Putting the Pieces Together

Lessons from Comprehensive School Reform Research

NCCSR

The National Clearinghouse for Comprehensive School Reform

Putting the Pieces Together

Lessons from Comprehensive School Reform Research

Christopher T. Cross
Editor

NCCSR
The National Clearinghouse for Comprehensive School Reform

The National Clearinghouse for
Comprehensive School Reform
2121 K Street, NW, Suite 250
Washington, DC 20037
Telephone 1 (877) 766-4CSR
Fax 1 (877) 308-4995
Email: **AskNCCSR@goodschools.gwu.edu**
Website: **www.goodschools.gwu.edu**

The mission of NCCSR is to collect and disseminate information that builds the capacity of schools to raise the academic achievement of all students.

NCCSR is funded by the U.S. Department of Education's Office of Elementary and Secondary Education and is operated by The George Washington University under Contract No. ED-99-CO-0137. The views expressed do not necessarily reflect the views of the U.S. Department of Education. The mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. government. Readers are free to duplicate and use these materials in keeping with accepted publication standards. NCCSR requests that proper credit be given in the event of reproduction.

Printed in the United States of America.

Table of Contents

Acknowledgements

A publication such as this requires a real team effort. The idea for doing this book came from the management team at the National Clearinghouse for Comprehensive School Reform (NCCSR), under the leadership of Arthur Gosling who has guided the Clearinghouse since its creation in 1999. That team has included David Huie, deputy director of NCCSR, Joel Gomez, principal investigator for NCCSR and Konyka Dunson and Lori Cavell, research associates at NCCSR. The final editing work was done by Greg Pearson, a research assistant at NCCSR who served as my right hand in this effort. Finally, several staff of the U.S. Department of Education, including Joseph Conaty and Margaret McNeeley, deserve special commendation for their support of this work.

Extremely important in the development of the book was the work of our editorial board. We were very fortunate in having a board that consisted of both experts in comprehensive school reform (CSR), as well as noted scholars in the field of education research. The board members were George Bohrnstedt, senior vice president of the American Institutes for Research; Susan Fuhrman, Dean of the Graduate School of Education at the University of Pennsylvania; Thomas Good, professor of educational psychology at the University of Arizona; Donna Harris, Mellon Postdoctoral Fellow at Wellesley College; and Mary Anne Schmitt-Carey, president of New American Schools. They are remarkable colleagues, each contributing significantly to the quality of this publication.

The authors whose work is published here have demonstrated their profound commitment to working in the CSR field. Each chapter reflects the thoughtful involvement of multiple authors who have collaborated with one another to present a thoroughly researched and thoughtfully analyzed paper. Our goal has been to present a comprehensive understanding of comprehensive school reform after five years of effort by thousands of teachers, principals, district leaders, and state officials, with the support and encouragement of officials in the U.S. Department of Education, as they have worked to improve learning opportunities for all children.

Finally, commendation goes to the NCCSR partner organizations that have worked together since 1999 to build the capacity of schools to raise the academic achievement of all students. Under the leadership of the Institute for Educational Policy Studies in the Graduate School of Education and Human Development at The George Washington University, NCCSR has been supported by partners the Institute for Education Leadership (IEL) and the Council for Basic Education (CBE) in bringing the best possible information to schools.

Foreword

As we as a nation have worked hard on reforming our education system so that all students will have the opportunity to succeed in school and in life, few of these efforts have attracted as much positive attention and have been sustained over so long a period of time as has the notion of comprehensive school reform.

As Brian Rowan and his colleagues note in chapter one of this book, the roots of comprehensive school reform (CSR) can be traced back several decades to the RDDU (research, development, dissemination, utilization) paradigm and to a 1975 report from the RAND Corporation, commonly known as the "change agent" study. However, what we today call CSR had its federal roots in two events in the 1990s. First was the creation in 1991 of the New American Schools Development Corporation; the second, the enactment in 1994 of the Comprehensive School Reform Demonstration program. Both of these efforts resulted in the investment of both intellectual and financial capital in substantially developing work in the field.

Our work in this volume is framed by the 11 components of comprehensive school reform that are contained in the federal law and explicated first in the Rowan, et al chapter. These 11 elements describe the characteristics that must be present to have an effective CSR program. They range from a reliance on the need for a research basis for student learning and support by school constituents to extensive staff development and clear evaluation of impact. These 11 components of CSR are the guiding principles that frame the work of the National Clearinghouse for Comprehensive School Reform (NCCSR).

The Congressional and Executive branch leaders who have paved the way for CSR also recognized that there must be a sustained effort to generate research to support CSR; research on everything from implementation to program effectiveness. Through its work, the NCCSR has worked diligently to advance the state of knowledge, to disseminate that work to the field and to gather together in one place the best information available for those in the field, especially school leaders faced with the daunting challenge of what must they do to improve the performance of all students in every school.

When this book was conceived, our goal was to make it comprehensive—to bring together in one place the very best total picture of CSR that was possible. To that end, we sought authors who would inform our readers about the history and context for the development of CSR, the challenges faced in implementing CSR programs, the evidence about the effectiveness of various models, and finally the perspectives of both a researcher and a practitioner who have been involved with this effort for well over a decade.

In the chapters that follow, the careful reader will learn a great deal about comprehensive school reform. In the chapter by Rowan, Camburn and Barnes, readers will learn of the history leading up to the present day, as well as about the important issue of CSR implementation. What is important? What factors in model

design impact implementation? How has implementation worked in practice?

In the chapter by Borman, Carter, Aladjem, and Le Floch readers will learn about what the authors term "the necessary conditions" for CSR, as well as what it takes to sustain this rather intense form of school improvement—an issue of ever-increasing importance to policymakers across the nation.

In the Borman, Hewed, Overman, and Brown chapter, the focus is on results. Here Borman and his colleagues look at the data on 29 different models, a chapter that will give any school official considering a CSR program extensive food for thought.

In chapter four, Gil and Ross present their views from the experience of a practitioner (Gil) and of a researcher (Ross) who have seen it all. Gil, as superintendent in Chula Vista, California, implemented CSR widely in that district. Ross was a close observer and research participant in the aborted effort to have every school in Memphis, TN adopt a CSR model. Their recommendations confirm much of the research that precedes this final chapter.

While each reader will take away a different set of information, often quite detailed, there are some messages that seem clear to this editor:

- There is no single best solution for school improvement. Different programs work in different places for often highly different reasons. However, without school-level support, failure is too often the outcome.
- There are many potentially positive effects of CSR on student achievement. Often, different programs work best for certain students in certain contexts. A school official seeking to adopt and implement a CSR program must study these factors carefully.
- There are many factors that tell the story of whether a CSR program will succeed. For example, having teacher support is not sufficient if the program is not supported at every level of the school and the district. Other factors, such as the availability of funding and sustained professional development, are vital elements of success.
- CSR has not emerged full-blown as a new idea. It has its roots in four decades of school improvement efforts. CSR builds on the successes and failures of the past.

Readers may be intrigued by the fact that two of our chapters have lead authors named Borman. No, it is not the same person. Dr. Kathy Borman of the University of South Florida is the mother of Dr. Geoffrey Borman of the University of Wisconsin—Madison. Both have made distinguished names for themselves in the field. Perhaps in a few years we will discover that there are indeed genetic characteristics that contribute to being an excellent educational researcher. If so, we will carefully study the Bormans.

When the No Child Left Behind Act (NCLB) became law in 2002, the education world was confronted with a set of issues and a range of accountability mechanisms like none yet created at the federal level. While many of these factors were present in the 1988 and 1994 reauthorizations of the Elementary and Secondary Education Act, NCLB has created a sense of urgency in, for example, seeing to it that students of every race are being fully educated. Surely, one of the answers that many districts are now discovering is that the approaches contained in a great many of the CSR programs are aimed at just that result. For those educators, as well as parents and policymakers, we trust that this book will be of significant value. Our common objective is to improve the future for all of our children by improving the education that they receive today. To that end, we dedicate this volume.

Christopher T. Cross
Danville, California
August, 2004

CHAPTER ONE

Benefiting from Comprehensive School Reform: *A Review of Research on CSR Implementation*

Brian Rowan
Carol Barnes
The University of Michigan
Consortium for Policy Research in Education

Eric Camburn
University of Wisconsin-Madison

This chapter examines what happens when schools engage in a process of comprehensive school reform (CSR). Although this process often begins with a decision by schools to adopt a research-based "model" or "design" for school improvement, decades of research on planned educational change suggest that simply adopting a model or design, in itself, will not guarantee successful utilization of that model inside schools. Instead, successful school improvement results from a confluence of circumstances that must and can be orchestrated by external change agents (like CSR model providers), district and school leaders, and teachers and students working in cooperation with one another to implement a process of whole-school reform. The purpose of this chapter is to give the reader a sense of the strategies used by schools that have successfully engaged in this process.

To address the question of how comprehensive school improvement works in schools, we review previous research on CSR program implementation and present some original analyses of data from our own research. We begin by discussing the emergence of CSR as a policy instrument supporting improvement in American schools. This discussion shows that current initiatives aimed at getting schools to adopt "research-based" practices for school improvement have a long history in American education, a history from which educational researchers, policymakers, and school personnel have learned a great deal about how to make planned educational change more successful. To illustrate what has been learned and how it applies to current efforts to promote CSR in schools, we next review the extensive literature on planned educational change in education, paying special attention to recent studies of CSR implementation. Our purpose here is to describe the many factors that impinge on adoption and implementation of externally-developed, research-based models of comprehensive school reform, and to lay out some of the factors that promote successful use of such models in practice. Finally, in order to illustrate how CSR works in practice, we present some early findings from a study of CSR model implementation that we are conducting under the auspices of the Consortium for Policy Research in Education. Our chapter concludes by formulating some general lessons that can be drawn from this work.

In discussing these issues, we will not sugarcoat the process of comprehensive school reform. As this chapter demonstrates, efforts at comprehensive school reform are time-consuming and difficult, and they proceed with uneven success across schools. However, our chapter shows that successful school change is possible and depends to a considerable extent on the actions taken by: (a) external providers of design-based, technical assistance; (b) local school personnel; and (c) district personnel who provide support to local school change efforts. In particular, our chapter suggests that the process of CSR will be most successful when external change agents work to produce clear, specific, and high-quality designs for change and provide extensive implementation support to local schools; when local school communities coalesce around the central aims of the research-based model of school reform they are trying to implement and actively learn over a period of years

how to utilize that model in their own context; and when district personnel provide a stable and supportive policy environment clearly aligned with the aims of the practices being developed.

As it turns out, these principles of successful change are not new. In fact, the role of external change agents (like CSR model providers) in stimulating and supporting planned educational change in schools, the centrality of local school personnel to the successful implementation of research-based practices, and the importance of district leadership in promoting successful institutionalization of instructional improvement efforts, have been remarked upon and studied for decades in the voluminous literature on planned educational change in the United States. What is new, however, is that educational researchers, policy makers, and successful education practitioners have begun to arrive at a more complex understanding of the specific steps that must be taken to assure successful implementation of whole-school change efforts. As a result, there is now a new generation of thinking about how to stimulate and support programs of comprehensive change in schools.

To describe this evolution in thinking, our chapter begins with a review of previous efforts to support planned educational change in American education, showing how two important ideas—the belief in utilization of research-based practices as a key to school reform and the gradual evolution of the federal government's Title I program toward a focus on schoolwide reform—served as catalysts for passage of the Comprehensive School Reform Demonstration Act in 1997, which emphasized the utilization of research-based, "whole-school" designs for promoting instructional improvement in American schools.

Catalysts in the Movement Toward Comprehensive School Reform

Since World War II, efforts to improve instruction and student achievement in American schools have largely revolved around two major ideas (for a review, see Firestone and Corbett, 1988: 322). One has been the goal of moving research-based models of practice out of experimental and demonstration contexts into local schools. Over time, many strategies have been developed to achieve this purpose, all of which involve attempts to integrate educational research, development, dissemination, and utilization (for a review, see Keeves, 1990). For convenience, we label these strategies as part of a larger "RDDU paradigm" in American education (i.e., the ***r***esearch, ***d***evelopment, ***d***issemination, ***u***tilization paradigm).

In its earliest form, the RDDU paradigm viewed planned educational change as a linear sequence of steps beginning with research on a practical problem, moving through the development of new research-based education practices in experimental and demonstration sites, and culminating in the dissemination and utilization of research-based practices in local schools. Since that time, however, many different approaches to moving research into practice have motivated successive waves

of federal education policy, beginning with the ambitious curriculum reforms of the 1960s, continuing through the development of various federal programs during the 1970s and beyond, and culminating in today's emphasis on the adoption and utilization of "research-based" practices in the No Child Left Behind Act of 2001. As we discuss below, the problem of moving research-based practices into schools has proven far more complex than early proponents of the RDDU model anticipated, and as a result, an enormous research literature on the dynamics of planned educational change has developed that is of direct relevance to the problem of comprehensive school reform.

A second idea shaping the contemporary emphasis on comprehensive school reform arises out of the federal government's evolving strategies for using the Title I program as a lever for change in America's high-poverty schools. In its initial stages, Title I simply provided local school systems with additional funds to operate programs for disadvantaged students, under the optimistic assumption that, given additional resources, local schools would not only target funds toward the disadvantaged, but also succeed in improving educational outcomes for such students. However, these early assumptions proved as nettlesome as those guiding the early RDDU model. For example, in the earliest days of the Title I program, some schools used Title I resources as a kind of general aid program, leading to increased regulatory guidance from the federal government about how to "target" Title I services toward specific groups of pupils within schools. But this led to the tendency for schools to use pullout and other supplemental programs that had little broad impact on school operations and did little to promote systemic change within schools. Therefore, over the years, the Title I program has begun to move away from an emphasis on targeted assistance and toward an emphasis on schoolwide change, especially in schools serving high proportions of economically disadvantaged students.

Today's movement toward "comprehensive" school reform has its roots in these two seminal, post-war ideas about how to stimulate instructional improvement in American schools. As a result, the CSR movement can benefit from decades of research on planned educational change and from decades of research on, and shifts in, the federal government's Title I program. To illustrate how these two seminal ideas about educational improvement emerged over several decades, we turn now to a brief history of the trends just discussed and to a discussion of how these trends contributed to current thinking about comprehensive school reform.

RDDU as a Strategy for Educational Change

Decades of experience with the RDDU model in education suggest that research-based innovations are not easily implemented inside schools. This problem was first recognized in evaluative studies of the ambitious curricular reforms undertaken in the 1960's. At that time, a number of pioneering studies of curriculum innovation came to the conclusion (unsurprising today, but stunning at the time) that

teachers' use of new, research-based curriculum materials was extremely varied. Darling-Hammond and Snyder (1992: 63), for example, cite a finding from one such study, which examined implementation of the innovative BSCS biology curriculum in schools. As they note, the study "revealed that teachers teaching the same lesson from the same [BSCS] course versions to classes of similar ability levels taught so differently that 'there really is no such thing as a BSCS curriculum…in the schools.'"

Around the same time, education researchers and policy makers started conducting large-scale, social experiments as a way of testing the effects of innovative educational programs on student outcomes (Cross, 2004: 46). In the early 1970's, for example, the federal government sponsored a set of ambitious experiments examining the effects of planned variations in the federal government's Head Start and Follow Through programs—two programs aimed at improving the educational experiences of pre-school and early elementary-aged children. Like the evaluative studies of curricular implementation, these massive studies came to the conclusion that planned educational programs were implemented quite variably in local settings, so much so, in fact, that many researchers began to doubt that faithful implementation of research-based practices would ever occur inside schools (see, for example, the essays in Rivlin & Timpane, 1975).

Perhaps the most significant study reaching this conclusion, however, was the RAND study of Federal Policies Supporting Educational Change, also known as the RAND "change agent" study (Berman & McLaughlin, 1975). This study, which looked at the implementation of several different federal education programs designed to spur educational innovation (including Title III of ESEA, the Right-to-Read program, Vocational Education, Part D, and Title VII, Bilingual Education) came to the conclusion that none of these programs was being implemented faithfully in local schools. So prevalent was this lack of faithful implementation, in fact, that the RAND researchers abandoned the notion of "high fidelity" implementation altogether in their report and instead discussed three outcomes of implementation efforts: non-implementation, co-optation (where an innovation is so completely adapted to the local context that it loses all distinctiveness), and mutual adaptation (where an innovation is adapted to the local setting but where the local setting also is adapted to the innovation). In this study, non-implementation and co-optation were found to be far more prevalent than mutual adaptation, with the result that many observers viewed the RAND study as signaling the near impossibility of faithfully implementing externally-designed innovations. Firestone and Corbett (1988: 324), for example, note that as a result of the RAND change agent study, "the proposition that centrally developed innovations [c]ould…[never]…be implemented locally became widely accepted and publicized by academics, policymakers, and even the popular press…and federal funding for such efforts dropped dramatically…In its place emerged a 'let a thousand flowers bloom' theory [of educational change] that stressed local invention…and the development of a built-in 'capacity' to improve at the school or district site."

As it turns out, however, the RAND change agent study was but one of many studies of planned educational change conducted in the 1970s and 1980s, and more importantly, it produced a unique and especially pessimistic set of findings. As Firestone and Corbett (1988: 324) note, the federal programs studied by RAND researchers were grants programs that provided relatively small amounts of funding with only very broad guidelines for use in local schools and districts, and local educators were expected to find their own innovations and secure technical assistance for program implementation on their own. Other studies of planned educational change conducted around the same time showed that more ambitious and clearly specified programs, offering much more intensive support for implementation to local educators, in fact achieved much higher levels of faithful implementation than did the programs studied by RAND (see, for example, Crandall & Loucks, 1983; Emrick et al., 1977; and Louis et al., 1981).

Moreover, buried within the massive, five volume report of the RAND change agent study were a number of insights about factors promoting more successful and lasting implementation of planned educational change efforts—insights that have been confirmed repeatedly in successive waves of research on educational change. For example, the RAND researchers, like many after them, found that in addition to the characteristics of the local setting, the implementation strategy used by external change agents and the scope of the innovative project affected implementation outcomes. In particular, where teachers received more training, had frequent and regular meetings associated with such training, were asked to make more (rather than less) extensive changes in practice, were provided with well-developed materials ready for use, and were given opportunities to participate in day-to-day decisions about program implementation, implementation outcomes were more favorable. Moreover, innovative projects were more likely to be continued in sites where project goals and district goals were aligned, where superintendents had more longevity, and where principals were more supportive of the innovation (Berman & Pauly, 1975). These findings are perfectly in line with the results of other studies of planned educational change conducted around the same time, studies that Firestone and Corbett (1988: 324) interpreted as showing that "centrally supported assistance strategies combining a mix of quality products and effective assistance in a manner responsive to local concerns [can] promote local change."

Evolution of the Title I Program

This brief review suggests that research on planned educational change provides one line of evidence about what is required to successfully implement research-based practices in schools. But another set of developments in American education suggests that these practices should be used as part of a comprehensive process of school improvement that is schoolwide in scope, ambitious in aims, and attentive to the alignment of many different aspects of a school's educational program.

This idea has been especially prominent in recent efforts to use the Title I pro-

gram as a lever for change in America's high-poverty schools. As originally conceived, the Title I program was intended simply as an additional revenue stream for schools serving high percentages of economically disadvantaged students. Cross (2004: 29-30) quotes Samuel Halperin, a federal education official, as saying, "In 1965, everyone had a naïve view of education. We felt…all you needed to do was give [educators] some tools and some dollars and good things would happen. They didn't need a lot of specifics…it was assumed the right thing would happen." Later evaluations of the Title I program, however, proved this assumption incorrect. In some cases, Title I monies were misused in local districts, and over time, after successive evaluations of the Title I program showed negligible program effects on student achievement, dissatisfaction with the program became endemic (Cross, 2004: 30).

While there have been many changes to the Title I program since its inception in 1965—including major provisions emphasizing high academic standards and accountability for student performance—the direction of change most relevant to this chapter is the movement away from Title I as a model of targeted assistance toward its use to stimulate schoolwide programs of instructional improvement. Ironically, the current emphasis on schoolwide change resulted from early interest in preventing the use of Title I funds as general aid to local school systems, which led the federal government to develop various guidelines and fiscal accounting practices encouraging school systems to target Title I instructional services through use of pullout and targeted in-class service delivery models. Over time, however, in-depth evaluations of the Title I program, and much writing about the delivery of compensatory education services to students, found shortcomings in this model of targeted service delivery (for a review, see Rowan & Guthrie, 1989). The problem, it appears, was not so much the use of pullout and in-class arrangements, as it was the lack of coordination that often resulted in schools when a targeted service delivery model was in place. Indeed, at least one study found that student achievement was higher in Title I schools that had higher levels of schoolwide coordination of curriculum, instruction, and remedial services.

Concerns about lack of coordination inside Title I schools, plus a growing literature on the characteristics of effective schools more generally, led to an important change in the Title I program with passage of the Hawkins-Stafford Amendments to Chapter 1 (now Title I) in 1988. This change made it much easier for schools to operate "schoolwide" models of Title I service delivery. In essence, schoolwide models allowed schools to blend Title I and other funds together in order to better coordinate compensatory and regular education instructional programs and to provide Title I services to all students in a school. In 1988, the threshold for operating such a program was set at schools with 75% of students qualifying for Title I assistance, but in 1994 this threshold was changed to 50%. By 1996, roughly 50% of all schools eligible to operate schoolwide programs under Title I were doing so.

The theory of action underlying Title I schoolwide programs was that schools using this model would develop more comprehensive (and less fragmentary) strategies for instructional improvement (Wong & Meyer, 1998). Building on research on

effective schools, it was assumed that schoolwide programs would stimulate an integrated set of changes to many different aspects of a school's educational program (Desimone, 2002). In one sense, the strategy was to "let a thousand flowers bloom" by letting each school work on its own to achieve these aims. Subsequent research, however, showed that the desired process of school reform was occurring in only some of the schools pursuing the Title I schoolwide option. Moreover, research on the effects of schoolwide programs on student achievement was inconclusive, with improvements in achievement occurring in some jurisdictions, but not others (Wong & Meyer, 1998).

Despite these uneven results, practicing educators, policy makers, and researchers were enthusiastic about the move toward schoolwide programs. To be sure, the results of the program were uneven. But practicing educators believed schoolwide programs allowed more flexible use of Title I funds and produced a better fit of Title I-funded activities with the rest of the school's educational program (Wong & Meyer, 1998). Meanwhile, some prominent researchers were arguing that the use of Title I funds on a schoolwide basis would give high-poverty schools more latitude to fund an array of proven, replicable, schoolwide improvement programs (Slavin, 1999). Finally, policy makers were proud that 90% of the schools initially operating schoolwide programs had attained the achievement gains required to maintain their status in the program (Wong & Meyer, 1998).

The Emergence of Comprehensive School Reform

Our discussion suggests that current ideas about comprehensive school reform have their roots in two main ideas in American education. One is the continuing quest to move research-based practices into schools; the second is the idea that this will occur best when undertaken in conjunction with a schoolwide process of change that integrates new practices with many other aspects of a school's educational program. In this section, we show how these two ideas came together with particular force in the 1990's as a result of two further developments—the founding in 1991 of the New American Schools Development Corporation, and the authorization by Congress in 1997 of the Comprehensive School Reform Demonstration program.

New American Schools

Arguably one of the most important developments in American education during the past decade was the founding of the New American Schools Development Corporation (NASDC) as a private, non-profit corporation in 1991. Spawned as part of President George H.W. Bush's *America 2000* initiative, NASDC (now New American Schools, hereafter NAS) developed under the leadership of David Kearns, Chairman Emeritus of the Xerox Corporation and former Deputy Secretary of

Education. This was an extraordinary private-public partnership that raised more than $130 million in contributions from the nation's top businesses and foundations to foster development of "a new generation of American schools."

As Berends, Bodilly, and Kirby (2002: xv) note, NAS founders believed that past reform efforts in schools were too often "programmatic" in orientation, that is, focused on particular subgroups of students, isolated subjects, or selected grade levels in a school. Like many before them, the NAS founders decried this fragmented approach to school improvement and wanted to invent more comprehensive approaches to school improvement. In the words of Berends et al. (2002: xv, emphasis in original), " NAS's core premise was that all high-quality schools possess, de facto, a unifying design that…integrates research-based practices into a coherent and mutually-reinforcing set of effective approaches to teaching and learning for the entire school."

To promote the proliferation and adoption of such designs, NAS decided in 1991 to fund the development of several new, "break the mold" designs for whole-school reform through a grants competition. After selecting 11 design teams from a competitive request for proposals that was responded to by over 600 applicants, NAS then began with a development phase of one year, during which time design teams created their whole-school designs; a demonstration phase of two years, during which time designs were implemented in a small number of demonstration sites; and a scale-up phase of five years in which the designs were implemented in a wider set of school districts that were chosen by NAS to cooperate in this phase of work.

Launched with a great enthusiasm, the NAS initiative met with the same uneven results as most previous efforts to reform American schools. Hatch (2000) provides a particularly interesting assessment of the initiative, acknowledging both its achievements and its disappointments. One interesting development, consistent with the process of venture capitalism after which NAS modeled its activities, was that 4 of the original 11 design teams failed to survive the arduous process of moving out of the development phase and into the scale-up phase. A second development was that only half the schools participating in the scale-up phase ended up implementing the designs at the rate anticipated by the design teams, confirming that even thoughtful attention to the problem of implementation provides only an uncertain guarantee of success (for more detail on these findings, see the report by RAND researchers Berends, Bodilly, & Kirby, 2002).

Despite the uneven successes, the creation of NAS and the bringing to scale of seven new designs had a major impact on American education. For one, this major effort spawned a whole new "theory of action" in the area of school improvement—due in no small part to the exemplary efforts of RAND researchers, who both articulated the NAS theory of action, and evaluated its results (for a review of the RAND team's work, see Berends, Bodilly, & Kirby, 2002). For example, the RAND team's work spawned a whole new language for thinking about school reform. It coined the idea of comprehensive "designs" or "models" for school

improvement; it put the term "design teams" into wider use to describe organizations that both create school improvement designs and provide technical assistance to help schools implement these designs; it further articulated the logical underpinnings of "comprehensive" or "whole-school" improvement; and it provided a substantial body of research showing the conditions under which such a theory of action seemed to work in practice. Moreover, all of this effort—by NAS, its design teams, the jurisdictions involved in the scale-up phase, and the RAND research team—led to a further expansion of CSR. By 1997, NAS-sponsored design teams were working with 685 separate schools, and their successes had led to the creation of even more "design teams" outside the umbrella of NAS, and to the adoption of CSR designs by even more schools across the country (Hatch, 2000).

The Comprehensive School Reform Demonstration Act

All of this was aided by a second important development—the 1997 passage by Congress of the Comprehensive School Reform Demonstration Act. Initially, this legislation authorized the expenditure of up to $145 million, providing individual schools up to $50,000 to implement comprehensive designs for school improvement (Hatch, 2000). Funding for the program was subsequently expanded, and with the passage of No Child Left Behind in 2001, the term "demonstration" was dropped from the act. At that point, authorized expenditures were raised to $310 million. Today, a database maintained by the Southwest Educational Development Laboratory shows that 5160 schools have received funding under the CSR program and that over 800 different "models" or "designs" have been implemented.

The emergence of a federal program did still more to articulate the logic of comprehensive school reform and hasten its adoption by schools across the country. In particular, the current law requires schools receiving CSR funds to demonstrate that they have addressed 11 attributes of a "comprehensive" school improvement program. For convenience, these attributes can be described as follows. First, schools using CSR funds must provide for meaningful parent and community involvement in the process of planning for, implementing, and evaluating CSR, and they must identify resources (in addition to the CSR program) that will be used to support and sustain their efforts. Second, schools using CSR funds must employ proven educational methods derived from scientifically-based research, methods that have a proven record (or strong evidence) of improving student achievement. Third, schools using CSR funds must make every effort to integrate these research-based methods into a comprehensive design with aligned components that has the support of teachers, administrators, and staff at the school and that includes measurable goals and benchmarks for student achievement. Fourth, schools using CSR must seek out and use an external partner with experience and expertise in schoolwide reform and work with that partner to get high-quality technical support and assistance for ongoing, high-quality professional development for teachers and staff. Finally, schools using CSR funds must make plans for evaluating progress toward implementation and student results, annually.

The stimulus this CSR legislation provided for comprehensive school reform is nothing short of remarkable. Today, we estimate that somewhere between 10% and 20% of all elementary schools in the United States have adopted an external model of CSR or are working with their own locally developed model. And this has stimulated the growth of technical assistance for CSR, as well as a growing research and practice literature on it. Today, many different agencies are working to provide schools with information and guidance about how to engage in CSR, including not only a growing list of "model providers," but also NAS, the National Clearinghouse for Comprehensive School Reform (NCCSR), all of the regional educational laboratories, the AFT, and a variety of non-profit research and development agencies such as the American Institutes for Research and RAND.

These organizations have published many guidebooks describing the main features of various CSR models, promulgating research evidence on the effectiveness of these models, and/or providing advice about how to select a CSR model. There is also a growing research literature describing how CSR unfolds in practice and charting the effects of various CSR efforts on student achievement (reviewed below). Most of this literature, it should be noted, focuses on the largest and most widely-disseminated CSR models (like the Accelerated Schools Program, America's Choice, the Coalition of Essential Schools, Co-Nect, Core Knowledge, Direction Instruction, Expeditionary Learning/Outward Bound, Modern Red Schoolhouse, the School Development Program, Success for All, and others). But it is important to remember that many, many other CSR models exist, some locally developed.

A Normative Model of the CSR Process

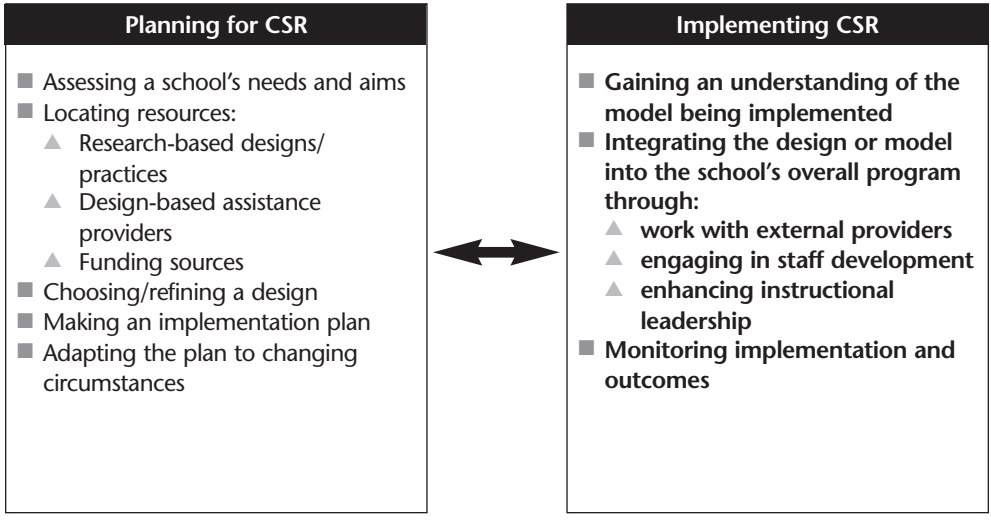
The previous section demonstrates that a new theory of comprehensive school reform emerged during the 1990s. Much of the writing about this theory, however, has been "normative" in perspective—that is, oriented to making recommendations about how to proceed when undertaking CSR. In this section, we review this normative literature, along with empirical research suggesting the various ways in which schools often fall short of implementing the theory in practice as well as studies confirming the efficacy of the model.

Figure 1 (next page) shows our version of the "normative" model of CSR. The figure is meant to portray CSR as a continuous process of improvement unfolding over time. In the model, a school enters the CSR process (on the left hand side of the figure) at the planning stage, assessing its needs and aims, scanning the environment for research-based practices or models of whole-school reform that meet these needs, and locating external assistance providers with proven records of providing support to schools. The initial planning phase concludes with the choice of a model and a plan and budget for implementing that model. The right hand side of Figure 1 describes an idealized implementation phase. Here, schools begin integrating their chosen model into their ongoing educational program by working with an external provider, engaging in staff development, and working to adjust

other elements of the school program in support of new practices. Finally, Figure 1 suggests that CSR is a continuous process. The two-way arrow running between the left and right hand boxes of Figure 1 is meant to suggest a feedback loop that feeds information about implementation and effects on student achievement back into the planning process to produce a new cycle of planning and implementation.

In our view, this "normative" model captures the intent of CSR legislation, as well as the large practice literature providing schools with guidance about how to engage in CSR. But, as we are about to see, many schools end up departing from this model in significant ways. Thus, the next two sections of this chapter discuss some of the research literature showing common ways that schools depart from the normative model in practice and also how such departures affect implementation outcomes.

Figure 1: "Normative" Model of CSR as a Continuous Process



The CSR Planning Process

There is no shortage of guidance about how to engage in the CSR planning process. The large literature on planned educational change, especially, emphasizes the importance of planning in any process of school change (e.g., Crandall et al., 1986; Louis & Miles, 1990; Fullan, 1991). Because of this, anyone undertaking a CSR planning process would be advised to consult one or more of the many guidebooks and planning tools now available to assist in planning for CSR. Among the guides that we have found helpful is Stringfield (1998). The reader also can consult the websites of any regional education laboratory or the National Clearinghouse for Comprehensive School Reform (NCCSR) to obtain a wealth of additional information on the CSR planning process, as well as links to even more planning resources.

In the paragraphs that follow, we discuss some of the most salient features of this literature.

As we shall see, the practice literature repeatedly emphasizes development of an inclusive planning process, one that that mobilizes a cross-section of parents, staff, teachers, and administrators at a school. Moreover, the literature suggests that planning should begin with a needs assessment targeting a school's improvement goals, proceed to careful research on alternative "designs" or research-based practices, including a search for external assistance providers, and then proceed to building a plan for implementation and evaluation of progress. The general literature on planning reminds us that this is best thought of as an evolutionary process in which initial plans are open to modification on the basis of serendipitous events and/or information gathered from monitoring of implementation progress (Louis & Miles, 1990). The practice literature also reminds us that in order to successfully "install" a new set of practices or a research-based design inside a school, school leaders will have to attend to the tricky problem of altering a school's organizational culture in ways that support new designs and practices (Fullan, 1991).

Due Diligence in Planning

The empirical literature on CSR shows how difficult this planning process can be. One problem is that CSR planning is often rushed in schools—probably because of a felt need to improve student achievement scores quickly, but also because of pressure from district offices. Yet hurried planning has been found to negatively affect CSR implementation (Datnow & Stringfield, 2000; Berends, Bodilly, & Kirby, 2002). The benefits of due diligence in planning are many. A slower and more deliberative process allows schools to consider more carefully the match between external designs and a school's current programs and practices, as well to consider the match of external designs to a school's organizational "culture" more broadly. For example, Bodilly (1996), Datnow (2000), and Datnow and Stringfield (2000) all have found that careful matching of external designs to school needs, culture, and existing programs increases implementation success. Equally important, a careful planning process allows teachers and administrators to better understand the demands of specific changes. Hall and Hord (1987), for example, describe the need for those working at change to understand the personal consequences of such efforts during the earliest stages of a reform initiative. Among the important concerns at this stage will be the amount of time required by the change effort, the specific activities to be undertaken, and other practical issues. Addressing these issues is especially important in schools with "crowded" reform agendas, where teacher and administrator overload can be a real problem (Desimone, 2002). In fact, careful consideration of teacher needs at the earliest stages of CSR has been found to increase implementation success (Nunnery et al., 1997).

Research on planned educational change also emphasizes the evolutionary nature of planning—the idea that planning is not simply a "front end" activity, but

is done constantly over the long haul. One reason for this emphasis is that change efforts—and especially complex efforts such as CSR—require significant lengths of time to achieve full implementation (upwards of 5-10 years by many accounts). As the RAND studies of NAS design implementation showed, a significant number of schools will achieve only partial implementation of complex designs after a period of three to four years (Berends, Bodilly, & Kirby, 2002). Implementation also can be slowed by teacher, student, and leader turnover (Muncey & McQuillan, 1996; Ross, Troutman, et al., 1997; Stringfield et al., 1998). Yet, in the current education policy environment, school systems often demand immediate payoffs from school reform, leading to occasions where complex change efforts are given insufficient time to succeed, and where planning veers from one direction to another without continuity. Successful change efforts, on the other hand, are continuous—not discontinuous. Moreover, they make realistic assessments of the time it takes to produce fundamental organizational change (Fullan, 1991).

Inclusiveness of the Planning Process

Another departure from the normative model occurs when planning takes place in a highly politicized context. Desimone (2000), for example, argues that a more inclusive process adds legitimacy to the CSR process, thus enhancing implementation success. Research evidence supports this view. Nevertheless, it appears that in many schools, broad inclusion of many different constituencies in the earliest stages of CSR is often absent. For example, we know that in many schools, administrative pressure, rather than broad participation by the school community, drives the reform process (Huberman & Miles, 1984; Datnow, 2000). To some extent, administrative pressure is an important element in stimulating change efforts, as Fullan's (1991) idea about the need for administrative pressure and support suggests. But as Datnow (2000) notes, asymmetrical power plays can negatively affect both the legitimacy of CSR within a school, and its subsequent implementation.

In fact, empirical research on CSR implementation confirms the importance of inclusive practices. Research on CSR implementation in Memphis, for example, showed that parent and community involvement in planning and implementation produced more positive implementation outcomes (Ross et al., 1997; Smith et al., 1997). Other studies point to a need for district representation at all stages of the CSR process. For example, Bodilly and Berends (1999), Cook et al. (1999), Muncey and McQuillan (1996), and Stringfield, Datnow and Ross (1998) have all reported that district support and guidance played a key role in successful implementation of CSR efforts. Districts can be especially important at the earliest stages of the CSR process by providing information about alternative CSR designs (Bodilly, 1998), and by providing input into the pace, direction, and form of change (Desimone, 2000).

Obviously, inclusion of teachers at early stages of CSR planning is important as well. The general literature on planned educational change places repeated empha-

sis on teacher "buy-in" as a key to successful school reform (e.g., Berman & McLaughlin, 1975; Fullan, 1991; Louis & Marks, 1998). The CSR literature also emphasizes this practice, emphasizing the role of teachers in choosing an external design for school improvement, usually through a faculty "buy-in" vote (Bodilly, 1996; 1998; Datnow et al., 1998; Ross, Henry et al., 1997; Stringfield et al., 1997). But Datnow (2000) has found that "buy-in" votes can be inauthentic expressions of teachers' real feelings, especially when power plays by administrators affect voting behavior. Moreover, there are many other ways in which teachers can and should be included in an authentic process of participation. Decades of research on planned educational change suggest the importance of teachers having full information on planned educational changes (Hall & Hord, 1987). At the earliest stages, especially, teachers need to understand the practical demands of a reform effort, how change efforts relate to their current values and work practices, the opportunities they will be given to learn new practices, the incentives and sanctions that will be brought to bear during the change process, and so on. All of this argues for inclusion of teachers at every step of the CSR process—but especially in the early processes of needs assessment, researching alternative designs for change, selection of a design to be implemented, and choice of external agents to provide implementation assistance.

Obtaining Information on CSR Designs and Design-Based Assistance

The CSR literature also calls for local schools to obtain rich information about alternative research-based models for change and about the organizations that provide design-based assistance to schools. In this section, we discuss this literature.

To begin, the CSR literature contains much practical guidance about how to analyze the fit of different "models or "designs" to a school's specific situation prior to model adoption. Attention to models fit within the local context makes sense, since the general literature on planned educational change demonstrates that successful innovations are classroom friendly, well-defined, practical, and relevant to teachers' needs and interests (Huberman, 1983). Of special note, however, is the emphasis placed in the CSR literature on finding out about the research base for CSR models prior to implementation. Because of this emphasis, several guidebooks have been written compiling the research evidence standing behind the most well-known CSR models (Hermann, 1999; Northwest Regional Educational Laboratory, 1998). In addition, Stringfield (1998) has offered useful advice about how to evaluate the research base of reform models.

There are formidable problems associated with evaluating the research basis for most CSR models, however. For one, only a handful of the hundreds and hundreds of such models now in use have been evaluated by "scientific" standards (Hermann, 1999; Borman et al., 2003). Moreover, practitioners usually do not have sufficient information or training to judge the quality of scientific evidence lying behind claims for effectiveness made by designers. On the other hand, analyses of

the state of research on particular CSR models might not make as much of a difference to the actual choice of CSR models as many analysts think, or CSR legislation requires, since a great deal of research suggests that even where strong scientific research exists, it is seldom the determining feature driving adoption of innovations by working professionals. Instead, it appears that professionals in applied settings filter research evidence through the lens of their own theories-in-use, so that the more consonant the evidence is with their existing preferences, the more likely it is to be accepted and put into use (Dunn, Holzner, & Zaltman, 1990).

None of this is meant to de-emphasize the importance of searching for information about models and practices. But it does suggest two things. First, those seeking to evaluate and build support for the adoption of particular CSR models in a given school setting would do well to do more than look for "scientific" evidence of effectiveness—which in most cases will be scarce or ambiguous. Stringfield (1998), for example, presents a number of strategies, not only for evaluating the evidentiary claims about a CSR model's effectiveness, but also for assessing its fit to a school, the experiences of prior adopters, the philosophical underpinnings of the model, the practical demands it will make on implementers, the costs and feasibility in the local setting, and so on. This is sound advice, suggesting that scientific evidence of effectiveness should serve as a necessary criterion for program selection, but that once a set of effective programs has been selected, other factors can be taken into account in selecting a final program to be implemented in a particular school setting.

Similar advice pertains to the choice of external providers of design-based assistance. Since assistance providers are usually closely associated with a particular CSR model, the adopter needs information about what services will be received, the track record of the assistance providers in similar contexts, and so on. A particularly useful practical guide to working with assistance providers was developed by NAS (2000). This document (available on the NAS website) lists guidelines for assuring the quality of design-based assistance, has a toolkit for engaging in a selection process, and even includes a standard contract that can be used in negotiating a service agreement.

Developing an Evaluation Strategy

Although the CSR literature places emphasis on gathering information about models and model providers prior to adoption, it is important to remember that teachers and school administrators will be in a much better position to evaluate the impact of a CSR model after it has been put into use in their school. As Hall and Hord (1987) note, after some period of use, the concerns of adopters usually evolve from practical concerns (How long will it take to make a change? What do I have to do in the process?) to concerns about the impact of the change on tasks and outcomes. This is another reason to stress the evolutionary nature of CSR planning, for sound judgments about the fit and impact of an innovation in a particular local

context are much better determined after the innovation has been put in use. Only after the innovation has been put into use can school professionals truly judge its fit and impact in the local setting and begin the process of mutual adaptation between the external model and local context that is the hallmark of comprehensive school reform.

An important part of this process is formative evaluation for continuous improvement. The CSR literature, for example, repeatedly recommends the establishment of benchmarks for implementation and student outcomes as part of any CSR effort. Schools that have adopted an externally-developed CSR model supported by a design-based assistance provider should expect the external provider to have developed such benchmarks, based on the provider's experience in many schools. The provider also should have a strategy for using those benchmarks in a process of formative evaluation. In many cases, progress toward benchmarks will be assessed by staff from the external provider, but as Ross (2000) notes, school improvement teams can also develop their own set of benchmarks and evaluation tools, either to supplement those used by external assistance providers, or if pursuing a "home-grown" model. The benchmarking process begins by outlining all of the components of a design or model to be implemented, along with a sense of how long it will take to achieve particular levels of implementation for each component, and rubrics describing different levels of implementation. Schools can then use a variety of methodologies described by Ross (2000) to assess progress toward the benchmarks—including questionnaires, focus group and individual interviews, and school and classroom observation protocols. In addition, schools should establish expectations about student outcomes, and by law, schools receiving CSR funds need to evaluate student outcomes on an annual basis. Again, Ross (2000) provides relevant guidance on this issue.

Of course, the process of formative evaluation is both time-consuming and useless if results are not used to plan for further action. Ross (2000) suggests that given the time commitments and challenges faced by teachers and leaders in CSR efforts, the process of formative evaluation might be accomplished better by external providers, third-party evaluators, or district personnel than by school staff. He also suggests strategies for feeding evaluation findings back to the school as a means of stimulating continuous improvement. The effectiveness of using results from formative evaluations is well established in the general literature on planned educational change (Fullan, 1991). When fed into naturally-occurring work groups such as grade level teams, school improvement teams, or other faculty groups, evaluative information on processes and outcomes directly addresses the concerns of innovation users about the practicality and consequences of their change efforts, and (if the process is working well) can enhance the commitment of school staff to the change process. As Crandall et al. (1986: 34) note, "The commitment of teachers increases as they simultaneously see themselves master the practice and perceive that their students are doing better." Moreover, information from formative evaluations is a critical part of what Miles (1992) called the development of problem-cop-

ing strategies within a process of evolutionary planning—a sustained "mindfulness" leading to further diagnosis and action-taking.

Resource Considerations

The final element of a CSR plan discussed in this chapter is budget development. Several studies of CSR have noted the obvious centrality of resources to implementation success (for a review, see Desimone, 2002). Reflecting on the NAS experience, Glennan (1998) found that design implementation is vitally affected by resource constraints, and that implementation is weaker or ends with lack of resources. Similarly, Berends, Bodilly, and Kirby (2002) note that lack of funding was the single most important reason teachers implementing NAS designs cited for dropping a design. The Memphis studies of NAS design implementation support these findings, showing that schools with the most implementation success were those that had both an organizing principle around which to allocate resources and more resources to work with (Ross, Troutman, et al., 1997; Smith et al., 1997). Desimone (2000) suggests that allocation of sufficient resources is taken by teachers as a sign of organizational commitment to the CSR process. This is supported by Berends' (2000) finding that teachers who reported having more resources also reported higher commitment to the CSR process.

Because of the importance of resources to design implementation, the CSR practice literature contains useful advice about how to allocate resources in support of CSR efforts. Odden's (2000) discussion of this issue is particularly helpful. Working from a set of budget assumptions, Odden demonstrates that the first-year costs of adopting several promising NAS-supported designs can be covered using a variety of funding strategies. But each of these strategies involves a change in the current staffing and resource allocation patterns used in schools. The strategies developed by Odden (2000:11), for example, closely follow recommendations of the NAS design teams that he analyzed, and they call for a significant restructuring of roles within schools. In particular, NAS designs often expand the role of classroom teachers to include some of the responsibilities previously performed by various categorical and student personnel specialists in schools, whom Odden recommends be dropped from school budgets. To offset the increased demands on teachers arising from these changes, NAS designs typically reduce class size, so that the expanded teacher role is accompanied by a reduced student load. Moreover, funds formerly devoted to various school specialists are now used to employ additional instructional leaders and to significantly increase staff development.

Overall, the strategies discussed by Odden (2000) can be difficult to implement in some settings. First, they require the loss (or at least reclassification) of some jobs. Moreover, his strategies work best in sufficiently-funded schools—not in schools that are resource starved and lack the funds that Odden's re-allocation strategy depends on. Moreover, it is quite possible that Odden's figures underestimate at least one important cost of school restructuring and CSR. That is the non-reim-

bursed expense of additional time for common planning and instructional coordination occurring outside paid professional development time. Common planning time has been found to be central to the process of successful school change in many studies (Bodilly 1998; Darling-Hammond & Miles, 1998; Muncey & McQuillan, 1996; Ross, Troutman, et al., 1997). So, if schools do not already have common planning time built into their schedules, this is an additional resource that must be obtained to implement CSR successfully.

