture the variance within the implementation stages, the Bodilly Scale has been modified to include one additional anchored rating.

Original Bodilly Scale

0 — **Not Implementing**. No evidence of the element.
1 — **Planning.** The school was planning or preparing to implement.
2 — **Piloting.** The element was being partially implemented with only a small group of teachers or students involved.
3 — **Implementing.** The majority of teachers were implementing the element, and the element was more fully developed in accordance with descriptions by the team.
4 — **Fulfilling.** The element was evident across the school and was fully developed in accordance with the design teams' descriptions. Signs of institutionalization were evident.

Adjusted Bodilly Scale

0 — **Not Implementing**. No evidence of the element.
1 — **Planning.** The school was planning or preparing to implement.
2 — **Piloting.** The research-based model or practice was initially being implemented with only a small group of teachers or students involved with a plan for scale-up in future semesters or years.
3 — **Partially Implementing.** The school is having difficulty maintaining the level of teacher or administrator buy-in necessary for full implementation, or is struggling with implementation for some other reason, resulting in only portions of the research-based model being implemented, or only portions fo teachers participating.
4 — **Implementing.** The majority of teachers were implementing the majority fo aspects of a research-based model or practice.
5 — **Fulfilling.** The research-based model or practice was evident across the school and was fully developed in accordance with the design teams' descriptions. Signs of institutionalization were evident.

OMB Clearance # 1875-0222
Expiration date: 01/31/2006

Classroom Observation Form
For the Longitudinal Assessment of CSRD Implementation and Outcomes

Instructions

Please complete the attached *Classroom Observation Form* during the 2003 field-based site visits. The classroom observation form is designed to guide your field-based data collection.

- **Prior to the observation period, researchers should discuss what they expect to see in the classroom within the context of the reform method.** While the categories may not exactly match what is happening in the classroom, the observer should attempt to classify the activities as best possible, and should use the comments section to clarify.

- **There is a possibility that researchers may not see the reform-related instructional strategies or techniques during every classroom visit.** For example, although teachers may frequently engage in team teaching or modeling, observers may not have the opportunity to witness method-related techniques during the short observation period.

- **The classroom observations are just one source of evidence that may provide concrete examples or details the site visit teams can use in their reports.**

- **There is space at the bottom of page one for observers to write brief descriptions of what is being observed.** This section should include a brief description of how the observed activities might relate to the reform effort, the level of student engagement, and any other important items observed. Site visit team members also may need to ask classroom teachers to provide context for the lessons or activities that were observed.

- **In the section on the "Observed use of time," observers can mark whether they observe the item in question at any time during the class period (as opposed to during each 10-minute block).** If other major items or activities are taking place, or if clarification is necessary, the observer should provide these details in the comments section or on the back of this sheet.

- **In the section on the "Observed use of time,"** Item 5 should be a percentage, reflecting the percent of time the majority of students are academically focused during the observation period, and Item 6 should be High, Medium or Low (*High:* Two or fewer students not participating; *Medium:* Less than 10 percent of students not participating, perhaps engaged in conversation, but not disruptive to majority of class; or *Low:* Many students doing their own thing).

OMB Clearance # 1875-0222
Expiration date: 01/31/2006

Classroom Observation Form
For the Longitudinal Assessment of CSRD Implementation and Outcomes

Date: _____
Time: _____
School: _____ Teacher(s): _____
Model: _____ _____
Subject: _____
Grade level: _____ Observer: _____

Implementation level, defined by school: _____
Implementation level, defined by observer
(to be completed following observation): _____

Directions: 45-minute observation period, and 10 minute follow-up interview with teacher to check some items if necessary

Observation Items

A. Number of students in class: _____

B. Number of students by racial/ethnic group:

Black _____
White _____
Hispanic _____
Asian _____
Other _____

C. Observed use of time (Check in ten-minute blocks)

Observers should complete this section for three 10-minute blocks within the 45-minute observation time. Observers should spend a short amount of time (5 minutes) familiarizing themselves with the classroom and curriculum prior to completing this section. Each ten-minute block may have more than one checkmark denoting the presence of an instruction strategy or method. (Comments are not necessary; can be used for individual observer's notes).

	10-minute blocks			
	1	2	3	Comments
1. Instructional Orientation				
a. Direct instruction with the entire class (lecture or teacher question/student response)				
b. Team Teaching				
c. Cooperative/Collaborative Learning				
d. Individual Tutoring				
2. Classroom Organization				
a. Ability Grouping				
b. Multi-age Grouping				
c. Work centers (for individuals or groups)				
d. Individual desk arrangement				
3. Instructional Strategies				
a. Instructional feedback to enhance student learning				
b. Integration of subject areas				
c. Project-based learning				
d. Use of higher-level thinking skills				
e. Teacher acting as coach/facilitator				
4. Student Activities				
a. Independent seat work				
b. Experiential, hands-on learning				
c. Systematic individual instruction (assignments geared to individual needs)				
d. Independent inquiry/research				
e. Student discussion				
5. Academically focused class time				
6. Level of student attention				

Adapted from "School Observation Form," *How to Evaluate Comprehensive School Reform Models, University of Memphis*, developed by S. Ross for New American Schools, 2000.

