CPRE

Consortium for Policy Research in Education

Study of Instructional Improvement

August 2009

School Improvement by Design

Lessons from a Study of Comprehensive School Reform Programs

Brian Rowan
University of Michigan

Richard Correnti
University of Pittsburgh

Robert J. Miller
University of Michigan

Eric M. Camburn
University of Wisconsin-Madison

Want to learn more about new and upcoming CPRE publications, project research findings, or where CPRE researchers are presenting? Please visit our website at **http://www.cpre.org** or sign up for our e-newsletter, *In-Sites*, at **insites@gse.upenn.edu**.

Be sure to bookmark this website: **http://www.icpsr.umich.edu/**. A Study of Instructional Improvement will be releasing its research data (collected from more than 115 elementary schools, 5,300 teachers, 800 school leaders, and 7,500 students) through the Inter-university Consortium for Political and Social Research (ICPSR). Researchers will be able to access and perform online analyses of the raw data files.

About the Consortium for Policy Research in Education (CPRE)

Established in 1985, CPRE unites researchers from seven of the nation's leading research institutions in efforts to improve elementary and secondary education through practical research on policy, finance, school reform, and school governance. CPRE studies alternative approaches to education reform to determine how state and local policies can promote student learning. The Consortium's member institutions are the University of Pennsylvania, Teachers College-Columbia University, Harvard University, Stanford University, the University of Michigan, University of Wisconsin-Madison, and Northwestern University.

In March 2006, CPRE launched the Center on Continuous Instructional Improvement (CCII), a center engaged in research and development on tools, processes, and policies intended to promote the continuous improvement of instructional practice. CCII also aspires to be a forum for sharing, discussing, and strengthening the work of leading researchers, developers and practitioners, both in the United States and across the globe.

To learn more about CPRE and CCII, visit our web sites at www.cpre.org and www.ccii-cpre.org.

Brian Rowan
University of Michigan

Richard Correnti
University of Pittsburgh

Robert J. Miller
University of Michigan

Eric M. Camburn
University of Wisconsin-Madison

Study of Instructional Improvement

CPRE
Consortium for Policy Research in Education

School Improvement by Design

Lessons from a Study of Comprehensive School Reform Programs

August 2009

Work on this paper was supported by grants from the Atlantic Philanthropies, 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), and the William and Flora Hewlett Foundation. All data presented, statements made, and views expressed in this report are the responsibility of the authors and do not necessarily reflect the view of the Consortium for Policy Research in Education, its institutional partners, or the funders of this study—the Atlantic Philanthropies, U.S. Department of Education, the National Science Foundation, or the William and Flora Hewlett Foundation.

Contents

Author Biographies

Brian Rowan is Burke A. Hinsdale Collegiate Professor in Education and a Research Professor at the University of Michigan. A sociologist by training (Ph.D., Stanford, 1978), Rowan's scholarly interests lie at the intersection of organization theory and school effectiveness research. Over the years, he has written on education as an institution, on the nature of teachers' work, and on the effects of school organization, leadership, and instruction practice on student achievement. His current work includes large-scale studies of teaching practice, research on instructional interventions, and the study of federal technical assistance efforts. Prior to joining the education faculty at the University of Michigan in 1991, Rowan was a senior research director at Far West Laboratory for Educational Research and Development, and Chairperson of the Department of Educational Administration at Michigan State University.

Richard Correnti is an Assistant Professor at the University of Pittsburgh's School of Education and a Research Scientist at the Learning Research and Development Center. His research interests include the measurement and determinants of instruction, how professional development influences instruction, and how instruction influences student learning. His work on SII includes program evaluations of CSR design effects on instruction and, separately, on student achievement; and the causal pathway from instruction to achievement.