The CSR Implementation Process

In this section, we turn from a review of the literature on planning for CSR to a review of research on CSR implementation. A growing body of research on this issue is available, but much it of consists of research on just a few of the CSR models currently operating. The best (and most highly-cited) research, for example, comes from the RAND study of the original NAS designs (Berends, Bodilly, & Kirby, 2002; Stringfield, Ross, & Smith, 1996) as well as research on a handful of other widely-disseminated CSR models such as the Accelerated Schools Program (Finnan et al., 1996); America's Choice (Supovitz, Poglinco, & Snyder, 2001), the Coalition of Essential Schools (Muncey & McQuillan, 1996), the Comer School Development Program (Cook et al., 1999; Haynes, 1998), Core Knowledge (Datnow et al., 1998), and Success for All (Datnow & Castellano, 2000). Care should be taken in generalizing from this limited body of research. One problem is that attention to just a few models potentially limits the external validity of research on CSR implementation. But equally important, CSR models—especially the most heavily-studied ones—are evolving constantly as their design teams continue to refine the models and change procedures used to support local implementation.

Despite these cautions, much has been learned about the process of CSR implementation through research. Moreover, research on CSR implementation is generally quite consistent with research on planned educational change. In particular, research on CSR implementation, like research on planned educational change, shows great variability in local implementation. Apparently, variability in CSR implementation results in part from differences in the characteristics of CSR models themselves, but implementation of the same CSR model has been found to vary across district and school contexts, and teachers implementing the same CSR model within the same school also show variability in implementation (see, for example, Berends, Bodilly, & Kirby, 2002). This section reviews research seeking to account for this variability and describes several factors that have been found in previous research to improve implementation outcomes.

The Effects of CSR Designs on Implementation

One factor affecting implementation is the overall "design" of CSR models, some of which appear to be more easily and faithfully implemented than others (Desimone, 2000; Berends, Bodilly, & Kirby, 2002). This is consistent with previous research on planned educational change, which found that the characteristics of an educational innovation affected its prospects for implementation (Firestone & Corbett, 1988). In particular, the literature on planned educational change suggests that innovations that are very clear and specific in the guidance they provide for implementation, and those that also provide technical assistance to support implementation, are the ones that end up being well implemented. As we shall see, a similar set of findings has emerged from research on CSR.

Many researchers have tried to develop a conceptual framework for describing the design or characteristics of various CSR models. In the sense used here, a "design" is a kind of blueprint for change that varies along two dimensions. First, the design of any CSR program will identify some targets of change. That is, it will delineate the particular features of schooling that are to be restructured as a result of the CSR process. Second, CSR models will have designs for bringing about change that include not only descriptions of what the targets of change should look like after successful implementation, but also specific ideas about the people and processes that will be involved in bringing about these changes.

Targets of Change. In thinking about targets of change, it is important to recall that the aim of CSR is to bring about comprehensive change in schools. For this reason, we should expect CSR models to have many targets of change, including changes in the areas of curriculum and instruction, school organization and management, family and community participation, and so on. As discussed earlier, the federal government's CSR legislation suggests 11 characteristics of any CSR program, and these areas can be seen as constituting a wide variety of targets for school change. As it turns out, however, CSR models differ in the number of elements they target for change. For this reason, various guidebooks have been developed to help practicing educators get a sense of changes targeted by particular CSR models. We have found two of these guides especially useful, one developed by the American Institutes for Research (Herman et al., 1999) and another by NWREL/NCCSR (1998-2003).

Overall, the guidebooks show that CSR models can differ in the extent to which they emphasize different targets of change in schools. Some designs primarily seek to change organizational processes within schools; for example, school planning processes and/or decision making structures. Often, the assumption is that these organizational changes will provide a springboard for later (but less well-specified) changes in curriculum and instruction. The Accelerated Schools program and the Comer School Development program are examples of this approach. Other programs focus directly on changes to the curriculum, but little else. The Core Knowledge program is a particularly striking example of this approach. It focuses

largely on changes in curriculum materials, providing only minimal guidance about changes to instructional practice or school organization and management. Still other models focus on curriculum and instruction and organizational and managerial arrangements. But here too, comprehensiveness can vary. For example, Success for All focuses intensively on a single curricular area (elementary school literacy instruction), but it does so comprehensively, by calling for changes in curriculum, instruction, grouping, assessment, staffing, and other elements of a school. Still other CSR models focus on multiple areas of the curriculum and seek comprehensive changes in schools; for example, Roots and Wings, Co-Nect, or Modern Red Schoolhouse.

One would expect the number or type of changes targeted by a particular design to affect the scope and pace of CSR implementation in schools. To some extent, the RAND studies of NAS design implementation support this view. For example, two of the most ambitious NAS designs (Co-Nect and Modern Red Schoolhouse) showed lower levels of implementation than schools implementing Success for All (Berends, Bodilly, & Kirby, 2002), although this might have resulted from more than the complexity of the designs, per se. Beyond this study, however, there is very little research on the issue of how targets for change affect CSR implementation.

An older generation of research on planned educational change does provide substantial evidence, however, that the complexity of innovations affects the implementation process. This literature has shown that simpler and less complex designs—i.e., those that require fewer changes—show smoother and faster implementation. However, simple innovations also can produce only modest changes in schools, especially if they target a single grade, a subset of teachers, a special program, and so on. By contrast, more complex innovations, when successfully implemented, appear to produce more fundamental change; but successful implementation of complex innovations seems to occur only when the implementation process is broken into separate parts that are carefully staged and sequenced (Firestone & Corbett, 1988: 382).

Specificity (or Clarity) of Designs. Designs differ not only in the number and kind of changes they target in schools, but also in how clearly the end state of the change process is described. Some designs, for example, present clear descriptions of how things should look after full implementation—for example, how instructional practices should look in well-implemented cases, or the specific organizational arrangements that are to be implemented. But other designs are less clear—often by intention. For example, the Accelerated Schools Program is very specific about how to organize decision making processes within schools to take stock of needs, create a vision for change, and govern the change process. But the design is much less specific about the kinds of instructional changes that schools are expected to make. Schools are told to aim toward "powerful learning," but that is defined broadly as instruction that is authentic, learner-centered, interactive, continuous, and inclusive. No curriculum materials are required by the program;

nor does the program mandate specific instructional practices. Instead, teachers are expected to engage in an active process of discovery to arrive at "powerful learning" within their own classroom. This approach stands in strong contrast to a highly-specified instructional program like Success for All, which gives teachers an explicit set of curricular materials and lesson scripts, complete with a daily and weekly schedule of lesson activities.

The literature on planned educational change suggests that innovations that are more clearly specified are implemented more quickly and with more fidelity than those that are less clearly specified, a finding that is confirmed in research on CSR implementation literature (for reviews, see Firestone & Corbett, 1988; Desimone, 2000). With respect to CSR, Bodilly (1996), Stringfield et al. (1997), and Smith et al. (1997) all found that more clarity in design—especially the presence of a more structured curriculum, specific guidelines for classroom practice, and more training—all promoted faster and smoother CSR implementation. Others have argued that a lack of specificity explains why locally developed CSR programs have experienced less successful implementation than externally developed models (Desimone, 2000).

These are important insights, but as we discuss below, it is important to remember that "clarity" of design is more than a pre-existing characteristic of a program. It also emerges during the implementation process. Thus, plans for changing instruction are more clear if, prior to implementation, design teams provide school personnel with concrete descriptions of desired instructional practices in the form of various print descriptions, video tapes, or lesson scripts; but clarity also emerges when design-based assistance providers model desired instructional practices for school personnel, observe teaching practices and provide feedback, and so on. Similarly, plans to change a particular segment of the curriculum become more clear when teachers have access to curriculum guides, textbooks, or model lesson plans prior to implementation; but clarity also improves as teachers receive training from design-based assistance providers about the philosophy and uses of materials during implementation.

Blueprint for Implementation Support. This brings us to a third feature of program designs—their plans for assuring strong implementation in local settings. By design, models differ not only in terms of targets for change and specificity of model descriptions, but also in plans for the timing and sequencing of implementation activities, and the amount and nature of on-site assistance to be provided to schools. Partly as a result of federal CSR legislation, more and more external providers are offering higher levels of technical assistance to local schools than they did in the past. Such assistance comes in varied forms, however. It includes not only more clearly specified plans for implementation and formative assessment, but also the provision of on- or off-site staff development for teachers and school leaders, site visits by design team members to local schools (to check implementation and/or to follow up on training provided), and sometimes the appointment of dedicated personnel to local schools to assist with implementation.

These elements of CSR design make good sense. Decades of research on planned educational change show that implementation is stronger when program developers provide on-site technical support for local implementation (Crandall & Loucks; 1983, Emrick et al., 1988; Louis et al., 1981). Moreover, the limited literature on CSR implementation confirms this point. As we discuss in more detail below, both the extent and quality of professional development received by teachers affects the extent and pace of CSR implementation in schools. So, the presence of plans for on-site technical assistance, and the creation by design teams of a dedicated and competent staff for providing such assistance, are keys to successful implementation of CSR models in schools (Slavin, 1999).

The Effects of External Assistance on Implementation

We have just seen that the "design" of a CSR effort consists of describing the targets of change for the CSR process, the timing and sequencing of that process, the people and processes that will be used to bring about such changes, and a description of how a school should look once it has fully implemented the change process. But a design—in itself—does not bring the change process to fruition. Instead, schools bring a design to fruition by working with external assistance providers to implement the design in practice. Thus, the CSR implementation process, which increasingly involves extensive work with external assistance providers, is very different from other models of planned educational change that simply ask schools to engage in change based on "information" found in papers, technical reports, and guides found in clearinghouses or professional publications.

Work with external assistance providers involves the formation of what Miles (1992) called "temporary systems" to support change. For example, most CSR providers work directly with schools, but only on a temporary basis (usually one to three years) and usually at decreasing levels of support over time. Thus, external providers are best conceived as temporary "linking agents" in the implementation process; that is, agents working to facilitate communication and information between the design teams who created CSR models and the schools who are implementing these designs (Havelock, 1971). There are, we should point out, additional ways of establishing linkages between design teams and schools that do not rely on the direct activities of linking agents. For example, many CSR model providers have created networks of schools implementing their model, and they hold conferences and other meetings for personnel working in these schools (Desimone, 2000). The goal here is to develop a "linkage system" that can be sustained after withdrawal of direct, on-site support by linking agents (Keeves, 1990).

The literature on planned educational change demonstrates the importance of linkage agents to implementation. For example, early studies found that simply sending schools information describing innovative "designs" did not lead to high levels of program implementation, while on-site assistance by linking agents did (Emrick & Peterson, 1977). This same literature has identified several different roles

that external providers can play in the change process (Hood, 2002). Linking agents rely on interpersonal communications to establish linkages, help local educators learn more about and make wise selections of research-based practices to implement, provide on-site technical assistance throughout the change process (including problem definition, needs assessment, planning and evaluating change efforts), provide direct training in and support of new practices, and provide feedback from local schools to design teams. Many have noted that this is a unique role in the education system that requires linking agents not only to be skilled practitioners, but also skilled change agents, knowledgeable about cutting edge research and its translation into practice (Hood, 1982).

Research on CSR implementation confirms the importance of linking agents to successful implementation of CSR models (for reviews, see Desimone, 2000; Berends, Bodilly, & Kirby, 2002). For example, the RAND studies of NAS design implementation showed that the quality of external assistance provided by linking agents was highly predictive of implementation success. When linking agents more clearly communicated the CSR program's design and purposes, provided higher-quality staff development, and engaged in on-site monitoring of implementation, schools were characterized by higher levels of implementation. But the RAND studies also found that many NAS design teams were stretched when it came to employing linking agents effectively. For example, Bodilly (1998) found that stability in the external assistance team was an important factor in implementation success, but stability in team membership was often missing. And Berends, Bodilly, and Kirby (2002) found much variability in the quality of assistance provided by linking agents—both across CSR providers, and across linking agents working for the same provider. In fact, in the RAND studies, a substantial proportion of teachers were unsatisfied with the quality of external assistance they received. Some teachers reported interacting only rarely with linking agents; others reported that assistance was not informative or helpful, and many reported valuing assistance from their local colleagues over assistance from design team representatives.

Our own research on three of the most widely-disseminated CSR models (the Accelerated Schools Program, America's Choice, and Success for All) suggests that these particular design teams have recognized the difficulties identified by RAND researchers and are working to improve their strategies of assistance to local schools. To be sure, these particular models still emphasize direct technical assistance from linking agents, but this goes beyond initial training to also include additional training in specific design components, observation in the context of site visits, assistance in interpreting state assessment results, and ad hoc assistance via e-mail and telephone. Moreover, these design teams also have developed additional assistance strategies that include formation of national and local networks that teachers and school leaders can join for implementation support; they have built into their designs explicit opportunities for collegial learning; they have enriched the instructional guidance provided in curricular documents, lesson plans, and so on; they have developed better descriptions of the teaching practices they want

teachers to implement (through development of vignettes, video tapes, and model classrooms); and they include within their designs a definite plan for expanding the amount and quality of instructional leadership exercised by a school's staff (Peurach, Glazer, & Gates, 2004).

The Effects of Staff Development on Implementation

Central to the package of services provided to schools by design teams is professional development. Of course, the importance of professional development to planned educational change is well-established (Fullan, 1991). Thus, it is not surprising that research on CSR implementation finds staff development to be one of the most critical ingredients in promoting successful implementation of CSR designs (for a review, see Desimone, 2002). To be sure, the quality of staff development has been found to vary—across CSR models, across the linking agents providing it, and across the schools where it occurs (Berends, Bodilly, & Kirby, 2002). For example, a common finding in the CSR literature is that teachers attribute implementation problems to a lack of training. However, combining the findings from a variety of studies, Desimone (2000) suggests that higher levels of CSR implementation are typically associated with more intensive staff training, taking place over longer periods of time, organized on a schoolwide basis, and reinforced by local facilitators.

Decades of research on planned educational change reinforce these observations. But research also suggests that staff development—especially in the CSR context—should include more than a series of workshops provided by linking agents or other design-based assistance providers. Rather, in CSR, staff development is part of a larger, organizational design that transforms schools undergoing CSR into learning organizations (Fullan, 1991). A general goal is to make the new practices embedded in a CSR design clear and practical for school staff, and to give staff the opportunity to learn how to implement new practices in a supportive and information rich environment. Workshops provided by linking agents can help in this process, but many CSR designs incorporate additional strategies to enrich the learning environment for school professionals.

These additional strategies include embedding learning opportunities inside instructional materials (i.e., in lessons plans, teachers' guides, curriculum standards, student assessments, etc.), setting up model classrooms inside schools, changing staffing patterns inside schools to assure the presence of on-site and locally-based instructional leadership, using staff meetings and common planning periods as opportunities for discussing and learning about new design-based practices, and strengthening collegial relationships through formation of study groups, classroom observation pairings, or participation in teacher networks (Peurach, Glaser, & Gates, 2004). Indeed, some authors suggest that those engaged in CSR think in terms of a system of professional development that is school-based, rich in collaboration and problem solving, aimed at promoting both practical and theoretical understanding of new practices, sustained over time, and integrated into a comprehensive process of school change.

The Effects of Leadership and Professional Culture on Implementation

To this point, we have focused on the role that program design teams play in promoting successful CSR implementation in schools. As the review to this point shows, when design teams work to produce clear, specific, and high-quality designs for change, and when they provide extensive technical assistance and implementation support to local schools, CSR implementation often proceeds successfully. But the process of comprehensive school change also requires positive actions on the part of the local school community as well, and in most treatments of this issue, the key community member is identified as the school principal.

The importance of the school principal to planned educational change has long been established (Fullan, 1991). Thus, it is not surprising to find that principal leadership has been identified as important to successful implementation of CSR efforts (for a review, see Desimone, 2000). For example, the RAND studies of NAS design implementation found that teachers' perceptions of principal leadership were among the most important predictors of implementation success (Berends, Bodilly, & Kirby, 2002), and many other studies, of a remarkably diverse set of CSR models, have discussed the centrality of principal leadership to CSR implementation (see, for example, Anderson & Shirley, 1995; Christenson, 1996; Cooper et al., 1998; Haynes, 1998; Smith et al., 1997; Smith et al., 1998). From all of this literature, a common core of findings has emerged about the leadership activities of principals that contribute to implementation success. Principals who lead successful change efforts have a clear vision of the short and long range goals of the change effort they are leading, are actively involved in decision making, directly or indirectly support the professional learning of teachers, aggressively seek resources for change, buffer school staff from unwarranted intrusions and distractions, and seek policy changes at the district level that support a school's change efforts (Rutherford et al., 1983; Hord & Huling, 1986; Desimone, 2000; Berends, Bodilly, & Kirby, 2002).

It would be a mistake, however, to assume that principal leadership alone can carry the day. For one, not all principals are willing to assume the complex leadership role just described. More importantly, the task of leading change inside a comprehensive school reform effort is simply too large and complex for a single individual—no matter how energetic, charismatic, and forceful. As a result, many CSR models deliberately restructure schools to provide additional leadership roles in schools. Camburn, Rowan, and Taylor (in press) have studied this process in some detail in a study of leadership teams in schools implementing three CSR models—the Accelerated Schools Program (ASP), America's Choice (AC), and Success for All (SFA). They found that schools implementing these models were characterized by different leadership configurations compared to schools not implementing the models. In particular, the CSR programs restructured schools by adding various instructional leadership positions to the school staff. For example, ASP added a

coach to the school, AC typically added a design coach and a literacy coordinator, and SFA added a reading facilitator. Interestingly, in the schools implementing CSR programs and in those not implementing one of the programs, leadership was exercised by small teams—ranging from three to seven people depending on school size. But, there was an apparent division of labor among team members. Principals tended to exercise high levels of instructional leadership, but also were highly concerned with general building management and external relations. Incumbents of CSR-specific roles, by contrast, appeared to specialize in instructional leadership, and to exercise more instructional leadership than was exercised by support staff in non-program schools, for example, Title I coordinators, mentor teachers, and so on. One explanation for this finding was that CSR programs provided school leaders with additional staff development in the area of instructional leadership, and this additional staff development boosted the attention school leaders gave to most instructional leadership functions.

In a related study, Taylor (2004) showed that the instructional leadership provided by these teams was insufficient—in itself—to enhance the interest and motivation of teachers to implement a CSR design. Instead, the effect of positive instructional leadership was conditioned by the larger professional culture of the school. In schools where teachers had formed a cooperative school culture, characterized by norms of support, innovation, trust, and collegiality, the effects of additional instructional leadership on teachers' clarity about the CSR process and motivation to participate in the reform effort was higher than it was in schools where the professional culture lacked these characteristics. Thus, as many observers have noted, the larger professional culture of the school is also an essential element in promoting school change (e.g., Fullan, 1991).

The Effects of Quality Assurance on Implementation

A final characteristic affecting implementation of CSR designs is quality assurance; that is, ongoing monitoring of implementation and student outcomes. Little is known about this process, although the efforts of NAS design teams during the scale-up phases have been cited by Berends, Bodilly, and Kirby (2002) as an important support for implementation. In particular, the NAS design teams were spurred by the demands of school people to develop design-specific "benchmarks" spelling out in very clear terms what schools at various phases of the implementation process were expected to have accomplished. As Berends, Bodilly, and Kirby (2002) note, design teams began to use these benchmarks to actively monitor implementation in school sites, and to actively communicate implication plans and schedules. All of this improved communication among external providers and school and district personnel, and it provided important feedback to design teams allowing them to improve their services and make them more effective. Again, this observation is consistent with the larger literature on planning for educational change, which stresses the evolutionary nature of educational planning and the need for continuous feedback about progress toward goals.

Context Effects on CSR Implementation

To this point, we have sketched out a process leading to successful implementation of the CSR process. As we have seen, successful implementation occurs with greater frequency in schools that exercise due diligence in planning—carefully searching for designs that match their needs, gaining the broad support of community members, addressing teachers' specific concerns about the change process, and carefully staging the change process, especially when implementing a complex design. Successful implementation also results from the positive efforts of design teams to formulate clear program designs, and when design teams provide high-quality technical assistance to schools. As we saw, the best designs include multiple and extended opportunities for professional learning, and they employ well-qualified and talented linking agents to work with schools. Conditions at the local school also are required to promote successful implementation. For example, schools with cooperative professional cultures and characterized by strong principal and staff leadership are more likely to benefit from the CSR process and to implement deeper changes in the school.

Unfortunately, research on CSR shows that in many cases, some or most of these conditions are missing, with the result that schools only partially implement or fail altogether to implement CSR efforts. The process of CSR is complex, and many things must come together simultaneously. But, even when conditions supporting change are positive—clear, practical, and well-supported designs are being implemented in schools that are characterized by strong leadership and a cooperative professional community—CSR efforts can proceed unevenly. That is because the process of CSR unfolds in a larger context of existing school, district, and state policies and practices. In this section, we discuss what has been learned about the effects of this larger context on the CSR process.

Coordinating CSR and Other Improvement Efforts

One problem impeding CSR implementation occurs when schools engage in multiple, inconsistent change efforts. One aspect of this problem is the adoption of very ambitious plans for change in school contexts that are already "overloaded" with change efforts. Another aspect of this problem is the potential lack of fit of any given change effort with other reform efforts. Along these lines, many studies have commented on the teacher overload that occurs in schools attempting too many uncoordinated changes. For example, Smith et al. (1997) found that successful implementation was higher in Memphis schools that were attempting less complex CSR designs. The explanation for this was that teachers were more easily able to attend to the change process in these schools given all else that was going on. Other analysts have argued that the adoption of CSR models can suffer from lack of "fit" in schools where many different changes are occurring (e.g., Bodilly, 1996; Datnow & Stringfield, 2000). Thus, it appears that the presence of too many change

efforts saps teachers' energies and creates a host of potentially inconsistent policies and practices, all of which impede successful implementation of coherent CSR designs (Berends, Bodilly, & Kirby, 2002).

Many studies of CSR implementation also suggest a need to coordinate CSR efforts with existing local, state, and federal reform efforts. For example, many CSR models have their own sets of curricular and instructional guidelines, as well as particular assessment instruments designed to measure students' achievement of model-specific learning objectives. However, this model-specific instructional guidance can be inconsistent with the instructional guidance provided by local districts, or by state assessments associated with accountability systems. Research on CSR implementation suggests that aligning these various forms of instructional guidance in support of a coherent set of curricular and instructional practices within schools is a major challenge of CSR (Glennan, 1998). In particular, several studies have shown that CSR implementation can be undermined in schools facing strong and high-stakes accountability systems, especially when the assessments involved in accountability programs are not closely aligned with the instructional objectives and learning outcomes promoted by particular CSR designs (Smith et al., 1997; Glennan, 1998; Bodilly & Berends, 1999; Desimone, 2002). Unfortunately, the CSR literature provides little concrete guidance about how to avoid such conflicts, except to stress the need for local school personnel to carefully assess the degree of "fit" or "alignment" between CSR models and local and state standards during the CSR planning stage. Failure to take this early step can negatively affect implementation at later stages.

School District Contexts

Other school district policies and practices also affect CSR implementation, as many studies have shown (Timar, 1989; Muncey & McQuillan, 1996; Ross, Troutman, et al., 1997; Wasley et al., 1997; Haynes, 1998; Stringfield, Datnow & Ross, 1998; Cook et al., 1999; Freidman, 1999; Berends, Bodilly, & Kirby, 2002; Desimone, 2002). The RAND research on NAS implementation, for example, found large between-district differences in CSR implementation (Berends, Bodilly, & Kirby, 2002). Findings from several other studies support these findings, leading to the conclusion that districts have an important role to play in CSR implementation. From the literature, it has been found that the district role in implementation begins during the CSR planning stage. At the outset of a CSR initiative, for example, districts can help schools choose wisely among alternative CSR designs, especially by assuring that schools receive adequate information from design-based providers and by helping schools choose designs that are consistent with ongoing state and district curricular, instructional, and accountability efforts (Wasley et al., 1997; Freidman, 1999; Berends, Bodilly, & Kirby, 2002; Desimone, 2002). Districts also support school-level implementation by making CSR central to the district's overall improvement agenda, by having leaders openly express support for CSR

efforts in schools, and by providing a steady stream of funding and other resources to support the work of design teams and school professionals as they work at CSR.

The importance of district resources to implementation success cannot be stressed enough. The RAND studies, for example, found that lack of resources was the single most powerful explanation for the failure of schools to institutionalize CSR reforms after initial implementation (Glennan, 1998). Moreover, the RAND studies suggested that district support must extend beyond simply paying fees to design teams, for such fees, it was found, typically accounted for only a portion of the costs associated with implementing CSR in schools. In a RAND study of NAS implementation in San Antonio, Texas, for example, fees paid to the NAS design teams covered only a third of the total cost of CSR in schools, leaving other important costs to be paid by districts, including costs incurred by district staff in providing information, technical assistance, staff development, and evaluation services to CSR schools (see also, Levin, 1995).

Other studies suggest that the political and financial support for CSR can be difficult to come by in many school systems. Timar (1998), for example, has argued that in large and complex school systems, the redefinition of administrative and teaching roles called for by many CSR programs, the decentralization of administrative control to accommodate schools' needs for flexibility, and the management of various forms of conflict that inevitably accompany change efforts can prove difficult. Moreover, large urban school systems face a great deal of turbulence—turnover in district leadership, district budgetary or political crises, conflicts in collective bargaining, and conflicts between design teams, district staff, and local school personnel. All of this can negatively affect CSR implementation (Berends, Bodilly, & Kirby, 2002; Desimone, 2002). Power struggles among district constituencies can be especially common, resulting in failures to achieve the requisite coordination among the multiple constituencies that is needed to support CSR implementation (Mirel, 1994; Bodilly, 1996; Muncey & McQuillan, 1996; Datnow, 2000; Desimone, 2002). As a result, it is clear that CSR implementation involves more than just resolving the "technical" problem of choosing the right design and providing schools with sufficient resources and technical assistance to assure proper implementation. Research also indicates that CSR occurs in a complex and dynamic political environment and succeeds in the long haul only when there is a continuous cooperation among many different political constituencies.

A Brief Look at the Study of Instructional Improvement

To further illustrate these findings on CSR implementation, we turn now to a brief discussion of our own research on the process of CSR implementation. Over the past four years, the authors of this paper have been working with colleagues at the Consortium for Policy Research in Education to study three of America's most widely-disseminated CSR programs—the Accelerated Schools Program (ASP), the America's Choice program (AC), and Success for All (SFA). Known as the *Study of Instructional Improvement* (and hereafter called "SII"), our research is being conducted in 116 elementary schools located in 45 school districts in 17 states across the United States. SII is designed as a quasi-experiment that follows four demographically-matched groups of schools as they engage in the process of school improvement over a four-year period. Three of these include schools implementing one of the three CSR models under study. A fourth group consists of "comparison" schools that are not implementing one of these CSR models.

Research Questions

In this paper, we will use data from SII to address three salient questions about CSR implementation. First, we will examine how the three CSR programs under study are designed; that is, the changes the model developers are trying to make in schools, and how they have decided to go about making these changes. Second, we will ask whether schools implementing these three different CSR models develop different patterns of school organization and culture, different processes for instructional management and coordination, or different patterns of classroom instruction as a result of working with these different CSR programs. Finally, we will examine how various contextual features of the schools under study affect the CSR implementation process. Here, for example, we will look at the extent to which factors such as district size and complexity, state and local policy environments, and school or district demographic conditions affect the scope, pace, and success of CSR implementation in schools implementing the different CSR models.

Sample

To address these research questions, researchers conducting the *Study of Instructional Improvement* constructed a purposive sample of 116 elementary schools located in 17 states across the United States. As part of this process, schools were stratified along two dimensions: the date when they began CSR implementation and the poverty levels of the neighborhoods they served. Working from comprehensive lists of schools provided by the CSR programs under study, and using U.S. Census data from 1990, SII researchers selected a final sample of schools that was distributed across the sampling strata shown in Table 1 (next page). Table 1 shows that nearly half (56 of

the 116) of the schools in the sample were located in America's highest poverty neighborhoods (i.e., neighborhoods at or above the 75th percentile of a measure of neighborhood poverty developed for this study). Table 1 also shows that the sample was nearly equally divided among schools that began implementing CSR programs during the school years AY 1998, AY 1999, or AY 2000.

Table 1: Stratification of Elementary Schools in the Study of Instructional Improvement by Year of Entry into CSR and Neighborhood Poverty Level

	High Poverty (75th percentile or below)	Medium Poverty (50th to 74th percentile)	Low Poverty (25th to 49th percentile)	Total Schools
AC				32
1998	1	3	0	4
1999	6	0	6	12
2000	11	4	1	16
ASP				28
1998	0	0	5	5
1999	6	4	3	13
2000	5	3	2	10
SFA				31
1998	7	1	4	12
1999	8	4	2	14
2000	0	2	3	5
Comparison	12	7	6	25
Total Schools	**56**	**28**	**32**	**116**

Table 2 (below) shows the average demographic characteristics of schools in the study, as well as the demographic characteristics of the school leaders and teachers working in these schools. On average, the elementary schools in the sample had about 500 students, were located in school districts enrolling over 100,000 students, and served neighborhoods where about 20% of households were in poverty. Like most elementary schools in the U.S., teachers in these schools tended to be predominantly female, from non-minority backgrounds, with around 12-14 years of experience. A substantial percentage of teachers also held graduate degrees. School leaders in the sample (principals, assistant principals, CSR program facilitators, and other teacher leaders) also tended to be females, from non-minority backgrounds, holding graduate degrees, and had served in their present leadership roles for around five years.

Table 2: Average Demographic Characteristics of Schools, Leaders, and Teachers in the Sample, by Program Group

Schools, N=144	**Total**	**ASP**	**AC**	**SFA**	**Comp**
School Enrollment	503.77	484.96	562.74	465.23	498.19
District Enrollment	117,40	20,544	171,867	99,772	177,359
Proportion-Household Poverty	.19	.14	.19	.23.	.22
Avg. Woodcoc-Johnson (LA)	99.28	97.68	102.32	94.15	103.31
Percent Minority Enrollment	79.17	69.59	88.20	82.07	75.37
Leaders, N=681					
Female	0.84	0.82	0.86	0.88	0.78
Hispanic	0.05	0.02	0.04	0.07	0.12
African-American	0.29	0.19	0.44	0.26	0.29
Asian	0.05	0.00	0.09	0.04	0.05
Graduate Degree Held	0.80	0.83	0.78	0.83	0.77
Yrs. Administrative Exp	5.75	8.17	4.99	6.10	7.10
Teachers, N=4,120	42				
Female	0.86	0.88	0.84	0.86	0.88
Hispanic	0.09	0.08	0.07	0.09	0.14
African-American	0.22	0.14	0.34	0.21	0.17
Asian	0.05	0.01	0.07	0.04	0.05
Graduate Degree Held	0.58	0.57	0.56	0.55	0.58
Subject Specialists	0.44	0.45	0.45	0.42	0.42
Yrs. of Experience	12.96	12.95	12.34	12.10	14.85

Data Collection in SII

Researchers conducting the Study of Instructional Improvement used a variety of data collection instruments to examine CSR implementation processes in these schools. For example, SII researchers distributed annual surveys to school leaders and teachers over a three-year time period (AY 1999–AY 2002) asking respondents to report on their work activities and perceptions of CSR design and implementation, and on many different aspects of school culture, climate, and organization. In addition, teachers completed detailed instructional logs at different points in the study that were used to develop measures of classroom instruction occurring in schools. The survey instruments used in the study can be found on the study's website (www.sii.soe.umich.edu). The measures used here are also described in more detail in various SII publications (Camburn, Rowan, & Taylor, 2003; Correnti, Rowan, & Camburn, 2004; Rowan, Camburn, & Correnti, in press; Rowan, Harrison, & Hayes, in press).

SII researchers also have conducted intensive case studies of 12 schools from the larger survey sample (three schools implementing each of the CSR programs under study, and three additional "comparison" schools). These case studies are derived from extensive, face-to-face interviews with district and school administrators, with CSR leaders working in the schools under study, and with teachers. The purpose of this work was to learn as much as possible about implementation dynamics in case study schools. In addition, SII researchers have been observing classroom instructional practices among a selected subset of teachers in each case study site to learn more about patterns of instructional change in the schools under study.

The CRS Programs Under Study

An important feature of SII's research design is that it examines implementation processes and outcomes in schools working with three very different CSR programs. As discussed below, the three programs studied here differed along several dimensions, including: (1) the number and kinds of changes they wanted to make in schools; (2) the nature and extent of instructional guidance they gave to teachers; (3) the number and type of school leaders they appointed and the roles they asked these leaders to play in schools; and (4) the kinds of school organization and climate they sought to foster.

In order to capture these differences succinctly, SII researchers developed a conceptual language grounded in organization theory. In essence, this conceptual scheme views CSR programs as systems for controlling activities within schools. Using this conceptual language, we argue that the Accelerated Schools Program uses a system of "cultural control" to produce instructional change in schools, that America's Choice uses a model of "professional control" to produce change, and that Success for All uses a model of "bureaucratic control" to promote change in schools. A further description of these ideas is presented next.

We begin by discussing the Accelerated Schools Program, which we view as using "cultural" controls to secure instructional improvement. ASP's approach to producing school change, as we shall see, revolves around promoting a normative commitment among school leaders and faculty to the program's vision or ideal of "powerful learning" for all students. From the outset, ASP linking agents use the staff development process to emphasize the program's commitment to this construct, and to define powerful learning as constructivist in nature, with an emphasis on authentic, learner-centered, and interactive forms of instruction. But ASP is not prescriptive in nature. The program does not target particular school subjects for improvement, nor does it provide teachers with a great deal of explicit guidance about curriculum objectives or teaching strategies. Instead, ASP linking agents help schools use a systematic process of organizational development to uncover a unique path toward powerful learning and to adopt the locally appropriate forms of instructional practice consistent with this approach. Moreover, classroom teachers play a key role in this process by working inside their classrooms to develop new teaching practices consistent with the ideal of powerful learning for all students.

America's Choice presents a contrasting approach, using what we call "professional controls" to stimulate instructional improvement. This program has its origins in the standards-based reform movement, and as a result, the program is built around some definite ideas about the curricular content that should be taught in schools and about methods of teaching inside classrooms, especially in the area of language arts. At the time of our study, for example, AC typically began its work in local schools by focusing on the school's writing program (moving only later to changes in reading and mathematics programs). Unlike ASP, however, AC typically provided teachers with a great deal of instructional guidance. For example, teachers in AC schools received a curriculum guide, were taught a set of recommended instructional routines for teaching writing (called "writers' workshop"), and worked with locally appointed AC coaches and facilitators to develop "core writing assignments" and clear scoring "rubrics" for judging students' written work. Thus, in the area of writing instruction at least, AC was trying to implement a well-specified, standards-based curriculum grounded in professional consensus about what constitutes a desirable instructional program.

AC also employed a sophisticated approach to professional development to spur instructional change in schools. Schools that adopted the program, for example, were expected to create two new leadership positions—a design coach and a literacy coordinator. Design coaches were expected to help principals plan school improvement activities, to organize opportunities for faculty to score and analyze students' written work, and to develop exemplary writing assignments for use in classrooms. AC literacy coordinators, by contrast, were expected to work with classroom teachers. Their role was to model AC-endorsed instructional practices and to observe and critique teachers as they learned to use these new practices. Of all the CSR programs under study, then, AC placed the most importance on expert coaching as a means to improving teachers' instructional work.

Success for All provides yet a third model for promoting instructional change in schools. Of the three programs under study, it gave schools the clearest and most highly-specified plan of instructional improvement for implementing a set of highly-specified instructional routines for the teaching of reading. In particular, the SFA program is built around a clear and well-defined reading curriculum. The program also provides teachers with a weekly lesson sequence, and within this sequence, each lesson is designed around a "script" intended to guide teaching activities through a 90-minute reading period. Moreover, in grades K-2, SFA provides schools with a set of curricular materials for use throughout the school. In this sense, the SFA reading program is highly routinized.

SFA schools also were more centrally managed than other schools in our study. For example, SFA schools were expected to appoint a full-time literacy coordinator, and this staff member was given substantial responsibility for schoolwide coordination of the reading program, including the task of constituting reading groups and making teaching assignments to these groups on a schoolwide basis every eight weeks. Moreover, instructional leaders in SFA schools and SFA linking agents were asked to supervise implementation of SFA instructional routines. Thus, while the SFA program includes a set of professional development workshops for teachers and leaders (that are roughly equivalent in terms of allocated time to the staff development provided to teachers and leaders by the other CSR programs under study), and while SFA leaders engage in modeling and coaching activities (much like AC leaders do), SFA is unique among the programs studied here in its emphasis on faithful implementation of clearly defined instructional routines in all classrooms in a school.

How Change Strategies Affect Program Implementation

Data from the *Study of Instructional Improvement* demonstrate that the different approaches to school improvement taken by the programs under study shape the processes of CSR implementation in schools. As a result, we turn now to a description of CSR implementation processes in the 116 schools in the SII sample, and how these processes differed across schools participating in the three CSR programs under study. The findings in this section come from published reports (see, for example, Camburn, Rowan, & Taylor, 2003; Correnti, Rowan, & Camburn, 2004; Rowan, Camburn, & Correnti, in press; and Rowan, Miller, Correnti, & Camburn, 2004) and from unpublished analyses conducted by SII researchers. Readers interested in more detail on the findings presented below can contact the senior author.

CSR Implementation in ASP Schools

As discussed above, ASP is viewed in our conceptual framework as using "cultural" controls to stimulate instructional improvement. Following the ideas of noted organization theorist William Ouchi (1980), we predicted that schools using this form of control would be characterized by a decentralized form of governance that granted staff members a great deal of discretion in their work activities, so long as that work was perceived as consistent with ASP's cultural ideal of "powerful learning." Put differently, organizations guided by cultural controls achieve unity of purpose and overall coordination, not through explicit monitoring and supervision of work activities, but rather by generating a sense of trust within the organization that all members are working on their own initiative toward achievement of the same, overarching ideal.

The use of cultural controls to stimulate instructional improvement can have both strengths and weaknesses. A major strength of the approach is that schools governed by cultural controls should be able to adapt easily and quickly to conditions in their external environments (Meyer & Rowan, 1978). For example, the fact that ASP's commitment to powerful learning does not include detailed guidance about curriculum and instruction should make it easier for local schools to "align" or "fit" their instructional work to state and district standards-based education reforms. Moreover, ASP's emphasis on teacher initiative and innovation as a means to instructional improvement (rather than administrative supervision) should increase the faculty commitment to school improvement efforts, especially since self-direction is a powerful motivating force within organizations. So, the emphasis on local discovery and initiative appear at first glance to be strengths of the ASP approach to comprehensive school reform.

However, there are possible weaknesses in ASP's approach as well. For one, there is no guarantee that ASP schools will develop a coherent focus on improving the core curriculum areas of reading and mathematics, for in ASP, schools can pursue any number of targets for instructional change. In addition, the lack of strong direction from ASP about specific targets for instructional reform, when coupled with ASP's emphasis on individual initiative and innovation, might lead to what has been called a "Christmas tree" approach to school improvement, where schools adopt many different—and not particularly coherent—school improvement initiatives.

SII survey and case study data confirmed these predictions. For example, the survey data showed that ASP schools were less likely to focus on improving reading and mathematics teaching than other schools in the SII sample; conversely, ASP schools were more likely than other schools in the sample to focus on improvements in other academic areas, to focus on improvements in school climate, and to strive for improved parent participation. Overall, in fact, SII survey data showed that ASP schools had more targets of school improvement than did the other schools in our sample, and that leaders in ASP schools were more worried than

leaders in other schools about the possibility of undertaking too many school improvement initiatives.

SII data also showed that ASP schools developed a distinctive form of professional culture, one that emphasized faculty innovation within a climate characterized by trust among colleagues. But these positive elements of the culture were accompanied by what appeared to be a "hands off" approach to instructional management in which teachers were given more latitude to enact distinctive instructional preferences than were teachers in the other schools in our study. That is not to say that ASP schools lacked direction in their improvement efforts. For example, staff at ASP schools did report more consensus about expectations for student learning than did staff in other schools in the study. But this consensus functioned as the primary means of instructional coordination in ASP schools, rather than more explicit coordination of instruction through development of clear and formal curricular standards, the use of mastery tests to regulate students' matriculation through the curriculum, or other formal means of promoting instructional coordination across classrooms and grade levels.

This lack of formal coordination was also reflected in the instructional leadership styles of ASP school leaders. Like the other CSR programs in SII, ASP called for the creation of additional leadership positions inside schools; in particular, the appointment of an ASP facilitator. But survey data reported by Camburn, Rowan, and Taylor (2003) showed that the addition of this new staff position rarely led to an overall increase in the total number of leaders in ASP (vs. comparison) schools. Moreover, the survey data suggested that principals, assistant principals, ASP facilitators, and other ASP school leaders gave less emphasis to instructional leadership than did leaders in all the other groups of schools in our sample. Our case study suggested that this "hands off" approach to leadership also extended to ASP's external linking agents. In two of our three case study sites, for example, faculty expressed dissatisfaction with how ASP linking agents performed their liaison role, and in all three of our case study sites, respondents reported only infrequent interactions with these linking agents.

Given the emphasis in ASP schools on individual initiative, and in light of the "hands off" approach to instructional leadership found in ASP schools, it is not surprising that both our case study data and our survey data showed ASP schools to be characterized by diverse (rather than uniform) teaching practices. For example, the case study data suggest that, lacking explicit instructional guidance from ASP program staff, teachers in ASP schools looked to state and district staff for instructional guidance. ASP teachers also reported that textbooks were a major influence on their teaching practice. Moreover, data from SII's teacher logs showed that less than half (44%) of the ASP teachers in our sample displayed a pattern of instruction that was distinctively ASP in form, with only 4 of the 28 schools in our sample providing log data having more than 70% of teachers thus classified. Indeed, both the case study and the survey data suggested that teachers in ASP schools tended to adopt a form of "bricolage" in which information and ideas were bor-

rowed from many different sources in order to construct a personalized pattern of teaching. For this reason, when we used SII log data to sort teachers into distinctive groups based on the proximity of their teaching practices to those endorsed by the different CSR programs in our study, ASP teachers ended up being classified into a variety of groups—including many ASP teachers whose instruction more closely resembled instructional forms endorsed by SFA or AC than ASP. All of this is consistent with ASP's emphasis on individual initiative, its lack of explicit instructional guidance, and the relatively "hands off" approach to instructional leadership that characterized this program's approach to stimulating instructional improvement.

CSR Implementation in America's Choice Schools

America's Choice demonstrates how an alternative strategy for promoting instructional improvement affects the CSR process. This program pursued what we have been calling a model of "professional" control in its work with schools. Firmly grounded in a set of professional standards for curriculum, teaching, and learning (especially in the area of literacy instruction), AC placed strong emphasis on having "expert" school leaders support model implementation though coaching inside classrooms. The strengths of this model are several. For one, teachers are given substantial curricular and instructional guidance—including curriculum guides, assessment exams and scoring rubrics for judging students' work, and exemplary lesson assignments for use in classrooms. In addition, the program's design calls for extensive coaching by principals, design coaches, and the AC literacy facilitator. But there is a catch in all of this explicit guidance. The professional culture in many U.S. schools grants teachers substantial discretion, allowing them to pursue instructional practices of their choice, all of which are supported by strong norms of privacy. AC's design for school improvement works against these ingrained tendencies in schools' professional cultures, and for it to work well, new norms supporting collaboration among teachers and leaders and the emergence of a critical discourse about teaching might be required for the coaching and support provided by AC leaders to serve as a stimulant for changes in classroom teaching.

The AC design has other salient features that could have a bearing on CSR implementation. AC's instructional design—which calls for changes in schools literacy and mathematics programs—is extraordinarily ambitious. For example, the program's literacy component requires 120 minutes of sustained language arts teaching at the elementary grades, about 30 minutes more than schools typically offer. Moreover, the literacy program requires many complex changes in teaching practice—the use of a writer's workshop with sustained writing assignments for students, the development of new rubrics for judging the quality of students' writing, the use of "leveled" texts and guided reading practices for reading instruction, and an emphasis on reform-oriented mathematics. Given the changes required in teaching, AC is careful to roll out new components slowly (e.g., at the time of our

study, two years were spent rolling out the writing program, and only then was the reading program phased in). So, AC's agenda for change is ambitious and complex, although carefully staged.

Another salient feature of the AC program is its clear set of academic standards. While this kind of clarity has many advantages, the "fit" or alignment of AC curricular goals and teaching practices to state and district standards could be a problem. However, the challenges of aligning AC's curricular and teaching standards to the external environment is balanced by the advantages the program derives from this clarity. Because AC is very clear about the curriculum it wants taught, and because it provides substantial support to help teachers teach this curriculum, there are many reasons to expect that AC will produce faithful implementation of its program across varied school contexts.

SII data confirm the strengths and weaknesses of AC's approach to school improvement. On a positive note, teachers and leaders in AC schools reported more clarity about the pace and direction of school improvement planning than did teachers in ASP and comparison schools (although about the same level as reported by personnel in SFA schools). Also, school leaders reported that school improvement plans in AC schools were focused more squarely on making improvements in the area of reading than did leaders in ASP and comparison schools (again, the clarity was about the same in SFA schools). Overall, then, the planning process in AC schools seemed to be characterized by high levels of clarity and focus.

AC schools also were characterized by strong instructional leadership. For example, the AC schools in our sample typically had about the same number of instructional leaders per teacher as SFA schools (about five teachers for every leader in AC schools) which was less than the number of teachers per leader in ASP and control schools (about nine teachers per leader). Moreover, all of the leaders in AC schools (including the principal), received extensive staff development on how to enact their role as instructional leaders. As a result, AC leaders were more likely than leaders in ASP and comparison schools (but not SFA schools) to report working directly with teachers, to report providing professional development to teachers directly, and to report placing an emphasis on monitoring improvement efforts. Teachers' survey reports reflected these high levels of instructional leadership. AC teachers were more likely than ASP and comparison teachers (but not SFA teachers) to report observing or being observed by a school leader to improve their teaching, and our case study data contained many reports from teachers remarking on the helpful support AC school leaders gave them in changing their classroom practice.

The emphasis on active instructional leadership was consistent with the professional culture inside AC schools. AC leaders and teachers reported a higher press to standardize instructional practices in their schools than did teachers and leaders in ASP and comparison schools (but not SFA schools), and they reported lower levels of teacher autonomy. But all of this appears to have diminished at least some of the established patterns of collaboration within AC schools. For example, AC teachers were less likely than teachers in ASP and comparison schools (but not SFA schools)

to report that there was support for innovation in their schools, and they were less likely to see their schools as characterized by strong norms of collaboration. These findings probably reflect the emergence of a different form of professional culture in these schools—one built around real standards of best practice and the emergence of expert leaders in schools, as opposed to the more typical norms of autonomy, discretion, and trust upon which conventional faculty cultures are more typically based.