D. EXAMPLES OF LINKAGES TO OTHER CSR COMPONENTS

Observers should indicate the presence of these actions or materials in the classroom at any time during the 45-minute observation period. Other observations that relate to the 11 components should be added to this form.

Component	Observed
1 and 11. Proven (scientifically-based) strategies and methods for student learning;	
a. Teacher using curriculum developed or endorsed by the developer	
b. Teacher using instructional strategies called for by the method	
c. Materials carrying message of method's concepts are displayed in the room	
d. Other: _____	
2. Integrates a comprehensive design:	
a. Copy of the school improvement plan present in the classroom	
b. Teacher observed using method's instructional strategies during all subjects	
c. Other: _____	
3. Provides high quality professional development:	
a. Copy of teacher's individual professional development plan present in the classroom	
b. Coaches or classroom observers in the classroom	
c. Other: _____	
4. Includes measurable goals and benchmarks for student academic achievement:	
a. Charts or graphs measuring student performance displayed on classroom walls	
b. Other: _____	
5. Is supported by school staff:	
a. Teacher actively using model curriculum or instructional strategies	
b. Model materials posted throughout the room	
c. Listen for positive comments about the lesson and school	
d. Other: _____	
6. Provides support for school staff:	
a. Class schedule allows teachers sufficient planning and meeting time	
b. Classrooms well equipped with instructional materials	
c. Other: _____	
7. Provides for the meaningful involvement of parents and the local community	
a. Parents or community volunteers present/assisting in the classroom	
b. Evidence of parent involvement in homework activities	
c. Other: _____	
8. Uses high quality external technical support and assistance	
a. Presence of conference or training materials in classroom	
b. Other: _____	
9. Includes plan for annual evaluation of the implementation and the student results;	
a. Other: _____	
10. Identifies other resources to coordinate and sustain reform	
a. Other: _____	

E. FOLLOW-UP INTERVIEW WITH THE TEACHER

Choose two or three questions per class to ask each teacher following the observation. Observers should have answers to all questions by the end of all observations.

1. Do you believe you receive adequate support from the school in order to implement the reform method? (Probe for examples.)

2. Do you think there is alignment between the reform method, the professional development or technical assistance you receive, and the assessments that are administered to students? (Probe for examples.)

3. Has the technical assistance you received met your needs and expectations? (Probe for examples.)

4. Are there any reasons you don't like to use the reform method, or have you experienced any problems in implementation? (Probe for examples, or reasons why other teachers might resist implementation.)

5. How are parents involved in your students' education? (Probe for classroom volunteers, homework assignments; ask for any documentation like sign-in sheets).

ADDENDUM TO SITE VISIT PROTOCOLS

1. **Add Component 11 to the Site Visit Protocol:**
"Scientifically-based research, or strong evidence that the school's reform program will significantly improve academic performance" [note difference between Components 1 and 11].

2. **Add "alignment" to Topics for District and State Officials:** Inquire specifically about the how the district or state know about the alignment among standards, assessments, and curriculum and instruction—both within the state or district's policies and between the district and state.

3. **Clarify the use of the Classroom Observation Form as follows:**
a. Randomly select classrooms from a classroom roster.
b. Prior to the classroom observations, identify five instructional characteristics that illustrate the use of the method(s). Note those characteristics observed in every classroom, an indicator of the extent of implementation.
c. If the method is content-specific, observe some classrooms that are illustrative of the method and others that are not. If the method is a comprehensive process model, then all classrooms are potentially illustrative of the model. Note on the form whether the classroom selected is method-specific or not.
d. Note any other methods being implemented in the school, besides the CSR-related method. Most important, note if the "other methods" observed in the classroom are "add-ons" or integrated into a comprehensive whole. Make notes on the white space in the Form and discuss the issue in the case study report.
e. Remember to enter the school code and teacher's name on each page of the classroom observation form, to aid tracking after the site visit. Also, note the name of the CSR method and other methods used in the classroom. Other handwritten comments are encouraged.

4. **Watch for the incidence and potential effects of "student transfers," as part of covering public school options**