Most importantly, both the case study data and our instructional logs strongly suggest that AC's press to implement specific teaching strategies worked well and produced significant change. However, this occurred mostly in the area of writing instruction, which was the focus of AC implementation for the first three years of the program in most schools in our sample. For example, data from the instructional logs showed that 73% of all AC teachers in our sample used instructional practices that looked more like AC instruction than the instruction characteristic of other programs in our study. Moreover, these high levels of implementation occurred in the majority of AC schools. For example, in 12 of 30 AC schools in our sample with log data, 70% or more of the teachers were classified as AC teachers, and in five of these, 100% of the teachers were thus classified.

Still, our case study data suggest that AC schools had some problems in implementation. Teachers typically were eager to implement AC's writing program and were frequently impressed by its results with students. But the amount of intensive staff development required to get this program component implemented, as well as the amount of time it took school leaders to work with all teachers in a school on an individual basis, worked against getting the reading and math components of the AC program implemented. Associated with this was the fact that teachers devoted more time to writing instruction than is typical in elementary schools. As a result of the emphasis on writing, AC teachers in our case study sites reported giving less attention to implementing AC's reading practices, feeling that they had little time for improvements in this area and because many teachers reported that staff development in reading was not as strong as it was in the area of writing.

Still, the AC program demonstrates quite clearly that it *is* possible to use the CSR process to produce real changes in classroom instruction, and that such changes can involve implementation of complex and ambitious forms of pedagogy. But, as the AC case also demonstrates, ambitious change is not easy. In AC schools, for example, successful change resulted from a clear plan for improvement, intensive staff development involving a great deal of face-to-face modeling, coaching, and support from school leaders, and a real commitment on the part of AC leaders to support the AC instructional model.

CSR Implementation in SFA Schools

SFA follows yet a third approach to comprehensive school reform, using what we call a "bureaucratic" approach to stimulating instructional improvement. The

defining attribute of SFA is its reading program, which is built around a set of lesson scripts that structure the content, pacing, discourse, and instructional activities in classrooms at all grade levels of an SFA school. In addition to this scripting of instruction, SFA also relies on centralized instructional management practices more than the other CSR programs in our study. Along these lines, for example, SFA reading facilitators are given authority to group and regroup students for reading on a schoolwide basis every eight weeks, and SFA school leaders and linking agents are encouraged by the program's central staff to monitor local activities to assure and stimulate faithful program implementation.

The strengths of SFA's "bureaucratic" model are many. The clear instructional guidance it provides to teachers, the emphasis it places on monitoring for faithful implementation, and the professional development received by school leaders and classroom teachers all focus CSR efforts around a clear target (reading) and should produce strong implementation outcomes. But there is a possible tradeoff in all of this. As we have seen, classroom teachers in American schools are used to substantial autonomy, and typically see it as confirming their expertise and professionalism. Thus, SFA's emphasis on following routines, and on administrative monitoring and supervision, could be a drawback for the program, especially if the consequent reduction in teacher autonomy leads to faculty resistance or decreases teachers' motivation to enact the program. Moreover, the very clarity of the program—which is a strength in terms of encouraging faithful implementation—could produce problems of "fit" to the local setting, especially if elements of the SFA reading program such as its curricular objectives, mandated texts, or prescribed teaching practices are inconsistent with state or district preferences in these areas.

Data from the *Study of Instructional Improvement* confirm these strengths and weaknesses in SFA's approach to school improvement. On a positive note, teachers and leaders in SFA schools reported that their school improvement plans were more focused on improvements in the area of reading than did teachers in ASP and control schools (but not AC schools)—confirmation that there was a real focus of improvement activities in schools adopting SFA. Moreover, SFA teachers and leaders reported that there was more clarity about the pace and direction of school improvement planning in their schools than did teachers in ASP and comparison schools (but not in AC schools). Overall, then, the planning process in SFA schools was characterized by high levels of clarity and focus.

SFA schools also were characterized by high levels of instructional leadership. As with AC schools, SFA schools in our sample had more instructional leaders per teacher than both ASP and comparison schools (about five teachers for every leader in SFA schools, versus about nine teachers per leader in ASP and comparison sites). And leaders in SFA schools were more likely than leaders in ASP and comparison schools to place a strong emphasis on monitoring improvement efforts in their schools. Furthermore, teachers' reports reflected this leadership emphasis. Just as in AC schools, SFA teachers reported being more likely to observe or be observed by a school leader as part of their improvement agenda, and case study data contained

many reports from teachers remarking on the frequency with which SFA leaders checked to see that SFA instructional routines were being followed and on the helpfulness of the feedback provided by these leaders.

SFA schools also appeared to be characterized by a distinctive professional culture. For example, teacher leaders and administrators in SFA schools were more likely to report that instruction was standardized across classrooms and grades than were leaders in other schools, and they also reported giving teachers less autonomy. In line with this, SFA teachers reported lower levels of collaboration in their schools than did ASP and control group teachers (although not AC schools), perceived less support in their schools for innovation and risk taking, and showed declining motivation to implement SFA instructional routines as time went on. Despite this, teachers in our SFA case study schools often reported that the changes they were making were successful and were leading to striking improvements in students' achievement. Still, teachers also sensed that participation in the SFA program was eroding their autonomy, and thus there was a consequent decrease in motivation to implement the program over time.

Still, both the case study data and our instructional logs showed that SFA's approach to standardizing instructional practices was working well in most schools. For example, data from the instructional logs showed that 83% of SFA teachers in our sample used instructional practices that looked more like SFA instruction than instruction characteristic of other programs in our study. Moreover, there were high levels of implementation in a large percentage of SFA schools. For example, in 16 of 28 SFA schools in our sample that provided log data, more than 70% of the teachers were classified as SFA teachers in our analyses. Thus, like the AC program, SFA demonstrates that it *is* possible to use the CSR process to produce real changes in classroom instruction, and that such changes can produce consistent forms of pedagogy. But, in the SFA case, this change appears to have been produced not only as a result of a clear plan for improvement and systematic staff development involving face-to-face modeling and coaching by school leaders, but also by sustained and vigorous efforts to monitor and supervise implementation using a bureaucratic model of control.

Contextual Effects on Implementation Processes in SII Schools

To this point, we have been discussing the effects of different CSR models on implementation in schools. We have seen that the different CSR programs under study are built around different instructional models, that they use different approaches to getting these instructional models implemented in classrooms, and that the different strategies used by the programs produced both different professional climates and different patterns of instructional practice in the schools under study. These effects, we should note, are not small (in statistical terms). Indeed, the standardized effect sizes of participation in a particular CSR pro-

gram on SII measures of professional climate were mostly in the range of a tenth to a third of a standard deviation, while standardized effect sizes of participation in one of the CDSR models on SII instructional measures were almost always in the range of a third to a half of a standard deviation. By the usual standards of social science research, then, program-specific effects on CSR implementation outcomes are strong.

However, SII survey and case study data also confirm what many previous studies of CSR implementation have found—that the process of planned educational change proceeds variably, even among schools working with the same CSR program. In this section, then, we turn to an examination of sources of variation (other than program design) that account for different implementation outcomes among schools in the SII sample. Following past research, we discuss how characteristics of districts and schools work in conjunction with CSR program designs to affect implementation outcomes.

District Characteristics and CSR Implementation

We begin with a discussion of the role that district contexts play in CSR implementation. Consistent with much previous research, SII survey data suggest that larger school districts present schools with a unique constellation of conditions that shape the CSR planning and implementation process. For example, SII schools in larger districts were generally located in inner city neighborhoods and served larger percentages of poor and minority students. In addition, these schools were generally housed in older facilities, and had higher enrollments than schools in smaller districts. One consequence of these characteristics was that schools in larger districts faced some special problems above and beyond the pressing need to improve reading and math instruction. For example, in surveys, leaders in these schools were more likely than leaders in other schools to report that their school improvement plans were focused on improving students' health, welfare, and attendance, on upgrading facilities (including facilities for instructional media), and on making improvements to school climate generally. Thus, it appears that schools in larger districts had many needs, and a more complex school improvement agenda.

In addition to this larger school improvement agenda, schools in larger districts also faced more complex curricular and instructional policy environments. Teachers in larger districts, for example, were more likely than teachers in smaller districts to report that policies about teaching and curriculum were inconsistent and that they were unsure about which policies to follow. One reason this occurred was that schools in larger districts were more likely to operate under centralized instructional guidance from districts, and to be under state-level scrutiny for achievement performance, all of which led to districts to initiate district-wide improvement initiatives that changed frequently and/or did not necessarily "fit" with the specific CSR initiatives being undertaken in schools. In one of our case study sites, for example, a CSR initiative that teachers were enthusiastic about was

dropped, largely because a district initiative supplanted the CSR effort and was mandated centrally. It is important to note, however, that the different CSR programs participating in SII were differentially affected by these problems. For example, our case study data suggest that schools working with ASP more easily adapted to district and state improvement directives, largely because the lack of instructional and curricular definition in the ASP design allowed schools to more easily build district-wide agendas into their own CSR efforts.

Although much previous research on CSR describes conflicts between districts and CSR schools, SII survey data suggest that the CSR processes often were accompanied by many positives in larger districts. For example, schools in larger districts often scored above schools located in smaller districts on SII measures of instructional leadership and coordination, and more importantly, there was an increase over time on most SII measures of professional culture in these schools. Thus, while schools in larger districts often started out lower than other schools in terms of faculty trust, academic press, and motivation to implement a CSR program, these same schools showed increases on these SII climate measures over time that exceeded those found in schools located in smaller districts. So, while the CSR process unfolds in a more complex environment in schools located in larger districts, and while these schools often begin the CSR process with conditions that are (in theory) less favorable to educational change, the evidence from SII suggests that district size and complexity are not an inevitable barrier to successful school change.

School Conditions Affecting Change

SII data also show that schools within the same district often vary in both the nature and success of their CSR efforts. This, of course, is a common finding in the CSR literature (see especially Berends, Bodilly, & Kirby, 2002) and results from the fact that the CSR process involves a complex set of interactions among school leaders, school faculty, and CSR linking agents. SII case study data suggest that all of these constituencies are important to the success of CSR efforts. For example, case study data show that CSR linking agents working in different schools can be perceived as knowledgeable and helpful (or not), and that even within the same school, different teachers view linking agents differently. Equally important, there are real differences in the professional background, preparation, and work activities of local school leaders. For example, SII case study and survey data show that principals are an important source of leadership for CSR, but that principals vary greatly in the ways in which they lead CSR efforts. Some principals, for example, are firmly behind a school's CSR efforts, but other principals (often new to a school), are more agnostic about the CSR program being implemented in their schools and thus less supportive. In addition, others occupying leadership roles within CSR schools (e.g., coaches, facilitators, and other teacher leaders) vary in the extent to which they emphasize instructional leadership. So, principals and other leaders make a real difference to the CSR process as well. Finally, SII case study and survey

data show that teachers play a central role in CSR success. But even inside the same school, teachers differ in their enthusiasm for and participation in CSR-related staff development, and in their capacity to implement program-specific changes in their teaching.

All of these actors must work together in very complex ways to produce successful CSR outcomes in a school. However, because of the number of actors inside a school (including the many teachers, school leaders, and CSR linking agents), and because of the varied professional backgrounds, experiences, and pre-dispositions of these many actors, it is often difficult to predict a priori just how well the CSR process will unfold in a given school. Thus, SII data suggest that strong instructional leadership (by the principal and others) is a necessary condition for implementation success, but SII also data suggest that leadership alone is insufficient to produce real change in teaching practices. For example, a study by Taylor (2004) using SII survey data found that strong leadership enhances teachers' motivation to participate in CSR efforts only when a school's professional culture is characterized by stronger norms of collegial support, innovation, trust, and collegiality. But even more importantly, SII log data show that teacher motivation of this sort is insufficient to produce real change in teaching practices. Instead, changes in teaching practice result from a complex process of professional learning that is produced by teachers' participation in staff development workshops, by the assistance and support teachers receive from school leaders, and by the individual capacities and dispositions of teachers themselves.

Given the complexity of schools as organizations, the various environments in which they operate, and the diversity of people who work in them, it is perhaps unsurprising to learn that schools vary in the extent to which they succeed in the process of educational change. Moreover, some of this variation among schools in implementation outcomes is a result of the fact that they work with differently designed CSR programs, some is the result of the state and district environments in which the schools are located, and some is due to the features of the schools themselves and the people who work in them. But none of this should obscure what is perhaps the important finding from SII data—that the percentage of schools in which participation in the CSR process is producing real changes is considerable. Thus, the point to take away from our discussion of context effects on the CSR process is this: CSR is more likely to lead to real changes in school organization, climate, and new patterns of instructional practice, when many circumstances converge. These include the presence of a CSR program with a well-specified instructional design, and a programmatic emphasis on instructional leadership, intensive staff development, and constant monitoring of implementation. To be sure, the contexts in which these conditions emerge affects the extent to which programs get implemented, but SII data show that well-designed programs that offer a clear instructional design and extensive support for change have a very strong chance of being implemented by a majority of teachers in a school—regardless of context.

Conclusion

We conclude this chapter with a brief review of the major lessons to be drawn about planned educational change derived from our review of research on CSR implementation in schools. This review demonstrated quite clearly that the process of comprehensive school reform can work to produce robust changes in instructional practices in a large number of schools. However, the chapter also showed that successful CSR depends to a considerable extent on the actions taken by: (a) external providers of design-based, technical assistance; (b) local school personnel; and (c) district personnel who provide support to local school change efforts. In particular, our chapter demonstrated that the process of CSR was most successful when external change agents worked to produce clear, specific, and high-quality designs for change and provided extensive implementation support to local schools; when local school communities coalesced around the central aims of the research-based model of school reform they were trying to implement and actively learned over a period of years how to utilize that model in their own context; and when district personnel provided a stable and supportive policy environment clearly aligned with the aims of the practices being developed. Thus, the process of CSR is a complex, cooperative process, involving multiple agencies and actors. However, when these agencies and actors work together to support the change process in local schools through careful planning, provision of implementation support, and through persistent efforts at change, the evidence shows quite clearly that the CSR process can lead to important changes in school organization and culture and to fundamental changes in teaching practice.

Postscript

Work on this paper was supported by grants from the U.S. Department of Education to the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania (Grant # OERI-R308A60003), the National Science Foundation's Interagency Educational Research Initiative to the University of Michigan (Grant #s REC-9979863 & REC-0129421), the William and Flora Hewlett Foundation, and the Atlantic Philanthropies. Opinions expressed in this paper are those of the authors, and do not reflect the views of the U.S. Department of Education, the National Science Foundation, the William and Flora Hewlett Foundation, or the Atlantic Philanthropies. We thank Charles Vanover, Beth Sanders, Robert Miller, and Richard Correnti for research assistance.

References

Anderson, L. W., & Shirley, J. R. (1995). High school principals and reform: Lessons learned from a statewide study of Project Re: Learning. *Educational Administration Quarterly, 31*(3), 405–423.

Berends, M. (2000). Teacher-reported effects of New American Schools design: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis, 22*(1), 65–82.

Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the challenges of whole-school reform. New American schools after a decade*. Santa Monica, CA: RAND.

Berman, P., & McLaughlin, M. W. (1975). *Federal programs supporting educational change: Volume 4. The findings in review.* Santa Monica, CA: RAND.

Berman, P., & Pauly, E. W. (1975). *Federal programs supporting educational change: Volume 2. Factors affecting change agent projects*. Santa Monica, CA: RAND.

Bodilly, S., Purnell, S., Ramsey, K., & Keith, S. (1996). *Lessons from New American Schools Development Corporation's demonstration phase*. Santa Monica, CA: RAND.

Bodilly, S. J., Keltner, B., Purnell, S.W., Reichardt, R. E., & Schuyler, G. L. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: RAND.

Bodilly, S. J., & Berends, M. (1999). Necessary district support for comprehensive school reform. In G. Orfield and E.H. DeBray (Eds.), *Hard work for good schools: Facts, not fads, in Title I reform* (pp. 111–119). Cambridge, MA: Harvard Civil Rights Project.

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

Camburn, E., Rowan, B., &, Taylor, J. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis, 25*(4), 347–374.

Christenson, G. (1996). Toward a new leadership paradigm: Behaviors of Accelerated Schools principals. In C. Finnan, E. St. John, J. McCarthy, & S. Slovacek (Eds.), *Accelerated Schools in action: Lessons from the field* (pp. 185–207). Thousand Oaks, CA: Corwin Press.

Cook, T. D., Habib, F., Phillips, M., Setterson, E. A., Shagle, S. C., & Degirmencioglu, S. M. (1999). Comer's School Development Program in Prince George's County, Maryland: A theory–based evaluation. *American Educational Research Journal, 36*(3), 543–597.

Cooper, R., Slavin, R. E., & Madden, N. A. (1998). Success for All: Improving the quality of implementation of a whole–school reform effort through the use of national reform network. *Education and Urban Society, 30*(3), 385–408.

Correnti, R., Rowan, B., & Camburn, E. (2004). *Variation in 3rd grade literacy instruction and its relationship to student achievement among schools participating in comprehensive school reforms*. Unpublished report, University of Michigan at Ann Arbor.

Crandall, D. P., & Loucks, S. F. (1983). *A roadmap for school improvement: Executive summary of the study of dissemination efforts supporting school improvement. People, policies, and practices: Examining the chain of school improvement* (Volume 10). Andover, MA: The Network.

Crandall, D. P., Eiseman, J. W., & Louis, K. S. (1986). Strategic planning issues that on the success of school improvement efforts. *Educational Administration Quarterly, 22*(3), 21–53.

Cross, C. T. (2004). *Political education: National policy comes of age*. New York, NY: Teachers College Press.

Darling-Hammond, L., & Snyder, J. (1992). Curriculum studies and traditions of inquiry: The scientific tradition. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 41–77). New York, NY: MacMillan.

Datnow, A. (2000). Power and politics in the adoption of whole-school reform models. *Educational Evaluation and Policy Analysis, 22*(4), 357–374.

Datnow, A., & Castellano, M. (2000). Teachers' response to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal, 37*(3), 775–779.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk, 5*(1–2), 183–204.

Desimone, L. (2002). How can comprehensive school reform models be implemented? *Review of Educational Research, 72*(3), 433–480.

Dunn, W., Holzner, B., & Zaltman, G. (1990). Knowledge utilization. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 725–733). New York, NY: Pergamon Press.

Emrick, J. A., & Peterson, S. M. (1977). *A synthesis of findings across five recent studies in educational dissemination and change*. San Francisco, CA: Far West Laboratory for Educational Research and Development.

Firestone,W. A., & Corbett, H. D. (1988). Planned educational change. In N. J. Boyan (Ed.), *Handbook of Research on Educational Administration* (pp. 321–340). New York, NY: Longman.

Friedman, L. B. (1999). *Launching the Comprehensive School Reform Demonstration Act in six Midwest states: Implications for schools, districts, and model providers implementing the program*. Oak Brook, IL: North Central Regional Education Laboratory.

Fullan, M. G. (1991). *The new meaning of educational change*. New York, NY: Teachers College Press.

Glennan, T. K. (1998). *New American Schools after six years*. Santa Monica, CA: RAND.

Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process*. Albany, NY: State University of New York Press.

Hatch, T. (2000). What does it take to "go to scale"? Reflections on the promise and the perils of comprehensive school reform. *Journal of Education for Students Placed at Risk, 5*(4), 339–354.

Havelock, R. G. (1971). *Planning for innovation through dissemination and utilization of knowledge*. Ann Arbor, MI: Institute for Social Research.

Haynes, N.M. (1998). Lessons learned. *Journal of Education for Students Placed at Risk, 3*(4), 87–89.

Herman, R., Aladjem, D., O'Malley, A. O., McMahon, P., Quinones, S., Masem, E., Reeve, A., et al. (1999). *An educators' guide to schoolwide reform*. Arlington, VA: Education Research Services.

Hood, P. (2002). *Perspectives on knowledge utilization in education*. San Francisco, CA: WestEd.

Hord, S.M., & Huling-Austin, L. (1986). Effective curriculum implementation: Some promising new insights. *Elementary School Journal, 87*(1), 97–115.

Huberman, A. M. (1983). Recipes for busy kitchens. *Knowledge: Creation, Diffusion, Utilization, 4*(1), 478–510.

Huberman, A. M., & Miles, M. B. (1984). Rethinking the quest for school improvement: Some findings from the DESSI study. *Teachers College Record, 86*(1), 34–54.

Keeves, J. P. (1990). Knowledge diffusion in education. Knowledge utilization. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 715–723). New York, NY: Pergamon Press.

Levin, H. M. (1995). Learning from Accelerated Schools. In J. H. Block, S. T. Evertson, and T. G. Guskey (Eds.), *School improvement programs: A handbook for leaders* (pp. 267–288). New York, NY: Scholastic Press.

Louis, K. S., Rosenblum, S., & Molitor, J. (1981). *Strategies for knowledge use and school improvement: A summary*. Washington, DC: U.S. Department of Education.

Louis, K. S., & Miles, M. B. (1990). *Improving the urban high school: What works and why*. New York, NY: Teachers College Press.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experience in restructuring schools. *American Journal of Education, 106*(4), 532–575.

Meyer, J. W., & Rowan B. (1978). The structure of educational organizations. In M. W. Meyer and Associates, *Environments and organizations* (pp. 78–109). San Francisco, CA: Jossey-Bass.

Miles. M. B. (1992). *40 years of change in schools: Some personal reflections*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mirel, J. (1994). *The evolution of the New American Schools: From revolution to mainstream*. New York, NY: Fordham Foundation.

Muncey, D. E., & McQuillan, P. J. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the Coalition of Essential Schools*. New Haven, CT: Yale University Press.

Nunnery, J. A., Bol, L., Dietrich, A., Rich, L., Kelly, S., Hacker, D., et al. (1997). Teachers' initial reactions to their pre-implementation preparation and early restructuring experiences. *School Effectiveness and School Improvement, 8*(1), 72–94.

Odden, A. R. (2000). The costs of sustaining educational change through comprehensive school reform. *Phi Delta Kappan, 83*(6), 433–438.

Ouchi, W. G. (1980). Markets, bureaucracies, and clans. *Administrative Science Quarterly, 25*(3), 129–141.

Peurach, D., Glazer, J., & Gates, K. (2004). *Supporting instructional improvement: teacher learning in comprehensive school reform.* Unpublished report, University of Michigan at Ann Arbor.

Rivlin, A. M., & Timpane, P. M. (1975). *Planned variation in education: Should we give up or try harder?* Washington, DC: Brooking Institution.

Ross, S. M. (2000). *How to evaluate comprehensive school reform models*. Arlington, VA: New American Schools.

Ross, S., Henry, D., Phillipsen, L., Evans, K., Smith, L., & Buggey, T. (1997). Matching restructuring programs to schools: Selection, negotiation, and preparation. *School Effectiveness and School Improvement, 8*(1), 45–71.

Ross, S. M., Troutman, A., Horgan, O., Maxwell, S., Laitinen, R., & Lowther, D. (1997). The success of schools in implementing eight restructuring designs: A synthesis of first-year evaluation outcomes. *School Effectiveness and School Improvement, 8*(1), 95–124.

Rowan, B., Camburn, E., & Correnti, R. (in press). Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement. *Elementary School Journal*.

Rowan, B., Harrison, D. M., & Hayes, A. (in press). Using instructional logs to study mathematics curriculum and teaching in the early grades. *Elementary School Journal*.

Rowan, B., & Guthrie, L. F. (1989). The quality of Chapter 1 instruction: Results from a study of 24 schools. In R. Slavin (Ed.), *Effective Programs for Students At Risk*. Boston, MA: Allyn and Bacon.

Rowan, B., Miller, R. J., Correnti, R., & Camburn, E. (2004). *How school reforms work: Early results from A Study of Instructional Improvement.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Rutherford, W. L., Hord, S. M., Huling, L., & Hall, G. E. (1983). *Change facilitators: In search of understanding their role*. Austin, TX: Research and Development Center for Teacher Education.

Slavin, R. E. (1999). How Title I can become the engine of reform in America's schools. In G. Orfield and E. H. DeBray (Eds.), *Hard work for good schools: Facts, not fads, in Title I reform* (pp. 86–102). Cambridge, MA: Harvard Civil Rights Project.

Smith, L. J., Maxwell, S., Lowther, D., Hacker, D., Bol, L.,& Nunnery, J. (1997). Activities in schools and programs experiencing the most, and least, early implementation success. *School Effectiveness and School Improvement, 8*(1), 125–150.

Stringfield, S., Millsap, M., Yoder, N., Schaffer, E., Nesselrodt, P., Gamse, B., et al. (1997). *Special strategies for educating disadvantaged children. Urban and suburban/rural. Final Report.* Washington, DC: U.S. Department of Education.

Stringfield, S., Datnow, A., & Ross, S. M. (1998). *Scaling up restructuring in multicultural, multilingual contexts*. Santa Cruz, CA: Center for Research on Education, Diversity, and Excellence, University of California.

Stringfield, S. (2000). *Choosing success*. Washington, DC: American Federation of Teachers.

Taylor, J. (2004). *Distributed Instructional Leadership and School Capacity for Improvement.* Unpublished doctoral dissertation, University of Michigan, AnnArbor.

Timar, T. (1989). The politics of school restructuring. *Phi Delta Kappan, 79*(4), 165–175.

Wasley, P. A., Hampel, R. L., & Clark, R. N. (1997). The puzzle of whole-school change. *Phi Delta Kappan, 78*(9), 690–697.

Wong, K. K., & Meyer, S. J. (1998). Title I schoolwide programs: A synthesis of findings from recent evaluations. *Educational Evaluation and Policy Analysis, 20*(2), 115–136.

CHAPTER TWO

Comprehensive School Reform and Achievement: *A Meta-Analysis*

Geoffrey D. Borman
University of Wisconsin-Madison

Gina M. Hewes
Laura T. Overman
Johns Hopkins University

Shelly Brown
University of North Carolina, Greensboro

In this meta-analysis, we review the research on the achievement effects of the nationally disseminated and externally developed school improvement programs known as "whole-school" or "comprehensive" reforms. In addition to reviewing the overall achievement effects of comprehensive school reform (CSR), we study the specific effects of 29 of the most widely implemented models. We also assess how various CSR components, contextual factors, and methodological factors associated with the studies mediate the effects of CSR. We conclude that CSR is still an evolving field and that there are limitations on the overall quantity and quality of the research base. The overall effects of CSR, though, appear promising and the combined quantity, quality, and statistical significance of evidence from three of the models, in particular, set them apart from the rest. Whether evaluations are done by the developer or by third-party evaluators and whether these evaluators use one-group pre-post designs or control groups are especially important factors for understanding differences in CSR effects. Schools implementing CSR models for five years or more showed particularly strong effects, and the models showed similar benefits across schools of higher- and lower-poverty levels. A long-term commitment to research-proven educational reform is needed to establish a strong marketplace of scientifically based models capable of bringing comprehensive reform to the nation's schools.

Introduction

The latter half of the 20th century was marked by recurring efforts at school reform and improvement in the United States. Yet, as Slavin (1989) observed, this cycle of reforms—like a pendulum swing—has continued to move from one fad to another with little evidence of national progress. As each new reform is widely disseminated and implemented, the research follows closely behind, sometimes weighing in on the issue only after the schools have moved on to the next apparent innovation. Recent national reform and policy movements, though, may halt this frustrating cycle. Indeed, for the first time Congress and other educational policymakers are making some funding sources available to only those schools that implement educational reforms with high-quality evidence of effectiveness. Most notably, the Comprehensive School Reform Program (CSRP)—formerly known as the Comprehensive School Reform Demonstration (CSRD) program—provides grants to schools to adopt proven comprehensive reforms. With the recent proliferation of externally developed comprehensive school reform (CSR) models, the simultaneous growth in the CSR research base, and the significant public and private financial backing this reform movement has received, the potential for generating a national wave of research-based educational innovation has never been greater.

In addition to their focus on research-based solutions for school improvement, current CSR initiatives help reconcile the two most important recent educational reform movements in the United States. Since the 1980s, competing, and often contradictory, reforms have combined top-down, centralized efforts to improve

schools and teaching with efforts at decentralization and school-based management (Rowan, 1990). The general spirit of today's reform efforts continues to articulate top-down standards, which dictate much of the changes in the content of schooling, but fundamentally leaves the process of school change up to the discretion of local educators. The problem is that the complex educational changes demanded by current standards-based reform initiatives, combined with an increasingly heterogeneous student population largely composed of students whom schools have traditionally failed, have pushed the technology of schooling toward unprecedented levels of complexity. In many ways, expecting local educators to reinvent the process of educational reform, school by school, is both unrealistic and unfair. Externally developed CSR models provide a type of top-down direction for designing and supporting the process of school reform. In this case, though, the top-down direction is not in the form of distant legislative mandates, but is, in theory, tangible and accessible support for school change rooted in research and literally packaged and delivered to each school.

In this exhaustive meta-analysis, we review all known research on the achievement effects of the most widely implemented, externally developed school improvement programs known as "whole-school" or "comprehensive" reforms. In addition to reviewing the achievement effects of CSR as a general strategy, we synthesize research on the specific effects of the 29 most widely implemented CSR models.[1] In quantifying the overall and specific effects of CSR models, we also assess how the methodological and contextual factors associated with the studies of CSR differ. In addition, we identify common components across reforms, such as whether the model specifies and includes a particular curriculum, or whether it specifies and provides a plan for the ongoing professional development of teachers. Using these methodological, contextual, and programmatic factors as predictors of effect size, we assess how they may influence the estimates of the models' effects. The resulting information allows us to examine:

- the general effectiveness of the CSR strategy;
- the effects associated with specific CSR model components;

1 Initially, we had identified 33 reform models for possible inclusion in this meta-analysis. Four of the models, though, had no quantitative data on their achievement effects from which we could calculate effect size estimates. These four CSR models were Foxfire Fund, League of Professional Schools, QuEST, and Ventures Initiative and Focus System. The 29 models remaining in the analyses were: Accelerated Schools Project, America's Choice School Design, ATLAS Communities, Audrey Cohen College: Purpose Centered Education, Center for Effective Schools, Child Development Project, Coalition of Essential Schools, Community for Learning, Community Learning Centers, Co-nect Schools, Core Knowledge, Different Ways of Knowing, Direct Instruction, Edison Project, Expeditionary Learning Outward Bound, High Schools That Work, High/Scope Primary Grades Approach to Instruction, Integrated Thematic Instruction, MicroSociety, Modern Red Schoolhouse, Montessori, Onward to Excellence, Paideia, Roots & Wings, School Development Program, Success for All, Talent Development High Schools with Career Academies, The Learning Network, and Urban Learning Centers.

- the effects of each of the 29 CSR models; and
- the extent to which differences in the methodological and contextual features of the studies mediate the estimates of effects.

What is CSR?

The "scale up" of CSR designs is happening at an unprecedented rate, as evidenced by the growing number of externally developed school reform designs (e.g., Accelerated Schools, Core Knowledge, High Schools That Work, Success for All) being implemented in thousands of schools, serving millions of students throughout the United States. CSR is expanding rapidly because many models have established development and dissemination infrastructures for replicating and supporting implementations across numerous schools. In other words, the developers can transport their CSR models to schools across the U.S., help local educators understand the tenets of the reform, and teach them how to implement the school organization and classroom instruction that the model suggests. In every case, the developers provide some type of initial training or orientation to help educators to, at least, understand the underlying philosophy of the model. In many circumstances, though, replication also involves a more specific "blueprint" for implementing and sustaining the model. Highly specified models, for instance, often prescribe new curricular materials, new methods of instruction, alternative staffing configurations, and a series of ongoing professional development activities.

The Policy Context for CSR

In addition to the replicable nature of many of the models, expansion of CSR has been fueled by a series of recent national developments: the movement toward systemic and standards-based reform; the establishment of the New American Schools Development Corporation; new federal legislation allowing the use of Title I funds—the primary source of federal assistance to at-risk students from high-poverty schools since 1965—to support schoolwide educational programs in high-poverty schools; and the federal CSRP legislation that provides hundreds of millions of dollars to support the costs of adopting externally developed reform models. Only since the mid-1990s has the idea of schoolwide reform emerged as a prominent strategy for helping improve the outcomes of at-risk students from high-poverty schools. Before then, the school-based services funded through Title I, and other categorical programs for at-risk students, targeted only those students with the lowest test scores. As a result, the vast majority of schools used the funds to develop specialized pull-out programs that provided remedial services to the subgroups of students with the greatest academic needs (Borman, Wong, Hedges, & D'Agostino, 2001).

Instead of the seemingly piecemeal and uncoordinated categorical targeted assistance programs that had served Title I schools since the mid-1960s, a growing belief developed that at-risk students and high-poverty schools could be better

served by schoolwide reforms. This belief was encouraged by informed opinion (e.g., Rotberg, Harvey, & Warner, 1993), by general findings from the effective schools research tradition (Edmonds, 1979; Teddlie & Reynolds, 2000), and by the concept of systemic reform (e.g., Smith & O'Day, 1991), more than by specific groundbreaking empirical studies. Inspired by the emerging vision of standards-based reform, the 1994 reauthorization of Title I called on states to raise academic standards, to build the capacity of teachers and schools, to develop challenging new assessments, to ensure school and district accountability, to ensure the inclusion of all children, and to develop coordinated systemic reforms. The new legislation encouraged schoolwide initiatives rather than targeted programs for all schools where at least 50% of the students were poor. These sweeping changes began the transformation of Title I from a supplemental remedial program to the key driver of the standards-based, schoolwide reform movement (Borman, 2000a).

During the 1990s, Title I schoolwide projects proliferated across the country. In 1991, only 10% of the eligible Title I schools operated schoolwide programs, but by 1996, approximately 50% of the eligible Title I schools had implemented them (Wong & Meyer, 1998). Rather than implementing the characteristic Title I pullout programs, educators were granted the flexibility to invent and implement their own reforms designed to upgrade the whole school. A number of studies from the 1990s showed that, in the short-term, these schoolwide efforts did not produce compelling evidence of positive achievement effects and, for the most part, did not result in the desired reforms (Wong & Meyer, 1998, 2001). Also during the 1990s, a more general review indicated that site-based management reforms failed to affect student outcomes positively in large part because the schools failed to develop coherent statements of beliefs or models for guiding the work and decision making of the school (Murphy & Beck, 1995). These outcomes, combined with new evidence from the Congressionally mandated Prospects study of the modest overall impacts of Title I services (Borman, D'Agostino, Wong, & Hedges, 1998; Puma et al., 1997), suggested that federal policies for improving education for at-risk students from high-poverty schools were in need of further retooling.

At the same time, the growing research base on several externally developed school restructuring efforts, such as the Comer School Development Program (Comer, 1988; Haynes, Emmons, & Woodruff, 1998) and Success for All (Slavin, Madden, Dolan, Wasik, Ross, Smith, & Dianda, 1996; Slavin & Madden, 2001), seemed to indicate hope for a high-quality education for at-risk students. In addition, the companion study to the national Prospects evaluation of Title I, the Special Strategies Study (Stringfield et al., 1997), indicated that whole-school, externally developed programs funded by Title I appeared more likely to have positive impacts on academic achievements than either traditional Title I pullout programs or locally developed reforms.

Along with the growing policy and research support, in 1991 then-President George Bush announced the creation of a private-sector organization called the New American Schools Development Corporation (NAS), which was intended to support

the creation of "break the mold" whole-school restructuring models for the next century (Kearns & Anderson, 1996). Using a business model, NAS turned to the marketplace for proposals for new models of American schools that would enable all students to achieve world-class standards in core academic subjects, operate at costs comparable to current schools after start-up funding, and address all aspects of a school's operation. After receiving nearly 700 proposals in February 1992, NAS chose 11, and provided funds for a three-year program of development and testing. Since 1995, NAS has continued to focus on "scaling up" seven of the models to thousands of schools nationwide. Providing more than $150 million over the past decade in financial and technical assistance to the reform developers, NAS has helped create a market for CSR and has helped scale up the CSR movement.

In response to the promise of the externally developed programs disseminated by NAS and by other independent model developers, the U.S. Congress also has encouraged individual schools to implement "scientifically based" whole-school reforms and to seek the assistance of external groups in developing their school reform plans. In 1998, Congress initiated the CSRP, which encourages schools to develop comprehensive plans for implementing "scientifically based" strategies for school reform. Through a competitive process, CSRP awards a minimum of $50,000 per year for three years to qualifying schools. Since first authorizing CSRP in fiscal year 1998 and allocating a total of $145 million, Congress has steadily increased its support. In fiscal year 2002, allocations for CSRD equaled $310 million. This figure includes $235 million set aside specifically for Title I schools and $75 million available to any school wishing to apply through the Fund for the Improvement of Education. This initiative, combined with Title I's continuing focus on schoolwide change and the efforts of NAS and other independent developers, has led to the continuing expansion of externally developed CSR models.

Previous Reviews of CSR

To date, there have been five major practitioner-oriented reviews, or "catalogs," of CSR models (see Herman et al., 1999; Northwest Regional Educational Laboratory, 1998, 2000; Slavin & Fashola, 1998; Traub, 1999; Wang, Haertel, & Walberg, 1997). Due to the rapid expansion of the CSR movement and the CSR research base, though, these reviews are quickly growing outdated. Although the reviews, most notably the Herman et al., and Slavin and Fashola publications, have provided some appraisals of the effects of the various CSR models, none has offered a comprehensive, quantitative synthesis of the overall effects of CSR or of the effects of the various CSR models. Rather, as Stringfield (2000) suggested, these publications are akin to *Consumers' Reports* guides for education, offering information that is important for educators to consider when "shopping" for a reform model. The reviews typically contain summaries of the general attributes of the CSR models, appraisals of the level of support that is provided by the developers, the costs associated with implementing the models, and various ratings of the strengths of the

research supporting each CSR design.

In addition to these reviews of CSR models, there have been several recent articles critiquing the research supporting particular designs and CSR in general. Most notably, these criticisms have suggested that some CSR research may be tainted by the fact that the developers are often also the evaluators (Pogrow, 2000; Walberg & Greenberg, 1999). Another source of controversy involves whether the use of a quasi-experimental, untreated control group design is really preferable to an analysis of pretest-to-posttest gain scores across a large number of sites (Pogrow, 1998; Slavin, 1999). In a sense, this debate has pitted the greater reliability of a large number of gain-score analyses against the greater internal validity of a relatively small number of matched control-group designs when attempting to judge whether an educational intervention has produced "exemplary" effects on achievement. Despite the controversy and debate, no empirical data from CSR evaluations have been systematically brought to bear on either question.

Beyond these methodological considerations, studies and reviews of CSR and the process of school change have identified several common, substantive factors that have a bearing on the success or failure of externally developed reforms. First is the rather straightforward observation that the quality of the CSR model implementation matters. A number of researchers have demonstrated a strong relationship between reform implementation and positive effects—both qualitative and quantitative—across a variety of reforms (e.g., Berman & McLaughlin, 1978; Crandall et al., 1982; Datnow, Borman, & Stringfield, 2000; Stringfield et al., 1997).

Second, although some reform models have been criticized because their prescriptive designs may suppress teacher creativity and require an inordinate amount of preparation time (Datnow & Castellano, 2000), externally developed reforms that are more clearly defined tend to be implemented with greater fidelity and, in turn, tend to have stronger effects on teaching and learning than reforms that are less clearly defined (Bodilly, 1996, 1998; Nunnery, 1998). Third, well-implemented reforms tend to have strong professional development and training components and effective follow-up to address teachers' specific problems in implementing change within their classrooms (Muncey & McQuillan, 1996; Nunnery, 1998). Finally, for external models of school change to make an important impact within schools, teachers and administrators must support, "buy into," or even help "co-construct" the reform design (Borman et al., 2000; Datnow & Stringfield, 2000). Although there have been no systematic analyses across a wide range of CSR models, it would seem that those models with clear components addressing each of these issues would tend to result in more reliable implementations and stronger effects than CSR models without such components.

Further, the federal government has detailed 11 clear characteristics (see Chapter 1, page 10) of what it views as a truly comprehensive approach to reform. Not surprisingly, some of these overlap with the components identified in the CSR and school change research literature, including high-quality technical support from the external CSR partner, continuous teacher and staff development and

training, and staff support or "buy-in" for the reform initiative. The federal recommendations, though, cite several other characteristics that may be associated with effective CSR models, such as evidence that the reform has been replicated successfully, measurable goals for student performance and benchmarks for meeting those goals, and the involvement of parents and the community in the governance of the school and the development of the school improvement plan.

Objectives and Hypotheses

The results from studies of CSR differ in many ways, including (a) who reported the findings (i.e., the developer or someone else); (b) the methods used (e.g., pretest-posttest comparison, experimental comparison, or nonequivalent control-group design); (c) the student and school context (e.g., high-poverty versus lower poverty settings); (d) actual characteristics of the CSR models (e.g., the costs associated with the model, or the level of support for implementation provided by the developer); and (e) indicators of the model's effectiveness (e.g., test scores from reading, math, science, or some other subject). Differences across studies such as these are commonly found in the social sciences, especially in the case of education.

Indeed, given the programmatic, methodological, and contextual diversity of the CSR literature, questions emerge concerning how, or if, we should proceed with a synthesis of its findings. As Borman (2000b) pointed out, there are varying perspectives on what the reviewer should do when confronted by such a variegated literature, in terms of overall research quality and other features, such as the research designs, samples, and the actual circumstances involved. On one hand, Glass (1976) stated, "It is an empirical question whether relatively poorly designed studies give results significantly at variance with those of the best designed studies" (p. 4). On the other hand, Slavin (1986) argued, "Far more information is extracted from a large literature by clearly describing the best evidence on a topic than by using limited journal space to describe statistical analyses of the entire methodologically and substantively diverse literature" (p. 7). Should the researcher combine studies that used varying methods and are characterized by varying substantive characteristics or should one focus only on the "best evidence?"

We believe that there are two important reasons to begin our analysis with a review of the complete CSR literature. First, as Glass (1976) suggested, by empirically examining a diverse range of studies, we may assess how and to what extent methodological differences across the studies are associated with differences in CSR effects. When outcomes are robust across studies of varying methodologies, one can be more confident in the conclusions. On the other hand, if studies differ in terms of both rigor and results, then one may focus on the subset of more rigorous studies when formulating conclusions. This analysis of the consequences of methodological variations for the estimation of CSR effects, which is unique to the enterprise of meta-analysis, allows methodologists and consumers of the research literature to recognize the biases in the literature and to understand empirically

both their frequency and magnitude.

Second, from a practical perspective, relatively little is known about what common components characterize effective CSR models. Well-intended federal policies have outlined the elements that should comprise a school reform that is truly comprehensive. These policies, though, have not benefited greatly from the cumulative knowledge of the CSR research base. By examining how effects vary across models and contexts, it is our hope to provide new evidence of both how and where CSR may make the biggest difference in achievement. It also may suggest some components or specific models that do not appear to be affecting student outcomes in meaningful ways.

Our meta-analysis begins by assessing these methodological, programmatic, and contextual variations across an extensive collection of all known studies of 29 of the most widely discussed and disseminated CSR models. This preliminary analysis shows how and to what extent the methodological, programmatic, and contextual factors shape our understanding of the overall achievement effects of CSR. Specifically, the preliminary analysis empirically identifies and quantifies the potential methodological biases in the literature, reveals the common characteristics of CSR programs that make a difference in terms of achievement, explores differences in achievement effects associated with varying contexts (e.g., the grade level or the subject area targeted by the reform), and, in general, characterizes the overall quality of the research evidence.

After characterizing the overall CSR research base, and after empirically identifying its potential methodological biases, our second objective is to assess the efficacy of each of the 29 CSR models. Rather than surveying the overall CSR research base and the methodological, programmatic, and contextual factors within it, this phase of our research develops standards for assessing the quality, quantity, and statistical significance of the models' effects on achievement. In short, we establish the extent to which each of the 29 models is supported by scientifically based research. We address this concern by focusing on only the subgroup of studies that provides the best evidence for evaluating the effectiveness of each of the 29 CSR models. We determine which studies provide the best evidence not by a priori judgments or by other potentially subjective criteria, but by our empirical analyses of the CSR literature's methodological biases.

Obviously, our hypotheses concerning the evaluation results require attention to a range of moderating influences that are model-specific, methodological, and contextual in nature. Model-specific influences include those that we identified in our literature review: how tightly prescribed the reform design is, especially as it relates to curriculum and instruction; the extent to which the developer provides ongoing technical support and professional development to address teachers' specific problems in implementing the reform; and the ways in which developers secure teacher support for the reform. They also include various foci suggested by the 11 components identified in the federal definition of CSR. These include having measurable goals for student performance and benchmarks for meeting those

goals, incorporating a strong parent-governance component, and providing evidence of successful replication of the model. Though relatively little quantitative research has linked these model-specific influences to achievement, we hypothesized that CSR models having specific components designed to address the areas identified in our literature review and the 11 federal characteristics of CSR would tend to be better implemented and more comprehensive reforms than CSR models without these components. In turn, we expected the better-implemented and more comprehensive models to yield the strongest effects on achievement.

The two primary methodological characteristics we identified are related to who is doing the research and the general strength, or internal validity, of the study design that the researcher chooses. We hypothesized that evaluations performed by the CSR developer would yield higher estimates of effects than evaluations done by others. In addition, we predicted that studies employing experimental or quasi-experimental treatment-control comparisons would yield lower effect estimates than studies based on analyses of CSR pre-post gain scores. Though imperfectly matched comparison groups could cause positive biases, it is more likely that effect estimates based on simple one-group pre-post designs will yield greater positive biases. Cook and Campbell (1979) note that threats to internal validity, including history, maturation, and regression-to-the-mean effects, are likely to make one-group pre-post designs among the weakest. Also, empirical results from a meta-analysis of Title I program effects by Borman and D'Agostino (1996) illustrated that analyses of pre-post gains resulted in positive biases, relative to studies employing quasi-experimental control group comparisons, of approximately one fifth of one standard deviation.

The contextual factors affecting CSR effects are largely unexplored and are, therefore, less predictable. Our analyses of the relative effects of CSR in reading, math, and other subjects, across various grade levels, and across varying poverty levels are unprecedented. Given the targeting of recent policies, most notably the CSRD program, on scaling up reform within high-poverty contexts, we hoped to find particular benefits for these schools.

Method

Selection of Comprehensive School Reform Models

The goal of our analysis was to synthesize the research on widely disseminated, externally developed CSR, or whole-school, reform models. To be considered for the current study, therefore, a reform model needed to meet four basic criteria: 1) it is a whole-school or schoolwide reform design; 2) it is the subject of at least one prior study, whether positive or negative, on which we could base our review; 3) it is a model disseminated by developers that is external from the schools; and 4) it has been replicated in 10 or more schools. Previous reviews and catalogs of reform models, including the fall 2000 edition of the Northwest

Regional Educational Laboratory's (NWREL) *Catalog of School Reform Models* (NWREL, 2000) and *An Educators' Guide to Schoolwide Reform* published by the American Institutes for Research (AIR) (Herman et al., 1999) used similar selection criteria.[2] At the time of our selection, these publications were the only known sources of information available to define the universe of CSR models meeting our criteria.

Therefore, our selection of models drew directly from the previous NWREL and AIR catalogs. Through these sources, we identified 33 CSR models, but only 29 of the models provided at least one report of their achievement effects from which we could calculate effect size estimates. The 33 models originally selected for the present research were implemented in 55.6% of the schools that received CSRP funds for externally developed models, as reported in the SEDL database, and the 29 models ultimately included in this review represented 53.4% of the CSRP implementations. The results of this review should generalize reasonably well to the population of schools implementing CSR models using CSRP and Title I program funds. The review, though, clearly does not represent schools that use these funds to implement "home-grown," non-externally developed CSR designs, or schools that package one or more externally developed, targeted, non-schoolwide interventions to develop their own CSR approaches. Finally, because we cannot review the research for CSR models with no research base, these models are not represented in this synthesis either.

Summary descriptions of each school reform model are presented in Table 1 and further descriptive information about the main features and costs of each model is presented in Table 3. The descriptive information in Table 3 is adapted from NWREL's *Catalog of School Reform Models* and is supplemented with a narrative description of each reform's research base.

2 The two catalogs' inclusion criteria were slightly different, but similar to our goal of including models that were nationally disseminated, externally developed, comprehensive school reform. The AIR catalog based its model selection on five criteria: 1) "promoted by their developers as a means to improve student achievement in low-performing schools;" 2) "mentioned by name in federal [CSRD] legislation;" 3) "used in many schools and districts;" 4) "have obtained national visibility in the educational and national press;" and 5) "there is some research evidence about their effects on students and/or implementation in schools." Any reform meeting the second criterion was included automatically, but other reforms had to meet at least three of the other four criteria (Herman et al., 1999, p. 7). Models for the first edition of the NWREL catalog were selected through an open application process. Any developer requesting an application packet was sent one. NWREL then chose from among the submitted applications, based on criteria similar to the Herman et al. (1999) set: "Criteria for selecting models included evidence of effectiveness in improving student academic achievement, extent of replication, implementation assistance provided to schools, and comprehensiveness" (NWREL, 2000). The selection process for NWREL's second edition was by invitation only. Developers of models that met criteria of adoption by five or more schools receiving CSRD funds, were nominated by a state or regional lab CSRD manager, or acknowledged by one of several national educational organizations, were asked to submit applications. Submitted applications were then reviewed based on the criteria outlined previously.

Literature Search Methods

Broad searches of the literature on comprehensive school reform and its effects on achievement were conducted using several approaches. The preliminary literature review involved computerized searches of the Education Resources Information Center (ERIC) database (1966-2001) and the PsychLit database. A second method used general World Wide Web searches (search engines such as Google), and specific searches of CSR developers' Web pages for references to research or any other published or unpublished studies or compilations of data. We also collected all studies referenced in the Herman et al. (1999) and NWREL (1998) reports.

After completing this initial review stage, we compiled separate lists of the references gathered for each of the 33 reform models. We then sent these lists to each of the developers for their review and feedback. All 33 developers responded, either to confirm that our list included all the references known to them, to make suggestions for further references, or to provide studies we were unable to obtain through other sources. The final phase of review involved exhaustive bibliographic reference chasing based on all reports obtained through the computerized databases, via the World Wide Web, and from the developers. After performing this series of search methods, we found no other available evaluations of comprehensive school reform and achievement outcomes.

The period of aggressive collection of studies began in fall 2000 and concluded at the end of that calendar year. After 2001 began, we no longer conducted an extensive literature search; we did, however, continue to contact reform developers as necessary and followed up with locating articles discovered in the previous round of literature searches and the review of references in articles as they arrived at our facility. Thus, the review includes studies completed through late 2001.

Inclusion Criteria

Liberal inclusion criteria were applied in the preliminary stages of the literature search. All study abstracts provided by the database searches and all evaluations of comprehensive school reform and achievement that were referenced in the documents were reviewed to ascertain whether any report of achievement data, in the form of test scores, may have been provided by the studies. If an abstract or study did not suggest these data were reported, the study was excluded from further consideration. More than 800 studies, abstracts, and summaries were read during this preliminary stage of the review process. The vast majority of these studies, though, were not considered beyond this stage, as they typically documented implementation outcomes or the theories supporting the reform model but provided no assessment of the model's achievement effects.

In the second stage, we focused on the subset of studies that provided some form of assessment of the model's effects on students' test scores. From these studies, we chose those that allowed us to generalize to the effects of externally developed whole-school reform models implemented in the United States. In other

words, the studies we selected had to help us answer the question: "What would be the expected achievement effects of the reform model if a school or district in the United States chose to contact the developer and arrange to implement the program as a schoolwide intervention?" More specifically, we deemed studies eligible for further consideration based on the following criteria:

- sufficient achievement data for reform participants, and, when applicable, comparison group students, were provided from which effect sizes could be computed;
- the study design involved some form of comparison from which an effect could be determined: either a one-group pre-post design involving treatment students only or a quasi-experimental or experimental treatment-control comparison;
- the sample or data provided were not duplicated in another study accepted for inclusion;
- the sample used in the evaluation was composed of students from a school in the United States; and
- the sample of students was from the school's regular education program.

Many studies reviewed did not meet these eligibility requirements. This was due, in large part, to insufficient information for calculating effect sizes. The most common reason for excluding studies was the failure to provide a standard deviation or information about the testing instrument from which a standard deviation could be imputed (imputation of data is discussed below in more detail). A substantial number of studies included samples or data that were reported in other studies accepted for inclusion, so they were eliminated. Many other studies used a non-U.S. sample, or a special population, such as special education students. In the end, 232 studies met all requirements and were selected for analysis.[3]

Moderator Variables

In addition to collecting the information necessary for calculating effect sizes and weights (e.g., achievement outcomes, standard deviations of the achievement outcomes, and the sample sizes), we coded a number of other characteristics that corresponded to two general areas: contextual information related to the particular

3 Despite all efforts to obtain the studies from traditional sources (e.g., libraries and ERIC), the model developers, and the authors of the studies, there were 10 publications that we did not obtain. Because we had no opportunity to review these studies, we were not able to establish whether they would have met our requirements for inclusion in the synthesis. These 10 studies were distributed across the following CSR models: Accelerated Schools Project (1); Co-nect Schools (1); Direct Instruction (2); High Schools that Work (1); Paideia (1); School Development Program (2); and Success for All (2). The full list of the 232 studies is available in the original *Review of Educational Research* article by Borman, Hewes, Overman, and Brown (2003).

implementation that was evaluated, and methodological variables related to the study design. Because studies often reported multiple outcomes from multiple contexts or multiple research designs, the contextual and methodological characteristics were coded at the level of the outcome rather than at the level of the study.

Contextual Variables. Contextual variables helped us examine potential differences in effect size related to the context in which the CSR model was evaluated. The contextual variables included:

- subject area tested;
- grade level evaluated;
- years of CSR model implementation for the results given; and
- the poverty level of the school served by the CSR model.[4]

We identified five major subject areas that were tested and evaluated in the CSR literature: language arts, math, science, social studies, and a general category. These were developed from a wider array of subject areas noted in the studies. Language arts included reading and other literacy-related subcategories such as comprehension, vocabulary, spelling, language, word knowledge, and writing. Math covered subcategories including computation, arithmetic, and math applications. Science included both science and health. Social studies included all social sciences and history. The general category typically consisted of composite scores across subjects or general ability measures. These mutually exclusive categories were coded into five indicator variables.

The grade level tested was a dichotomous variable, where "0" represented elementary school grades (K-5), and "1" represented all other grades (i.e., 6-12 and mixed across levels). If a study listed a range of grades associated with an achievement outcome such that grade levels were mixed across the elementary/middle school break, that outcome was assigned to the "1" middle/high/mixed grades category. For example, if a study provided outcome data for students in grades 4-6, the outcome was assigned "1" on this variable.

A smaller subgroup of studies identified the number of years that the CSR model had been implemented at the school and indicated the poverty level of the school. In all cases, we indexed poverty level by the percent of students at the CSR school who were eligible for the free lunch program. The number of years that the CSR model had been implemented at the school site ranged from 1 school year to 14 school years, with an average of 2.96 years.

4 Perhaps the most important contextual information, the level or quality of the model's implementation, was rarely provided. This is one of the most important deficits in the research literature on CSR.

Methodological Variables. The methodological variables describing the evaluations included the following:

- type of effect data provided (i.e., correlational, categorical, or mean difference);
- type of research design (i.e., randomized experiment, quasi-experimental matched school design, quasi-experimental covariate-adjusted design, quasi-experimental match to a "similar" school, quasi-experimental comparison to state or district outcomes, or one-group pre-post design);
- whether the study used a longitudinal design or not; and
- whether the study was conducted by the reform developer or not.

Each of these methodological characteristics was represented by an indicator code (0, 1) in our analyses. For type of effect data, we coded an outcome as one that provided correlational effect data when it showed a simple correlation between participation in the reform model and achievement. Categorical effect data included outcomes that provided a binary achievement result, such as pass/fail or met standard/did not meet standard. The final type of effect data came from mean treatment-control achievement differences or pre-post differences for the treatment group.

We coded six types of research designs, including those that used (a) true random assignment of schools or students to the CSR and control conditions; (b) a quasi-experimental design that included explicit matching of the CSR school (or students) with a comparison school (or students) based on prior achievement levels and student demographics; (c) a covariate-adjusted comparison between the CSR school (or students) and non-CSR school (or students) based on prior achievement levels and, occasionally, student demographics; (d) a comparison of the CSR school (or students) to a non-CSR school (or students) stated to be "similar" based on unspecified criteria; (e) a simple comparison of the CSR school (or students) to all other schools (or students) in the district or state; and (f) a one-group design examining pre-post changes in the CSR school's (or students') achievement outcomes. For our main analyses, we contrasted the one-group pre-post analyses to all of the other designs, which used some form of comparison group.

Third, we coded an indicator variable as "1" for studies that used a true longitudinal design, which tracked the achievement outcomes for the same group of students over time. True longitudinal designs included all outcomes for which there were two or more time points, including simply a pretest and posttest, for the same sample of students at each time point. All other outcomes, including those that contrasted the results for one grade cohort of students in one year to the results for the same grade cohort in a subsequent year, and those that included a simple cross-sectional, posttest-only comparison, were coded as "0." Our original coding scheme provided more detail on the research design, including several distinct types of

cohort studies. In analyses not shown, however, all of the non-longitudinal comparisons were found to yield similar effects, or were simply too few in number to stand alone. Consequently, all research involving non-longitudinal designs was pooled and contrasted to true longitudinal designs.[5]

The final methodological characteristic that we coded contrasted evaluations by the CSR developer to those done by others. Those studies that included among its authors the name of any of the CSR model's original developers were coded as "1," and all other studies were coded "0."

Reform Attributes. Separately from the data entry and coding for each study, each of the 29 CSR models was coded for its components by two or three independent coders as to whether each reform model possessed each of the following characteristics:

- required a set of specific curricular materials;
- required replicable pedagogical practices;
- required a faculty vote with at least 75% approval before the reform could be adopted;
- required a specific and replicable component designed to engage parents and the community in the governance of the school and the planning and implementation of the school improvement process;
- required a set of replicable student performance assessment methods and benchmarks that schools may use to track students' progress; and
- required ongoing teacher and staff professional development and training.

Also, for each of the 29 reforms, we documented the number of schools in which the reform had been replicated, the level of technical support the developer provided to schools, and an estimate of the full marginal cost for the first year of implementing the model.

The information for coding these reform attributes came from the Herman et al. (1999) report, the NWREL (1998) catalog, the developers' websites, documents from the developers, and in some cases, direct contact with the developers. The coding relied on interrater agreement among two or three coders, who independ-

5 The separate types of cohort designs initially coded included: a) comparing the outcomes for one grade level (e.g., third graders) in one year to the outcomes for the same grade level (e.g., third graders) in a subsequent year; b) comparing the outcomes for one grade level in one year (e.g., third graders in 1999) to the outcomes for the same student cohort in a subsequent year (fourth graders in 2000); or c) comparing the outcomes for several grade levels (e.g., third through fifth graders) in one year to the outcomes for the same grade levels (third through fifth) in a subsequent year. "True" longitudinal designs are distinguished from all of these in that they track the same sample of students across each time point. In contrast, the cohort designs have different, but often overlapping, samples of students at each time point.

ently coded the first six (bulleted) attributes. Where the coders did not agree, consensus was met by discussing the reasons for the selected response. If, after this process, there was still no consensus, the CSR developers were contacted to clarify. A single coder derived all cost information, the level of developer support, and the number of replicated schools for the reform models.

Seven of the nine attributes were coded "yes" or "no" for each reform. One attribute, ongoing access to technical support and assistance from the developer, was adapted from the Herman et al. (1999) report, in which it was presented as a scale ranging from 0-4. On this scale, a score of 0 indicated that the developer provided no on-site or other assistance to help schools implement the model, essentially no contact with the school after CSR implementation, and no benchmarks or other useful tools for helping schools assess the progress of their implementation. A score of 4 suggested that the developer provided on-site and other assistance to help schools implement the model, maintained frequent contact with the school after CSR implementation, and provided useful benchmarks and tools for helping schools assess their progress. For reforms not rated in Herman et al. (1999), we used these same criteria to develop ratings on the same scale. There was little variation in reforms' ratings on the 0-4 scale. Most reforms were 3 or 4, with only one being coded as a 2 and none as a 1 or 0. Thus, we recoded this information into an indicator variable where "1" represented the highest support rating of 4 and "0" indicated all other lower ratings.

The number of schools at which the reform was replicated was a continuous variable. This variable was based on the most recent information available regarding the number of schools being served by each of the CSR model developers. The NWREL (2000) catalog provided this information and the date associated with it. When the information was missing, or if the date was before NWREL's most recent update (May 1, 2001), developers were contacted directly for up-to-date information.

Both the full first-year marginal personnel and non-personnel costs for implementing each reform model were estimated. Non-personnel costs included the amount a school would be expected to pay for all materials and services provided by the developer and any additional costs associated with computers, furnishings, and other items demanded by the reform model but not provided through the developer. Personnel costs included the costs of hiring any new staff associated with the reform (e.g., tutors, full-time facilitators, or coaches). In essence, these marginal cost estimates provided a "worse-case scenario" for the costs of the reform. They estimated the total dollar amount of all resources that are demanded by the CSR model, regardless of whether schools could reallocate existing resources to the CSR implementation. For 21 reform models, the total marginal costs were estimated based on information provided in the Herman et al. (1999) report.[6] For the other eight models, the costs were estimated from information from the developers. All costs were based on a school of 500 students and 25 teachers, and were separated into the two distinct personnel and non-personnel cost variables.[7]

Data Imputation

To make use of the maximum number of studies possible, we imputed estimates for sample size and standard deviation under a limited range of circumstances. In all cases, outcomes for which data were imputed were flagged with a dummy code. These two imputation dummy codes, for sample size and standard deviation, were included as covariates in our final analyses of effect size.

Sample Size. If the student sample size was not provided, we estimated the number of students involved in the study based on national averages obtained from the National Center for Education Statistics' 1998-1999 Common Core of Data. In addition, this procedure relied on information in the study indicating the grade level of the sampled students and the number of schools included in the analysis. For example, if an evaluation reported data for second graders in one school, but not the actual sample size, we estimated the sample size to be 75, which is the average size of a school's second grade cohort based on national data from the 1998-1999 Common Core of Data.

For studies that used a district or state as the comparison group, we imputed the comparison group sample size as the treatment sample size rather than using the true district or state sample size. We employed this method to avoid dramatically inflating the weights assigned these studies and conferring a level of precision to these results that was not appropriate.

Standard Deviation. If we were not able to obtain the pooled standard deviation from the study, we imputed a standard deviation in one of two ways. First, if the test was a national standardized test, we consulted available norming data from the test developer to obtain a standard deviation. Falling into this first category are situations where Normal Curve Equivalent (NCE) scores were presented without sample standard deviations. In these cases, we imputed the population standard deviation of 21.06 and flagged the case. Second, if the test was a state or local assessment, for which the state or district maintained a Web page, we used the overall state or local standard deviation reported for the test, grade, and year that corresponded to our data. These strategies of using national, state, or local population standard deviations are akin to methods outlined by Hedges (1981) for computing effect sizes, namely Cohen's *d* or Hedges's *g*, based on the average, or pooled, standard deviation.

6 To achieve greater consistency between the cost estimates provided by select developers during 2001 and the cost estimates for other models based on data in the Herman et al. (1999) report, we adjusted the latter cost estimates to constant 2001 dollars using gross domestic product implicit price deflators.

7 By assuming the same number of students and teachers for each model, we were able to gain greater consistency in the cost estimates. Nevertheless, the estimated marginal costs of implementing the reform models may vary widely by school, depending on a variety of other factors. Rather than relying on these general estimates to project costs for implementing a reform in a particular school, we suggest contacting the developer directly to obtain specific cost estimates.

Independence of Observations

There were several situations that threatened the assumption of independence of observations, which is central to most forms of hypothesis testing. The most obvious of these were reports of duplicate samples, which could arise in three ways: a) when researchers included the same sample in multiple studies; b) when researchers presented multiple analyses of the same sample in one or more studies by using somewhat different sets of covariates, for example; and c) when researchers duplicated a sample across a series of studies of multi-year outcomes, for example, by reporting first-year results in a preliminary report, and repeating in later reports (along with the outcomes for the second and subsequent years of implementation) analyses of the first-year sample as originally presented, or as the remaining longitudinal sample. In the first two situations, we accepted the first or main analysis of the sample and rejected subsequent reports of duplicate samples: the study with the earliest date, whether published or unpublished, was used for analysis. In the third situation where longitudinal samples were involved, we used only the most recent outcomes for a given sample of students. In this way, we focused on the achievement effects from the longest exposure to the model by the school and students.[8]

Samples were further duplicated when results were reported for both a full student sample and for some clearly defined subsample, such as for a separate racial/ethnic group or for those who were low achievers at baseline. In these cases, only the full sample was included for our main analyses. These samples best supported our analysis of the schoolwide effects of CSR. The final way in which independence of observations was threatened involved multiple outcomes within a single achievement domain (e.g., language arts) or across two or more achievement domains (e.g., reading and math) for a distinct sample of students. These situations were resolved by taking the mean effect size across all outcomes and/or domains for the main analysis. For example, if the same student sample had outcomes for reading comprehension, reading vocabulary, and math, the mean effect size across the three areas served to represent a single effect size for that sample. For our subanalyses of the outcomes for the separate subject areas, effect sizes for the various achievement domains were disaggregated and were estimated independently as subject-specific CSR effects.

8 We did not include long-term effects of the models that are sustained after discontinuation of the program. We confronted one such example for Success for All, which has been shown to have sustained effects through the end of eighth grade (Borman & Hewes, 2003). This analysis, though, estimates the sustained effect beyond the discontinuation of this elementary school program in fifth grade. This type of analysis, though highly important, offers a different type of information than that offered by the other analyses that are the focus of this review.

Characteristics of the Selected Studies

From the 232 studies that met all inclusion criteria, 1,111 independent observations were defined. Each of the 1,111 observations was for a distinct CSR model and sample of students from which an effect size was computed. The school was the primary unit of analysis for the meta-analytic findings. It was selected because CSR is designed to affect whole schools and because schools were typically the unit of analysis reported in the primary studies. Key contextual characteristics, including the poverty level and years of CSR implementation, were also school-level features that we hoped to explore as predictors or moderators of effect size.

Reported within-school student sample sizes varied considerably, though. For example, some studies reported achievement data for an entire school, other studies reported data for a single grade level within a school, and still others reported data for a smaller sample of students within a grade level or school. As a result of these differences, we chose to weight all observations based on the student sample. Table 1 presents the number of studies, observations, and treatment and control students involved in the evaluations of all 29 CSR models. This table and Tables 2 and 3 summarize, respectively, the methodological characteristics of the studies and the coded attributes of each of the CSR models. The tables, which list the reforms alphabetically, reveal the diversity of the reform models and studies in the meta-analysis.

The contextual characteristics presented in Table 1 reveal that the number of studies and observations varied widely by reform model, from a low of one study with one observation for Audrey Cohen to a high of 49 studies with 182 outcomes for Direct Instruction. The median number of studies was four and the median number of observations was 23. Overall, these studies involved 145,296 students participating in the CSR schools and 77,660 comparison students. The mean years of implementation across all reforms and studies was 2.96, and, on average, 65.06% of the students in the CSR schools were eligible for the free or reduced price lunch program.

Methodological characteristics are presented in Table 2. Nearly half of the outcomes were derived from one-group pretest-posttest study designs. Over 40% of the observations were from studies conducted by the developers, and about one third were from studies using true longitudinal sample designs. Outcome data were presented as means for most observations, followed by categorical data, and mixed outcome data. Less than 1% of the outcomes relied on correlational data. About three of four outcomes were based on elementary school samples.

The CSR model attributes presented in Table 3 show that there is considerable variety among the 29 models in terms of their general characteristics and the components that they require in typical implementations. For example, 10 of the 29 reforms required specific curriculum materials (34%), and 12 required specific instructional practices (41%). Forty-five percent required a 75% faculty vote; 21% required a parent involvement program; 38% required student assessments and benchmarks; and 34% required ongoing professional development. More than half of the models received the highest rating for ongoing technical support. The number of replication sites varied widely from a low of 15 schools to a high of 1,800. First-year, worst-case scenario costs also varied widely: for personnel, from no cost to $208,361 for Roots & Wings and Success for All, with a median of $13,023; and for non-personnel costs, from $14,585 for Accelerated Schools Project to $780,000 for Montessori, with an overall median of $72,926. Edison Project was assigned the median values for personnel and non-personnel costs because this reform works within a school's given budget.[9]

9 In some cases, Edison pays additional start-up costs that are above and beyond the district's or school's per-pupil allowance. Because these are not marginal costs incurred by the schools or districts, they are not included in our estimate of the cost of implementing Edison.

Table 1. Means for the Contextual Moderator Variables by CSR Model.

	Number of Studies	Number of Observations	
Accelerated Schools Project	6	50	
America's Choice School Design	2	27	
ATLAS Communities	3	8	
Audrey Cohen College	1	1	
Center for Effective Schools	1	26	
Child Development Program	2	2	
Coalition of Essential Schools	3	6	
Community for Learning	1	3	
Community Learning Centers	5	17	
CoNECT Schools	5	42	
Core Knowledge	6	58	
Different Ways of Knowing	2	3	
Direct Instruction	49	182	
Edison Project	5	209	
Expeditionary Learning Outward Bound	6	40	
High Schools That Work	45	64	
High/Scope Primary Grades	4	23	
Integrated Thematic Instruction	2	2	
MicroSociety®	3	32	
Modern Red Schoolhouse	6	23	
Montessori	2	7	
Onward to Excellence	4	13	
Paideia	4	5	
Roots & Wings	6	14	
School Development Program	10	25	
Success for All	42	173	
Talent Development High Schools	1	2	
The Learning Network	3	38	
Urban Learning Center	3	16	
All CSR Models	**232**	**1111**	

	Number of Students		Years of Implementation	% Free Lunch
	Tx N	Control N		
	4,127	1,231	2.87	63.25
	8,922	5,655	2.17	
	2,289	509	2.43	
	115	125	6.00	59.00
	2,269	0	2.15	
	318	262	8.00	
	666	115	4.75	31.63
	1,166	0	2.00	
	163	11	2.35	62.59
	5,487	4,927	2.58	59.83
	3,591	3,220	3.84	49.75
	1,205	870	1.67	28.00
	22,562	20,056	2.92	69.46
	18,955	5,616	2.74	59.49
	3,136	1,927	3.48	33.16
	20,471	1,740	3.65	66.85
	1,220	1,217	3.39	76.73
	49	45	4.50	
	6,115	650	5.6	87.00
	3,979	2,213	2.65	81.00
	914	922	3.00	
	3,530	0	4.31	80.00
	1,442	643	6.00	
	3,154	3,074	2.46	72.42
	4,865	3,840	3.70	53.50
	18,523	17,820	2.56	81.46
	803	870	1.67	
	3,741	102	2.00	64.64
	1,524	0	1.19	93.50
	145,296	**77,660**	**2.96**	**65.06**

Table 2. Means for the Methodological Moderator Variables by CSR Model.

	One-Group Pre-Post Design	Study by Developer	Longitudinal Sample	
Accelerated Schools Project	0.86	0.12	0.08	
America's Choice School Design	0.52	0.52	0.41	
ATLAS Communities	0.75	0.00	0.00	
Audrey Cohen College	0.00	0.00	1.00	
Center for Effective Schools	1.00	0.00	0.00	
Child Development Program	0.00	1.00	0.00	
Coalition of Essential Schools	0.83	0.00	0.83	
Community for Learning	1.00	1.00	0.00	
Community Learning Centers	0.94	0.00	0.94	
CoNECT Schools	0.14	0.05	0.00	
Core Knowledge	0.00	0.00	0.31	
Different Ways of Knowing	0.33	0.33	0.00	
Direct Instruction	0.17	0.08	0.63	
Edison Project	0.75	0.69	0.06	
Expeditionary Learning Outward Bound	0.65	0.18	0.53	
High Schools That Work	0.94	0.95	0.00	
High/Scope Primary Grades	0.04	0.52	0.52	
Integrated Thematic Instruction	0.50	0.00	1.00	
MicroSociety®	0.84	1.00	0.00	
Modern Red Schoolhouse	0.39	0.52	0.04	
Montessori	0.00	0.00	0.57	
Onward to Excellence	1.00	1.00	0.23	
Paideia	0.40	0.00	0.80	
Roots & Wings	0.21	0.64	0.07	
School Development Program	0.28	0.64	0.32	
Success for All	0.06	0.45	0.75	
Talent Development High Schools	0.00	1.00	0.00	
The Learning Network	0.79	0.13	0.29	
Urban Learning Center	1.00	1.00	0.13	
All CSR Models	**0.46**	**0.41**	**0.34**	

	Imputed Data		Type of Outcome Data				Middle or High School Outcomes
	N	SD	Categ.	Corr.	Means	Mixed	
	0.92	0.12	0.00	0.00	1.00	0.00	0.04
	1.00	0.00	0.56	0.00	0.37	0.07	0.33
	0.88	0.00	0.88	0.00	0.13	0.00	0.75
	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	1.00	0.00	1.00	0.00	0.00	0.00	0.58
	0.00	0.00	0.00	0.00	1.00	0.00	0.50
	0.17	0.83	0.00	0.00	1.00	0.00	0.83
	0.00	1.00	0.00	0.00	1.00	0.00	0.00
	0.06	0.00	0.06	0.00	0.94	0.00	0.35
	0.98	0.07	0.45	0.00	0.55	0.00	0.31
	0.05	0.78	0.67	0.00	0.33	0.00	0.00
	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	0.34	0.11	0.30	0.02	0.69	0.00	0.11
	0.86	0.60	0.37	0.00	0.55	0.09	0.30
	0.40	0.63	0.13	0.00	0.83	0.05	0.83
	0.94	0.00	0.58	0.00	0.31	0.11	1.00
	0.00	0.48	0.00	0.00	1.00	0.00	0.00
	0.00	0.50	0.00	0.00	1.00	0.00	0.00
	0.47	0.16	0.38	0.00	0.63	0.00	0.31
	0.74	0.00	0.91	0.00	0.09	0.00	0.09
	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	1.00	0.69	0.31	0.00	0.69	0.00	0.54
	0.00	0.20	0.00	0.00	1.00	0.00	0.40
	0.93	0.36	0.43	0.00	0.57	0.00	0.00
	0.60	0.60	0.16	0.00	0.84	0.00	0.60
	0.38	0.08	0.15	0.00	0.84	0.01	0.05
	0.00	0.00	1.00	0.00	0.00	0.00	1.00
	0.05	0.26	0.50	0.00	0.47	0.03	0.13
	0.88	1.00	0.00	0.00	1.00	0.00	0.50
	0.56	**0.29**	**0.34**	**0.00**	**0.63**	**0.03**	**0.27**

Table 3. Summary of CSR Model Attributes.

	Number of Replication Sites	First-Year Costs[1]		Specific Curriculum	
		Personnel	Non Pers		
Accelerated Schools Project	1300	$13,543	$14,585	0	
America's Choice School Design	450	$104,181	$93,763	0	
ATLAS Communities	105	$8,334	$93,763	0	
Audrey Cohen College	16	$78,135	$89,595	0	
Center for Effective Schools	1000	$52,175	$55,000	0	
Child Development Program	165	$95,675	$65,000	1	
Coalition of Essential Schools	1000	$1	$250,000	0	
Community for Learning	118	$78,135	$85,428	0	
Community Learning Centers	15	$1	$61,700	0	
CoNECT Schools	198	$1	$612,582	0	
Core Knowledge	1020	$1	$58,341	1	
Different Ways of Knowing	600	$1	$87,512	1	
Direct Instruction	300	$52,090	$202,110	1	
Edison Project	136	$13,023	$72,926	1	
Expeditionary Learning Outward Bound	93	$1	$84,386	0	
High Schools That Work	1300	$1	$50,007	0	
High/Scope Primary Grades	100	$1	$135,435	1	
Integrated Thematic Instruction	1434	$1	$61,235	0	
MicroSociety®	200	$52,175	$67,450	0	
Modern Red Schoolhouse	110	$1	$223,988	0	
Montessori	1000	$159,723	$780,000	1	
Onward to Excellence	1000	$12,502	$62,508	0	
Paideia	100	$52,090	$100,013	0	
Roots & Wings	200	$208,361	$72,926	1	
School Development Program	600	$13,543	$33,338	0	
Success for All	1800	$208,361	$72,926	1	
Talent Development High Schools	35	$31,254	$28,129	1	
The Learning Network	200	$52,175	$32,188	0	
Urban Learning Center	29	$10,418	$165,647	0	
All CSR Models[2]	**200**	**$13,023**	**$72,926**	**0.34**	

	Specific Instruction	75% Faculty Vote	Parent Governance	Student Assessment Benchmarks	Ongoing Professional Development	High Developer Support
	0	1	0	0	0	0
	0	0	0	1	0	1
	1	0	1	0	1	0
	0	0	0	1	1	0
	0	0	0	1	0	1
	1	1	0	0	1	1
	0	1	0	0	0	0
	1	0	0	1	1	1
	0	0	1	1	1	0
	0	1	0	0	1	1
	0	1	0	0	0	0
	1	0	0	0	1	1
	1	0	0	1	0	0
	1	0	1	1	0	0
	1	1	0	1	1	1
	0	1	0	0	0	1
	1	0	0	0	0	1
	0	0	0	0	0	1
	0	1	1	0	0	1
	0	1	0	1	0	1
	1	0	0	0	0	0
	0	0	1	0	0	1
	1	1	0	0	0	0
	1	1	0	1	0	1
	0	0	1	0	1	0
	1	1	0	1	0	1
	0	1	0	0	0	1
	0	0	0	0	0	0
	0	0	0	0	1	0
	0.41	**0.45**	**0.21**	**0.38**	**0.34**	**0.55**

Note. [1] Median applied for Edison Project. [2] All Reforms summary line is reform-level, not observation-

Results

Distribution and Measures of Central Tendency for Effect Size

Our analysis of the effect size data began with an inspection of the distribution of the 1,111 unweighted effect sizes.[10] Applying Tukey's (1977) definition, we identified as statistical outliers any effect sizes that were more than three interquartile ranges above the 75th percentile or below the 25th percentile. Of the 1,111 independent observations, 19, or 1.8%, met this definition.

Similarly, we identified statistical outliers from the distributions of treatment and control sample sizes, with 132, or 12%, of the 1,111 independent treatment student samples meeting the Tukey (1977) criterion for statistical outliers. Of the control sample sizes, 75, or 13%, of the 598 independent samples met the criterion.

Statistical outliers may exert an overly strong influence on the results. Outliers on the dependent variable, effect size, are especially problematic, but outliers on sample size also are of concern. Because sample size plays an important role in weighting each effect size, unusually large samples may have an exceedingly large influence on the outcomes of our analyses. Therefore, we chose to Winsorize both effect sizes and sample sizes that were statistical outliers. In both cases, we set the value for the effect size or sample size to equal the value at three interquartile ranges beyond the 75th percentile or below the 25th percentile. Because some observations had multiple outlier values on these three variables, only 153 cases (13.7%) were involved in the Winsorizing. The 153 Winsorized cases were spread across 20 of the 29 reforms.[11]

Based on the 1,111 unweighted mean effect sizes, an overall weighted effect size was computed. The unweighted average of the 1,111 effect sizes was .15 and the overall weighted value for d was also .15. The average weighted effect size, which is equivalent to a pre-post gain or CSR-control difference of 3.16 NCEs, was greater than 0, $Z = 13.11$, $p < .001$. The standard error of the weighted effect size, which is the square root of v.*, was .01. This standard error was employed to calculate a 95% confidence interval for the average weighted effect size. The calculation resulted in a confidence interval of .13 to .18, or 2.74 to 3.79 NCEs. However, as

10 Please see the Appendix for a detailed description of the formulas and approaches used for effect size calculation and weighting.

11 We also ran the analyses with the original non-Winsorized values and obtained similar results. In the regression analysis, there were some minor changes in the magnitudes of coefficients, but not in the direction or level of statistical significance of the results. In the reform-specific analyses, again, all changes were inconsequential, except for three models whose effect size estimates were somewhat larger with the non-Winsorized values. The three models were Direct Instruction, whose estimated effect sizes were 0.06 greater using non-Winsorized values; Paideia, 0.03 greater for all cases and 0.05 greater for comparison-group-only cases and for third-party comparison-group cases only; and Expeditionary Learning Outward Bound, 0.03 greater for comparison-group-only cases.

Shadish and Haddock (1994) warned, due to the heterogeneity of the effect estimates, the average weighted effect size should not be interpreted as an estimate of a single population effect parameter, but rather simply as describing the mean of the 1,111 observed effect sizes.

Regression Analysis of Weighted Effect Sizes on Mediating Variables

To explain the heterogeneity of the effect sizes, we performed a modified weighted multiple regression analysis using an SPSS macro, METAREG.SPS, provided by Lipsey and Wilson (2001). This macro modifies the output that would result from a regular weighted least squares multiple regression and provides the correct standard errors, confidence intervals, and significance tests for meta-analysis. The modified weighted least squares multiple regression analysis for random effects was performed using weighted effect size as the dependent measure and the moderator variables as predictors. As explained previously, an estimate of the residual variance component was computed as the random-effects variance plus the estimation variance, and weights were defined by the reciprocal of the residual variance component. Table 4 presents the results of the regression analysis.

All moderator variables accounted for 8% of the variance in the weighted effect sizes. Full descriptions of the variables entered into the regression model are provided in the Method section. First, the comparison group indicator contrasted those observations based on a single-group pre-post design to observations that were based on quasi-experimental non-equivalent control-group designs and true randomized designs. The positive coefficient indicated that the one-group comparisons yielded relatively larger mean effect sizes. The magnitude of the coefficient suggested that, after controlling for the other variables in the model, comparisons using control groups produced effect size estimates .08, or about 1.7 NCEs, less than estimates generated by one-group, pre-post analyses of treatment effects.

Second, as expected, the model indicated that effect sizes produced by developers' evaluations were greater than those produced by other researchers' evaluations. The coefficient suggested that, after statistically taking into account the other moderators, evaluations by developers produced effect size estimates .16, or 3.4 NCEs, greater than those produced by external evaluations. Third, use of a longitudinal sample produced a larger effect size than use of other sample types, about .09 greater. This suggests that when researchers measure CSR effects over time on the same longitudinal sample of students the results tend to show stronger achievement effects than when researchers track effects across successive cohorts of students. Fourth, those outcomes that were estimated with imputed standard deviations had smaller effect sizes than those that were based on actual, reported standard deviations.

Table 4. Summary of Weighted Least Squares Regression Analysis for Moderator Variables Predicting Effect Size.

Variable	β	SE	Z
Constant	-0.21	0.28	-0.75
Methodological Variables			
Longitudinal Sample	0.09	0.04	2.40*
Means-Based Effect Size Formula	0.06	0.03	1.89
Standard deviation was imputed	-0.09	0.03	-2.72**
Sample size was imputed	-0.03	0.03	-1.52
Evaluation was conducted by the developer	0.160	.03	4.87***
Study design was one-group pre-post	0.08	0.03	2.35*
Middle School or High School Outcome	0.03	0.03	1.12
Reform Attribute Variables			
Reform requires			
Specific curriculum materials	-0.06	0.06	-0.86
Specific instructional practices	0.01	0.06	0.24
75% faculty vote for adoption	0.05	0.04	1.34
Parent governance	-0.09	0.0	-2.32*
Goals and benchmarks	0.01	0.04	0.21
Ongoing professional development	-0.09	0.05	-1.71
Replication	0.00	0.00	-1.22
High developer support rating	-0.01	0.04	-0.20
Cost of reform: Personnel	0.00	0.00	1.17
Cost of reform: Non-personnel	0.02	0.02	0.95

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Finally, only one reform model component was a statistically significant predictor of effect size, and the relationship was in an unexpected direction. Namely, models that required a component designed to involve parents in school governance and improvement had smaller effects on achievement than models that did not require this form of parent participation.

School Poverty Level and Years of Implementation as Moderators of Effect Size

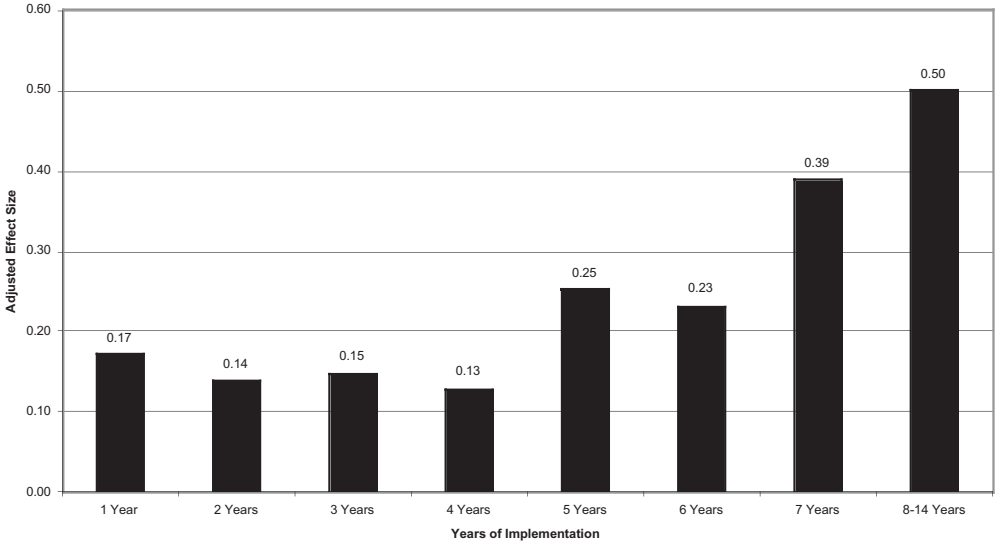
A subset of studies had complete data indicating the CSR school's free or reduced-price lunch participation rate. Of the 1,111 independent observations, 630 (57%) had complete data indicating the poverty level of the CSR school. Similarly, a subset of 975 of the 1,111 observations, or 88%, had complete data indicating the number of years that the CSR model had been implemented at the school.

After regressing weighted effect size on the methodological moderator variables, we obtained the residuals from the regression and added the mean weighted effect size to each observation. In this way, we calculated effect sizes that were statistically adjusted for all of the methodological variables. These adjusted effect sizes became the outcome measures for our subanalyses of the relationship between school poverty and years of implementation and CSR effects.

The weighted regression model using poverty level to predict adjusted effect size revealed that a school's poverty level was not a statistically significant predictor of effect size ($Z = .12$). In other words, across the range of school poverty levels, which tended to be relatively high, CSR was equally effective in relatively lower- and higher-poverty schools.

In a separate weighted regression model, years of implementation was a statistically significant predictor of effect size, with a coefficient of .02 ($Z = 2.82$, $p < .01$). Figure 1 displays the relationship between years of implementation and effect size. This figure shows that the CSR effect size, .17, was relatively strong during the first year of implementation. During the second, third, and fourth years of implementation, though, the effect declined slightly but, essentially, remained the same. After the fifth year of implementation, CSR effects began to increase substantially. Schools that had implemented CSR models for five years showed achievement advantages that were nearly twice those found for CSR schools in general, and after seven years of implementation, the effects were more than two and a half times the magnitude of the overall CSR impact of $d = .15$. The small number of schools that had outcome data after 8 to 14 years of CSR model implementation achieved effects that were three and a third times larger than the overall CSR effect.

Figure 1. Adjusted Effect Size by Years of CSR Model Implementation



Analysis of Subject Area as Moderator of Effect Size

A different level of aggregation of the outcome data was used to analyze the effects for different subject areas. In previous analyses, to retain independent samples of students, we took the mean outcome for students tested across more than one area. For instance, in studies of students attending a CSR school who took both math and reading tests, we aggregated the effects across both subjects and generated a single effect size for the student sample. Our analyses by subject area, though, maintained independence of observations by analyzing the effects in each subject area separately. All cases had information regarding the subject area evaluated, though some cases presented data for mixed subjects or for more general achievement outcomes. In all other ways, the database used in this analysis was similar to those used for the main analyses and for the subanalyses of school poverty and years of implementation.

The data for these analyses included 1,017 independent samples for reading, 679 for math, 229 for science, 138 for social studies, and 95 cases that could not be grouped into the other subject areas, either because the original research reported results with subjects grouped, or because the achievement test was more general in focus. With a mean effect size of .13 ($Z = 10.81$, $p < .001$) for reading, CSR had a statistically significant effect that was somewhat lower than the effect size found for CSR overall. The CSR effect size for math was essentially the same as the overall CSR effect, and slightly higher than the effect for reading, $d = .15$ ($Z = 9.86$, $p < .001$). The CSR effect on science outcomes was somewhat lower than the effects for math

and reading, $d = .09$ ($Z = 3.79$, $p < .001$), but was also statistically significant. CSR did not have a statistically significant effect ($Z = 0.72$) on social studies outcomes. Finally, the cases with outcome data for the general subject area revealed a relatively large CSR effect, $d = .20$, but also a high standard error (.05) and a wide 95% confidence interval, $d = .10$ to $d = .31$. This confidence interval, though, did not include 0 and the result was statistically significant ($Z = 3.86$, $p < .001$).

Evidence of Effectiveness for the 29 CSR Models

Table 5 presents the weighted mean effect size, *d*, the associated significance test, *Z*, and 95% confidence intervals, which represent the expected range of effects, separately by CSR model. There are three sets of columns in Table 5. The set of columns farthest to the left displays all available evidence concerning the achievement effects of each of the 29 models, regardless of the nature or quality of the study designs. The second set of columns presents results for only those cases that used some form of control group, and the final set of columns shows results for only those cases that were third-party, control-group studies. In the latter two, more restrictive presentations of the data provide the best evidence for evaluating the effects of the reform models, in that our prior regression analysis demonstrated that studies performed by the developer and those that used one-group pre-post designs yielded potential biases relative to third-party and control-group comparisons.

The names of the CSR models are listed along the left hand side of Table 5 and are grouped into four categories:

- *Strongest Evidence of Effectiveness;*
- *Highly Promising Evidence of Effectiveness;*
- *Promising Evidence of Effectiveness; and*
- *Greatest Need for Additional Research*

The four categories were established based on a combination of three criteria:

1) *Quality* of the evidence: Does the CSR model have research evidence from the highest-quality studies: control-group studies and third-party control group studies?
2) *Quantity* of the evidence: Does the CSR model have a relatively large number of studies and observations from which one may generalize the findings to the population of schools in the U.S. that are likely to adopt and implement CSR models? (For instance, we used ten or more studies overall and five or more third-party control-group studies as the, arguably arbitrary, standards necessary to be in the top category).
3) *Statistically significant and positive* results: Does the evidence from control-group studies show that the effects of the reform on achievement are positive and statistically greater than 0?

Table 5. Weighted Mean Effect Sizes by CSR Model.[1]

Model Name [2]	All Cases						
	N of Studies	N of Obs	*d*	SE	95% CI Lower Bound	95% CI Upper Bound	*Z*
Strongest Evidence of Effectiveness [3]							
Direct Instruction	49	182	0.21	0.02	0.17	0.25	11.61**
School Devlpment Prog	10	25	0.15	0.03	0.10	0.20	5.48**
Success for All	42	173	0.18	0.01	0.16	0.21	16.57**
Highly Promising Evidence of Effectiveness [4]							
Exp Lrning Outwrd Bnd	6	40	0.19	0.03	0.13	0.25	5.93**
Modern Red Schlhouse	6	23	0.26	0.03	0.20	0.31	9.25**
Roots & Wings	6	14	0.38	0.04	0.29	0.46	8.43**
Promising Evidence of Effectiveness [5]							
Accelerated Schools	6	50	0.09	0.02	0.05	0.12	4.79**
America's Choice	2	27	0.22	0.01	0.20	0.24	20.84**
ATLAS Communities	3	8	0.27	0.02	0.24	0.31	14.76**
Montessori	2	7	0.27	0.04	0.19	0.35	6.55**
Paideia	4	5	0.30	0.12	0.06	0.53	2.51*
The Learning Network	3	38	0.22	0.02	0.17	0.26	8.67**
Greatest Need for Additional Research [6]							
Audrey Cohen	1	1	-0.13				
Cnter for Effective Schs	1	26	0.13	0.01	0.11	0.15	11.86**
Child Development Proj	2	2	0.12	0.06	0.00	0.24	1.90
Coalition of Esstial Schs	3	6	-0.09	0.04	-0.17	-0.01	-2.21*
Community for Learning	1	3	0.15	0.01	0.13	0.16	16.49**
Commnty Learning Ctrs	5	17	0.03	0.03	-0.03	0.08	1.02
Co-nect	5	42	0.04	0.02	0.01	0.08	2.48*
Core Knowledge	6	58	0.03	0.02	-0.01	0.06	1.53
Diffnt Ways of Knowing	2	3	0.00	0.02	-0.04	0.04	0.04
Edison	5	209	0.06	0.01	0.05	0.08	8.68**
High Schools That Work	45	64	0.30	0.01	0.28	0.33	25.60**
High / Scope	4	23	-0.02	0.04	-0.10	0.07	-0.37
Integratd Thematic Instr	2	2	0.24	0.19	-0.13	0.62	1.27
MicroSociety	3	32	0.29	0.03	0.24	0.34	11.11**
Onward to Excellence II	4	13	0.25	0.03	0.19	0.30	8.35**
Talent Devmnt High Sch	1	2	0.14	0.03	0.09	0.19	5.38**
Urban Learning Centers	3	16	-0.03	0.03	-0.08	0.02	-1.06
All CSR Models	**232**	**1111**	**0.15**	**0.00**	**0.14**	**0.16**	**33.26****

Comparison Studies							Third-Party Comparison Studies						
				95% CI							95% CI		
N of Studies	N of Obs	*d*	SE	Lower Bound	Upper Bound	*Z*	N of Studies	N of Obs	*d*	SE	Lower Bound	Upper Bound	*Z*
40	151	0.15	0.02	0.12	0.19	8.40**	38	146	0.15	0.02	0.11	0.18	7.82**
9	18	0.05	0.03	-0.01	0.12	1.57	5	7	0.11	0.04	0.04	0.18	3.23**
41	162	0.18	0.01	0.16	0.20	15.32**	25	85	0.08	0.02	0.05	0.11	5.08**
4	14	0.51	0.07	0.38	0.64	7.50**	3	7	0.02	0.06	-0.09	0.14	0.39
4	14	0.17	0.04	0.10	0.25	4.48**	3	11	0.21	0.05	0.11	0.30	4.12**
5	11	0.35	0.05	0.25	0.46	6.68**	4	5	0.77	0.07	0.64	0.90	11.83**
3	7	0.21	0.07	0.07	0.36	2.93**	2	5	0.13	0.10	-0.06	0.31	1.33
1	13	0.25	0.01	0.23	0.27	20.75**	1	13	0.25	0.01	0.23	0.27	20.75**
2	2	0.40	0.03	0.33	0.46	12.15**	2	2	0.40	0.03	0.33	0.46	12.15**
2	7	0.27	0.04	0.19	0.35	6.55**	2	7	0.27	0.04	0.19	0.35	6.55**
3	3	0.57	0.18	0.22	0.93	3.16**	3	3	0.57	0.18	0.22	0.93	3.16**
1	8	0.33	0.04	0.25	0.41	8.18**	1	8	0.33	0.04	0.25	0.41	8.18**
1	1	-0.13					1	1	-0.13				
0							0						
2	2	0.12	0.06	0.00	0.24	1.90	0						
1	1	0.24					1	1	0.24				
0							0						
1	1	-0.06					1	1	-0.06				
5	36	0.00	0.02	-0.04	0.03	-0.24	4	34	-0.01	0.02	-0.04	0.03	-0.31
6	58	0.03	0.02	-0.01	0.06	1.53	6	58	0.03	0.02	-0.01	0.06	1.53
1	2	-0.04	0.03	-0.10	0.01	-1.46	1	2	-0.04	0.03	-0.10	0.01	-1.46
3	53	-0.13	0.02	-0.17	-0.10	-7.65**	3	53	-0.13	0.02	-0.17	-0.10	-7.65**
4	4	-0.06	0.02	-0.09	-0.03	-3.44**	0						
3	22	-0.05	0.04	-0.14	0.03	-1.22	2	11	-0.04	0.05	-0.15	0.06	-0.83
1	1	0.92					1	1	0.92				
1	5	0.13	0.07	0.00	0.26	1.95							
0							0						
1	2	0.14	0.03	0.09	0.19	5.38**	0						
0							0						
145	**598**	**0.12**	**0.01**	**0.11**	**0.14**	**16.87****	**109**	**461**	**0.09**	**0.01**	**0.07**	**0.10**	**10.59****

The notes to Table 5 provide more detailed information about the criteria used to evaluate the quantity of evidence for each of the four categories. Within each of the four categories, the models in Table 5 are listed alphabetically.