SCHOOL-LEVEL LOOSE ENDS
FOR FIELD STUDY TEAMS FOR 2004–05

Topics	Protocol Item to be Referenced
1. Was the CSR grant scaled-up or was it implemented all at once at the beginning of the grant period? Variability in start-ups, Timing for CSR.	1.2
2. Where are the funds and who controls the funds? (Survey) That data could be collected during site visits in the principal and district interviews. "Where are the funds?" "Who makes decisions about how the CSR and other funds are used? [this could be addressed on the one-page principal process influence document or in principal's interview]	10.1
3. Burkett mentioned Schoolwide Title I (in Fed Register on July 2, 2004) Field staff may need to know about it before going to the field.	2.1
4. Scope of Change (Maggie)—How demanding is the intervention, whether comprehensive or subject-based? Could be included in Component 1 (Research-based Method)	1.1
5. If principal change, does principal deviate from adopted method or facilitate CSR grant? (Principal behavior, not whether a change occurred or not. Burkett)	B
6. How to define PD? (Stephanie concerned about survey) SV protocol is complete. Might add across and among grade planning and training periods and structural PD, such as ongoing small learning communities to protocol. Expand definition of professional development (PD): The inventory/protocol addresses a wide variety of types of PD–school level. Site visitors should be reminded of the variety of topics that could be considered in PD, such as the number of hours spent in vertical and horizontal teacher planning time.	3.2.3
7. Effects of Voluntary Public School Choice (VPSC): Describe student mobility in your school. (Probe: Has VPSC been a factor in student mobility?)	ADD AS RIVAL
8. Burkett, "Ask teachers if they participate in decisions." [Do in the Leadership Team meeting in the prototype field study agenda.]	5
9. How are CSR funds being spent?	D
10. Has this school been identified as low-performing? If so, has it been sanctioned because of low performance?	A
11. Did the faculty formally vote to adopt the current reform in your school?	5.1
12. What opportunities do teachers have for making management decisions?	B
13. Do teachers have a common planning period in the day?	3.5
14. Do any of the following occur in ;your school? (Probe: Looping, grouping by houses or families; student groups stay with teachers for more than one year; interdisciplinary teaching, and paired or teacher teaming)	School Operations
15. If a secondary school, what kind of student groups exist to achieve particular instructional goals?	School Operations
16. Who organizes the scope and sequence of the curriculum at this school?	B

TOPICS FOR DISTRICT AND STATE OFFICIALS FOR FIELD STUDY TEAMS, 2004–05

District CSR Contact:

1. *Administration of CSR Program as a Whole:* Define and update district involvement in CSR program (targeting and selection process, number of awardees, funding process, and goals for how whole-school improvement will be achieved).

2. *Technical Assistance and Implementation Support:* Describe state technical assistance and other support provided to districts and schools that might be related to CSR.

3. *Relationships between CSR and District Initiatives.* Describe the relationships, if any, between CSR and:

   a) district reform initiatives (e.g., standards, curriculum requirements, use of research-based methods, new report card forms)
   b) district administration of Title I and schoolwide Title I
   c) district requirements for PD

4. *Relationships between CSR and NCLB and Other Federal and State Initiatives.* Describe the relationships, if any, between CSR and other initiatives, such as *NCLB* and public school choice (federal) and state accountability and reform (state)—e.g., any relationships between CSR and state standards and assessments, and the implications for CSR implementation.

5. *Monitoring and Evaluation:* Describe whether or how the district monitors and evaluates CSR implementation and outcomes—if possible focusing specifically on the two site visit schools.

State CSR Contact:

1. *Administration of CSR Program as a Whole:* Define and update state involvement in CSR program (targeting and selection process, number of awardees, funding process, and goals for how whole-school improvement will be achieved). Describe state priorities for CSR.

2. *Technical Assistance and Implementation Support:* Describe state technical assistance and other support provided to districts and schools that might be related to CSR.

3. *Relationships between CSR and State Initiatives.* Describe the relationships, if any, between CSR and:

   a) state reform initiatives (e.g., standards, curriculum requirements, use of research-based methods)
   b) state assessments and accountability initiatives
   c) state requirements for PD, teacher certification, or related policies

4. *Relationships between CSR and NCLB and Other Federal Initiatives.* Describe the relationships, if any, between CSR and other federal initiatives, such as *NCLB* and public school choice.

5. *Monitoring and Evaluation:* Describe whether o how the state monitors and evaluates CSR implementation and outcomes—if possible focusing specifically on the two site visit schools. Obtain a copy of the state CSR evaluation report (there may be several such reports, each completed annually).

APPENDIXES REFERENCES

Coalition for Evidence-Based Policy (December 2003). *Identifying and Implementing Education Practices Supported by Rigorous Evidence: A User Friendly Guide.* Retrieved October 15, 2006, from http://www.ed.gov/rschstat/research/pubs/rigorousevid/index.html.

Comprehensive School Reform Quality Center (2006a). *CSRQ Center Report on Elementary School Comprehensive School Reform Models.* Washington D.C.: American Institutes for Research.

Comprehensive School Reform Quality Center (2006b). *CSRQ Center Report on Middle and High School Comprehensive School Reform Models.* Washington D.C.: American Institutes for Research.

Gill, B. P., Hamilton, L. S., Lockwood, J. R., Marsh, J. A., Zimmer, R. W., Hill, D., and Pribesh, S. (2005). *Inspiration, Perspiration, and Time: Operations and Achievement in Edison Schools*. Santa Monica, Calif.: RAND Corporation. Retrieved March 10, 2007, from http://www.rand.org/pubs/monographs/MG351/.

Hausman, J. A. (1978). "Specification Tests in Econometrics." *Econometrica*, *46*(6), 1251–1271.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Mass.: MIT Press.

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access.