Strongest Evidence of Effectiveness. CSR models in this category include those that have a large number of studies and observations from schools and students across the United States, such that their outcomes have been replicated in a number of contexts and are reasonably generalizable to the population of U.S. schools

Notes. * $p < .05$; ** $p < .01$.

1 Effect sizes are presented as Cohen's *d*.

2 Abbreviated Model Names: Cnter for Effective Schs = Center for Effective Schools; Child Development Proj = Child Development Project; Coalition of Esstial Schs = Coalition of Essential Schools; Commnty Learning Ctrs = Community Learning Centers; Diffnt Ways of Knowing = Different Ways of Knowing; Exp Lrning Outwrd Bnd = Expeditionary Learning Outward Bound; Integratd Thematic Instr = Integrated Thematic Instruction; Modern Red Schlhouse = Modern Red Schoolhouse; School Devlpment Prog = School Development Program; Talent Devmnt High Sch = Talent Development High School.
We established CSR models' placements in the four categories, *Strongest Evidence of Effectiveness, Highly Promising Evidence of Effectiveness, Promising Evidence of Effectiveness, and Greatest Need for Additional Research,* on the basis of three criteria: (a) the *quality* of the reforms' evidence (i.e., research evidence from control-group studies or from third-party control group studies); (b) the *quantity* of their evidence (i.e., the number of studies and observations and the extent to which the findings may be generalized to the population of schools in the U.S. that are likely to adopt and implement CSR models; and (c) whether the expected achievement effects of the reform are *statistically significant and positive* (i.e., achievement effects are positive and are statistically greater than 0). Within each of the four categories, the models are listed alphabetically. More information regarding the nature of the reform models along with narrative descriptions of the supporting research base for each of the CSR designs may be found in the Appendix.

3 Models in the *Strongest Evidence of Effectiveness* category have ten or more studies total and five or more third-party comparison-group studies from across the U.S., such that their outcomes have been replicated in a number of contexts and are reasonably generalizable to the population of schools in the U.S. that is likely to adopt and implement CSR models. The models in this category also have *statistically significant and positive* achievement effects based on evidence from studies using comparison groups or from third-party comparison designs.

4 Models in the *Highly Promising Evidence of Effectiveness* category are those that had *positive and statistically significant results* from comparison or third-party comparison studies, but did not have research bases that were as broad and generalizable as those of the models in the top category. These models have five or more studies total and three or more third-party comparison-group studies.

5 Models in the *Promising Evidence of Effectiveness* category were reforms that had two or more studies total and at least one third-party comparison-group study. Though too few studies to generalize from their results with confidence, all of these CSR models had *statistically significant positive* effects from comparison or third-party comparison studies.

6 Models in the category *Greatest Need for Additional Research* were reforms with no more than one study total or those that did not have evidence of *statistically significant positive* achievement effects from comparison or third-party comparison studies.

that are likely to adopt and implement CSR models. These models also have *statistically significant and positive* achievement effects based on evidence from studies using comparison groups or from third-party comparison designs. Three reforms—Direct Instruction, School Development Program, and Success for All—met the criteria for this category.

Direct Instruction has an overall effect size of $d = .21$ ($Z = 11.61$, $p < .01$), with a 95% confidence interval of $d = .17$ to $d = .25$. The confidence interval expresses the degree of accuracy of the effect size estimate and suggests a range of effects that are likely to be found in similar implementations and studies of the reform model. In this case, similar implementations and studies of Direct Instruction are likely to reveal effects between $d = .17$ and $d = .25$. The effects for Direct Instruction estimated from comparison and third-party comparison designs were somewhat lower than the overall effects, but still positive and statistically significant, $d = .15$ ($Z = 8.40$, $p < .01$) and $d = .15$ ($Z = 7.82$, $p < .01$), respectively.

The School Development Program is another model meeting the highest standard of research evidence, with an overall effect size of $d = .15$ ($Z = 5.48$, $p < .01$) and a 95% confidence interval of $d = .10$ to $d = .20$. As with Direct Instruction, the effect of the School Development Program drops considerably when looking at effects only for comparison or third-party comparison studies: $d = .05$ ($Z = 1.57$, n.s.) and $d = .11$ ($Z = 3.23$, $p < .01$), respectively.

Success for All is the third model in the *Strongest Evidence of Effectiveness* category, with an overall effect size of $d = .18$ ($Z = 16.57$, $p < .01$) and a 95% confidence interval of $d = .16$ to $d = .21$. The effects are essentially the same when considering only Success for All comparison studies, $d = .18$ ($Z = 15.32$, $p < .01$), as most Success for All evaluations use a comparison group design. The effect estimate from Success for All third-party comparison studies, $d = .08$ ($Z = 5.08$, $p < .01$), is considerably less but still statistically significant.

Highly Promising Evidence of Effectiveness. Models in this category are those that had positive and statistically significant results from comparison or third-party comparison studies, but did not have research bases that were as broad and generalizable as those of the models that met the highest standard. Three reform models met the criteria for this category: Expeditionary Learning Outward Bound, $d = .19$; Modern Red Schoolhouse, $d = .26$; and Roots and Wings, $d = .38$.

Promising Evidence of Effectiveness. Models meeting this standard of evidence were reforms that had more than one study, but still too few to generalize from their results with confidence. All of these CSR models, though, had statistically significant positive effects from comparison or third-party comparison studies. The reforms in this category were: Accelerated Schools, with an overall effect size of $d = .09$; America's Choice, with an effect size of $d = .22$; Atlas Communities, $d = .27$; Montessori, $d = .27$; Paideia, $d = .30$; and The Learning Network, $d = .22$.

Greatest Need for Additional Research. The Greatest Need for Additional Research category included reforms with only one study or those that did not have evidence of statistically significant positive achievement effects from comparison

or third-party comparison studies. Seventeen of the 29 CSR models fell into this category.[12] Nearly all of the reforms in this category were there because too few studies have been done to establish statistically reliable and generalizable results. Four of the 17 models had no evidence from either comparison or third-party comparison studies, and another four models lacked evidence from third-party comparison studies. Finally, four CSR models had only a single effect estimate from both comparison and third-party comparison studies. On the other hand, though, there are a number of models, including the Center for Effective Schools, Community for Learning, Co-nect, Core Knowledge, MicroSociety, Onward to Excellence II, and Talent Development High Schools, that have promising early data but need several more rigorous evaluations to establish a stronger research base.

Two CSR models in this category presented unusual cases that are worthy of discussion. First, the High Schools that Work model has a large research base, composed, almost entirely, of one-group pre-post evaluations performed by its developer. The magnitude of the effect size from these studies, $d = .30$, is relatively large but the effect size from the one comparison-group study of High Schools that Work actually revealed a statistically significant *negative* effect of the model, $d = -.06$. This model has been widely replicated and studied and, in many ways, appears to be a promising high school intervention. That the model has been replicated with such success, has been so well supported by the developer, and has accumulated a large number of one-group pre-post evaluations are, indeed, laudable accomplishments. For many schools, this type of evidence may be sufficient to convince decision makers that the model is worthy of adoption. To meet even higher standards of research evidence, though, more research using control groups is needed to help clearly establish the model's apparent benefits.

Second, though only five studies of the Edison Project have been conducted, they have evaluated the reform in a large number of schools. Taking all of the evidence, Edison appeared to have a statistically significant positive effect size, $d = .06$. When examining the reports of third-party evaluators using comparison groups, though, the results revealed a statistically significant *negative* effect, $d = -.13$. Again, additional studies using comparison groups are needed, from both the developer and from third-party evaluators, to help reconcile these differences.

Discussion

CSR and the CSRP are at the forefront of the national movement to base educational policy and practice on solid research evidence. The recent reauthorization of the Elementary and Secondary Education Act of 1965 and the

12 Although four CSR models, Foxfire Fund, League of Professional Schools, QuEST, and Ventures Initiative and Focus System, were dropped from our study for lack of research evidence amenable to analysis, they could be considered among the models for which there is the greatest need for additional research.

federal government's single largest investment in America's elementary and secondary schools, the *No Child Left Behind* Act, have similarly required practices based on high-quality research for everything from the technical assistance to schools to the choice of anti-drug abuse programs. Like a mantra, the *No Child Left Behind* Act repeats phrases such as "scientifically based research" more than 100 times (Olson & Viadero, 2002). This legislation, urging the use of research-based educational practices and procedures in schools receiving federal CSRP and Title I funding, has the potential to revolutionize school improvement in some of the most challenging contexts in the United States.

Does the quantity and quality of the CSR literature provide the scientifically based evidence needed to identify the proven programs and practices that these new policies demand? Our research has sought to understand the CSR research base in various ways. We have described the overall characteristics of the diverse literature; we have identified its biases; and we have empirically established the best evidence that researchers, policymakers, and practitioners should apply to understanding the effects of CSR models. We have estimated the overall effects of the most widely used, nationally disseminated, externally developed CSR models and have gained insight into the overall effects of CSR as a national policy movement. We have also established that there is considerable variation in these effects that is explained by the models themselves, methods used in evaluating the models, and the circumstances in which the programs were implemented. Looking across the 232 studies of CSR and our various analyses of them, the evidence supports six primary findings.

The Characteristics of the CSR Research Base

First, *CSR is still an evolving field and there are clear limitations on the overall quantity and quality of studies supporting its achievement effects.* Only 12 reform models are supported by five or more studies of their achievement effects. Only four models have been the subject of five or more third-party studies that used comparison groups. Over 40% of the analyses of CSR effects have been performed by the developers, and about half of the analyses have used some type of quasi-experimental control group. Only seven studies of three CSR models, or about 3% of all studies of the achievement effects associated with CSR, have generated evidence from randomized experiments. These reform models and studies include: the School Development Program (Cook, Habib, Phillips, Settersten, Shagle, & Degirmencioglu, 1999; Cook, Hunt, & Murphy, 1999); Direct Instruction (Crawford & Snyder, 2000; Grossen & Ewing, 1994; Ogletree, 1976; Richardson, Dibenedetto, Christ, Press, & Winsbert, 1978); and Paideia (Tarkington, 1989). In addition to these shortcomings, many of the studies did not present sufficient detail to allow for replication of the findings. For instance, substantial numbers of reports contained no information about student sample sizes and did not provide standard deviations for the outcome measures.

Many of these problems, though, are to be expected given the recent emergence of CSR, in general, and many of the CSR models, in particular. Some models are at an early stage of program development that has not yet demanded third-party evaluations and more costly and difficult control-group comparisons. On the other hand, there are some models that have had relatively long histories, have been replicated in many schools, and should have accumulated this evidence. Still other CSR models are on their way to establishing a strong research base. Three models, in particular, have accumulated enough evidence to meet our highest standard of research evidence. Taken as a whole, there is a sufficient number of reasonably high-quality studies of CSR to evaluate its overall effects and to inform policy.

The Overall Effects of CSR

Second, the *overall effects of CSR are statistically significant, meaningful, and appear to be greater than the effects of other interventions that have been designed to serve similar purposes and student and school populations. Overall, students from CSR schools can be expected to score one eighth of a standard deviation, or 2.5 NCEs, higher on achievement tests than control students in non-CSR schools. Because the method of resource reallocation allows high-poverty schools to implement reform models at little or no extra cost, CSR can be a very cost-effective strategy for improving achievement outcomes*. Our various analyses suggest that students attending CSR schools can be expected to score between nearly one-tenth and one-seventh of a standard deviation, or between 1.9 NCEs and 3.2 NCEs, higher than control students on achievement tests. The low-end estimate represents the overall effect size of $d = .09$ for third-party studies using comparison groups and the high-end estimate represents the effect size of $d = .15$ for all evaluations of the achievement effects of CSR. Using a metric devised by Cohen (1988), U_3, the effect size of $d = .12$ for all studies using control groups tells us that the average school implementing a CSR program outperformed about 55% of similar control schools that did not implement a CSR model.

How should we interpret this overall effect? Cooper (1981) has suggested a comprehensive approach to effect size interpretation that uses multiple criteria and benchmarks for understanding the magnitude of the effect. First, and most generally, we may compare the overall CSR effect size to Cohen's (1988) definitions of a small effect within the behavioral sciences, $d = .20$, and a large effect, $d = .80$. Second, and more specifically, Cohen (1988) pointed out that the relatively small effects of around $d = .20$ were most representative of fields that are closely aligned with education, such as personality, social, and clinical psychology. Similarly, Lipsey, and Wilson's (1993) more recent compendium of meta-analyses concluded that psychological, educational, and behavioral treatment effects of modest values of even $d = .10$ to $d = .20$ should not be interpreted as trivial.

Finally, and even more specifically, how do CSR effects compare to previous national efforts to help close the achievement gap and improve the outcomes of large numbers of high-poverty and low-achieving students and schools? The most

obvious comparison to the effect of CSR is the effect of traditional Title I programs, which have historically funded targeted remedial interventions, such as pullout programs, and schoolwide programs designed to assist at-risk students. These programs were the subject of Borman's and D'Agostino's (1996) meta-analysis of the achievement effects of Title I programs, which synthesized the results from all federal evaluations conducted between 1965 and 1994. During these years, rather than schoolwide programs and CSR models, the primary methods for upgrading the educational programs of at-risk children were through specialized pullout programs and other targeted assistance. Borman and D'Agostino estimated that the average effect size associated with these efforts was $d = .11$. The Title I evaluations, though, were almost exclusively based on the less-preferred one-group pre-post design and may overestimate the true Title I effect. Borman and D'Agostino did make an adjustment for regression to the mean effects for all Title I outcomes from one-group pre-post designs. The comparison to this benchmark is suggestive, but because the primary studies and meta-analyses used different methodologies than those reported here, the comparison is imperfect.

A better comparison between CSR and conventional Title I programs may be drawn directly from the current study by estimating the CSR effect size from comparison-group studies in schools of 50% poverty or more. In most of these cases, the comparison schools have such high poverty rates that it is highly likely that they received federal Title I funds. In most cases, these schools implement Title I targeted or schoolwide programs and, in most cases, are not implementing other CSR models. These studies, therefore, provide a relatively good indication of the value-added effects of CSR, above and beyond the effect of traditional Title I programs. Across 346 such comparisons, the effect size, statistically adjusted for methodological characteristics, was $d = .12$. In other words, despite the fact that the vast majority of these control schools provided their students with extra resources and programs provided through Title I, the average CSR school still outperformed 55% of the Title I schools.

Are these benefits worth the seemingly high costs associated with implementing many of the CSR programs? On average, CSR programs have first-year costs of approximately $85,000, including both personnel and non-personnel expenditures, which include items such as training and materials. However, some developers have argued that schools with concentrations of poor children generally are able to garner sufficient resources to implement CSR models by simply reallocating existing supplemental funds and personnel from federal and state Title I programs, special education, desegregation settlements, and other sources (Slavin, et al., 1994). In this way, many schools can afford even the high-priced CSR models by simply trading in their largely remedial approaches of the past, most often represented by federal and state Title I programs, for new designs that will enable them to implement research-based schoolwide reform programs. As Odden & Archibald (2000) argued, this method of "resource reallocation" can make implementations of CSR programs essentially "costless." With a free-lunch participation rate of over

65%, the average school in this meta-analysis would be a clear candidate for the reallocation approach.

There are, indeed, clear challenges in determining the relative costs and benefits of CSR models (Levin, 2002), but if one assumes that implementations in high-poverty schools generally have few additional costs, the benefits we have found are obviously well worth these modest investments. There is some research evidence to suggest that even if one does not assume that CSR implementations are "costless," high-quality models are capable of yielding cost-benefit ratios that equal or exceed those found for other noted educational interventions, including the Tennessee Student/Teacher Achievement Ratio (STAR) class-size reduction effort (Borman & Hewes, 2003). Further, the analyses of Borman and Hewes revealed that a CSR model that focuses on early intervention and prevention actually may save schools the investments in the costly remedial practices of special education referrals and retentions in grade, which can alone offset the costs of implementing CSR models. Though this evidence is important, much more cost-effectiveness research is needed for a wider range of CSR models, and for a broader array of educational interventions in general.

These conclusions regarding our analyses of the overall effects of the CSR models are valuable for understanding general outcomes. These overall effects, though, are highly variable and should be viewed as averages found across a wide array of reform models and schools that were evaluated in a variety of ways. The overall effect size is a good indicator of the expected effects of CSR across a large number of schools. For instance, we can say with some confidence that policymakers may expect to find CSR effects of between $d = .09$ and $d = .15$ across similar studies of national or large district-wide samples of CSR schools. The effects for individual schools and the effects for individual reform models are likely to vary more widely. Our regression analysis and the specific effects of the 29 reform models reveal many reasons for the diverse findings, but a considerable amount of variability is left unexplained.

Explaining Differences in CSR Effects

Third, *the heterogeneity of the CSR effect and the fact that few of the general reform components helped explain this variability suggests that the differences in the effectiveness of CSR are largely due to unmeasured program-specific and school-specific differences in implementation*. Our regression analysis suggested that whether or not a CSR model, in general, requires the following components explains very little in terms of the achievement outcomes the school can expect: a) ongoing staff professional development; b) measurable goals and benchmarks for student learning; c) a faculty vote to increase the likelihood of model acceptance and buy-in; and d) the use of specific and innovative curricular materials and instructional practices designed to improve teaching and student learning. Similarly, the frequency with which the CSR models have successfully replicated their approaches in schools with diverse

characteristics; the overall level of external technical support and assistance from the developer, and the general cost of the model do not help us explain a substantial amount of the variability in the CSR effect.

The one reform attribute that was a statistically significant predictor of effect size suggested that CSR models that require the active involvement of parents and the local community in school governance and improvement activities tend to achieve worse outcomes than models that do not require these activities. Taking strong actions to encourage parents to play significant roles in school governance and reform may help the school grow as an institution, but these activities are not likely to have strong impacts on achievement (Epstein, 1995). In contrast to school-based efforts aimed at helping families enrich their children's learning opportunities outside of school, which are far more likely to help individual children succeed with specific academic goals, the focus on parent involvement in school governance could sidetrack schools if the immediate priority is to improve achievement.

The general lack of explanatory power for the required reform characteristics suggests at least two possible interpretations. The first is that these components are not important for promoting achievement in CSR schools and, therefore, there is no relationship. The second interpretation is that knowing whether or not a CSR model required schools to implement a given component tells us little about whether or not the component actually was implemented. This latter interpretation suggests that some or all of these components may make a difference in terms of achievement, but school-specific and model-specific differences in the ways that the components are actually implemented explain considerably more than simply knowing whether or not the CSR developer requires them. Consistent with research that has linked the success of school reform to the level and quality of implementation (Berman & McLaughlin, 1978; Crandall et al., 1982; Datnow, Borman, & Stringfield, 2000; Stringfield et al., 1997), the coordination and fit of the model to local circumstances, and the relationship between the CSR developer and the local school and school district (Datnow & Stringfield, 2000), we contend that knowing more about these largely unmeasured and unreported differences in implementation, across both schools and CSR models, would also enrich our understanding of the variability in the CSR effects.

Fourth, *rather than the general programmatic components of the CSR models, the methodological differences across the studies themselves tell us far more about the effects that we could expect to find.* Studies performed by the developer yielded considerably stronger effects than studies performed by others. Does this suggest, as Pogrow (2000) and others have implied, that the developers, to use a metaphor, have their thumbs on the scale and are consciously manipulating the evaluation to make the outcomes appear more favorable? This interpretation may have some merit in a few cases, but is probably not a reasonable explanation of the overall trend. Perhaps equally likely is that some third-party researchers may seek to taint a model due to a personal grudge or professional dislike for its particular orientation. A more plausible source of developer bias is a variant of the so-called "file-drawer" problem,

which involves the tendency for researchers to publish or otherwise disseminate their statistically significant findings but consign their nonsignificant findings to a dusty filing cabinet. In this case, CSR developers may selectively report the positive outcomes for their models and file away null and negative findings.

Rather than overt bias or selective reporting, another explanation for the stronger outcomes we find for the developers' studies is that when developers are more actively involved in the study of their models, they also are more likely to be actively involved in studying a high-quality implementation. After all, why would developers want to study half-hearted implementations of their models? Further, if developers found that they were studying half-hearted implementations, they would be in the best position of anyone to help the school improve the quality of its implementation. Many of the studies performed by developers may represent what Cronbach et al. (1980) termed the "superrealization" stage of program development. Before broad field trials, interventions are often studied under optimal conditions as assessments of what the program can accomplish at its best. The extent to which the developers' studies and results may generalize across broader implementations of their CSR models, though, is of some concern.

The second key methodological finding was that studies using a one-group, pretest-posttest design produced larger effect sizes than studies using control groups. This is a clear methodological bias that should be addressed in future CSR research. Ideally, evaluations should include randomized designs, which assign schools at random to CSR and control conditions. As Borman (2003) pointed out, innovations should not be forced on schools through random assignment. Schools should be partners in the process of experimentation and should be supportive of the CSR model under study. The only clear trade-off in such studies is that some schools will receive the innovation now and others assigned to the control condition will receive the program later, if it proves to be worthwhile and effective.

High-quality, quasi-experimental control-group designs are also desirable. When comparing directly randomized experiments and quasi-experiments that were designed to answer the same research questions, Lipsey and Wilson (1993) found that quasi-experiments are more highly variable in the results that they produce. As a result, although quasi-experiments may be less expensive than true experiments to conduct in the short run, they are less efficient in the long run because one needs many more of them to arrive at the same conclusion as a randomized experiment. If randomized or matched control groups are not possible, even a comparison of the CSR school's outcomes to district averages will provide some understanding of the value-added effects of the model.

Fifth, *the models meeting the highest standard of evidence, Direct Instruction, the School Development Program, and Success for All, are the only CSR models to have clearly established, across varying contexts and varying study designs, that their effects are relatively robust and that the models, in general, can be expected to improve students' test scores.* As the results in Table 5 demonstrate, the outcomes vary considerably by reform model. In most cases, however, the research base for each CSR model is still

too small to generate reliable estimates of the models' expected effects. For instance, it is certainly premature to conclude that the Audrey Cohen CSR model is likely to have a negative effect on achievement of d = -.13 when replicated in schools. It is also too early to say that Integrated Thematic Instruction will likely have a relatively strong positive effect of d = .24 when implemented in other schools. In some cases, promising and highly promising models are emerging. Expeditionary Learning Outward Bound, Modern Red Schoolhouse, and Roots & Wings are all on the brink of establishing strong research bases. The models meeting the standard for the *Strongest Evidence of Effectiveness* category are distinguished from these models and others by the quantity and generalizability of their outcomes, the quality of this evidence (for instance, six of the seven randomized experiments and many high-quality quasi-experimental control-group studies have been conducted on the models achieving the highest standard of evidence), and the reliable effects on achievement.

Sixth, turning to contextual differences that we studied, *the number of years of model implementation has very important implications for understanding CSR effects on achievement. The strong effects of CSR beginning after the fifth year of implementation may be explained in two ways: a potential cumulative impact of CSR or a self-selection artifact.* Specifically, schools may be experiencing stronger effects as they continue implementing the models, or it may be that the schools experiencing particular success continue implementing the reforms while the schools not experiencing as much success drop them after the first few years. Both explanations seem to have some credence. Nonetheless, it is of considerable significance that the average school across all studies reviewed had implemented its CSR model for approximately three years. The typical study in this meta-analysis, therefore, may underestimate the true potential of CSR for affecting change in schools and for improving achievement.

We explored the significance of two other important contextual variables for understanding differences in achievement outcomes. The poverty level of the school in which the CSR model is implemented and the subject area that is tested for CSR effects do not explain large differences in the observed effects. All schools, regardless of poverty level, appear to benefit from CSR and most subject areas tested reveal similar reform impacts. Because federal funds for implementation of CSR models target high-poverty schools, this finding is of importance. It suggests that the schools from the most challenging high-poverty contexts are benefiting just as much from CSR as are schools from more advantaged circumstances.

Conclusion

Historically, teaching has been fraught with what Lortie (1975) called "endemic uncertainties." Moreover, Cook and Payne (2002) argued that the dominant perspectives on evaluation and improvement in education suggest that the context of each district, school, and classroom is so distinctive that only highly specific change strategies mapped to site-specific circumstances are likely to modify and improve their central functions. The continued growth and early success of CSR, which has advanced the application of replicable technologies that are based on scientific knowledge, provides a clear contrast to these long-standing theories and beliefs about schools, educational change, and evaluation.

The successful expansion of CSR shows that research-based models of educational improvement can be brought to scale across many schools and children from varying contexts. There are adaptations that are sensitive to context—for instance there is a Spanish version of the Success for All program, *Éxito Para Todos*, for English language learners—but the general models of school improvement also include well-founded and widely applicable instructional and organizational components that are capable of being brought to scale across a large number of schools. The increasing marketplace of CSR models and the proven replicability of many of the programs are important developments. To further advance CSR, though, policymakers and educators must demand clear evidence that the reforms will make a difference.

The models meeting our highest standard of evidence have been well researched and have shown that they are effective in improving achievement across reasonably diverse contexts. These models certainly deserve continued dissemination and federal support through CSRP and Title I. All CSR models—even those achieving the highest standard of evidence—would benefit from more federal support for the formative and summative evaluations that are necessary to establish even more definitively what works, where, when, and how. Rather than approving programs on the basis of the 11 requirements (e.g., parent outreach program, clear goals and benchmarks) that make a model "comprehensive," we suggest that schools and policymakers pay even stronger attention to the models' outputs.

Clear research requirements, ample funding for research and development, and a focus on the CSR models' results may support the transformation of educational research and practices in much the same way that it has helped transform medical research and treatment. Like the series of studies required in the Food and Drug Administration's premarketing drug approval process, a similar set of studies might guide the research, development, and ultimate dissemination of educational programs (Borman, 2003). Once a CSR program has met a standard of evidence, then its implementation using federal funds, namely those from CSRP and Title I, should be approved. Before programs have accumulated such evidence, some concern should be shown for the ethics of supporting educational programs with unknown potentials. In medicine, Gilbert, McPeek, and Mosteller (1977) noted that only half of the new treatments subjected to randomized clinical trials actually showed benefits beyond the standard treatments patients would have received. Without the benefit of high-quality evaluation, many widely disseminated educational practices may simply waste the time of teachers and students or, potentially, do harm.

At the same time, we do not suggest that schools and policymakers dismiss promising programs before knowing their potential effects. Instead, we challenge the developers and the educational research community to make a long-term commitment to research-proven educational reform and to establish a market place of scientifically based models capable of bringing comprehensive reform to the nation's schools. Similar to Donald Campbell's (1969) famous vision of the "experimenting society," we must take an experimental approach to educational reform, an approach in which we continue to evaluate new programs designed to cure specific problems, in which we learn whether or not these programs make a difference, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available.

Postscript

This chapter was written under funding from the Office of Educational Research and Improvement, U.S. Department of Education (Grant No. OERI–R-117-D40005). However, any opinions expressed do not necessarily represent positions or policies of OERI. The authors thank all 33 of the reform developers for their help in accessing studies and information about their CSR models. A previous version of this chapter appeared as Borman, Hewes, Overman, and Brown (2003), copyright 2003 by the American Educational Research Association. Adapted by permission of the publisher.

References

Berman, P., & McLaughlin, M. W. (1978). *Federal programs supporting educational change, vol. VIII: Implementing and Sustaining Innovation*. Santa Monica, CA: RAND.

Bodilly, S., Purnell, S., Ramsey, K., & Keith, S. (1996). *Lessons from New American Schools Development Corporation's demonstration phase*. Santa Monica, CA: RAND.

Bodilly, S. J., Keltner, B., Purnell, S.W., Reichardt, R. E., & Schuyler, G. L. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: RAND.

Borman, G. (2000a). Title I: The evolving research base. *Journal of Education for Students Placed At Risk, 5*(1–2), 27–45.

Borman, G. (2000b). The effects of summer school: Questions answered, questions raised. *Monographs of the Society for Research in Child Development, 65*(1), 119–127.

Borman, G. (2003). Experiments for educational evaluation and improvement. *Peabody Journal of Education, 77*(4), 7–27.

Borman, G., & Hewes, G. (2003). The long-term effects and cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis, 24,* 243–267.

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

Borman, G., Rachuba, L., Datnow, A., Alberg, M., MacIver, M., Stringfield, S., et al. (2000). *Four models of school improvement: Successes and challenges in reforming low-performing, high-poverty Title I school* (CRESPAR Report No.48). Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed At Risk.

Borman, G. D., Stringfield, S. C., & Slavin, R. E. (2001). *Title I: Compensatory education at the crossroads*. Mahwah, NJ: Erlbaum.

Borman, G. D., D'Agostino, J. V., Wong, K. K., & Hedges, L. V. (1998). The longitudinal achievement of Chapter 1 students: Preliminary evidence from the Prospects study. *Journal of Education for Students Placed At Risk, 3*(4), 363–399.

Borman, G. D., Wong, K. K., Hedges, L. V., & D'Agostino, J. V. (2001). Coordinating categorical and regular programs: Effects on Title I students' educational opportunities and outcomes. In G. D. Borman, S. C. Stringfield, & R. E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 79–116). Mahwah, NJ: Erlbaum.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24,* 409–429.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Comer, J. P. (1988). Educating poor minority children. *Scientific American, 259*(5), 42–48.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston. MA: Houghton Mifflin Company.

Cook, T. D., Habib, F., Phillips, M., Setterson, E. A., Shagle, S. C., & Degirmencioglu, S. M. (1999). Comer's School Development Program in Prince George's County, Maryland: A theory–based evaluation. *American Educational Research Journal, 36*(3), 543–597.

Cook, T. D., Hunt, H. D., & Murphy, R. F (2000). Comer's School Development Program in Chicago: A Theory-Based Evaluation. *American Educational Research Journal, 37,* 535–597.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 150–178). Washington, DC: Brookings Institution Press.

Cooper, H. (1981). On the effects of significance and the significance of effects. *Journal of Personality and Social Psychology, 41,* 1013–1018.

Crandall, D. P., Loucks-Horsley, S., Baucher, J. E., Schmidt, W. B., Eiseman, J. W., Cox, P. L., et al. (1982). *Peoples, policies, and practices: Examining the chain of school improvement* (Vols. 1–10). Andover, MA: The NETWORK.

Crawford, D. B., & Snider, V. E. (2000). Effective mathematics instruction: The importance of curriculum. *Education and Treatment of Children, 23*(2), 122–142.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., et al. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco, CA: Jossey-Bass.

Datnow, A., Borman, G., & Stringfield, S. (2000). School reform through a highly specified curriculum: A study of the implementation and effects of the Core Knowledge Sequence. *The Elementary School Journal, 101,* 167–192.

Datnow, A., & Castellano, M. (2000). Teachers' response to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal, 37*(3), 775–779.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed At Risk, 5*(1–2), 183–204.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–24.

Epstein, J. L. (1995). School/family/community partnerships: Caring for the children we share. *Phi Delta Kappan, 76,* 701–712.

Gilbert, J., McPeek, B., & Mosteller, F. (1977). Statistics and ethics in surgery and anesthesia. *Science, 198,* 684–689.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8.

Grossen, B., & Ewing, S. (1994). Raising mathematics problem-solving performance: Do the NCTM teaching standards help? *Effective School Practices, 13*(2), 79–91.

Haynes, N., Emmons, C., & Woodruff, D. (1998). School Development Program effects: Linking implementation to outcomes. *Journal of Education for Students Placed At Risk, 3*(1), 71–86.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 106–128.

Herman, R., Aladjem, D., O'Malley, A. O., McMahon, P., Quinones, S., Masem, E., Reeve, A., et al. (1999). *An educators' guide to schoolwide reform*. Arlington, VA: Education Research Services.

Hirsch, E. D., Jr. (1995). *Core Knowledge Sequence*. Charlottesville, VA: Core Knowledge Foundation.

Hirsch, E. D., Jr. (1996). *The schools we need*. New York: Doubleday.

Kearns, D. & Anderson, J. (1996). Sharing the vision: Creating New American Schools. In S. Stringfield, S. Ross, & L. Smith (Eds.), *Bold plans for school restructuring* (pp. 9–23). Mahwah, NJ: Erlbaum.

Levin, H. M. (2002). *The cost effectiveness of whole school reforms* (Urban Diversity Series 114). New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209.

Lipsey, M. W, & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Press.

Lortie, D. C. (1975). *Schoolteacher*. Chicago: University of Chicago Press.

Muncey, D. E., & McQuillan, P. J. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the Coalition of Essential Schools*. New Haven: Yale University Press.

Murphy, J., & Beck, L. (1995). *School-based management as school reform: Taking stock*. Newbury Park, CA: Corwin Press.

Northwest Regional Educational Laboratory. (1998). *Catalog of school reform models: First edition*. Portland, OR: Author. (ERIC Document Reproduction Service No. ED439510)

Northwest Regional Educational Laboratory & National Clearinghouse for Comprehensive School Reform. (n.d.). *Catalog of school reform models*. Portland, OR & Washington, DC: Author. Retrieved August 6, 2004, from http://www.nwrel.org/scpd/catalog/index.html

Nunnery, J. (1998). Reform ideology and the locus of development problem in educational restructuring: Enduring lessons from studies of educational innovation. *Education and Urban Society, 30,* 277–295.

Odden, A., & Archibald, S. (2000). *Reallocating resources: How to boost student achievement without asking for more*. Thousand Oaks, CA: Corwin Press.

Ogletree, E. J. (1976). A Comparative Study of the effectiveness of DISTAR and eclectic reading methods for inner-city children. Chicago, IL: Chicago State University. (ERIC Document Reproduction Service No. ED146544).

Olson, L., & Viadero, D. (2002). Law mandates scientific base for research. *Education Week, 21*(20), pp.1, 14–15.

Pogrow, S. (1998). What is an exemplary program, and why should anyone care? A reaction to Slavin and Klein. *Educational Researcher, 27*(7), 22–29.

Pogrow, S. (2000). Success for All does not produce success for students. *Phi Delta Kappan, 82,* pp. 1, 67–81.

Puma, M. J., Karweit, N., Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). Prospects: *Final report on student outcomes*. Bethesda, MD: Abt Associates, Inc.

Raudenbush, S.W. (1994). Random effects models. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York, NY: Russell Sage Foundation.

Richardson, E., Dibenedetto, B., Christ, A., Press, M., & Winsbert, B. (1978). An assessment of two methods for remediating reading deficiencies. *Reading Improvement, 15*(2), 82–95.

Rotberg, I. C., Harvey, J., & Warner, K. E. (1993). *Federal policy options for improving the education of low-income students. Vol. I: Findings and recommendations*. Santa Monica, CA: RAND.

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

Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design of schools. In C.B. Cazden (Ed.), *Review of research in education* (pp. 353–389). Washington, DC: American Educational Research Association.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York, NY: Russell Sage Foundation.

Sizer, T. R. (1992). *Horace's school: Redesigning the American high school*. New York, NY: Houghton Mifflin Company.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15(*9), 5–11.

Slavin, R. E. (1989). The PET and the pendulum. *Phi Delta Kappan, 70,* 752–758.

Slavin, R. E. (1999). Rejoinder: Yes, control groups are essential in program evaluation: A response to Pogrow. *Educational Researcher, 28*(3), 36–38.

Slavin, R. E., & Fashola, O. S. (1998). *Show me the evidence!* Thousand Oaks, CA: Corwin Press.

Slavin, R. E., & Madden, N. A. (2001). *One million children: Success for All*. Thousand Oaks, CA: Corwin Press.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S. M., & Smith, L. M. (1994). Whenever and wherever we choose: The replication of Success for All. *Phi Delta Kappan, 75,* 639–647.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S., Smith, L., et al. (1996). Success for All: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76.

Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing, Politics of Education Association yearbook, 1990* (pp. 233–267). London: Taylor & Francis.

Stringfield, S. (2000). A synthesis and critique of four recent reviews of whole-school reform in the United States. *School Effectiveness and School Improvement, 11,* 259–269.

Stringfield, S., Millsap, M., Yoder, N., Schaffer, E., Nesselrodt, P., Gamse, B., et al. (1997). *Special strategies for educating disadvantaged children. Urban and suburban/rural. Final Report.* Washington, DC: U.S. Department of Education.

Tarkington, S.A. (1989). *Improving critical thinking skills using Paideia seminars in a seventh-grade literature curriculum.* Unpublished doctoral dissertation, University of San Diego, CA.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer Press.

Traub, J. (1999). *Better by design? A consumer's guide to schoolwide reform.* Washington, DC: Thomas B. Fordham Foundation.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

U.S. Department of Education. (2002). *Comprehensive School Reform (CSR) Program Guidance*. Retrieved August 6, 2004, from http://www.ed.gov/programs/compreform/guidance/index.html

Walberg, H. & Greenberg, R. (1999). The Diogenes factor. *Phi Delta Kappan, 81,* 127–128.

Wang, M. C., Haertel, G. D., & Walberg, H. (1997). *What do we know? Widely implemented school improvement programs*. Philadelphia: Temple University Center for Research in Human Development and Education.

Wong, K., & Meyer, S. (2001). Title I schoolwide programs as an alternative to categorical practices: An organizational analysis of surveys from the Prospects study. In G. D. Borman, S. C. Stringfield, & R. E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 195–234). Mahwah, NJ: Erlbaum.

Wong, K. K., & Meyer, S. J. (1998). Title I schoolwide programs: A synthesis of findings from recent evaluation. *Educational Evaluation and Policy Analysis, 20,* 115–136.

Appendix

Computation of Effect Sizes

Differences in the nature of the outcome data required nine separate methods for computing effect sizes. The nine methods were of three general types: 1) those that used means and standard deviations (six); 2) those that used frequency distributions (two); and 3) those that used correlations (one). For the first and second types, there was a further distinction between effect sizes based on treatment-control comparisons or one-group pre-post designs.

The nine different formulas were all algebraically equivalent, and yielded estimates of the standardized mean difference or common effect size index known as Cohen's *d* or Hedges's *g* (Lipsey & Wilson, 2001). This equivalence was of importance, as we intended to combine the effect estimates from the various formulas in our analyses. Three of the six means-based effect size calculations relied on variations of the common formula

$$d = (M_T - M_C)/S,$$

where $(M_T - M_C)$ is the difference between the CSR participants' and non-participants' group means, and S represents the pooled standard deviation. A variation of the formula for *d* involved adjusting for group differences on the pretest. If the two groups were shown to be similar at pretest, or there was some other statement of pre-intervention similarity, or the posttest group means were presented in the report as having been adjusted for pretest differences, then we simply used this common formula. For cases where there were pretest differences between participants and non-participants, but adjusted posttest means were not presented, we adjusted the posttest means ourselves using the pretest group means and the correlation between pretest and posttest.[13]

A second variation of the formula for *d* used participants' and non-participants' gain scores as estimates of means. If a comparison-group design was not used, another variation of this basic formula utilized only the participants' mean gain score in the numerator. In this variation, the participants' pretest in effect serves as the comparison. For both of these variations, the denominator was the pooled or population standard deviation on the posttest itself and not the standard deviation of the gain scores. Finally, three other methods for calculating an effect size used the test statistics *t* and *F* or used a *p* value when the actual group means were not presented in the study.

13 When the correlation between pretest and posttest was not provided, we imputed a pre- post correlation of 0.08. These cases were so few that we did not include a flag to indicate an imputed value.

We used two methods for calculating effect sizes based on categorical outcomes. When results from a χ^2 analysis with $df = 1$ and total sample size (N) were presented, we used these data to estimate an effect size directly. In other cases, we approximated an effect size (d) based on an arcsine transformation of the proportion (p) of successes for each group

$$d = arcsine\ (p_1) - arcsine\ (p_2).$$

Lipsey and Wilson (2001) stated that the arcsine transformation generally produces a more conservative estimate than the probit transformation and suggested that if effect sizes based on frequency distributions are to be included with other effect sizes based on means and standard deviations, as in the present research, a sensitivity analysis should be conducted to determine which to use.

Our sensitivity analysis showed that the arcsine and probit transformations produced similar overall means, but the probit transformation produced longer tails at both ends of the effect size distribution. Furthermore, the effect sizes based on a calculation of means and standard deviations from the actual grouped frequency distributions produced much higher estimates of d than either the arcsine or the probit transformation. For these reasons, we used the arcsine transformation for the cases where the outcome variable was non-continuous.

The final method of effect size calculation used correlational data and applied only one formula, which used the correlation between group membership and the outcome variable. Again, this formula produced an effect index that was algebraically equivalent to an effect size based on means and standard deviations.

Computation of Variance Components, Weights, and Weighted Effect Sizes Within a Random-Effects Model

From the outset, it was presumed that a random-effects model was most appropriate for the analysis of CSR effects for two reasons. First, the large number of potential methodological, programmatic, and contextual moderators, which were outlined earlier in the introduction, underlies the concept of a study's true effect size as random (Raudenbush, 1994). Second, this set of potential moderators was not considered to be exhaustive. The qualities of instruction in the schools and the characteristics of local implementations, among other program attributes, were all assumed to contribute to the variation in the estimated effect sizes. Thus, it was hypothesized that various reforms, across programs and schools, would not yield the same fixed population effect.

To test whether the true effect size varied, in addition to the variability introduced by sampling variance, or estimation variance, a homogeneity test of the weighted effect-size estimates was performed. Because the value of 10,777.03 for the homogeneity test statistic, Q, exceeded the upper-tail critical value of χ^2 at 1110 degrees of freedom ($p < .001$), the observed variance of the effect sizes was significantly greater than that which would be expected by chance if all observations shared the same population effect size. This statistical test confirmed the a priori assumption of a random-effects model specification.

The random-effects variance estimates, v^*_i, for the effect sizes for control group comparisons were computed based on the formulas

$$v_i = ((N_T + N_C) / (N_T * N_C)) + (d^2 / (2(N_T + N_C)) \text{ and}$$

$$v^*_i = \sigma_\Theta^2 + v_i,$$

where v_i represents the within-study variance component, and σ_Θ^2 is the between-studies or population variance component, which was calculated based on the method-of-moments procedure explained by Raudenbush (1994). Given that there were no control students for the one-group, pretest-posttest outcomes, the variance formulas were

$$v_i = (1/ N_T) + ((d^2)/(2* N_T)) \text{ and}$$

$$v^*_i = \sigma_\Theta^2 + v_i,$$

Finally, the formula for the computation of the weights, for each observation, i, under the assumptions of a random-effects model was

$$w_i = 1/v^*_i.$$

CHAPTER THREE

Challenges for the Future of Comprehensive School Reform

Kathryn M. Borman
University of South Florida
Kevin Carter
American Institutes for Research
Daniel K. Aladjem
American Institutes for Research
Kerstin Carlson LeFloch
American Institutes for Research

The first two chapters of this book detailed the impact and unique benefits of comprehensive school reform (CSR). This chapter describes the challenges faced by policymakers advocating CSR and practitioners implementing CSR. Challenges to implementing and sustaining whole-school reform are bound to be legion in number, given the complexity of the task. In this chapter, we will explore these complexities, focusing on two main components: challenges inherent in creating necessary conditions for CSR and challenges associated with sustaining CSR.

In our review of "necessary conditions," we will discuss policies at all levels supporting the implementation of CSR. Challenges facing policymakers and practitioners include negotiating CSR reforms within a federal policy system, building capacity for ambitious school improvement, ensuring faculty acceptance, obtaining parental support, maintaining fidelity to a research-based plan, and supporting special populations, in particular English language learners, students with disabilities, and racial and ethnic minorities. The second portion of the chapter focuses on sustainability of CSR—challenges facing schools throughout the implementation process over time. In that section, we discuss the challenges associated with understanding the costs of CSR, developing capacity, and continuing CSR research using varied methodologies.

Necessary Conditions for CSR

Nesting CSR within Federal, State, and District Policies

Over the past decade, states, districts, and schools have been attracted to the promise of comprehensive school reform. However, constraints placed on schools by state legislatures, state education agencies, and district administrators can derail reform. District policies may be undercut by state-mandated requirements, and both districts and states are required to address federal legislative controls. This hierarchy of educational authorities can introduce complications into the process of reform, but may also offer supportive mechanisms that enhance implementation and sustainability. In the following sections, we consider how each level of this system—federal, state, and district—may support or inhibit comprehensive school reform efforts.

U.S. education policy represents a typical embodiment of American federalism. In the textbook version of American federalism, each level of government—federal, state, and local—is responsible for distinct functions, with authority and responsibility devolved to the lowest possible level. Historically, the federal role in education policy was quite limited. However, over the past half-century, the federal role expanded to include issues related to supporting the service of traditionally underserved populations. The federal role changed so dramatically with the enactment of the No Child Left Behind (NCLB) Act of 2001 that the U.S. Department of Education now delves into policy areas previously considered beyond its purview. For example, NCLB's requirements in the area of teacher qualifications provide a

good example of how federal policy can extend into the classroom, affecting school and district-level hiring decisions.

In general, the federal role in education is growing, but CSR itself does not seem to be garnering much attention or support at the federal level. During the late 1990s, the Comprehensive School Reform Demonstration (CSRD) program (also known by its legislative name, the Obey-Porter program) acted as a catalyst for further development of CSR models, but more recent funding programs associated with NCLB are less explicitly supportive of CSR. The curricular and programmatic emphases in the new law appear to be primarily on mathematics and English/language arts achievement for students rather than on schoolwide program improvements. Funds that carry stipulations requiring particular program emphases at the expense of others can result in schools being unable to devote attention to areas targeted by CSR models. Thus, if schools cannot find permissible ways to work around such incompatibilities, they may be forced to forego certain funding streams.

The state role with regard to CSR is somewhat more complex. Because state education agencies (SEAs) were responsible for administering the federal Obey-Porter funding program, states had the opportunity to shape CSR implementation. SEA involvement in CSR was not limited to the management of this grant—the SEA takes on multiple roles in reform implementation, acting as a grants administrator, reform initiator, technical assistance provider, and gatekeeper for the selection and adoption of specific reform models in their state.

Recent research on SEA's CSR strategies suggests that states vary greatly in the ways they take on these roles. For example, when first administering CSRD funds, the Maine SEA worked to ensure alignment between state policy goals and the use of federal funds: Maine restricted the CSRD program to high schools, linked it to a statewide reform effort known as *Promising Futures,* and assigned dedicated staff in a newly-created Center for Inquiry in Secondary Education to oversee Obey-Porter and related programs. The resulting policy coherence helped to ensure successful implementation and sustainability of CSR in Maine schools, even after the Obey-Porter funding cycle had terminated (Hamann, Lane, & Johnson, in press).

California also sought to integrate CSR with other reform efforts, but did so by merging the CSRD funding stream into the public school accountability system. To support low-performing schools, California established the Immediate Intervention/Underperforming Schools Program (II/USP), through which schools with low Academic Performance Index (API) scores could apply for improvement grants.[1] One type of grant made available through II/USP was the CSRD grant under the Obey-Porter program. Hence, in California, federal CSR funds became inextricably linked with accountability, imposing a set of policy mechanisms, both positive (in some cases, an additional year of funding) and negative (punitive sanctions for schools that failed to improve) (O'Day & Bitter, 2003). California's approach, especially in association with No Child Left Behind, may be a harbinger of things to come more generally, boding well for the adoption of CSR reforms.

The series of court decisions resulting from the *Abbott v. Burke* cases in New Jersey provide another example of ways in which state governance structures—in this case, the judicial system and the governor—may intervene in school reform. In New Jersey, "whole-school reform" efforts (as they were referenced in the Abbott rulings) became intertwined in a decades-long battle to address funding disparities among New Jersey school districts. The 1998 *Abbott v. Burke* ruling (known as "Abbott V") required the state commissioner of education "to implement whole-school reform...as expeditiously as possible" in 30 of the most impoverished districts, and the Success for All (SFA) model was specifically cited as "the recommended version of whole-school reform for elementary schools." Other requirements of Abbott V included the establishment of full-day kindergarten and new technology programs and substantial new funds for school facilities. The Commissioner wrote that, "for the first time in more than a quarter century, there is consensus on an approach that clearly has demonstrated great potential to provide the state's poorest school children with the kind of education they deserve." (Klaghoz, 1998).

However, consensus regarding the potential efficacy of whole-school reform for Abbott districts dissipated in the years following the landmark ruling, resulting in spring 2002 in a freeze on new funding for schools affected by the Abbott ruling. The following year, Governor James McGreevey formally asked the court to lift the order to implement whole-school reform in the Abbott districts. Ultimately, the Abbott parties entered into mediation, resulting in a new set of regulations, *Improving Learning and Literacy in Abbott Districts: Implementing Standards-Driven Instruction, Reforms, Programs and Services under Abbott v. Burke,* issued by the state Commissioner in September 2003. Under this set of regulations, elementary schools are still required to maintain a contract with an approved whole-school reform provider, with exceptions made for high-performing schools and those with an evidence-driven case that they would be better served by alternative reform strategies.

In her ongoing assessment of reform effects on outcomes in the Abbott districts, Erlichson notes that while both Abbott districts and wealthier New Jersey districts are currently posting gains in student achievement, Abbott districts are making greater gains in language arts proficiency, gaining on average 24 points in contrast to the non-Abbott average of 15 points over the same two-year period (Erlichson & Goertz, 2001). Erlichson and others have nonetheless been critical of New Jersey's

1 The Academic Performance Index (or API) was established by the Public Schools Accountability Act of 1999 (PSAA) and is the cornerstone of California's accountability system. The API is a school-level composite indicator calculated on a scale of 1000, in which a score of 800 represents proficiency. The tests currently used to establish each school's API include the California Standards Tests, the CAT/6 Survey, and (when applicable) the California High School Exit Examination. API growth targets are established for each school based on their distance from the 800 target. Various types of rewards, sanctions, and technical assistance were also established through PSAA, to be linked to schools' API scores and provided depending on whether they achieve their growth targets.

implementation process, noting that promised funds from the state to support CSR implementation have been extremely slow in getting to Abbott schools. This delay (and other factors) compromised both the faculty buy-in process and sustainability of reform in these schools. The New Jersey Department of Education ultimately argued that whole-school reform proved less effective than expected and sought to have the court order lifted. Through mediation, the parties agreed in 2003 to allow the highest-performing schools to abandon their CSR models.

Hence, the degree to which SEAs elect to integrate CSR initiatives with other policies and programs is an important variable. Lane and Gracia (2004) point to other key state issues, including the level of resources available, SEA leadership capacity, and SEA support strategies. With regard to resources, states opt to target their CSRD funds differently; while some awarded the minimum grant of $50,000 to as many schools as possible, other states chose to give larger grants to fewer schools. In New York, for example, the average grant awarded during the period 1999 to 2004 was $141,236, while many states' averages are much closer to the minimum level. The average grant size is increasing nationally, suggesting that states are modifying their resource allocation strategies (Lane & Gracia, 2004).

Meanwhile, the stability of CSR leadership at the state level appears to vary greatly. While many states have employed three or more different coordinators of the Obey-Porter grant program, Hamann, Lane and Johnson (in press) note that Maine has had the same primary CSR coordinator since 1998. This may contribute to higher levels of program expertise, familiarity with CSR programs and implementing schools, and more effective support for those schools. As in other areas of education policy, leadership stability is central to a cohesive, sustained implementation effort. State leaders who accumulate years of experience also build their own capacity, knowledge, credibility, and familiarity with local issues.

In the area of support strategies, some states take a limited, reactive stance, restricting support to assistance in the grant-writing process or responding to specific requests from schools. Others engage in more active strategies, including the establishment of school networks, annual meetings, and provision of on-site technical assistance. As states work to establish systems of support for low-performing schools (as required under NCLB), researchers will doubtlessly devote more time to gauging the level of coordination with ongoing CSR support activities. State education policies are increasingly formulated in response to NCLB guidelines, a fact that likely has long-term implications for the implementation of CSR. Models that focus on aspects of the schooling process not measured by accountability mechanisms—such as school governance or parent involvement—may be abandoned in favor of approaches explicitly linked to reading and mathematics achievement. Indeed, in a recent study of CSR implementation in three states, Datnow (2001) notes that accountability systems are the main driver of reform, not the federal CSR program or model developers.

A parallel set of issues frames the ways in which the nested nature of schools within districts affects CSR implementation. Like states, districts may allocate funds

to support reform, maintain stable leadership, provide coordinated systems of support, and integrate policies for reform. This last aspect in particular can have a substantial impact on CSR implementation. In California, for example, district mandated implementation of the Open Court instructional program has greatly constrained individual schools' efforts to sustain CSR models with a contrasting focus, such as Different Ways of Knowing (O'Day & Bitter, 2003).

Examples at the state and district levels illustrate how federal, state, and district policy complexities can make adoption and implementation of CSR problematic. The difficulty of negotiating various demands of policymakers presents an enormous challenge to developers, implementation teams, and school and district personnel alike and can threaten the success of whole-school reform.

Building Capacity

One factor that researchers and policymakers regularly cite as critical to the success of school improvement efforts of any type—but especially CSR—is capacity. Confusing the situation is a lack of consensus as to a definition of capacity. There is, nonetheless, substantial overlap of various perspectives (Marsh, 2000; McLaughlin & Talbert, 2003; Spillane & Thompson, 1997). By "capacity," we mean here the *potential or ability of districts, schools, and teachers to undertake transformative school improvement efforts*. Capacity is not evenly distributed across districts and schools—it is not coincidental that schools such as Palo Alto High, Boston Latin, the Bronx High School of Science, and La Jolla High graduate high numbers of successful students while nearby (less affluent) schools struggle just to get students to attend. In this section, we address those factors and conditions that either enable or prevent the arrival at a threshold level of capacity by districts, schools, and teachers.

District Capacity

Researchers currently examine district capacity from a multitude of perspectives as districts play a variety of roles in supporting school improvement generally and CSR specifically (Elmore & Burney, 1997; Massell & Goertz, 1999; Spillane & Thompson, 1997). These roles are administrative (e.g., financial, compliance-related, human relations-related), operational (e.g., transportation-related), and fundamentally educational. It is on the latter that we focus. Specifically, we review findings of studies focused on characteristics of districts that have successfully supported instructional reform efforts. Each of these studies addressed in some form the central challenge posed by districts' lack of capacity, as defined by Fullan (2004). Fullan maintains that district offices traditionally have been organized into functional and operational "silos." District activities are organized around important functions, but those functions may not include teaching and learning. Curriculum and instruction "silos" may be present, but the work done and decisions made within them do not drive what most districts do.[2]

Using multi-level survey data and four-year case studies of three districts in California, McLaughlin and Talbert (2003) described how districts effectively support reform. Their findings, while not focused on CSR, are still instructive. The findings suggest how to address Fullan's "silo effect" without necessitating wholesale and impractical changes to the organization of district offices. McLaughlin and Talbert found that sharing responsibilities across levels within district offices promotes effective reform. Additionally, turnover in the superintendent's office need not be disruptive if proven communication strategies are ingrained in the district's culture. The study found that successful districts:

- *Took a systems approach to reform.* Districts undertaking successful school reform see themselves as the unit of change, not individual schools. The reforming district must develop a clearly articulated program of change, a reform theory that encompasses "all elements of [its] policy environment." While firmly establishing the parameters of district-school relations, the district must also endow its reform effort with open communication and system-wide collaboration. To this end, system-wide planning is encouraged, a practice that involves teachers, administrators, and others at all levels of the district and engenders a sense of ownership of district reform goals. By involving many stakeholders in the development of their reform strategy, districts build a wider commitment to goals and increase shared accountability.

- *Created a learning community in the district office.* Reforming districts acknowledge the reluctance of teachers and principals to enact change, both for personal and professional reasons. Administrators in reforming districts support risk-taking on the part of teachers and principals so long as those efforts are attempts to achieve effective reform. Successful districts also strive for an atmosphere of trust and honesty and encourage teachers and principals to participate in an open dialogue without worrying about being penalized for their input. In addition, district administrators themselves engage in an ongoing process of improvement, and must be willing to change their own methods of support to meet the needs of the district's schools. A district must learn to provide effective support for reform while improving its process of inquiry—that is, its tracking of schools' progress, reallocation of resources, analysis of data, and the other means by which it ensures a coherent focus on teaching and learning.

2 Of course, to talk about districts as though they are monolithic is a bit simplistic. Districts' diversity in terms of political organization, structure, and size is overwhelming (LeFloch and Kirshstein, 2003). Observers might well expect that the concentration of CSR schools within districts would affect district capacity to support CSR implementation. Early findings from the National Longitudinal Evaluation of Comprehensive School Reform (NLECSR), conducted by the authors of this chapter, do in fact suggest that the variation in district size and organization affects capacity.

- *Focused on teaching and learning.* Aggressively implementing a system-wide focus on instruction and establishing clear, specific instructional goals provide the best chance for improved student outcomes. It is then critical that districts align curricular materials and professional development activities with these same well-articulated, measurable goals. Resource allocation and the alignment of funding sources must be customized to support the district's goals for instruction, even if this entails foregoing funding with stipulations that could hinder the district's reform efforts. A carefully crafted focus on teaching and learning also allows the district to withstand potentially disruptive initiatives and accountability measures from the state and federal governments.
- *Supported professional learning and instructional improvement.* Improving district responsiveness is critical for successful reform. Being responsive to a reforming school's needs includes, among other things, providing "cutting-edge," often site-based professional development activities for both principals and teachers, hiring needed staff, and respecting a school's need to establish and follow its own unique agenda (while still staying within the framework of the district's overall reform strategy). Doing so allows district administrators to gain the trust of school-level staff. Districts should be aggressive in adjusting both their administrative organization and their budgets to improve responsiveness to schools.
- *Used data to support accountability.* Reforming districts collect and examine data to focus their reform efforts and improve reform practices. Open communication about student outcomes as well as other indicators of instructional progress is critical for a constructive system of accountability. Making data available to staff throughout the district improves communication, eases anxiety about performance outcomes, provides information that helps guide resource allocation, and increases districts' preparedness for handling state accountability measures. Improved data collection and analysis also strengthens communication and overall relations between districts and local policymakers and parents.

Marsh (2000) also found that many of these factors affect district capacity, although capacity is seen as human, social, and physical capital and understanding in her work. Arguing that districts matter to the efforts of public schools (Massell & Goertz, 1999) and can be "agents of serious instructional improvement" (Elmore 1999), Marsh points to the district role in improving student performance (Murphy & Hallinger, 1986, 1988). Restructuring districts also have the opportunity to create new and stimulating roles for teachers and administrators to support school reform (David, 1990). A school district culture may undermine or support teacher attitudes and commitment to reform (Rosenholtz, 1991). Finally, Marsh cites

Goldring and Hallinger (1992) on the influence of district coherence on performance. Knowing and understanding reforms on the part of district administrators influences how districts allocate resources for reform and attenuates Fullan's "silo" effect (Spillane, 1998; Price & Ball, 1995).

School Capacity

School capacity refers to the ability or potential of schools to fulfill successfully their central mission of teaching and learning. Whether an adopted CSR reform model is one with a highly prescriptive design (such as Direct Instruction or Success for All) or a less prescriptive and more philosophically oriented design (such as Accelerated Schools Project), the CSR model implementation process places demands on schools unlike those of reforms that do not target the *whole* school. When fully realized, CSR changes virtually every aspect of a school's organization, processes, and activities. These changes affect all instructional staff and often most, if not all, non-instructional staff. As such, even without models that emphasize devolution of management functions to teachers and other school-based stakeholders, schools implementing CSR can face substantial obstacles in making it work.

Newmann and colleagues (2000) offer a framework for understanding school capacity, particularly as capacity influences instructional quality. They define school capacity as the entire staff's collective ability to improve student achievement school-wide, organizing this concept as five interactive components. The first of these components is the knowledge, skills, and disposition of individual staff members; that is, the degree to which staff hold high expectations for student achievement and are professionally competent in instructional and assessment practices and in implementing a curriculum appropriate for their particular students. The second component, a school-wide professional community, itself comprises four elements: the staff's ability to agree on clear goals for student learning, willingness to collaborate and accept collective responsibility for achieving these goals, professional inquiry into problem solving, and opportunities to influence school activities and policies. Third, program coherence— the extent to which school programs are coordinated, focused on clear goals, and sustained over time— is a factor in capacity. The fourth dimension is the school's technical resources, including a high-quality curriculum, instructional and assessment materials, and adequate physical facilities. Last, effective principal leadership influences a school's capacity through the principal's authority to affect the other four components either positively or negatively.

Wohlstetter has addressed the issue of decentralized management from several perspectives (Wohlstetter et al., 1994; Wohlstetter et al., 1997). Most recently, she studied the role of school networks in supporting reform, including their impact on school capacity (2003). Wohlstetter and her colleagues applied Lawler's (1991) high-involvement management model to a study examining the effects of school networks created as part of the $53 million Annenberg Challenge grant program to

support school reform in Los Angeles.[3] Twenty-eight school networks ("families" in the parlance of the Los Angeles program) were created. These school networks were, conceptually speaking, analogous to both networks of schools created by some CSR model developers and also to what CSR model developers could provide. Specifically, Wohlstetter confirmed the impact on school capacity of decentralization to networks of power, information, rewards, and knowledge and skills. While the potential for school networks to increase capacity for CSR seems promising, it remains to be understood whether model developers themselves can or will consistently support such networks.

Teacher Capacity

Newmann (2000) underscores the importance of teacher capacity—knowledge and skills—as crucial to building overall school capacity and views professional development as the primary tool to develop it. The literature on teacher professional development is broad in scope and rich in lessons. The seminal work of Garet et al. (2001), although not specific to CSR, outlines the state of knowledge regarding effective and high-quality professional development. Garet and his colleagues conducted the first large-scale empirical study of the impact of teacher professional development on instructional practice. They found that three "core" features of professional development have positive effects on classroom practice and the accrual of knowledge and skills. These "core" features include: 1) focus on content knowledge; 2) opportunities for active learning; and 3) coherence with other learning activities. They also found that three "structural" features of professional development activities, when present, had positive impacts on teacher learning. These three "structural" features are: 1) the form of the activity; 2) the collective participation of teachers from the same school, grade, or content area; and 3) the activity's duration. The challenge for districts and schools implementing CSR is to provide high-quality professional development activities focused both on best practices for a given content area and on model-specific instructional strategies.

Ensuring Faculty Acceptance (Buy-in)

An important and recurring challenge in CSR implementation is ensuring faculty endorsement of CSR in general and of a chosen model in particular. If those directly responsible for CSR implementation are skeptical about reform strategies, the model's potential is unlikely to be realized and efforts at implementation unlikely to be sustained. In the early years of CSR, model developers and schools learned valuable and often painful lessons about factors related to faculty "buy-in." The les-

3 Lawler's model provides a framework for thinking about how to design organizations in a way that devolves authority and responsibility down to employees in service of greater organizational performance. The framework calls for the devolution of four resources: *information, knowledge and skills, power,* and *rewards*. In short, the model posits that organizations that provide employees with the *information* related to their organization's environment, the *knowledge and skills* to do their jobs, *power* to affect work processes, and *rewards* for productivity and quality will experience higher performance.

sons learned during these early phases of CSR implementation are relevant for those engaged in reform today.

The current literature on buy-in suggests that three elements are critical to promoting buy-in and commitment by teaching staff. These include 1) the process by which teacher support for reform is obtained; 2) alignment of teachers' perceptions of the school's main problems and the strategies proposed to address them; and 3) the amount of professional development and technical support provided by the developer.

The critical role of the reform adoption process has been investigated in studies by RAND (Bodilly et al., 1996) that clearly show how important decisions about model adoption and support for adopted reform strategies are to the reform's ultimate success. Ideally, a district's targeting of schools and initial interactions with developers should help ensure that school-level staff understand the design vision and can be committed to it from the outset. Securing an initial commitment ensures that schools (and teachers) do not feel forced into a program of change and will thus be less likely to undermine reform efforts through apathy or resistance. Early studies of the adoption process indicated that in many cases schools did not make a well-informed choice and firm commitment to the design was never obtained (Bodilly et al., 1996).

As is now obvious to many model developers, it is essential that a "critical mass" of teachers along with their principal endorse a prospective model, preferably through active participation in the selection process (Datnow, McHugh, Stringfield, & Hacker, 1998; Ross, Henry, et al., 1997). Most model developers require that at least 80 percent of the faculty support model adoption and encourage those not interested in participating in the model's implementation to transfer to another school—often an impossibility given the difficulties of negotiating transfers and hiring teachers in many districts.

Despite widespread acknowledgement of the importance of teacher participation in the adoption process, districts and schools struggle to include teachers in the process. In a four-year study of 22 schools, Datnow (2000) concluded that power relations within schools and districts were a key factor in the adoption of reform models. In some cases, pressure came from principals pushing reform to gain district resources. Teachers frequently noted that in circumstances such as these the voting process was essentially meaningless ("It was a vote, but we voted until they got their way;" Datnow, 2000, p. 362); some did not recall voting at all. In other cases, districts mandated or strongly "encouraged" reforms, using incentives or threats to guarantee adoption. In general, the adoption process in many schools reflected little if any receptivity to teacher concerns about or ideas for reform.

More recent studies echo these findings. The ongoing National Longitudinal Evaluation of Comprehensive School Reform (NLECSR) reveals that weak commitment to a design model can result from a flawed selection process. For example, both NLECSR survey results and qualitative study data suggest that principals and

teachers often do not agree about critical aspects of the model adoption process, including whether or not they voted to adopt a model. Survey data from principals and teachers indicate that while only 9.5 percent of principals said their school did not engage in a voting process prior to adoption, 46 percent of teachers do not recall having voted. The disparity between principal and teacher recollections casts suspicion on efforts to engage teachers in the adoption process.

Teachers participating in NLECSR focus groups asserted that they had approved model adoption only after experiencing some degree of coercion, generally in the form of threats from the principal that professional development services would be deliberately withheld or warnings that students' welfare and academic progress would be compromised if a model were not adopted. Even though research has documented the problems associated with insufficient faculty buy-in, widespread endorsement remains difficult for many reforming schools to achieve.

Second, the fit of the proposed model with teacher perceptions of their school's most pressing problems can also affect the level of faculty endorsement. In an evaluation of the implementation of the Coalition of Essential Schools model, Muncey and McQuillan (1996) observed that implementation was hampered when teachers disagreed about necessary changes for improving student achievement. Further, site visit data gathered by NLECSR researchers suggest that both principals and teachers were more likely to endorse models consistent with the school's prior practices. As they explained to NLECSR researchers, teachers and principals often chose models because they were "close to what we already do." Likewise, Datnow (2000) found that teachers were more likely to support models with strategies resembling those already in use at the school and with conceptions of appropriate instructional practice matching those of most faculty. One principal explained that his school selected the ATLAS model because it was the "least intrusive" (p. 361). There is a downside to this phenomenon, however. While ideological compatibility is liable to bring about teacher buy-in, schools are less likely to undertake the aggressive changes needed to improve student achievement if a model whose components too closely mesh with existing schemes is selected.

In addition to the adoption process and the model's ideological compatibility, there is a third factor important to the buy-in process—the amount and type of professional development provided. Provision of high-quality professional development can bolster faculty buy-in during the initial stages of adoption. In a study of the implementation of pilot instructional programs in mathematics and English developed by the College Board, researchers at the American Institutes for Research concluded that high-quality professional development was a catalyst for teacher buy-in, even among teachers who initially were reluctant to support the program (AIR, 2003).[4] In their highly detailed study of National Science Foundation reforms in mathematics and science and the impact of these reforms on schools in four major districts, Borman and her colleagues (K. Borman et al., 2004) determined not

4 The College Board programs mentioned here were "Textual Power" and "Mathematics with Meaning."

only that professional development promoted buy-in, it also sustained the implementation of the reforms and contributed to student achievement in mathematics as measured by high stakes tests in the two districts—Chicago and Miami—where student level data were made available to the researchers.

This aspect of CSR often depends on the capacity of a model developer to provide ongoing support during the course of model implementation. Designers of locally developed models may not have the ability to provide extensive professional development, sophisticated technological tools, and tailored technical assistance. Some research suggests that a model developer's flexibility and willingness to allow adaptation of the model's professional development activities in response to teacher feedback (or negative outcomes such as poor student academic performance) may be an important factor in building support during the implementation process. Evaluation work by Howard Bloom and his colleagues (2001) provides an example of the importance of developer flexibility in implementation. Bloom's team conducted a case study of the Accelerated Schools Project (ASP) that specifically examined the impact of ASP on student achievement outcomes under commonly occurring detrimental conditions such as high student mobility. By monitoring achievement over time, the researchers were able to notify developers that an overemphasis on governance as opposed to "powerful learning" was sabotaging ASP reforms, based on third-grade achievement scores (the outcome measure used by the evaluators). By making "operational refinements" that involved increasing the emphasis on curriculum and instruction in staff training carried out by model developers, ASP schools saw increased academic performance. The developer's flexibility underscored its commitment to the school's individual needs, and strengthened teacher buy-in. The formative nature of the ASP study permitted developers to meet the challenges of faculty buy-in and model refinement.

Moreover, data from NLECSR case studies emphasize the importance of sustaining support on the part of teachers. Preliminary findings suggest that teachers who join the faculty of a school already implementing a CSR model must be promptly included in training designed to introduce them to the various model components, classroom instructional content, and ways that pedagogy is aligned with key model features. Without this training, new teachers are unlikely to understand or support the reform effort fully. Hence, appropriate technical assistance and professional development are critical to sustaining teacher buy-in throughout many school-year cycles of the implementation process.

The research conducted by Bodilly and her colleagues (1996) reflects the themes reviewed in this discussion of faculty buy-in. In their formative evaluation of the ten original New American Schools models, they differentiated among three types of CSR model designs including those that were team-developed, extensively developed at the site, and more limited in development at the site. Team-developed models (Accelerated Schools and Success for All/Roots and Wings) were implemented faster and more effectively than others. However, implementation was influenced by a number of factors, many beyond the control of developers. They included:

- Struggles to combine existing organizations with strong cultures into a new partnership;
- Difficulties in translating abstract notions into specified design concepts and processes for implementation;
- Difficulties in quickly building the staff needed to fully develop the design and assist the sites in implementation; and
- Problems caused by inadequately developed designs resulting in confusion at the sites in the first year (Bodilly et al., 1996).

Developers of at least one of the designs (SFA/Roots and Wings) had the opportunity to interact with personnel at their sites fully one year in advance of other developers' entry into their schools. These findings draw attention to the importance of investing time and human resources into the process of providing on-site support to those in the reforming schools.

Assuring faculty buy-in is critical to the successful implementation of comprehensive school reform models. Faculty input into choice of the model is a significant first step. Teachers who feel coerced or who believe that they have no voice in deciding whether to adopt and which model to choose are more likely to resist active participation in implementation. Further, models whose various components are in sync with teacher perceptions of school needs may be easier to implement if adoption has been openly negotiated. Finally, the importance of providing adequate and sustained technical assistance and professional development must not be overlooked.

Garnering Parent Support

Researchers, practitioners, and policymakers endorse the general view that parent involvement is important for students and that parent support is a key element of effective schools. The Hawkins-Stafford Amendments of 1988 mandated increased attention to parental involvement in Title I schools and the No Child Left Behind Act of 2001 calls for a continued emphasis on the role of parents in the educational process. NCLB's provisions for parental notification and choice for students attending "failing" schools are based on assumptions that parents can be important actors in the process of stimulating change in schools.

Studies on parental involvement, parental support, and community engagement are often contested on the grounds that they lack rigor or sometimes include findings that do not lend themselves readily to constructive policy recommendations. In their review of 41 evaluation studies of parental involvement, Mattingly, Prislin, McKenzie, Rodriquez, and Kayzar (2002) found only limited empirical support for the claim that parental involvement in schools leads to increased student achievement. Most disappointing was the finding that a majority of the measured outcomes did not show a significant improvement in studies that used the most

stringent criteria, a statistically significant trend, "suggesting that the purported effectiveness of parent involvement programs is an artifact of weak evaluation methods" (Mattingly et al., 2002).

Much recent research on parent involvement in schools is couched in a framework articulated by Epstein (1996, 2001). Epstein identifies six important types of interaction between parents, schools, and the community, including: activities that support parenting, strategies that support communication, volunteer activities, activities that encourage learning at home, opportunities for involvement in decision making, and collaboration with the community. Of these, studies most frequently identify parent support for learning at home as the interaction type with the most positive consequences for student achievement, attendance, and classroom behavior (Epstein, Simon, & Salinas, 1997; Jordan, Snow, & Porche, 2000; Shaver & Walls, 1998). Much of the existing literature also focuses on attitudes of school staff and how staff can be encouraged to overcome barriers to parental involvement (Delgado, 1991; Epstein, 1990; Nord, 1998).

Despite conflicting opinions about research "quality," a loose consensus argues that parents' involvement in their children's schooling improves student achievement, especially when such involvement is centered around supporting children's academic (intellectual) growth. Increasingly, schools are expected to reach out to their communities with programs intended to intensify parental involvement. In fact, increasing parental involvement and support is a specific element of many CSR models. ATLAS Communities, Modern Red Schoolhouse, Turning Points, and Urban Learning Centers are all models that emphasize a central role for families in schools and the educational process. Yet questions remain about how to engage parents in ways that optimize children's success in school and sustain school improvement efforts.

CSR model developers generally expect parents and the community to play an important role, not only in supporting individual student achievement but also in supporting reform efforts more broadly. Indeed, evidence from several studies suggests that the participation of adults in classrooms contributes to the overall success of CSR. Positive outcomes occurred in large part because adults reinforced teachers' efforts either by being engaged as learners themselves or by assisting the teacher in carrying out instructional tasks. In their policy recommendations, these researchers urged the continued development of professional networks of teacher-adult colleagues as an essential first step in creating a strong sense of community among all school participants. The recommended objective is to create a school community with a shared vision, values, and purpose; trust and teamwork; respect and recognition; and, perhaps most important, incorporation of diversity (Stringfield & Rossi, 1995). While these objectives are desirable, it is not clear how schools or specific reform models can easily achieve them.

Virtually all CSR models stipulate that an increased emphasis be placed on community and/or parental involvement, although the level of "prescriptiveness" varies greatly among models. For example, the Accelerated Schools Project (ASP)

emphasizes effective school governance and the involvement of key community stakeholders in instigating whole-school change. ASP is anchored around three guiding principles (unity of purpose, empowerment plus responsibility, and building on strengths), addressed through a collaborative approach to the schooling process—one that includes parents. As a more process-oriented model, ASP did not mandate a set of pedagogical techniques or curricula but rather urged schools to reinforce constructive relationships with family and community members so that students benefit from a wide variety of resources and opportunities. Over time, the developers' instructional emphasis shifted to "powerful learning" and more rigorous academic standards, although the involvement of all school community members in "taking stock of their current status and setting priorities for action" through a shared governance structure has remained a central feature of the model (Bloom, Rock, Ham, Melton, & O'Brien, 2001). Hence, while some model aspects shifted to meet evolving school needs, the central role of parents and the community remained unchanged.

On the other hand, Success for All/Roots and Wings, a far more "prescriptive" model, relies on a school-based facilitator to coordinate virtually all school functions, including building relationships with families and communities. SFA schools are also expected to create a Family Support Team. These teams include members of the school administration, the SFA facilitator, school social workers/counselors, and others such as attendance monitors, teachers, and volunteers. Developers identify four areas in which family involvement can support SFA reform efforts and achievement: 1) assuring students' regular attendance in school; 2) enhancing the efficacy of school-based interventions; 3) providing support for learning (such as reading at home); and 4) coordinating the educational, health, and social services of schools and other municipal, state, and private agencies. SFA schools are also encouraged to recruit community members who might have specialized knowledge and skills in particular areas or vocations (auto mechanics, for example) to visit schools and talk with students (Borman & Burke, 2003). Preliminary analysis of NLECSR site visit data suggest that of the eight models being studied, schools implementing SFA exhibited the most consistent evidence of parents' awareness of SFA program components and school involvement.

As noted earlier, much of the parent involvement literature discusses the importance of helping students learn at home, and research that focuses specifically on parent involvement in the CSR context appears to support this general trend. NLECSR survey data point to the importance of parent involvement in the academic process (as opposed to more administrative processes within a school): preliminary analytic models show a significant relationship between measures of stakeholder understanding of CSR and scales that gauge parent involvement in academic activities. The study's survey scales distinguish between parent involvement in academic activities and parent involvement in more general terms; only the former scale exhibits statistically significant relationships.

This preliminary finding is similar to results obtained in the Borman (2003)

meta-analysis of CSR effectiveness, to be discussed in detail in the next section. The Borman study's findings are clear and conclusive with regard to parental involvement:

> The one reform attribute that was a statistically significant predictor of effect size suggested that CSR models that require the active involvement of parents and the local community in school governance and improvement activities tend to achieve worse outcomes than models that do not require these activities. In contrast to school-based efforts aimed at helping families enrich their children's learning opportunities outside school, which are far more likely to help individual children succeed with specific academic goals, the focus on parent involvement in school governance could sidetrack schools if the immediate priority is to improve achievement. (Borman et al., 2003)

The Borman analysis underscores the importance of focusing parent involvement on specific academic activities rather than channeling their efforts toward other components of the Epstein framework, such as parenting, volunteering, or decision making. These findings, as well as those of Bloom and colleagues (2001), K. Borman and Burke (2003), and preliminary NLECSR survey results frame a central challenge for CSR: Should those implementing CSR models continue efforts to engage parents in school governance structures, as many encourage, or should they focus on engaging parents in students' academic lives exclusively? As CSR models continue to evolve, those concerned with their implementation must determine the most appropriate role for parents and community in the reform and schooling processes.

Fidelity to a Research-Based Plan

A defining characteristic of CSR is the primacy given to the use of research-based approaches to every aspect of the schooling process. This section briefly describes developers' and schools' challenge of establishing fidelity to a research-based plan.

The 2001 No Child Left Behind Act defines CSR models as reform programs possessing 11 key components that must be present for a reform model to be considered both comprehensive and research-based (please see Chapter 1, page 10) for a discussion of these components). These 11 components are important to review here because NCLB calls for using them as indicators to measure the programmatic adequacy of reform models. In other words, this list defines the minimum requirements for every CSR model currently launched in U.S. schools. Implicitly, NCLB requires almost all schools (certainly most urban and high-poverty

schools)—whether or not they receive federal CSR funds—to judge their performance against the requirements of these components. Reforming schools receiving CSR funds in particular must meet the criteria set forth within this list to be in compliance with federal law.

It is an open question, however, whether every CSR school could in fact demonstrate that all these elements are present. A recent meta-analysis of research studies on CSR's effectiveness in boosting student achievement carried out by Borman, Hewes, Overman, and Brown (2003) reveals a shortage of evidence permitting developers to make even modest claims about the performance of their models. Moreover, it is clear that there is wide variation in models' effectiveness in cases where evidence can be mustered. Both Borman's work and Herman's (1999) review suggest that among many CSR models in operation, very few qualify as research-based. Thus a major problem facing CSR and policymakers is the lack of a proven research base.

The research base that does exist presents a number of problems. In their analysis of findings from studies on the 29 most widely disseminated CSR models (see the G. Borman findings in Chapter 2, page 80), Borman's team of investigators was frustrated by the scarcity of reliable and useful research. They took into account several factors in determining each study's quality: 1) who reported the findings (i.e., the model developer or someone else); 2) the methods used (e.g., pre-test/post-test comparisons, experimental comparisons, or nonequivalent control-group designs); 3) the school- and student-level context (e.g., high-poverty versus lower-poverty settings); 4) actual characteristics of the CSR models being studied (e.g., the costs associated with a model or the developer's level of support for implementation; and 5) indicators of the model's effectiveness (e.g., test scores in reading, math, science, or some other subject) (Borman, Hews, Overman, and Brown, 2003). In addition to weighing methodology, research design, and data quality, the team also considered model-specific influences, including such aspects as whether or not a given model was highly prescriptive or lacked specificity.

Results of the meta-analysis illuminate the extent to which high-quality evidence for evaluating CSR outcomes is lacking. Only a handful of models met the researchers' most stringent tests—showing strong evidence of effectiveness in improving student performance based on research studies that themselves met high scientific standards. The CSR models showing "strongest evidence of effectiveness" included Direct Instruction (DI), the School Development Program, and Success for All (SFA). Two of these—DI and SFA—are highly prescriptive designs, while the third, the School Development Program (also known as the Comer Schools program), tries to be highly integrative, meaning to develop the physical, cognitive, psychological, language, and social and ethical dimensions of each child. An additional set of models, Expeditionary Learning/Outward Bound, Modern Red Schoolhouse, and Roots and Wings, present "highly promising evidence of effectiveness" based on the research. Most of the 29 models, however, were plagued by a lack of useful studies and therefore lacked a strong research base.

Inclusion of Special Education Students

Comprehensive school reform means not only a reengineering of all school processes and activities, but also making challenging instruction available to all students in all grades, including general and special education students. Two groups of special education students historically have not been served by reform initiatives: English language learners and students with disabilities. The "philosophies" of most comprehensive school reform models presuppose that all children, regardless of any disabling conditions, should learn in the same environment and be expected to meet the same standards as all others, an assumption also made by No Child Left Behind. At the same time, most popular models fail to specifically address the needs of or strategies for these students, (Hamann, 2001).

The overarching question for educators and policymakers is whether including special education students in CSR is or is not an effective approach, and whether certain conditions imposed by the district or school can make this approach effective in the absence of specific strategies espoused by model developers. This section first looks briefly at research on the effects of CSR model exposure on ELL or ESL students (students whose first language is one other than English) and second considers the effects of CSR model exposure on students with disabilities.

Some CSR models do give attention to the treatment of special education students (for example, Success for All), but most allow schools to make independent decisions about the extent to which these students will be included in model-specific instruction and activities. Because there is variance in the ways that special education students are involved in reform activities, there are varying outcomes for these students in CSR schools (RMC Research Corporation, 2003). A review of state applications for federal CSRD funds by Hamann and colleagues (2001) concluded that the CSR movement in general overlooks ESL students and that state educational agencies (SEAs) bidding for CSR funds "promise only procedural compliance rather than active investigation and resolution of problems [that] ELLs encounter." This same study observed that seldom, if ever, do states depart from the traditional approach to inclusion—designing or adopting a reform model and then adapting it however possible to the needs of ESL students.

Few studies are available on the impact of CSR models for special education subpopulations. Recent research by Datnow and associates (2003) investigated whether certain well-known CSR models are more effective than others at improving the achievement of students in linguistically or culturally diverse settings and whether certain policies at the state, district, or school level enhance the effectiveness of CSR models in these diverse schools. Datnow and her collaborators undertook a mixed-methods study of 13 urban public elementary schools to pursue these questions. A qualitative component consisted of over 300 interviews with teachers, administrators, model representatives, parents, and students to capture a realistic impression of how CSR model guidelines were practiced at the classroom level. A quantitative component compared achievement scores of students in CSR schools

with non-CSR comparison schools over a four-year period. The data were disaggregated to determine whether CSR model schools experienced more significant gains than others and whether ESL students also posted higher achievement gains as a result of their involvement in CSR.

The study's findings were organized around three key areas: 1) state and district polices that either helped or hindered reform and its sustainability; 2) school-level adaptations made by teachers that made the reforms more helpful in meeting the needs of ESL students; and 3) assessment outcomes for reforming schools and their subpopulations of ESL students. In general, for schools with high proportions of ESL students, including such students in the reform is difficult because many state- and district-level policies dictate specific instructional practices for these students. Requirements posed by districts frequently came into direct conflict with requirements of model developers. For example, teachers were required by the district to participate in a certain number of hours of ESL professional development, which precluded participation in an adequate number of model-specific professional development sessions. In another instance, district requirements that students participate in pull-out classes for ESL instruction meant that students missed classroom time devoted to coverage of topics prescribed specifically by the model curriculum.

Some schools were able to make successful adaptations to reform activities that enabled them to reach ESL students ordinarily beyond the reach of the model's scope. One school took advantage of the Core Knowledge Model's flexibility in implementing curricula to translate materials into Spanish. Still other schools struggled to make use of mandated materials written in English when a majority of students spoke another language as their first.

Implementing CSR yielded mixed results as to whether teachers experienced increased sensitivity to multicultural, multilingual students. Generally, teaching a rich reform curriculum that introduced students to a variety of cultures and world traditions resulted in teachers becoming more aware of their students' heritage and historical background. On the other hand, some teachers were discouraged that the reform developers prohibited certain activities they had undertaken independently to introduce students to Native American history, for example. Ultimately, the study concluded that teachers' attitudes toward multicultural students and students' abilities affected the extent to which students made progress under a reform model.

Although assessment data reflected no significant increase in achievement test results for ESL students in reforming schools, the study concluded that better model implementation correlated with better outcomes for ESL students overall. Predictably, the researchers also found that turnover in district leadership—and changes in instructional priorities—resulted in decreased support for CSR and "decisions to abandon reforms coincided with…changes in the broader policy context" (Datnow, 2003). A number of schools abandoned reform efforts as a result of decreasing support for CSR at the district level, emphasizing the importance of aligning district resources to support CSR schools and ensuring that districts are

capable of making the systemic changes necessary to aid reform.

In general, even for those models that specifically address inclusion of English language learners, there is inadequate research on whether the reform models result in increased achievement for such students. This is due in part to the fact that there is no consensus on the most reliable means of assessing such students (Menken, 2001). Because of their limited proficiency in English, students who are administered English language tests are tested primarily on their capacity to comprehend the meaning of test items included in these measures as opposed to subject matter knowledge.

If research on inclusion of ESL students is scarce, there appears even less research on inclusion of students with disabilities in reform model activities. A study by Purnell and Claycomb on Success for All, one of the more popular and widespread CSR models and thus the subject of a significant proportion of CSR research, concluded that although SFA developers did not make specific provisions or provide specific professional development related to students with disabilities, teachers overcame ingrained habits of isolating and tracking disabled students and tried to include them in model activities. However, in doing this, they often diluted materials or decreased the difficulty of assignments and thus failed to practice the model's standards for rigor. SFA disavows separate special education programs, but the study revealed problems that inevitably result from implementing a reform model designed with general education students in mind. At the outset, special education teachers were unsure about their inclusion in implementation efforts. Ultimately, these teachers came into conflict with general education teachers over how special education students would divide their time between participation in model activities and more focused efforts to teach them as a special population.

Since research suggests that whole-school reform benefits all students, comprehensive school reform holds great promise for students with disabilities (McLaughlin et al., 2001). Past approaches to special education instruction (along with other, equally significant factors) have resulted in the disproportionate representation of minority students in such programs, a problem likely to be exacerbated as American schools grow more diverse. As McLaughlin and her colleagues assert, "This problem will only be solved through rethinking the notion of culture that informs programs and reforms and through the design of a new and more flexible system of services focused on helping each student reach his or her highest level of achievement."

Inclusion of Racial and Ethnic Minorities

Students belonging to minority racial and ethnic groups (in particular African-American, Latino, and Native American students, as well as students in immigrant families) are affected in relatively large numbers by CSR because of their prevalence in urban, low-performing schools where reform models are frequently adopted. Nonetheless, few mainstream CSR models specifically address instructional needs of these students as a group, and few research studies have measured CSR's impact on minorities in more than a very small set of schools. Just as many model developers believe that comprehensive reform should engage all students equally, regardless of disabling conditions or language proficiency, others do not address race and ethnicity as factors in designing and implementing instructional strategies. Cooper and Jordan explain that "there is a widespread philosophy among school reformers that effective education should be culturally neutral, [as if] the broad issues pertaining to school reform and improvement... [are] devoid of cultural implications" (Cooper & Jordan, 2003).

On the contrary, however, for CSR models to be truly effective in improving education for all students, developers and school staffs implementing CSR must 1) be cognizant of the varied cultural backgrounds and values of racial and ethnic minorities and 2) work to change prejudicial beliefs about and low academic expectations for minority students. Prior to CSR reforms in the late 1980s, most targeted reform programs that benefited racial minorities did so by supporting their achievement at basic proficiency levels rather than offering them challenging and engaging coursework (The College Board, Feb 2000). CSR, by applying the same rigorous instructional strategies and programs to all students within a school, potentially can bring more minority students into the upper tier of high achievers so long as those strategies are accompanied by high expectations by teachers.

While current research does not in every case call specifically for multicultural curricula, efforts by teachers to understand their students' cultures and interact meaningfully with students from different backgrounds other than their own are important in making CSR successful (Cooper & Jordan, 2003). The previous section of this chapter pointed out that non-English speaking students struggle in classes where the teacher is poorly informed about those students' cultural backgrounds and/or provides little instructional content that introduces students to multicultural, multilingual concepts. Similarly, some researchers argue that African-American, Latino, and other minority students perform poorly because the school curriculum contains few or no lessons that relate to their own lives and culture (Banks, 1992). Although others contend that culturally-based curricula simply distract from the problems of inner-city poverty that undermines African-American achievement (e.g., Ginwright, 2000), some studies indicate that aligning aspects of the curriculum with students' cultural backgrounds results in academic benefits for minority students (Boykin, 1996 and 2000). There is growing evidence that most students, but particularly Latino students, benefit from instructional strategies that are stu-

dent-centered and organized to encourage students to work together to solve problems in math and science and to work together collaboratively (K. Borman et al., 2004). At the very least, helping minority students understand their own culture by discussing role models, customs, and historically relevant literature can remove bias and strengthen relationships between teachers and minority students.

Many researchers claim that CSR cannot be successful in the first place—regardless of any curricular focus on minority students—if districts and schools do not work to overcome entrenched bias against students of color and those belonging to ethnic minorities. Among these researchers, Fine (1991), Lipman (1998), and Anyon (1997), all argued to one extent or another that educational leaders, as part of privileged institutionalized hierarchies, too often ignore the systemic poverty and alienation of minority students and their families and fail to address some teachers' prejudicial beliefs about the ability of African-American students as well as students from first- or second-generation immigrant families. The legacy of segregation and racial oppression still affects the confidence level of minority students. When districts and schools implement CSR, they must use the restructuring process to raise student expectations. In other words, restructuring must involve changing "norms" that hold minority students themselves responsible for low achievement in cases where their failure may be the fault of "cultural aversion"—teachers' reluctance to address issues such as equality, discrimination, and social justice (Cooper & Jordan, 2003) and to provide assistance to students they believe are incapable of thriving academically.

CSR holds promise for students who are members of racial and ethnic minority groups to the extent that reform models institute broad-based changes that potentially improve schools and districts both in terms of instruction and governance. When CSR can raise teacher quality, increase parent involvement in students' academic lives, and more sensibly distribute financial resources, it is strengthening those factors that research shows improves minority student achievement. For CSR to have an impact on students of racial and ethnic minority status, CSR model developers must impress upon district and school staff the need to be aware of students' cultural backgrounds and increased sensitivity to frustration resulting from racial and ethnic prejudice. Too often, classification of minority students as "at-risk" results in a lowering of expectations on the part of teachers accompanied by a belief that socioeconomic status predetermines a child's academic outcomes (Flaxman et al., 1998). CSR can help close the achievement gap when it encourages schools to incorporate aspects of minority students' cultures in instructional content and change long-standing perspectives on student achievement through which these children are expected to fall well short of academic standards.

Sustainability of CSR

Apart from meeting challenges associated with launching CSR, districts and schools face challenges in sustaining implementation efforts over a period of several years. We next consider three aspects of sustaining CSR: costs, capacity, and continuing CSR research using varied methods.

Understanding the Costs

Providing an education to America's 48 million public school students is a costly endeavor for the schools and school districts that serve them (NCES, 2003). Funding comprehensive school reform (CSR) models becomes particularly expensive for schools and districts in urban and low-income areas where models are most effective (Borman et al., 2003). Schools undertaking reform face the daunting task of selecting from among dozens of models from both nonprofit and commercial developers requiring the purchase of new curricular materials and customized technical assistance as well as reallocation of staff and existing resources. Wholesale changes such as these can all involve considerable expense.

Rather than adopt one of many externally developed models, some school districts create their own reform model unique to that district's needs and capacity for reform. While these types of reforms seem less expensive than independently developed programs, comparing externally developed and locally developed comprehensive reform programs is problematic. Although upfront, "startup" costs may be lower for locally developed programs, the overall amount may equal to or actually exceed that of programs purchased from external developers. Unlike discrete, piecemeal reform efforts, which might involve relatively affordable changes to one particular aspect of the school enterprise, comprehensive school reform involves other costs commensurate with the magnitude of change being undertaken. Schools and districts must estimate the cost of such an undertaking before deciding to commit staff and resources to implementation of a comprehensive school reform model.

Crucial questions facing educational practitioners include how to estimate the costs associated with comprehensive reform and how to secure funds needed to pay for it. This section discusses attempts by education researchers to estimate costs associated with comprehensive school reform. It also highlights recommendations by these researchers on how adequate funds can be procured by administrators based on how schools and districts have successfully funded comprehensive school reform in the past. Finally, it closes with an examination of some theoretical problems dogging the question of comprehensive reform costs.

While costs of CSR models vary from one developer to another, certain core costs remain roughly the same. Allan Odden, in a widely-disseminated analysis of comprehensive school reform implementation, studied the costs of implementing CSR models developed under the aegis of the New American Schools (Odden,

2000).[5] Odden identified two types of costs among these models: ongoing costs, recurring year to year throughout the implementation of a reform model; and "other" costs, often one-time or "startup" costs that do not necessarily recur annually. Ongoing costs, which include salaries and benefits for teachers and principals, can be covered without additional funds from outside sources—schools can simply reallocate existing funds used to cover these expenses. Other types of costs supporting features unique to each reform model can be significantly higher.

Odden provided a specific estimate of what comprehensive school reform implementation will cost in an "average" school. (In Odden's analysis, an average school is one with 500 students, 1 principal, and 20 teachers).[6] The "core" costs of any reform program are salaries for principals and classroom teachers, and in Odden's model total $1.07 million. Additional costs, including salaries for instructional facilitators, education specialists, intervention strategies, professional development, parent outreach, and technology result in a "base total" of $1,605,000. The addition of tutors for high-poverty schools (required by the Roots and Wings model), an additional facilitator, and a family health team raise the total to $1,955,000. These last two features are required by some, but not all, of the NAS models. One of Odden's conclusions is that the costs of model-specific features vary widely and can be prohibitively expensive for many schools.

Odden's analysis, while thorough, contains some assumptions that warrant closer examination. For one, he presupposes that education specialists—individuals teaching art, music, physical education, languages, and other similar "special" subjects, are not crucial to a school's overall instructional strategy and could be reassigned indiscriminately for the purpose of meeting staffing demands of various comprehensive school reform models. This may well be possible in some jurisdictions, but may not be in others. In an effort to assuage model developers who insist that such positions should not be eliminated to make room for a CSR model, Odden adds the cost of four education specialist salaries to his estimated cost structure, a total of $200,000 for each model. The result is the high cost of implementation of certain models above and beyond the "core" costs.

Another important point not addressed in Odden's framework is the fact that comprehensive school reform model developers often make use of varying pricing and servicing strategies; in effect, they will negotiate with individual schools and districts to arrive at a total cost within the school's or district's budget for reform. For example, some models adjust the duration and frequency of training in the

5 Although dozens more have been developed and marketed, NAS endorses ten comprehensive school reform models. They are: Accelerated Schools Project, ATLAS Communities, Co-nect Schools, Different Ways of Knowing, Expeditionary Learning Outward Bound, The Leonard Bernstein Center for Learning, Modern Red Schoolhouse, The National Institute for Direct Instruction, Turning Points, and Urban Learning Centers.

6 This model is based on a 1 to 500 principal-to-student ratio and a 1 to 25 teacher-to-student ratio. Odden arrived at these ratios in consultation with developers of models sponsored by the New American Schools Development Corporation.

model's content and approach provided to school staff to realistically reflect a school or district's budget. This is especially critical in cases where schools or districts are relying on federal CSR grants to fund reform. Although most models have a preferred amount or level of training, a model's staff can simply scale back professional development sessions to make them affordable.

Similarly, some models do not mandate costs above and beyond the price of core reading programs or other necessary materials. Once schools have purchased the required materials, they are free to use the remainder of any money earmarked for comprehensive school reform for other expenditures. Such negotiations can dramatically affect the overall cost to schools and districts of implementing comprehensive school reform.

Schools needing funds above and beyond those available prior to pursuing comprehensive school reform have several options for obtaining additional money. Probably the most vital funding source is the Comprehensive School Reform Program, formerly known as the Comprehensive School Reform Demonstration Program, passed by Congress in 1997. The program's purpose is to "stimulate schoolwide change covering virtually all aspects of school operations." The program was modified under the No Child Left Behind Act of 2001 and its total funding increased to $310 million, more than double the original amount of $145 million in 1998. Grants are awarded to states for disbursement to school districts adopting a federally-approved comprehensive school reform model with components consistent with federal guidelines. These models are expected to meet stringent criteria set forth in the CSR program requirements, including provision of high-quality technical assistance delivered by individuals with expertise in instructional and schoolwide improvement strategies.

Schools often assume that monies received under other federal grants can be transferred to comprehensive school reform efforts. This is seldom the case, and such transferability is prohibited under the 2001 No Child Left Behind Act. While funds can be transferred from certain programs to certain other programs, this option is not available to schools and districts implementing CSR. Schools must seek other sources for funds if they expect to depend on moving funds from non-Title I programs to a CSR effort.

Reallocation of existing resources is a crucial consideration for schools and districts undertaking comprehensive school reform. Wholesale restructuring of schools necessitates restructuring financing schemes as well, and schools often are able to uncover funds being used ineffectively or inefficiently on current school operations. District administrators play crucial roles in these types of efforts to expedite the implementation of comprehensive school reform. Besides making changes to district-level spending plans and shifting district-level resources to help schools, it is also important for them to allow schools some latitude in meeting district goals and aligning their academic priorities with those of the district. Further, districts must often give more control to principals and school leaders to best facilitate implementation. Karen Miles (2000) identifies three crucial strategies that dis-

trict administrators can take to ensure successful implementation: 1) re-align spending levels and patterns to support comprehensive school reform designs and academic priorities; 2) support and encourage schools in restructuring their resources to support CSR and academic priorities; and 3) redesign district practices to give schools and principals more control over the use and organization of their resources (Miles, 2000).

Miles recommends that districts take steps to secure comprehensive school reform funding. Districts characterized by disproportionate spending on schools, districts struggling to provide a basic level of high-quality education, and districts with spending patterns out of sync with strategies espoused by CSR models may need to postpone adoption and implementation to prevent false starts or budget crises. Districts also must help schools restructure their instructional schemes. Miles makes six specific recommendations involving staffing reassignments, increased focus on literacy, increased use of part-time and adjunct teachers, and reduced reliance on pull-out programs. Most of these recommendations are called for by one or more of the CSR models sponsored by New American Schools. Last, and perhaps most importantly, districts should allow schools autonomy in two critical areas: hiring and budget. Schools should also be permitted to "recruit and retain the right people for their design and school" and to reallocate funds from staffing to materials or training as necessary. Miles also observes that reallocating funds to schools in support of comprehensive school reform requires two steps, both reflected in Odden's analysis: districts must find funds to cover the "startup" costs of CSR (including purchase of materials, etc.) and then ultimately must revise budgets to reflect the cost of sustaining CSR over a period of several years.

There is evidence that districts are providing the necessary flexibility, at least with respect to budgeting, for schools undertaking CSR. Surveys completed by administrators in school districts with CSR implementation underway revealed the extent to which resource and funding adjustments occur in support of CSR. In year one of the National Longitudinal Evaluation of Comprehensive School Reform (NLECSR), 59 percent of district administrators surveyed indicated they had "redesigned budget systems to provide school sites with budget autonomy." Seventy-eight percent indicated that they had "encouraged schools to redeploy dollars from federal and state categorical programs" and 69 percent revealed they had "realigned district spending levels to better support CSR." In year two of data collection, the figures were 65 percent, 73 percent, and 63 percent, respectively.

As Miles suggests, securing abundant resources is not always as important as ensuring that existing resources are allocated in such a way as to maximize their potential. Research demonstrates that the use of funds impacts student achievement in direct and indirect ways. This section concludes with an examination of two theoretical approaches to the study of funding in schools.

From the mid-1960s with James Coleman's 1966 Equality of Educational Opportunity Survey (EEOS), until recently, production function analysis was the primary tool used to investigate education finance. Production function analysis in

education involves the application of an economic model to schools and school resources. Inputs—the things purchased by schools to be used in the teaching process—directly relate to outputs; in this case, student achievement, and the relationship between inputs and outputs can be discerned through the application of correlation or multivariate analyses (regression analysis).

Production function analysis became discredited over the course of the 1980s and early 1990s. Few studies using this approach arrived at statistically viable conclusions about whether the production function could reliably correlate education inputs with outputs. Opponents of the methodology argued that schools were individually unique and idiosyncratic and such an approach lacked trustworthiness. In a 1989 review of more than 100 production function studies completed over a 20-year period, Eric Hanushek concluded that no reliable connection existed between student achievement gains and a series of pre-determined "inputs." Because the production function method was created for the purpose of predicting profits resulting from tangible inputs (capital and labor) in industry, it was not useful for predicting outcomes that depended on non-material things, such as teacher quality (Hodas, 1993). Although a meta-analysis conducted by Larry Hedges (1994) and a similar study by Faith Crampton (1995) partially resurrected the approach, production function research has received less attention over the past ten years.

At the heart of the debate over the viability of the production function approach was the question of whether or not increased education expenditures had any effect on student outcomes (Picus, 1995). In other words, did more public money spent on schools always, sometimes, or never bring about better student achievement gains? This debate continues to play out in education research circles, with a growing focus on the allocation and specific uses of monetary resources to improve teaching and learning.

As a successor to the production function approach, Cohen, Raudenbush, and Ball (2001) developed a new model for assessing education resource outcomes whereby resources, monetary and otherwise, are "amplifiers" or "constraints" for instructional regimes. Put another way, school resources only indirectly affect student outcomes. Put to wise and effective use in designing and implementing instructional strategies, resources can positively affect achievement gains. The authors argue that how money is spent—specifically, the quality of materials purchased, the experience and skill of the teachers hired, and the effectiveness of professional development—crucially affect the extent and depth of student learning. Coordinating teaching and learning is less difficult in environments with coherent organization and guidance for instruction, and more difficult in those that lack such coherence" according to the authors. This paradigm seems a fitting vehicle for analyzing the results of comprehensive school reform model implementation.

Comprehensive school reform can be a high-cost endeavor for schools, but results reveal that the effort is worthwhile. As the research base on CSR outcomes grows, evidence accumulates that whole-school reform brought about by CSR has strong positive effects on student achievement scores. Funds for CSR implementa-

tion may be elusive, but they are available, and CSR can have powerful effects when districts allow schools autonomy in establishing budgets and setting goals. School and district leaders need to embark on a careful consideration of the costs involved, and ensure before making a commitment to CSR that the prospects for cooperation with and support for schools are such that successful, sustained implementation will be possible.

Developing Capacity

Earlier, we described factors associated with building capacity at the district and school levels. In this section we address factors that promote the maintaining of capacity at the district and school levels. Recall that this chapter defined capacity as the potential or ability of districts and schools to undertake transformative school improvement efforts. Developing capacity refers to districts' and schools' keeping that potential or ability replenished over a long period. Clearly, there are many challenges to overcome in developing capacity. We identify four interrelated ones here: 1) surmounting the instability of model designs; 2) securing commitment to reform (by districts, principals, and teachers); 3) navigating an evolving and varied context (multiple reforms) and appreciating limits on what districts can support; and 4) determining and/or negotiating the extent of a developer's support and/or capacity to work with schools.

Although policymakers and researchers often use the language of biomedical researchers to describe research into and evaluation of social and educational programs including CSR, there are important differences between the implementation and evaluation of CSR and the implementation and evaluation of biomedical treatments. One important difference is that CSR model developers are constantly redesigning their models. As they work with schools and districts, they adapt and retool their models, as would be expected. These modifications, though, make expanding the scope of CSR while still serving (and "scaling up") existing schools difficult. Additionally, models evolve not only because of developer redesign, but because they "mutate" to adapt to local conditions. Implementation of CSR (and other social and educational interventions) is far more complex than taking a pill once a day. While many patients fail to take medications precisely as prescribed, they are not responsible for adapting and modifying the biochemical pathways through which cholesterol-reducing medications work, for instance. District administrators, principals, teachers, and students, on the other hand, are active participants in "constructing" as much as "implementing" CSR (Pressman & Wildavsky, 1984; Lipsky, 1980; Berends et al., 2002).

Ensuring initial acceptance of reform by teachers, as discussed above, is only part of the challenge of building capacity. Commitment to reform by the district, principals, and teachers needs to be maintained beyond the initial implementation phase (Berends et al., 2002). Mutation and adaptation will occur even with extensive buy-in. As commitment and funding wane, however, implementation will fal-

ter. Principal leadership—both actual behaviors and the perception of principals as leaders among teachers—can be a powerful factor in developing capacity and sustaining CSR.

A particularly perplexing question about ensuring ongoing teacher commitment has been raised by Little and Bartlett (2002). Studying teachers engaged in CSR at the high school level, they found that teachers engaged in ambitious or large-scale reform such as CSR initially find such activities to be rewarding and stimulating. Early enthusiasm leads to fulfilling professional growth and learning opportunities. Over time, however, such enthusiasm is not sustained and leads to burnout, conflict, and decreasingly active engagement in reform. Causes of decreased enthusiasm include disappointment over unavailable resources, colleagues' waning commitment, sheer exhaustion from the scope and scale of reform activities, and disagreement over the specifics of implementation. Little and Bartlett suggest that organizational structures to support teacher collaboration and leadership can strengthen resolve. At the district level, as reform unfolds, additional reforms are often (usually) added to complement, supplement, and/or supplant CSR. It is an open question how many competing reforms districts can adequately support and integrate (Hatch, 2002; Berends et al., 2002).

Similarly, a key role for model developers (as discussed above) in supporting schools and building capacity requires establishing and maintaining networks of schools implementing their particular model. There are clear limits on the capacity of developers to serve adequately multiple, large, and geographically disparate networks of schools (Berends et al., 2002).

Continuing CSR Research Using Varied Methodologies

The challenges discussed in this chapter pose considerable barriers to further growth of CSR. A crucial goal for the research and policymaking communities is motivating ongoing research on CSR that uses varied scientific methods. The first step is for education researchers to agree on the question(s) that need answering. Research questions themselves ought to drive the selection of research methods.

We argue that in the case of CSR, there are three questions that should focus a research agenda. These questions, which can be stated a variety of different ways, can be summarized as follows:

- How does CSR work? What impacts implementation of CSR? Are some models or some components of CSR harder/easier than others to implement? Is there an optimal "bundle" of CSR components? Are there differential effects across school settings and among student populations?
- How can districts and developers best support model implementation?
- What works? To what extent do students in CSR schools outperform students in other schools? Do some models have greater effects than others?

Each of these questions centers around processes and outcomes involved in adopting, implementing, and sustaining CSR and each lends itself to a different research method or set of methods. Arguably, none of the questions have been fully answered (see Herman et al., 1999 and Borman et al., 2003, both discussed above, on the quality of research addressing the "what works?" question).

The first two questions allow for a wide range of research methods as they do not (necessarily) imply that there are causal links to be drawn. Rather, these implementation questions can shed light on the processes by which CSR operates. Mixed methods approaches are especially well suited to answering these questions. Quantitative survey methods can yield broad and generalizeable findings while more focused, qualitative research can provide rich descriptions that answer many "how?" questions. Several studies, some quite large, currently employ mixed methods to answer these questions. The NLECSR, Longitudinal Assessment of CSR Implementation and Outcomes (LACIO), the Study of Instructional Improvement (conducted at the University of Michigan), the Longitudinal Research on Whole School Improvement through CSRD and Other Reform Efforts project (conducted by the University of Southern California and Johns Hopkins University), and the Study of District Strategies for Comprehensive School Reform (conducted by Policy Studies Associates (PSA)) all address issues of implementation. The AIR, Michigan, and PSA studies also look explicitly at the district role in supporting reform, while the NLECSR and RAND studies look more directly at the role of developers.

The third question, the "what works?" question, calls for a randomized controlled experiment to establish or disprove causality. There are a number of impediments to conducting a randomized study of CSR effects (see Cook, 2002 for one discussion of these), chief among them that the scale of the study would need to be larger than that of virtually all previous CSR research while the cost would be prohibitive, approaching that of national studies such as Project Follow-Through. While some researchers have attempted small-scale trials (fewer than 30 schools) of single models, a study examining multiple models would require on the order of 20-25 schools per model in addition to controls. Therefore, a study of only five models would require more than 200 schools to detect statistically significant effects of relatively large size. In an ideal such study, schools and models would be nested within districts as well, meaning a sample frame would need at least two (and preferably three) models within each district and approximately 80 schools. Conducting a simple three-level model would require 20 districts, or a total of 1,600 schools, with each to be randomly assigned to treatment or control groups. Since a study like this would also need to be conducted across states, student assessments would need to be administered to ensure a uniform outcome measure, with active parental consent obtained for every tested student. Feasibility is sometimes dismissed as a research consideration, but the financial cost and manpower required preclude outright the undertaking of a study like the one described above.

It is worth noting the difficulties in gaining consent for participation in a randomized field trial alone. The communities of greatest interest for a randomized

study of CSR would be predominantly urban, high-poverty communities with high proportions of African-American, Latino, and other minority groups. Past research on economically marginalized groups during the twentieth century, however, has generated tremendous distrust and fear of researchers among these populations. This distrust was sown primarily through experiments conducted by biomedical researchers because participants were either not given the option of informed consent or were kept uninformed about the potential health risks posed by the study. In some studies of experimental (often hazardous) treatments for fatal diseases, alternative treatments known to be effective were knowingly withheld from participants.

The most notorious of these was the Tuskegee Syphilis Study, a project undertaken to examine the progression of untreated syphilis in 400 African-American males in rural Alabama (Thomas & Quinn, 1991). From 1932 to 1972, physicians with the U.S. Public Health Service told these men they were being treated for "bad blood" and denied them treatment for their condition. Even when penicillin was demonstrated to be effective against syphilis, researchers denied their subjects access to the drug and took steps to prevent them from obtaining treatment elsewhere. As many as a hundred men died, and many likely passed the disease on to sexual partners and children. In the 1990s, researchers cited the distrust caused by the Tuskegee Study as a contributing factor in the low participation of African-Americans in clinical trials, organ donation efforts, and preventive care.

In all probability, a large-scale, quasi-experimental study will have to suffice as an acceptable alternative to a randomized field trial. Several such studies are currently underway: the National Longitudinal Evaluation of Comprehensive School Reform (NLECSR, conducted by AIR, the University of South Florida, and the National Opinion Research Center at the University of Chicago), the RAND Evaluation of Comprehensive School Reform, and the LACIO (conducted by WestEd). All three are funded by the U.S. Department of Education.[7]

Conclusion

This chapter has outlined major challenges facing CSR. These challenges include those inherent in the American federal system of education, a system that by its very nature creates complexities in relationships between and among schools, districts, states and the federal government. Additional challenges reside in the very nature of engaging whole-school reform in a context—the school—with shifting patterns of resource allocation, high teacher and student mobility, and other vexing difficulties. Although problems beset any reform agenda that must negotiate complex and often contradictory relationships, avenues leading to successful whole-school reform do exist in the current policy context. For example, it is possible for states, districts, and schools to coordinate their efforts in such critically important areas as building capacity. We suspect that under the

7 See http://www.goodschools.gwu.edu/csrl/csrip.pdf for details on these and other studies cited.

provisions of the No Child Left Behind Act, this possibility will accelerate, becoming a mandate, with implications for CSR, a reform that flourishes in district contexts that allow for considerable individual school autonomy and site-based management, a condition that appears to be receding under the press of NCLB. Ironically, taking all levels of the education system into account as we do in the first sections of this chapter, we see that generally, the most effective coordinated efforts in implementing and sustaining CSR reforms occur as a result of judicial decisions, legislative action, or both, that dictate top-down change. It is also clear that accountability systems currently drive education reform and the district, school, and classroom levels. What may be necessary, then, is that to meet the challenges discussed in this chapter, it will be necessary for agencies that have not often worked together in the past to do so now. New alliances among schools, districts, model developers, and private organizations have begun to occur that may magnify the capabilities of key players and institutions to exploit the landscape of comprehensive school reform.

What this chapter consistently reveals is the importance of aligned policies as well as coordinated and sustained effort in building capacity for reform over time. When district and school personnel are invested in the schools, agree on issues and problems to be addressed in undertaking reform, and create learning communities for the study by all teachers of best practices for promoting students' academic growth, it is likely that successful processes of both implementing and sustaining reform will occur. Consider the example of the state of New Jersey discussed in this chapter. In some respects, New Jersey is an extreme case. Many of the most disadvantaged schools in New Jersey began to undertake whole-school reform under duress, through a mandate attached to the court ruling in the Abbott V decision. The result was an intensive effort to enhance academic outcomes for students in the state's most under-served and high-poverty districts. Some things did not change: the superintendent of Newark's schools left, was replaced by a popular individual who in turn was replaced despite protestations by the union and other vocal supporters, high inner city teacher and student mobility rates did not decline, etc. Nonetheless, despite this turmoil and policy churn and despite the persistence of a host of negative conditions including poverty, high rates of teacher and student mobility and the like, CSR continues to be seen as the most important strategy for enhancing teaching and learning. Capacity in large part was built on the good will and enthusiasm of those inside and outside the schools as well as upon the continuing stream of financial support Newspapers regularly reported on the progress of Abbott district schools, and researchers regularly released results of their ongoing research in the public media. Thus, the press of accountability and interest combined with the (often) coordinated effort of a hierarchical array of agencies created the opportunity for comprehensive school reform to take hold in the schools.

An array of challenges faces educators and administrators who seek meaningful reform. Among these is the ongoing development of capacity within schools and districts. A consistent finding across studies has been the importance of facul-

ty support for CSR, and ensuring their support is the first step in CSR implementation. Almost as important to the long term sustainability of CSR is the ongoing support of parents as well as of community-based institutions of all kinds. The engagement in promoting their children's academic success of a majority of parents whose students attend a given school remains among the most difficult challenges facing whole-school reform. In addition, maintaining fidelity to a research-based plan is highly problematic given the dearth of research establishing the scientific basis for many models. Meeting the needs of English language learners and students with disabilities is a challenge for all schools; however, schools implementing CSR models face an additional challenge since a fundamental principle underlying CSR is that school improvement must be designed to reach all students in a school. As we have seen, CSR model requirements can be at odds with the requirements schools and teachers face in addressing the needs of their special education students. As well, costs are often cited as a challenge to schools implementing CSR. Preliminary analysis of case study data in the ongoing NLECSR study suggests the importance in maintaining professional development related to the implementation of the model. Professional development is costly and can only be maintained if monies flow for that purpose through CSRD grants (or other sources) to individuals schools. While many principals and teachers equate the implementation of a given model with the flow of resources to sustain professional development, very few schools appear to invest in CSR professional development, instead earmarking other funding streams for this purpose.

The challenge of securing additional resources should not be understated, especially in the context of addressing problems faced by schools with large numbers of immigrant and special needs students. However, creative negotiations with external providers and appropriate district support can lessen the challenge faced by schools in this regard. Model developers under the aegis of a new coalition formed very recently including model developers representing 11 CSR designs, and are planning to work with districts and schools to coordinate their efforts. Presumably, one outcome of this coalition's work will be the opportunity for schools and districts to purchase customized versions of a given model's program.

While we have mentioned a large number of challenges that we believe will confront CSR in the near term, two major challenges overwhelm the others. The first is the impact of the policy context on implementation; the other is the lack of a strong research base for all but a handful of models. Unfortunately, CSR models are more often than not implemented in policy environments that are frequently incoherent and unstable. Somehow in the midst of this incoherence, schools must create an instructional climate that promotes student learning, maintains teacher enthusiasm, and sustains the difficult process of building capacity. Schools must learn to negotiate conflicting jurisdictions and sources of authority to implement CSR successfully. Districts and schools must also build capacity with respect to money, non-monetary resources, staff buy-in, and parent support. The full impact of policies governing schools and districts under the No Child Left Behind Act has

yet to be realized, but is certain to have an impact on CSR implementation and sustainability. One likely outcome is that district and state policies will become more tightly aligned as policies regarding support for low- and high- performing schools are determined. If funds are provided to support the implementation of CSR, a scenario that we see as unlikely because of CSR's emphasis on *whole*-school reform, CSR models are likely to flourish. On the other hand, the more likely outcome—one that seems to be anticipated by the coalition of model developers—is that districts and schools will opt to determine district-wide strategies that may or may not be compatible with aspects of some CSR models.

The research base on CSR is uneven at best. In light of No Child Left Behind and the call for practices that claim a strong research base, those CSR models with inadequate research evidence will be ignored by districts and schools seeking strategies to improve student achievement. While a number of research studies, including NLECSR, are underway and will provide useful information on what works under specific circumstances, still many questions remain about which strategies work best in the case of special populations including racial and ethnic minorities, English language learners, and special education students more generally. Similarly, the research evidence is thin in particular areas including best strategies for select populations such as middle school students, in addition to particular subject matter areas such as, for example, adolescent literacy. We know very little about how best to undertake at the school level the sustained development of challenging high level course work for all students, etc.

The fundamental challenge facing CSR, however, lies with its aim: to reengineer all school processes around a central vision of providing a high-quality, research-based, and standards-driven education for all children. George H. W. Bush described New American Schools' objective as creating "break-the-mold" schools. The effort by NAS to develop models of "break-the-mold" schools could have been conceived as a step away from the then almost century-old quest to create "the one best system" (Tyack 1974). The effort to scale up comprehensive school reform has instead become another effort to create the one best system. The fundamental grammar of schooling as Tyack and Cuban (1995) call it, has remained remarkably constant. The real challenge facing CSR is the persistence of the ways teachers teach and schools are organized. While the more prescriptive models tinker around the edges of what schools do and require fairly conventional teaching methods, ironically, they hold the most promise for successful large-scale implementation. More progressive reforms face much more daunting prospects.

Postscript

The authors gratefully acknowledge the research assistance of John Quinn Kerrigan and Andrea Boyle.

References

American Institutes for Research (2003). *Evaluation of mathematics with meaning and textual power pilot programs, Final report to the College Board.* Washington DC: Author.

Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform.* New York, NY: Teachers College Press.

Banks, J. (1992). African-American scholarship and the evolution of multicultural education. *Journal of Negro Education, 61*(3), 272–286.

Berends, M., Bodilly, S., & Kirby, S. (2002). Looking back over a decade of whole-school reform: The experience of New American Schools. *Phi Delta Kappan, 84*(2), 168–175.

Bloom, H., Ham, S., Melton, L., & O'Brien, J. (2001). *Evaluating the Accelerated Schools approach: A look at early implementation and impacts on student achievement in eight elementary schools*. New York, NY: MDRC.

Bodilly, S., Purnell, S., Ramsey, K., & Keith, S. (1996). *Lessons from New American Schools Development Corporation's demonstration phase*. Santa Monica, CA: RAND.

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

Borman, G., Hewes, G., Overman, L., & Brown, S. (2004). Comprehensive school reform and achievement: A meta-analysis. In C. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 53-108). Washington, DC: National Clearinghouse for Comprehensive School Reform.

Borman, G., Stringfield, S., & Rachuba, L. (2000). Advancing minority high achievement: National trends and promising practices and programs. New York, NY: The College Board.

Borman, K. M. & Burke, M. (2003). *How do you spell parent involvement? Answers from parents and other stakeholders*. Paper presented at the American Anthropological Association annual meeting, Chicago, IL.

Borman, K. M., Kersaint, G., Boydston, T., Lee, R. Cotner, B., Uekawa, K., et al. (2004). *Meaningful Urban Education Reform: Confronting the Learning Crises in Mathematics and Science.* Albany, NY: State University of New York Press.

Cohen, D., Raudenbush, S., & Ball, D. (2001). Resources, instruction, and research. In R. Boruch & F. Mosteller (Eds.), *Evidence matters: Randomized trials in education research* (pp. 80–119). Washington, DC: The Brookings Institution.

Cook, T. (2002). Randomized experiments in education: Why are they so rare? *Educational Evaluation and Policy Analysis, 24*(3), 175–200.

Cooper, R., & Jordan, W. (2003). Cultural issues in comprehensive school reform. *Urban Education, 38*(4), 380–397.

Consortium for Policy Research in Education. (1988). *Consortium for Policy Research in Education. CPRE Policy Brief*. New Brunswick, NJ: Author.

Crampton, F. E. (1995, March). *Is the production function dead? An analysis of the relationship of educational inputs on school outcomes*. A presentation to the Annual Conference of the American Education Finance Association.

Datnow, A. (2001, April). *The sustainability of comprehensive school reform models in changing district and state contexts*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA. Retrieved August 9, 2004, from http://www.usc.edu/dept/education/cegov/publications/Sustainability.pdf

Datnow, A., Borman, G., Stringfield, S., Overman, L., & Castellano, M. (2003). Comprehensive school reform in culturally and linguistically diverse contexts: Implementation and outcomes from a four-year study. *Educational Evaluation and Policy Analysis, 25*(2), 25–54.

Datnow, A., McHugh, B., Stringfield, S., & Hacker, D. (1998). Scaling up the Core Knowledge sequence. *Education and Urban Society, 30*(3), 409–432.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk, 5*(1–2), 183–204.

David, J. (1990). Restructuring in progress: Lessons from pioneering districts. In R. Elmore (Ed.), *Restructuring schools: The next generation of education reform* (pp. 209-250). San Francisco, CA: Jossey–Bass.

Delgado, G. (1991). Involving parents in the schools: A process of empowerment. *American Journal of Education,* 100, 20–46.

Elmore, R., & Burney, D. (1997, August). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2,* New York City. Paper prepared for the National Commission on Teaching and America's Future. New York City, August 1997. Retrieved August 6, 2004, from http://www.iadb.org/int/DRP/ing/Red4/Documents/ElmoreAbril4-5-2002ing.pdf

Elmore, R., & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond and G. Sykes (Eds.), *Teaching as a learning profession: Handbook of policy and practice*. San Francisco, CA: Jossey-Bass.

Epstein, J. (1996). Perspectives and previews on research and policy for school, family, and community partnerships. In A. Booth & J. Dunn (Eds.), *Family-school links: How do they affect education outcomes?* (pp. 209–245). Mahwah, NJ: Erlbaum Press.

Epstein, J. (2001). *School, family and community partnerships: Preparing educators and improving schools*. Boulder, CO: Westview Press.

Epstein, J., Simon, B., & Salinas, K. (1997, September). *Involving parents in homework in the middle grades. PDK research bulletin no. 18.* Bloomington, IN: Phi Delta Kappan International. Retrieved on August 6, 2004, from http://www.pdkintl.org/edres/resbul18.htm

Erlichson, B. A., & Goertz, M. (2001). *Implementing whole school reform in NewJersey: Year two.* New Brunswick, NJ: Rutgers University.

Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban public high school*. Albany, NY: State University of New York Press.

Flaxman, E., Schwartz, W., Weiler, J., & Lahey, M. (1998). *Trends and issues in urban education, 1998*. New York, NY: Teachers College, ERIC Clearinghouse on Urban Education.

Fullan, M. (2004, April). *Whole system reform.* Paper presented for the "From Whole School to Whole System Reform" conference, Chicago, IL.

Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–45.

Ginwright, S. (2000). Identity for sale: The limits of racial reform in urban schools. *Urban Review, 32*(1), 87–104.

Goldring, E., & Hallinger, P. (1992). *District control contexts and school organizational processes.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hadderman, M. (1998). *School Productivity* (ERIC Digest No. 119). Eugene, OR: ERIC Clearinghouse on Educational Management.

Hamann, E., Lane, B., & Johnson, S. (in press). *School portfolios, critical collegiality, and CSR.* Providence, RI: The Education Alliance at Brown University.

Hamann, E., Zuliana, I., & Hudak, M. (2001). *English language learners, the Comprehensive School Reform Demonstration Project, and the role of state departments of education.* Providence, RI: The Education Alliance at Brown University.

Hatch, T. (2002). When improvement programs collide. *Phi Delta Kappan, 83*(8), 626–634, 639.

Hedges, L., Laine, R., & Greenwald, R. (1994). Does money matter? A meta-analysis of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Herman, R., Aladjem, D., O'Malley, A. O., McMahon, P., Quinones, S., Masem, E., Reeve, A., et al. (1999). *An educators' guide to schoolwide reform*. Arlington, VA: Education Research Services.

Hodas, S. (1993). *Is water an input to a fish? Problems with the production-function model in education.* Education Policy Analysis Archives, 1(12). Retrieved August 6, 2004, from http://epaa.asu.edu/epaa/v1n12/

Ingersoll, R. (2001). Teacher turnover and teacher shortages. *American Educational Research Journal, 38*(3), 499–534.

Jordan, G., Snow, C., & Porche, M. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly, 35*(4), 524–546.

Lane, B., & Gracia, S. (2004, April). *State-level support for comprehensive school reform: Implications for policy and practice*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Lawler, E. (1991). *High-involvement management*. San Francisco, CA: Jossey-Bass.

LeFloch, K., Kirshstein, R. (2003, April). *Dancing around district diversity: The federal view and its implications for reform*. Paper presented to the annual meeting of the American Educational Research Association, Chicago, IL.

Lipman, P. (1998). *Race, class and power in school restructuring*. New York, NY: State University of New York Press.

Lipsky, M. (1980). *Street-level bureaucracy: the dilemmas of the individual in public services*. New York, NY: Russell Sage Foundation.

Little, J., & Bartlett, L. (2002). Career and commitment in the context of comprehensive school reform. *Teachers and teaching: Theory and practice, 8*(3–4), 345–354.

Marsh, J. (2000). *Connecting districts to the policy dialogue: A review of literature on the relationship of districts with states, schools, and communities*. Center for the Study of Teaching and Policy. Seattle, WA: University of Washington.

Massell, D., & Goertz, M. (1999, April). *Local strategies for building capacity: The district role in supporting instructional reform*. Paper prepared for the annual meeting of the American Educational Research Association, Montreal, Canada.

Mattingly, D., Prislin, R., McKenzie, T., Rodriquez, J., & Kayzar, B. (2002). Evaluating evaluations: The case of parent involvement programs. Review of Educational Research, 72(4), 549–576.

McLaughlin, M, & Talbert, J. (2003). *Reforming districts: How districts support school reform*. Seattle, WA: Center for the Study of Teaching and Policy.

McLaughlin, M., Artiles, A., & Pullin, D. (2001). Challenges for the transformation of special education in the 21st century: Rethinking culture in school reform. *Journal of Special Education Leadership, 14*(2), 51–62.

Menken, K. (2000). *What are the critical issues in wide-scale assessment of English language learners?* (Issue Brief No. 6). Washington, DC: National Clearinghouse for Bilingual Education.

Miles, Karen H. (2000). *Money matters: Rethinking school district spending to support Comprehensive School Reform. Issue brief*. Arlington, VA: New American Schools.

Muncey, D., & McQuillan, P. (1996). *Reform and resistance in schools and classrooms*. New Haven, CT: Yale University Press.

Murphy, J., & Hallinger, P. (1988). Characteristics of instructionally effective school districts. *Journal of Educational Research, 81*(3), 175–181.

Murphy, J., & Hallinger, P. (1986). The superintendent as instructional leader: Findings from effective school districts. *The Journal of Educational Administration, 24*(2), 213–236.

National Center for Education Statistics (2003). *Overview of public elementary and secondary schools and districts: School year 2001-02. NCES Statistical Analysis Report.* Washington, DC: U.S. Department of Education.

Newmann, F., King, M., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education,* 108, 259–299.

Nord, C. (1998). Father involvement in schools (ERIC Digest Report No. EDO-PS-98-3). East Lansing, MI: National Center for Research on Teacher Learning.

O'Day, J., & Bitter, C. (2003). Evaluation study of the immediate intervention/underperforming schools program and the high achieving/improving schools program of the P.S. Accountability Act of 1999. Palo Alto, CA: American Institutes for Research.

Odden, A. (2000). The costs of sustaining educational change through comprehensive school reform. *Phi Delta Kappan, 81*(6), 433–438.

Picus. P. (1995). Does money matter in education? A policymaker's guide. In W. Fowler, Jr. (Ed.), *Selected Papers in School Finance 1995* (pp. 19–35). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Pressman, J., & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland; Or, Why it's amazing that federal programs work at all, this being a saga.* Berkeley, CA: University of California Press.

Price, J., Ball, D., & Luks, S. (1995). Marshaling resources for reform: District administrators and the case of mathematics. East Lansing, MI: National Center for Research on Teacher Learning.

Purnell, S., & Claycomb, C. (2001). *Implementing reform: What Success for All teaches us about including students with disabilities in comprehensive school restructuring*. Alexandria, VA: National Association of State Boards of Education.

Rosenholtz, S. (1991). *Teacher's workplace: The social organization of schools*. New York, NY: Teachers College Press.

Ross, S., Henry, D., Phillipsen, L., Evans, K., Smith, L., & Buggey, T. (1997). Matching restructuring programs to schools: Selection, negotiation, and preparation. *School Effectiveness and School Improvement, 8*(11), 45–71.

Ross, S. M., Troutman, A., Horgan, O., Maxwell, S., Laitinen, R., & Lowther, D. (1997). The success of schools in implementing eight restructuring designs: A synthesis of first-year evaluation outcomes. *School Effectiveness and School Improvement, 8*(1), 95–124.

Rossi, R., & Stringfield, S. (1995). What we must do for students placed at risk. *Phi Delta Kappan,* 77, 73–76.

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

Rutgers School of Law. (1998). State Supreme Court Syllabus. Raymond Abbott, et al. v. Fred G. Burke, et al. (A-155-97). Retrieved on August 6, 2004, from http://lawlibrary.rutgers.edu/courts/supreme/a-155-97.opn.html

Shaver, A., & Walls, R. (1998). Effect of Title I parent involvement on student reading and mathematics achievement. *Journal of Research and Development in Education, 31*(2), 90–97.

Spillane J., & Thompson, C. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis, 19*(2), 185–203.

Spillane, J. (1998). State policy and the nonmonolithic nature of the local school district: Organizational and professional considerations. *American Educational Research Journal 35*(1), 33–63.

Supovitz, J., & May, H. (2003). *The relationship between teacher implementation of America's Choice and student learning in Plainfield, New Jersey.* Philadelphia, PA: Consortium for Policy Research in Education.

Thomas, S., & Quinn, S. (1991). The Tuskegee Syphilis Study, 1932-1972: Implications for HIV education and AIDS risk programs in the black community. *American Journal of Public Health, 81,* 1503.

Tyack, D. (1974). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.

Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia*. Cambridge, MA: Harvard University Press.

Wohlstetter, P., Smyer, R., & Morhman, S. (1994). New boundaries for school-based management: The high-involvement model. *Educational Evaluation and Policy Analysis, 16*(3), 268–286.

Wohlstetter, P., Van Kirk, A., Robertson, P., & Mohrman, S. (1997). *Organizing for successful school-based management.* Alexandria, VA: Association for Supervision and Curriculum Development.

Wohlstetter, P., Malloy, C., Chau, D., & Polhemus, J. (2003). Improving schools through networks: A new approach to urban school reform. *Educational Policy, 17*(4), 399–430.

CHAPTER FOUR

The Past and Future of Comprehensive School Reform: *Perspectives from a Researcher and Practitioner*

Steven M. Ross
The University of Memphis
Libia Gil
American Institutes for Research

In this culminating chapter, we will specifically examine, from a researcher and practitioner perspective, the key lessons and practical implications learned from our years of experiences with comprehensive school reform (CSR). Selecting what is "key" is a lofty and obviously subjective task given the volumes of literature (both old and new) on the subject of whole-school reform. As we describe below, we have built in some checks and balances in trying to extract the ideas that have received the strongest support.

In planning this chapter we deliberated on whether it would be best to present our views separately or combined. As we quickly discovered, because our agreements far outweighed our differences, writing independent sections would result in much redundancy. But in choosing the unified approach, we encountered the problem of clarifying voice in relating our personal experiences and caveats. We attempt to resolve this in two ways. One is to use separate, appropriately labeled sections to indicate that one of the two perspectives (researcher-Ross; practitioner-Gil) is presented. The second, applied within unified sections, is to convey consensual views using the collective pronoun, "we."

Putting CSR into Practice

Framework for Analysis

In formulating our perspectives for the section, we drew on several sources of information. A primary source, as would be expected, was the contents of this book, which provide many insightful findings and interpretations of CSR outcomes from national experts. A second source was Ross' experiences as a CSR researcher in other contexts (e.g., Ross, 2001; Ross, Nunnery, et al., in press; Ross, Sanders, et al., 2001a, 2001b) and Gil's experiences as CSR practitioner in the Seattle Public School District and the Chula Vista Elementary School District. A third source, primarily tapped by Ross, was published technical reports and journal articles on CSR by a variety of authors. These include the recent broader reviews and discussions of CSR research (e.g., Desimone, 2002; Datnow, Hubbard, & Mehan, 2002) as well as contemporary reports of specific research projects (e.g., Stringfield et al., 1997; Berends, 1999; Bodilly, 1996; Datnow, Stringfield, McHugh, & Hacker; 1998; Nunnery et. al, 1997; Yonezawa & Datnow, 1999). A fourth source, primarily tapped by Gil, was more recent work as a New American Schools educational consultant in school districts, particularly the Abbott Districts in New Jersey.

In the interests of objectivity and accuracy, we reflected on our sources of information, taking notes on what we perceived as their main implications. Using an informal qualitative-type analytical process, we then tried to extract major categories from the separate entries. It became immediately apparent from the resultant categories that different, and sometimes opposite, arguments for certain CSR orientations (e.g., locally vs. externally developed models) could be made depending on audience and context. For example, the practical advice offered to a disad-

vantaged urban school district may be quite different from that given to a single school located in a middle-class suburb. Trying to address diverse situations would be unwieldy. We therefore decided to concentrate on criteria that best fit the original intended emphases of CSR, namely, (a) schools that primarily serve disadvantaged or historically underserved children and (b) conditions that foster scalability as opposed to isolated applications in small numbers of schools.

Researcher and Practitioner Experiential Frameworks

Reading about school reform is not as instructive about practical issues as actually experiencing it. In the following sections, each of us reviews selected events and experiences that most shaped our thinking about CSR. The purpose is to communicate the personal frameworks from which our views and possible biases emerge.

Researcher Experiences: Two Influential Studies[1]

My most influential experiences were the research studies I conducted in two urban school districts—Memphis City Schools and Toledo Public Schools.

Memphis Restructuring Initiative (MRI). In March of 1995, Memphis City Schools was selected as one of 10 "scale up jurisdictions" by New American Schools (NAS).[2] In fall 1995, an initial cohort of 34 schools ("Cohort I") began its first year of implementation. A second cohort of 14 schools (12 elementary) was added during the summer of 1996. Based on a perception of success with the initial cohorts, the superintendent, Gerry House, mandated in 1997 that all 161 schools in the system adopt by the next year either an externally developed or locally developed reform model.[3] By the fall of 1998, all MCS schools were implementing one of 18 different whole-school reform models. In spring 2001, the new superintendent, Johnnie B. Watson, perceiving the reforms to be ineffective (Edmondson, 2001; Ross, 2001) on the basis of an internal study, abolished the entire MRI.

Research findings painted a more complex picture of successes and failures. Encouragingly, most schools' implementation of their designs was relatively strong in the early years (Bodilly, Keltner, Purnell, Reichardt, and Schuyler, 1998). Positive changes in instruction, indicating greater use of student-centered methods, were also observed (Bol et al., 2000; Ross, Troutman et al., 1997; Smith et al., 1998; Stringfield, Datnow, Herman, & Berkeley, 1997). Most critically, compared to matched control schools and all other elementary schools in the district, the restructuring schools demonstrated significantly greater gains on the state-mandated achievement after only two years, and continuing to the third and fourth years (Ross, Sanders, et al., 2001a, 2001b). In the fifth year of the reform, however, the

1 This section and first-person references refer to "research" experiences by co-author, Ross.

2 For a more detailed report, see Ross (2001).

same year that Superintendent House resigned and left the district, both the absolute test scores and gains scores plummeted to levels markedly below those of previous years and the entire state.

Thus, although the MRI had been touted nationally during its heyday as a success (Desimone, 2002), ultimately it proved a failure by being completely expunged after only five years. In analyzing the reasons for its rise and fall, we formulated several conclusions (Ross, 2001). Specifically, positive factors contributing to early success were: (a) strong central office support; (b) strong buy-in from the early cohorts of restructuring schools; (c) support from multiple community partners; (d) additional funding provided by NAS and the community; (d) and strong focus by the district and design teams on providing professional development. On the other hand, negative factors identified were (a) lack of capacity by the district to support 18 different designs in over 160 schools; (b) lack of buy-in (and often active resistance) by schools "forced" to adopt designs; (c) difficulty for some schools of implementing models that were largely conceptual and process-oriented rather than prescriptive; (d) limited use of formative evaluation data by schools to diagnose problems and guide improvement planning; and (e) perceptions of limited success in raising student achievement to proficient levels, despite the higher than average gain scores overall and in individual subject subtests in Years 2-4.

Toledo Public Schools Reforms. The focus of this 2002-03 research study was the effects of two reform programs, Direct Instruction and Success for All, which had been implemented in eight high-poverty schools (Ross et al., in press). Aside from comparing the program schools to matched control schools on student achievement, we also examined outcomes reflecting school climate, teacher reactions, observations by the research team and model experts, and interviews with multiple stakeholder groups such as principals, site facilitators, teachers, school board members, community members, district administrators, and teacher union leaders.

Results failed to show significant differences between model and control schools on reading achievement gains in any grade from 1999 to 2001. On the various school climate indicators, the DI schools scored similarly to control schools, whereas one SFA school scored significantly higher and the other significantly lower than control schools. The climate indicators at the latter school reflected a wide variety of problems (e.g., in discipline, faculty collaboration, trust, expectations of student success) requiring much deeper probing to identify and understand fully. The general implication for CSR, as will be discussed in later sections, is that with negative climate and low teacher buy-in, the school is not ready to take on

3 In separate personal communications with both authors, Gerry House maintains that such a "mandate" was never issued. Although there is apparently no written documentation confirming either interpretation, interviews by Ross (2000) and associates (Sterbinsky & Ross, 2003), with many former stakeholders (e.g., principals, design liaisons from the Teaching and Learning Academy, board members, and executive staff members) have ascertained a consistent, strong impression that a mandate, whether formally labeled that or not, was actually in effect. That is, all schools were expected to adopt a CSR model or provide a convincing rationale for not complying. As far as principals were concerned, the meaning of that policy was essentially the same as a "mandate."

the demands and changes of whole-school reform. Other results, in brief, indicated that: (a) teachers and principals were mostly favorable toward their school's model; (b) community members continued to voice strong support of the reforms, particularly so in the case of DI; (c) the teachers union, the school board president, and most district leaders, on the other hand, expressed concern about the unimpressive DI achievement outcomes; and (d) site visits indicated weak implementation of DI.

The overall conclusion reached from the research mirrored that from the MRI – the reform effort to date had not been successful. The major limiting factor for DI appeared to be poor implementation of the model components, due mostly to insufficient resources and professional development. Of consequence, union policies protected teachers' freedom to choose whether or not to participate in professional development activities occurring outside of regular school hours (e.g., summer or weekend training). Further, teacher buy-in to the models, while still moderately high, was showing signs of erosion. We concluded that the dominant influences of diverse authority sources (i.e., community investment, teacher union policies, district support) served to weaken some schools' ownership of the models and accountability for their success. Unlike the systemic reform efforts of San Antonio (Berends, Chun, Schuyler, Stockley, & Briggs, 2002) and Memphis (Ross, 2001), the district role here was more piecemeal (several schools at a time) and oriented to shifting responsibility to the individual schools increasingly over time.

Synthesis. In both Memphis and Toledo, we noted generic weaknesses related to effective reform practices. In many respects, "CSR" applies a new label to what educational reform experts have been advocating for many decades. About 25 years ago, principles highly consistent with today's CSR were formally espoused in Ron Edmonds' classic article, "Effective Schools for the Urban Poor," in the 1979 issue of *Educational Leadership*.[4] Drawing from school effects research, Edmonds extracted what he felt were the critical characteristics of schools that could bring disadvantaged children to the same level of basic skills as middle class children. Abbreviated versions of what has since become well known as the "Effective Schools correlates" are provided in Table 1.

4 Edmonds' work was selected as the theme for an invited paper by Ross in The Journal for Effective Schools. More recent works that reflect parallel "quality indicators" can be found in Stringfield and Tedlie (1989, 1990).

Table 1: Evaluation of the realization of Correlates of Effective Schools in the Memphis and Toledo Reform Programs

Correlate	Reform Programs	
	MRI	Toledo
Clearly Stated Mission	Strong	Strong
Safe and Orderly Environment	Strong	Weak
High Expectations	Moderate	Weak
Instructional Leadership	Weak	Weak
Opportunity to Learn	Moderate	Moderate
Monitoring of Progress	Moderate	Strong
Enhanced Communication	Weak	Weak

Weak Moderate Strong

For the purposes of this chapter, I have two reasons for introducing this table. One is to show that the correlates, while more restricted, have much in common with the current 11 CSR criteria. The second reason is to review a retrospective analysis that I recently conducted to determine the degree to which the Memphis and Toledo reforms incorporated the correlates (Ross, 2003). As summarized in Table 1, my conclusion was that most correlates were weakly addressed if at all. Although all schools had clearly stated missions (as defined by the reform models), consistent weak areas in both Memphis and Toledo were (a) instructional leadership from both teachers and principals, and (b) communication and partnering between the school, district, home, and the community. Although the models called for "high expectations" by administrators and teachers for themselves and for student attainment of mastery, such beliefs were not the reality at many schools. Teachers saw the students as at-risk of failure and set the bar low for academic achievement. In both contexts, there was fairly regular monitoring of student progress, but practically no internal monitoring of program implementation. For the most part, teachers and administrators viewed the latter type of evaluation as the purview of the model developer or the school district.

Thus, in general, although the two reform initiatives did appear to foster usage of the Correlates of Effective Schools, inculcation of specific ideals and practices occurred slowly and below a "threshold" of strong impact and potential sustainability. In Memphis, for example, once the reform models were expunged, most of

the schools appeared to drift back rather quickly to traditional administrative and pedagogical orientations. The latter outcome naturally raises the question of whether there was a real buy-in process and ownership by the schools.

Practitioner Experiences: Two Influential Roles-School District Administrator and Educational Service Provider[5]

Reflections from my district restructuring efforts and the implementation of whole-school change models in the Seattle Public School District and additional experiences with student-based decision making and school change models in the Chula Vista Elementary School District provided a strong base for analysis and further discussions. Subsequent experiences in multiple districts in New Jersey support early observations and concerns regarding effective implementation of CSR models.

Seattle Public Schools (SPS). In the 1980s and early 1990s, SPS engaged in a series of attempts to restructure the organization by shifting centralized decision-making processes to a school-based management model.

Within the context of restructuring efforts, SPS also welcomed a variety of programs and models for change. Dr. Hank Levin inspired many school leaders with his basic message and challenge to treat all students as gifted. The basic principles of creating a unity of purpose, building on strengths and empowerment with collaboration resonated with many staff and community members. Accelerated Schools Project became the first CSR model embraced by several schools. The first middle school component was also created in the school district.

Following a short period of time, Dr. James Comer visited the school district and similarly inspired many supporters, particularly schools then known as Chapter One schools, which recognized the value of a specific social and emotional support structure with the Comer model. The focus on students, understanding pathways to provide support, and the no blame principles integrated with community outreach and clear governance structure appealed to additional schools.

Initial introduction of the program models were coordinated and presented by the central office administrators with follow up visits to interested school sites. In all cases, the principal leadership position determined access to the school site. In other words, school visits only occurred with expressed principal interest. The next level of buy-in occurred at the staff level with a design team presentation and discussions.

One interesting observation made during this orientation process was that in some situations there was an enthusiastic school leader with limited staff support and in a few cases, there was limited leadership commitment with higher staff interest. On this issue, we learned over time that effective change implementation

5 This section and first-person references refer to "practitioner" experiences by co-author, Gil.

must include high interest and commitment from both the school leadership and staff.[6]

Chula Vista Elementary School District (CVESD). During my nine-year tenure as superintendent of CVESD, I was fortunate to work with a dedicated board that demanded a change process to raise expectations for student learning. It was not by accident that one of the first tasks I initiated was taking stock and developing a shared vision to establish a unity of purpose through an 18-month community and staff engagement process. Building on strengths of all stakeholders through inclusionary practices and transferring authority and responsibility for student learning to the school level mirrored the Accelerated Schools design process and resulted in a major shift in power and control roles at the central office.

A core belief established in the CVESD was that the goals for students are nonnegotiable. There was no compromise on setting high expectations, adhering to rigorous academic standards, and promoting powerful learning for all students. In recognition that students and communities had unique strengths and needs, the strategies for reaching the goals were flexible. An early conclusion reached is that there are no single solutions for all schools and that the most significant attribute for the successful implementation of any program appeared to be commitment and a belief in the model's efficacy for student learning.

The initial implementation of an Accelerated Schools model was easily adopted by five schools; four others adopted the Comer School model. Ultimately, Direct Instruction, MicroSociety, the Edison Project with Successful for All and other local models with standards-based instruction were also adopted by various schools. Five additional charter schools with locally designed change models quickly surfaced to take advantage of staffing selection flexibility and to design and customize their own school change models. Approximately half of the 39 schools adopted an externally designed model and others created locally designed models based on whole-school change elements similar to CSR.

The longitudinal data patterns based on multiple measures for student achievement were inconsistent in relation to particular models adopted. Efforts to correlate student outcomes to specific models yielded little conclusions about model efficacy. Early reports indicated that Accelerated Schools project schools made greater student test gains, which were sustained only at some schools in later years. Student mobility rates, the percent of English language learners, and the percent of free/reduced lunch participants were much better predictors of testing outcomes than a particular model. Leadership changes in many of the schools contributed to a disruption of commitment and changes in state accountability systems, which made it difficult to compare and extrapolate accurate longitudinal achievement patterns.

6 I was the Assistant Superintendent for Curriculum and Instruction when I left the district in the early 1990s, so the issue of sustainability and impact on student learning remains a question mark.

Despite the paucity of positive data, most CSR schools reported high levels of satisfaction and commitment to pursue their direction. There were multiple examples of low-performing schools using a model to coalesce their staff and community with a renewed vision of hope and confidence. In many cases, adopting an externally developed model was a vehicle for reorganizing the school environment. However, there was limited quantitative evidence to support a direct impact of CSR model implementation on student achievement gains. In fact, many locally designed models, including charter schools, made consistent gains in student achievement exceeding those with CSR adopted models.

Abbott School Districts in New Jersey. The history of three decades of court intervention mandates and the New Jersey Department of Education (NJDOE) implementation mandates have created an interesting situation of low trust, low confidence, and continued low performance in most schools and districts. Despite the infusion of resources to address equity in the original 30 low-wealth districts, the pattern of poor academic performance for most students was and still is pervasive.

In hopes of improving student performance, adopting whole-school reform models was a mandated intervention for all Abbott districts. The focus on school-level change was driven by the common belief that central offices were the problems and the simple solution was to bypass that component. Unfortunately, that belief did not culminate in achieving the goals of the reform.

Under the new leadership of Gordon MacInnes, Assistant Commissioner of Education, new insights and direction for supporting school districts have been identified. MacInnes recognizes that the central office can play a significant role in supporting and sustaining improvement efforts. A plan to rebuild the role of the Department of Education and to strengthen relationships at the central office level and view schools as part of a system remains to be developed and established. Utilizing external change facilitators to work collaboratively with district leaders, redefining roles and responsibilities of central office staff, data collection, retrieval and analysis to drive decision making, and an alignment of initiatives to focus on instruction are the key components of multiple strategies to challenge and support school districts.

The identified barriers to effective CSR implementation in New Jersey included the lack of processes to establish buy-in and ownership at all levels of the school systems. In most cases the lack of capacity for expansion and quality control on the part of model providers resulted in dissatisfaction with training experiences which, in turn, limited implementation quality and results.

Commonalities across districts. Based on these experiences, we have found the following common occurrences to limit CSR impacts in the three districts:

- Fragmented support services by a multitude of providers (including design teams), which did not further the overall vision for the district.
- Cumulative feelings by teachers, school leaders, and administrators of distrust and credibility gaps in response to any intentional school improvement initiatives. In many cases, hostility and entrenched cynicism toward change processes may be better descriptors of emotions expressed.

Putting CSR into Practice: Recommendations

In this section, we present specific recommendations, each supported by a rationale and further defined through associated "corollaries." Table 2 lists all Recommendations and Corollaries for convenient referencing.

RECOMMENDATION 1: Schools and providers should ensure that there is adequate teacher support before initiating CSR.

There is extensive research literature supporting the key role of teachers in the success of school reform (e.g., Berends, 2000; Borman, Aladjem, Carter, and Le Floch (2004); Datnow et al., 2002; Ross, Henry et al., 1997; Tyack, 1990). According to Rowan et al. (2004), the involvement of teachers needs to occur at the earliest stages of CSR planning rather than after a principal or leadership group has identified the favored reform options.

In an optimal situation, teachers enthusiastically participate in selecting the reform approach and demonstrate strong desire to implement it. Consequently, they will be open to spending the extra effort required to learn and adapt to new methods and structures. The opposite extreme is when teachers reject the reform and actively resist its implementation. Knowing what to do in these two contrasting situations is obvious (i.e., focusing on implementation vs. buy-in). But what constitutes "adequate" teacher support in typical, less extreme cases? The answer seems to largely depend on the primary impetus for the reform. A suggested Corollary to Recommendation 1 is therefore:

Corollary 1a. *The degree of teacher buy-in required to ensure success is situation-specific and needs to be gauged accordingly.*

Contrast the following situations. At one school, the leadership team identifies a particular CSR approach ("Reform Model A") as having strong potential to help the school raise student achievement. However, there is no explicit requirement or pressure for the school to adopt that specific reform. Unless teachers were highly involved in selecting it and enthusiastic about implementing it, they may see the reform model as an "add on"—something foisted on them by an over-zealous principal or faculty committee. Negative teachers may perceive that because the reform is only a site-based initiative, accountability for resisters will be limited.

In the second case, the district mandates that all schools in "corrective action" under NCLB adopt Reform Model B. Accordingly, if teachers are to remain at these schools, there is no option but to move forward with the reform. Even so, as was found in Memphis (Ross, 2000) and San Antonio (Berends et al., 2002), some teachers may not feel personally accountable to the mandate and will ignore or actively obstruct implementation. The principal and leadership team therefore will still need to work on creating adequate buy-in as implementation proceeds. To the extent that the district works with schools to facilitate the transfer of oppositional teachers to other sites, the chances for developing a "critical mass" of supporters should increase (Muncy & McQuillan, 1996). In Memphis and San Antonio, trans-

fer opportunities were limited, thus keeping the same resisters in the ranks over time. Eventually, many of these groups influenced others (especially new teachers) and prevailed in weakening the reform effort. Another corollary thus becomes:

Corollary 1b. *Districts, schools, and unions should collaborate in developing policies that facilitate transfer of teachers who are overly resistant to implementing a chosen reform design.*

RECOMMENDATION 2: Schools and providers should actively solicit and encourage district support.

Although there are numerous examples of schools successfully pursuing CSR independently of district involvement (e.g., Sterbinsky et al., 2003), research has shown that conflicts with district policies and other types of power struggles can seriously impede the progress and sustainability of site-based reforms (Bodilly, 1998; Datnow et al., 2002; Desimone, 2002; Yonezawa & Stringfield, 2000). Datnow et al. (2002), in their book on extending educational reform, describe numerous examples of schools resisting district-supported reforms and districts inhibiting school-supported reforms due to having different philosophies, financial or resource incentives, or preferred curricula or teaching methods. There are several ways that garnering district support for CSR is likely to increase the chances for schools' success:

- NCLB has put more pressure on districts to raise achievement and help schools attain Adequate Yearly Progress (AYP). Consequently, there appears to be a greater emphasis than in the past on schools using a consistent "standards-based reform environment" as opposed to site-based curricula (Desimone, 2002). Schools adopting or developing CSR models may need help from the district in aligning the various model components to the wider district and state plan (Rowan et al., 2004). At the very least, communicating with the district to demonstrate that such alignments exist should be highly important politically.
- It is often difficult for individual or small numbers of schools to obtain the resources needed to support independently all the components of CSR models.
- Unless the CSR model adopted is the primary district model, its components may conflict somewhat with regular district policies concerning professional development time, teacher transfers, curriculum and instruction, and administrative structures. In Chapter 3 of this book, K. Borman et al. (2004) address these very issues in relationship to special education and ESL subgroups. Not only is explicit accommodation for these groups lacking in most CSR models, district-level policies can be highly directive and in conflict with model curricula or emphasized instructional practices.

- Understanding of the CSR approach by district leaders is important in accommodating implementation needs (e.g., increased technology, extended day, reduced class size) and evaluating progress.
- To the extent the CSR approach is strongly supported by the district, its longevity at the school following principals or teacher changes should increase. Similarly, although belied by the events described in Memphis and San Antonio, sustainability should also be greater with a change in superintendents.

The above considerations and the research literature on CSR suggest several corollaries to Recommendation 2.

Corollary 2a. *CSR providers should include explicit district involvement as an integral part of their implementation plan.*

Corollary 2b. *CSR providers should require (or strongly encourage) the school district to sign a "Memorandum of Agreement" (MOA) that specifies and confirms acceptance of its roles.*

Corollary 2c. *CSR providers and schools should, where feasible, formally involve district leaders and staff in the reform through such activities as participating at planning meetings or school events, providing expertise or support (e.g., professional development, allocating resources, supporting accountability, and assuming advisory roles).*

The above three corollaries attempt to solidify and operationalize district involvement as formal agreements and concrete participation roles. In agreement with K. Borman et al. (2004), we encourage district relationships to extend beyond merely the superintendent. While these recommended strategies can't guarantee the sustainability of CSR at individual schools, they should certainly help to increase it. Recommendations 3 and 4 below, however, assume that CSR providers will be prepared to work actively with districts and schools in forming effective roles and collaborations. In many past situations, as our experiences and the literature attest (e.g., Berends et al., 2002; Datnow et al., 2002), adequate preparation and activity have not been evidenced.

Our recommendations and corollaries also generally apply to the involvement by schools and CSR providers with state education departments. It is most evident from the New Jersey experience that the state role in education reform initiatives can hinder or facilitate desirable change. State policymakers often contribute unintentionally to the overwhelming series of disjointed and fragmented mandated interventions for districts and schools to implement. One conclusion from field experiences is that current state departments have limited capacity to provide technical expertise for school change. A recommended remedy is to redesign the roles and functions of the departments—shifting the role of a compliance agency to a supportive role as a critical friend who can facilitate inquiry processes and establish a collaborative problem-solving model. The impact of a shared responsibility strategy is an approach that remains to be tested.

RECOMMENDATION 3: Before contracting with schools, CSR providers should ensure that schools have the capacity to implement the models.

Research on CSR uncovered numerous instances of schools that had good intentions of implementing CSR but struggled due to lack of readiness (Berends, 1999; Datnow et al., 2002; Smith et al., 1998). One type of readiness has already been discussed under R:1 as having adequate teacher support. But even where teachers are enthusiastic and committed to the reform model, their ability to implement it may be limited by several factors. In Memphis, one factor was understanding CSR models that were highly conceptual and theoretical during an era when high-stakes testing demands were emphasizing the teaching of basic skills (Ross, 2001; Smith et al., 1998). Despite the possibility of teacher backlash toward overly prescriptive reforms, K. Borman et al. (2004) and Rowan et al. (2004) generally see advantages of more specified CSR models for fostering buy-in and implementation quality.

Another readiness need includes having the resources needed for implementation (Fullan, 1991; Glennan, 1998; Rowan et al., 2004). Again in Memphis, for example, we observed schools that were unable to fully implement either the Conect or Modern Red Schoolhouse design due to lack of technology support (Ross, 2002). Other common examples include lacking the curricular, instructional, or planning materials that the design requires. Aside from materials, the CSR design may require space or staffing resources that are not presently available. In Toledo, we visited a SFA school that did not have a full-time facilitator, an essential program component (Ross et al., in press). Not surprisingly, implementation quality was perceived to be suffering as a result.

Means of reducing these types of problems are discussed in Chapter 1 (Rowan et al., 2004) in association with the "CSR planning process." Important elements of this process include involvement by a cross-section of diverse stakeholders (i.e., "inclusiveness"), an initial needs assessments targeting the school's improvement goals, and an "evolutionary" nature that involves flexibility in and continuance in planning "over the long haul" (Rowan et al., 2004, p. 14). To reinforce the importance of adequate resources to the success of CSR, corollaries to Recommendation 3 follow:

Corollary 3a. *CSR providers need to document specific resource requirements and classify them in terms of importance.*

Corollary 3b. *Schools need to perform a thorough study (needs assessment) of their capacity to allocate the necessary resources.*

Corollary 3c. *CSR providers and schools need to evaluate collaboratively the school's preparedness to adopt (or continue) the reform model based on the resources available.*

Corollary 3d. *If the school is judged to have inadequate resources to support successful implementation, CSR providers and schools should strongly consider not continuing with the particular reform.*

The last corollary, in my opinion, needs special emphasis. CSR providers are typically motivated by organizational goals and economic pressures to expand the

number of program schools. But, there is little point in encouraging a school to pursue a reform for which it lacks readiness or capacity. There are certainly numerous instances of schools securing the resources needed and achieving success completely independently of district assistance. But, in the vast majority of cases, insufficient capacity to implement the reform will lead to frustration for the teachers and poor results for both the school and the chosen reform model. Where justified by inadequate conditions, both the schools and CSR providers need to "say no." This suggestion leads to an additional corollary:

Corollary 3e. *CSR provider teams and schools lacking readiness for CSR should work collaboratively to determine which model elements can presently be used to increase readiness and success.*

The rationale for the above corollary is based on the idea that CSR does not have to be "all or nothing." Schools having the least readiness for CSR implementation are probably those most in need of change and able to benefit from the selective implementation of reform model components. Many CSR providers today are shifting from requiring full model implementation to providing more targeted services that address high-priority school needs (e.g., interpreting test data, improving professional development, involving parents, etc.). Requiring fidelity to the original, full model is thus becoming de-emphasized.

RECOMMENDATION 4: *Before contracting, provider teams should ensure that they have the capacity to provide the services needed for schools to be successful.*

Research on CSR underscores the challenges that provider teams face in supporting implementation of their models (Hatch, 2000; Slavin & Madden, 1998). Following the launch of NAS' 1995 "scale-up" phase, the original NAS teams rather quickly discovered that moving from local pilot sites to geographically dispersed schools created extraordinary, unanticipated demands on staffing, budgets, and adaptability of services to fit district and state policies (Ross, Alberg, & Nunnery, 1999). In my own research in Memphis and other contexts (Ross, 2001; Smith et al., 1998), concerns expressed by schools and districts frequently included the following:

- Professional development was insufficient or poor quality.
- Coaches or facilitators had a poor understanding of local culture or policies.
- Coaches or facilitators were poorly prepared to work with teachers.
- Design elements were not fully developed or adequately supported by materials.

The above considerations have little to do with the effectiveness of given designs, but rather with the decision of where and to what extent scale-up can be supported. A one-size-fits-all approach probably won't work with many schools given differences in local needs and conditions (Datnow et al., 2002). For example, district policies may require reforms to address certain needs in curriculum offer-

ings (such as a research-based reading program), staffing structures, or parent involvement. CSR models that are weak in these areas will have a difficult time realizing success. Although the quality and range of CSR services have advanced significantly in the past five years, the typical CSR provider still faces challenges in meeting the diverse needs of schools and districts nationally (Ross, 2002). Drawing from this rationale several corollaries to Recommendation 4 become:

Corollary 4a. *Provider teams should assess particular schools' needs to be successful in view site and district conditions.*

Corollary 4b. *Provider teams should assess their capacity to provide the services and model elements needed by each school (C: 4a). When capacity is low, the provider team should decline the opportunity to contract with the school.*

With regard to C:4b, Bodilly (2001) identifies qualities that give CSR provider teams increased chances of successful implementation. These include: (a) having a stable team with qualified personnel; (b) effectively communicating the design to schools; (c) effectively marketing the design to, and gaining resources from, districts; (d) emphasizing the core elements of schooling across the design (e.g., curriculum, instruction, student assignment, assessments, and professional development); and (e) fully supporting implementation. Further, in this book, K. Borman et al. (2004) describe the experiences of the Accelerated Schools Project in shifting their model's emphasis from establishing school governance to promoting "powerful learning in order to raise students' academic achievement. Flexibility and adaptability thus become important to the success and sustainability of CSR models.

RECOMMENDATION 5: Provider teams and schools should place strong emphasis on ensuring that adequate initial and continuing professional development is available to teachers and principals.

Higher student achievement seems unlikely to occur without concomitant improvements in teaching effectiveness. It is noteworthy, for example, that in examining CSR program effects on the Tennessee Comprehensive Assessment Program (TCAP), Ross et al. (2001a, 2001b) found teacher effects to be far more powerful than program effects in explanation variance in test scores. Although few CSR schools have the luxury of recruiting "master teachers" to implement and inspire the reform process, all have the potential of increasing the expertise of their existing staffs. Professional development to help teachers enhance their skills thus becomes critical to stimulate and sustain change. Contrary to this goal, a frequent finding from CSR research was weaknesses in the establishment and implementation of a comprehensive professional development program (Bol et al., 1998; Meyer & Wong, 1998; Ross, Troutman et al., 1997; Smith et al., 1997). By "comprehensive," we are referring to support provided at all stages of the reform implementation—beginning, developing, and full.

There is a strong tendency by CSR providers to concentrate professional development services at the beginning phases of implementation (Ross, 2002). This pri-

ority seems logical given the newness of the reform to all teachers. But, it has also come at the expense of services to developing (say, 1-3 years-old) and experienced (4 or more years-old) programs. As research shows, what frequently happens over time at CSR schools is that the quality of implementation erodes as teachers drift back to traditional methods (Berends et al., 1999, Datnow et al., 2002). As originally trained teachers leave their schools to retire or relocate, their replacements are very unlikely to have received much if any training in the particular CSR model. This problem was greatly exacerbated in Toledo by the union policy of allowing teachers choice in whether or not to participate in professional development occurring outside normal school hours (Ross et al., in press).

Based on the above rationale, we present the following corollaries to Recommendation 5:

Corollary 5a. *Ensure that strategies are in place to provide quality professional development to new teachers.*

Corollary 5b. *Ensure that strategies are in place to provide continual refresher and new professional development to all staff.*

Corollary 5c. *Develop capacity by the school to assume increasing responsibility for providing professional development over time.*

The final corollary is based on the reality (see Recommendation 7 below) that few schools will have the luxury of being able to purchase external CSR services forever (see K. Borman et al., 2004). The standard funding period for CSR grants was only three years. Once the grant expires, schools may find themselves unable or unwilling to contract with provider teams. If internal capacity to provide professional development is lacking, implementation of the CSR model will rapidly weaken. In this regard, Rowan et al. (2004) emphasize the need for professional development to become embedded inside the schools via integration with learning materials and lesson plans, establishment of model classrooms, usage of on-site and local instructional leadership, scheduling of staff meetings and planning periods, and participation by teachers in study groups, peer observations, and networks.

Research has strongly pointed to the critical role of principals in supporting CSR (Bodilly, 1998; Datnow et al., 2002; Muncey & McQuillan, 1996; Rowan et al., 2004). Merely being a groundbreaker or cheerleader may not have much impact in solving problems of implementation, buy-in, district interactions, and stakeholder involvement. However, as Rowan et al. (2004) point out, principal leadership responsibilities may be stretched too thin to have much impact on day-to-day instructional improvement. Thus, two additional professional development corollaries are:

Corollary 5d. *Ensure that strategies are in place to provide principals with initial and ongoing professional development.*

Corollary 5e. *Add instructional leadership positions to school staff where resources permit.*

RECOMMENDATION 6: Provider teams and schools should place strong emphasis on ensuring that site-based collection and usage of formative evaluation take place.

Although most CSR models include formative evaluation as a main component, the degree to which schools actually use such information for improvement planning is open to question (Bodilly, 1996; Rowan et al., 2004; Ross, 2003). Seemingly, in most cases where formative evaluation is used, it will be initiated and controlled by the provider team with the school taking a more passive role as a possible (and often reluctant) recipient. Such applications may, for example, take the form of "implementation" site visits, classroom observations, and surveys or interviews conducted with school administrators, teachers, and facilitators (Fuhrman, Clune, & Elmore, 1998). The CSR provider team then analyzes the data for use by its staff (e.g., coaches or trainers) and typically as feedback to the school.

Despite the immediate usefulness of such evaluations to the provider, we believe that they may have only limited impact on the school for the following reasons:

- conditioning for viewing evaluations as negative or punitive;
- lacking expertise to interpret the results;
- concerns by principals that the results reflect their personal ability or success;
- viewing the evaluation as being "owned" by the provider team or district rather than the school;
- lacking experience with sharing and discussing results as a school community;
- lacking experience with using data for improvement planning; and
- limited linkage between the formative evaluation results and the standard district or state improvement planning process.

Although overcoming all of these barriers may not be feasible, making a systematic effort to involve schools in data-driven improvement planning should certainly help to improve the quality and sustainability of the reform effort. Above all, as Rowan (2004) indicates in this book, "…the process of formative evaluation is time-consuming and useless if results are not used to plan for further action" (p. 17). Accordingly, we suggest the following corollaries:

Corollary 6a. *Provider teams and schools should make formative evaluation a primary component of the CSR model.*

Corollary 6b. *Provider teams and schools should develop strategies to increasingly shift responsibility for collecting the formative evaluation to the schools.*

Corollary 6c. *Provider teams and schools should collaborate in developing an improvement planning process that directly incorporates the formative evaluation results.*

Corollary 6d. *Provider teams and schools should work with the district in exploring strategies for directly incorporating the improvement planning strategies (C 6:d) into the formal district-mandated improvement plan.*

RECOMMENDATION 7: CSR provider teams should orient their services to shift increasing responsibility and autonomy to schools over time.

In an ideal world, where money and resources for schools are plentiful, it would be both beneficial and logical for schools to contract with CSR providers for full services year after year. But realistically, especially today with the added demands of NCLB, such resources are scarce. As Glennan (1998) found in his analysis of New American Schools' reform efforts, many schools discontinued their CSR program when the external funding was no longer available (also see in this book, K. Borman et al., 2004). If schools view a CSR model as an "all or nothing" proposition, lack of sufficient funding may leave no choice but to discontinue using it. In our view, CSR models should provide *frameworks* for establishing site-based reforms that are increasingly "owned," refined, and managed by schools and districts to meet local needs. This orientation involves CSR providers focusing on the question, "what needs to be done to ensure that the key reform components will be continued for many years once the formal CSR contract expires." Based on this rationale, we propose the following corollaries to Recommendation 7:

Corollary 7a. *Although fidelity to the CSR model is highly important (especially in early implementations), CSR providers should support adaptations that would make the reform more consonant with local needs.*

Corollary 7b. *Provider teams should categorize and prioritize external services according their value to the particular school and for sustaining the reform in general. At the completion of the CSR grant, the contracted services should become increasingly adapted to these areas.*

Provider teams cannot reasonably expect that a "one size fits all" mentality will accommodate schools' needs or their budgets. Measures of a provider's success might well entail how effectively it has adapted services to the individual schools and, in addition, built the school's capacity for terminating those services.

RECOMMENDATION 8: CSR provider teams and schools should actively encourage desired types of parent involvement.

In the wake of the new Memphis superintendent's decision in 2000 to expunge all CSR models, a particular outcome was most unexpected by the researchers (Ross, 2002). That was the complete lack of response by the parent community. The majority of parents expressed uncertainty or even complete ignorance of what the CSR model was at their child's school. It became obvious that most CSR providers and schools forfeited the opportunity to garner parent support and to benefit from their active contributions to the reform components.

In Chapter 3 of this book, K. Borman et al. (2004) discuss how parent involvement in activities that support student learning (e.g., via support of student attendance, homework completion, etc.), are likely to increase academic success. In this regard they cite the Borman et al. (2003) meta-analysis which associated parent

involvement in governance activity with lower student achievement, whereas the reverse pattern occurred with family involvement in enriching students' learning outside of school. In addition to the latter benefits, parent and community involvement is obviously important to the sustainability of the reform. If parents and school staff work together to create teamwork, community, participation, respect and recognition, and perhaps most important, incorporation of diversity (K. Borman et al., 2004), support for the reform initiative should be high, assuming of course, that benefits for achievement are seen.

RECOMMENDATION 9: CSR provider teams and schools should actively encourage and solicit scientifically based research on the effectiveness of the reform programs.

Despite the thousands of CSR grants awarded to schools since 1999, remarkably, there is only a paucity of rigorous research studies on the effectiveness of different reforms (Herman, 1999; Rowan et al., 2004). As Borman et al.'s (2003) meta-analysis (also see G. Borman et al., 2004), and other reviews indicate (e.g., Herman, 1999), the literature is flooded with "pre-experimental" and descriptive studies, along with numerous anecdotal reports, often published by the program developer, describing successes at individual schools. The number of rigorous quasi-experimental, matched treatment-control, or randomized studies is much fewer. Encouragingly, from these small numbers of studies, the overall effect size yielded for CSR in the G. Borman et al. (2003; 2004) research is positive, approximating +0.14, which exceeds the level associated with the previous Title I pull-out programs. True program effects might even be higher but accurate detection requires valid experimental-type research.

Seemingly, judging by the rising popularity of unproven models (Viadero, 2004), many consumers and policy makers regarded the effectiveness outcomes for CSR as "too little and too late." But the nature of CSR is changing, from reliance on intact "whole-school" models to integration of separate reform components, even as we write this chapter. In this sense, the time is still "now" to investigate via scientifically based research the effects of both traditional and emerging CSR approaches.

Conclusion

Having completed our recommendations (Table 2), we want to remind readers, in the same way that we just advised CSR providers above, there is no formula or set of prescriptions in CSR that will guarantee success. As Datnow et al. (2002) strongly emphasize, CSR is a co-constructed process involving participation by many different stakeholders along with influences of uncontrollable events. In many situations, most or all of the Recommendations and Corollaries will prove beneficial; in others, only certain combinations will. Conducting a careful needs assessment of each CSR context is probably the best fundamental advice that anyone can give. A shared judgment made by CSR providers and schools/districts on the assessment findings is the most crucial initial action to determine readiness and appropriate matches for an effective partnership to address student achievement.

There is often a tendency to confuse means and ends in education reform initiatives. Accordingly, CSR implementation is a strategy to reach the goal of raising student achievement; it is not the goal in and of itself. In a popular phrase: "the model is not the program." For this reason, locally developed CSR designs should also be supported and encouraged to incorporate the most appropriate CSR elements into their action plans.

In considering the current and future direction of CSR, we recognize that sustainable school change must include a systems approach. The focus on systemwide impact demands a synchronized change in central office roles and responsibilities to support school-level efforts. Deliberately connecting and building relationships for sharing best practices between schools are also essential strategies to organized learning and growth. In addition, both states and federal accountability plans now require district responses and responsibilities for student achievement.

The growing interest in charters, choice, and options within the public education system is another indication of systems focus. The implementation of CSR models lends itself to a borrowed business term, "Portfolio Management," as a label for the various models and designs within a single school district. An extension of traditional practices and embedded in this concept is the ability to select models of choice by staff and families. Both Memphis and Chula Vista School Districts were recognized at the New Schools Summit in 2004 for their early "Experiments in Portfolio Management." Despite its promising beginning, the original "model-oriented" CSR was and is not the panacea for closing the achievement gap and decreasing high school drop out rates for poor and historically underserved students of color. With the many lessons learned, CSR and associated reform models have built a strong foundation and knowledge base to progress to the next level of reform—redesigning "systems" (schools and districts) to achieve sustainable success in educating all students.

References

Berends, M. (1999). *Assessing the progress of New American Schools: A status report.* Santa Monica, CA: RAND.

Berends, M. (2000). Teacher-reported effects of New American Schools design: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis, 22*(1), 65–82.

Berends, M., Chun, J., Schuyler, G., Stockley, S., & Briggs, R. J. (2002). *Challenges of conflicting school reforms: Effects of New American Schools in a high-poverty district.* Santa Monica, CA: Rand.

Bodilly, S., Purnell, S., Ramsey, K., & Keith, S. (1996). *Lessons from New American Schools Development Corporation's demonstration phase.* Santa Monica, CA: RAND.

Bodilly, S. J., Keltner, B., Purnell, S.W., Reichardt, R. E., & Schuyler, G. L. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: Rand.

Bodilly, S. J. (2001). *New American Schools' concept of break the mold designs: How designs evolved over time and why.* Santa Monica, CA: Rand.

Bol, L., Nunnery, J., Lowther, D., Dietrich, A., Pace, J., Anderson, R., Bassoppo-Moyo, T., & Phillipsen, L. (1998). Inside-in and outside-in support for restructuring: The effects of internal and external support on changes in the New American Schools. *Education and Urban Society, 30*(3), 358–384.

Bol, L., Nunnery, J., Stephenson, P., & Mogge, K. (2000). Changes in teachers' assessment practices in the New American Schools restructuring models. *Teaching and Change, 7*(2), 127–146.

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

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2004). Comprehensive school reform and achievement: A meta-analysis. In C. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 53-108). Washington, DC: National Clearinghouse for Comprehensive School Reform.

Borman, K. M., Carter, K., Aladjem, D. K., & Le Floch, K. C. (2004). Challenges for the future of comprehensive school reform. In C. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 109-150). Washington, DC: National Clearinghouse for Comprehensive School Reform.

Datnow, A., Stringfield, S., McHugh, B., & Hacker, D. (1998). Scaling up the Core Knowledge sequence. *Education and Urban Society, 30*(3), 409–432.

Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many.* New York, NY: Routledge-Farmer.

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*(3), 433–480.

Edmonds, R. R. (1979, October). Effective schools for the urban poor. *Educational Leadership, 37*, 15–18, 20–24.

Edmondson, A. (2001, June 19). Watson kills all reform models for city schools. *The Commercial Appeal*.

Fullan, M. G. (1991). *The new meaning of educational change*. (2nd ed.). New York, NY: Teachers College Press.

Fuhrman, S. H., Clune, W., & Elmore, R. (1998). Research on educational reform: Lessons on the implementation of policy. *Teachers College Record, 90*(2), 237–257.

Gil, L. (2000). Comprehensive school change: Perspective of a superintendent. *Benchmarks, 1*(3) 1–3.

Gil, L., & Doyle, D. (2002). Embracing diversity: One school district's efforts to address the changing face of America's school children. *The State Education Standard, 3*(1), 20–24.

Glennan, T. K. (1998). *New American Schools after six years*. Santa Monica, CA: RAND.

Hatch, T. (2000). What does it take to "go to scale"? Reflections on the promise and the perils of comprehensive school reform. *Journal of Education for Students Placed at Risk, 5*(4), 339–354.

Herman, R., Aladjem, D., O'Malley, A. O., McMahon, P., Quinones, S., Masem, E., Reeve, A., et al. (1999). *An educators' guide to schoolwide reform*. Arlington, VA: Education Research Services.

Meyer, S. J., & Wong, K. K. (1998). Title I schoolwide programs: A synthesis of findings from recent evaluations. *Educational Evaluation and Policy Analysis, 20*(2), 115–136.

Muncey, D. E., & McQuillan, P. J. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the Coalition of Essential Schools*. New Haven, CT: Yale University Press.

Nunnery, J. A., Bol, L., Dietrich, A., Rich, L., Kelly, S., Hacker, D., et al. (1997). Teachers' initial reactions to their pre-implementation preparation and early restructuring experiences. *School Effectiveness and School Improvement, 8*(1), 72–94.

Ross, S. M. (2000). *Telling the story of comprehensive school reform: Using third-party evaluators in school districts*. Washington, DC: New American Schools.

Ross, S. M. (2001, December). *Creating critical mass for restructuring: What we can learn from Memphis. Policy brief*. Charleston, WV: AEL, Inc.

Ross, S. M. (2002). *Developing capacity for scaling-up comprehensive school reform models: Aggregate results for provider teams receiving capacity-building grants*. Report prepared for the Office of Educational Research and Improvement, U.S. Department of Education. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Ross, S. M. (2003). Effective schools correlates as indicators of educational improvement: An examination of three urban reform initiatives. *Journal for Effective Schools, 2*(2), 67–81.

Ross, S. M., Alberg, M., & Nunnery, J. (1999). Selection and evaluation of locally developed versus externally developed schoolwide programs. In G. Orfield and E. H. DeBray (Eds.), *Hard work for good schools: Facts not fads in Title I reform* (pp. 147–158), Cambridge, MA: Harvard Civil Rights Project.

Ross, S., Henry, D., Phillipsen, L., Evans, K., Smith, L., & Buggey, T. (1997). Matching restructuring programs to schools: Selection, negotiation, and preparation. *School Effectiveness and School Improvement, 8*(1), 45–71.

Ross, S. M., Nunnery, J. A., Goldfeder, E., McDonald, A., Rachor, R., Hornbeck, M., & Fleischman, S. (in press). Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success For All in an urban district. *Journal of Education for Students Placed At Risk*.

Ross, S. M., Sanders, W. L., Wright, S. P., Wang, L. W., Alberg, M., & Stringfield, S. (2001a). *Fourth-year achievement results on the Tennessee value-added assessment system for restructuring schools in Memphis*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Ross, S. M., Sanders, W. L., Wright, S. P., Stringfield, S., Wang, L. W., & Alberg, M. (2001b). Two- and three-year achievement results from the Memphis Restructuring Initiative. *School Effectiveness and School Improvement, 12*, 323–346.

Ross, S. M., Troutman, A., Horgan, O., Maxwell, S., Laitinen, R., & Lowther, D. (1997). The success of schools in implementing eight restructuring designs: A synthesis of first-year evaluation outcomes. *School Effectiveness and School Improvement, 8*(1), 95–124.

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

Slavin, R. E., & Madden, N. A. (1998). Disseminating Success For All: Lessons for policy and practice, Revised technical report. Baltimore, MD: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University Press.

Smith, L. J., Maxwell, S., Lowther, D., Hacker, D., Bol, L., & Nunnery, J. (1997). Activities in schools and programs experiencing the most, and least, early implementation success. *School Effectiveness and School Improvement, 8*(1), 125–150.

Smith, L., Ross, S., McNelis, M., Squires, M., Wasson, R., Maxwell, S., Weddle, K., Nath, L., Grehan, A., & Buggey, T. (1998). The Memphis restructuring initiative: Analyses of activities and outcomes that impact implementation success. *Education and Urban Society,* 30(3), 296–325.

Sterbinsky, A., Ross, S. M., & McDonald, A. J. (2003). *School variables as determinants of the success of comprehensive school report: A quantitative and qualitative study of 69 inner-city schools*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Stringfield, S., Datnow, A., Herman, R., & Berkeley, C. (1997). Introduction to the Memphis restructuring initiative. *School Effectiveness and School Improvement, 8* (1), 3–35.

Stringfield, S., & Teddlie, C. (1989). The first three phases of the Louisiana School Effectiveness Study. In B. Creemers, T. Peters, & D. Reynolds (Eds.), *School effectiveness and school improvement.* Amsterdam: Swets and Zeitlinger.

Stringfield, S., & Teddlie, C. (1990). School improvement efforts: Qualitative and quantitative data from four naturally occurring experiments in Phases III and IV of the Louisiana School Effectiveness Study. *School Effectiveness and School Improvement, 1*(2), 139–161.

Tyack, D. (1990). "Restructuring" in historical perspective: Tinkering toward utopia. *Teachers College Record, 92*(2), 170–191.

Viadero, D. (2004). Reform programs backed by research find fewer takers. *Education Week, 23*(32), 1, 18.

Yonezawa, S., & Datnow, A. (1999). Supporting multiple reform designs in a culturally and linguistically diverse school district. *Journal of Education of Students Placed At Risk, 4*(1), 101–125.

Yonezawa, S., & Stringfield, S. (2000). *Special strategies for educating disadvantaged students follow-up study: Examining the sustainability of research based school reforms*. Baltimore, MD: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

About the Authors

Chapter 1

Brian Rowan is a Professor of Education at the University of Michigan at Ann Arbor School of Education and a Senior Researcher at the school's branch of the Consortium of Policy Research in Education. Dr. Rowan has published numerous books and articles about education as an institution, the nature of teachers' work, and the effects of school organization, leadership, and instructional practice on student achievement. His current work includes a large-scale, longitudinal study of the design, implementation, and effectiveness of three of America's largest comprehensive school reform initiatives.

Eric Camburn is an Assistant Professor at the University of Wisconsin—Madison and a member of the team of investigators conducting the Study of Instructional Improvement. Dr. Camburn's research interests include school improvement and the improvement of leadership and instructional practice. His recent publications focus on the measurement of instruction and on the ways in which comprehensive reform programs use school leadership to improve instructional capacity.

Carol Barnes is an Assistant Research Scientist and Associate Director of the Study of Instructional Improvement at the University of Michigan at Ann Arbor School of Education. Dr. Barnes has over 20 years of experience in public policy—as a policy advisor at state and federal levels of government and as a university-based researcher. Her research interests include school renewal and instructional improvement in high-poverty settings and, most recently, the pedagogical aspects of program implementation and the role of learning as a lever for complex change.

Chapter 2

Geoffrey D. Borman is an Associate Professor at the School of Education at the University of Wisconsin—Madison, a Senior Researcher with the Consortium for Policy Research in Education, and a principal investigator in the Center for Research on the Education of Students Placed at Risk at Johns Hopkins and Howard Universities. Dr. Borman's main research interests revolve around social stratification and the ways in which schools can overcome inequality. His work is regularly highlighted in media outlets such as *The New York Times, The Washington Post, USA Today, Education Week,* and National Public Radio's *Talk of the Nation.*

Gina M. Hewes is an Associate Research Scientist at the Center for Social Organization of Schools (CSOS) at Johns Hopkins University, where she works with the Systemic Support for School Reform research group. Dr. Hewes' projects at CSOS include an investigation of the relationship between the availability of standards-based instruction and student and school risk factors; a meta-analysis of the research literature on the Success for All comprehensive school reform program; and a study on causes and consequences of retention in grade on elementary school students.

Laura T. Overman is a Research Assistant at the Center for Social Organization of Schools at Johns Hopkins University. Her research interests include identifying factors that contribute to urban school success and factors that foster academic resilience in urban school children.

Shelly Lynne Brown is an Assistant Professor of Sociology at the University of North Carolina at Greensboro (UNCG). Prior to teaching at UNCG, she worked as a Postdoctoral Fellow at the Center for Social Organization of Schools at Johns Hopkins University. Dr. Brown's major research interests are in the area of race and education, with a particular focus on the influence of race, poverty, and school composition effects on the achievement of students in pre-kindergarten through twelfth grades.

Chapter 3

Kathryn M. Borman is a Professor of Anthropology and Associate Director of the David C. Anchin Center at the University of South Florida. Dr. Borman works with the American Institutes for Research and NORC on the OERI-funded National Longitudinal Evaluation of Comprehensive School Reform, where she directs the study of 40 schools in five school districts. Dr. Borman served as the editor of the AERA journal, *Review of Educational Research,* and is author of the forthcoming book for SUNY Press, *Meaningful Urban Education Reform: Confronting the Learning Crisis in Mathematics and Science.*

Kevin R. Carter is a Research Associate at the American Institutes for Research and has participated in the National Longitudinal Evaluation of Comprehensive School Reform since its inception in 2000. He received a B.A. from Wake Forest University and will receive an M.A. from Johns Hopkins University in 2005.

Daniel K. Aladjem is a Principal Research Scientist at the American Institutes for Research and a Project Director for the National Longitudinal Evaluation of Comprehensive School Reform, the federal government's largest study of CSR. In addition to CSR, his areas of expertise include school reform and state and local policy.

Kerstin Carlson Le Floch is a Senior Research Analyst with the American Institutes for Research and specializes in school reform, accountability, and state education policy. Dr. Le Floch is currently involved with the National Longitudinal Evaluation of Comprehensive School Reform and directs the Study of State Implementation of Accountability and Teacher Quality under No Child Left Behind.

Chapter 4

Steven M. Ross is the Director of the Center for Research in Educational Policy and the Lillian and Morrie Moss Chair of Excellence in Urban Education at The University of Memphis. Dr. Ross received the distinguished Board of Visitors Eminent Faculty Award at the University of Memphis and is the author of four books and over 120 journal articles in the areas of at-risk learners, school program evaluation, educational technology, and individualized instruction. He has been the Editor of *Educational Technology Research and Development* since 1993.

Libia S. Gil recently joined the American Institutes for Research to continue her work as the former Chief Academic Officer for New American Schools (NAS). Dr. Gil provides senior counsel on leadership development initiatives and assists states and districts in developing strategies for improving student achievement by bridging research evidence with practice. Prior to joining NAS, Dr. Gil was Superintendent of the Chula Vista Elementary School District, where she fostered the successful implementation of numerous school change models.

2121 K Street, NW, Suite 250
Washington, DC 20037
Tel: 1 (877) 766-4CSR
Fax: 1 (877) 308-4995
www.goodschools.gwu.edu