Robert J. Miller is an Associate Research Scientist at Texas A&M University-College Station, and he serves as the Associate Director of Surveys for the Education Leadership Research Center. Prior to joining Texas A&M University, he was an assistant research scientist and associate director of surveys for A Study of Instructional Improvement (SII) at the University of Michigan. Dr. Miller's main fields of interest are education policy, organization theory, and the analysis of school effectiveness. His research has appeared in leading peer-reviewed journals and he is currently an associate editor for Educational Administration Quarterly. His research work includes program evaluation of Comprehensive School Reform organizational designs, and the estimation of effects of teacher involvement in school governance on student achievement outcomes. Miller earned his Ph.D. in Educational Administration and Policy from the University of Michigan-Ann Arbor.

Eric Camburn is an Assistant Professor of Educational Leadership and Policy Analysis at the University of Wisconsin-Madison. His current research centers around understanding efforts to improve educational practice in urban schools, including programmatic efforts to improve school leadership and instruction; the organizational factors that support such improvement efforts; and the impact such change efforts have on leadership practice, instruction, and student achievement. His research has appeared in numerous journals, including the American Educational Research Journal, Educational Evaluation and Policy Analysis, and the Elementary School Journal. Camburn has nearly 20 years of experience in designing and conducting large-scale studies.

Preface

We are pleased to be disseminating this reprint of a chapter that originally appeared as Chapter 49 of the *Handbook of Education Policy Research*, edited by Gary Sykes, Barbara Schneider, and David N.Plank and published for the American Educational Research Association by Routledge Publishers in 2009. The reprinted chapter presents key findings from *A Study of Instructional Improvement*, a study that was conducted under the auspices of the Consortium for Policy Research in Education and directed by Brian Rowan, David K. Cohen, and Deborah Loewenberg Ball (all at the University of Michigan). This study examined the design, implementation, and instructional effectiveness of three of America's most widely disseminated comprehensive school reform programs (the Accelerated Schools Project, America's Choice, and Successful for All) over a four year period that encompassed the school years 2000–2001 through 2003–2004. During the course of the study, data were collected in 115 elementary schools in every region of the United States, with more than 5,300 teachers, 800 school leaders, and 7,500 students and their families participating.

Although the focus of A Study of Instructional Improvement was on three comprehensive school reform programs, the goal of the study was much larger. Indeed, our aim in studying the Accelerated Schools Project, America's Choice, and Successful for All was less to evaluate the effectiveness of these specific programs and more to produce some larger insights into a process that we call "school improvement by design."As discussed in this report, design-based school improvement occurs when blueprints for organizational and instructional practice are developed by organizations operating outside of schools and are then used to improve classroom teaching and student learning inside schools. In point of fact, the numerous comprehensive school reform programs that have been operating in the United States since the late 1990's are but one of many instantiations of a larger trend toward

design-based instructional improvement that has been present in American education since at least the 1960's.

With this in mind, our goal in presenting this study of three comprehensive school reform programs is to provide readers with some key insights into the larger workings of design-based school improvement. As the reader will see, the argument that we develop in this report does not suppose that school improvement by design always works to improve instruction and student achievement in schools. Rather, we seek to identify various design principles that either improve or detract from the chances that an externally-developed program will make a difference to instruction and student achievement in schools. In particular, we argue that design-based school improvement tends to work best—not when the process encourages local educators to invent instructional and organizational solutions to the practical problems of teaching and learning that they face—but rather when it helps teachers learn how to use a well-specified set of practices through extensive supports. To the extent that our study further clarifies these design principles, we believe we will have made an important contribution to research on school improvement, especially school improvement in America's most academically challenged schools.

The study reported here would not have been possible without the cooperation of the many teachers, students, parents, and administrators who participated in it. Nor could we have progressed very far without the cooperation of the leaders of the CSR programs we studied—Gene Chasin of the Accelerated Schools Project, Marc Tucker of America's Choice, and Robert Slavin and Nancy Madden of Success for All. We extend our thanks to these participants.

We also thank the sponsors of this study. The lead sponsor was the Atlantic Philanthropies, and Harvey Dale, Joel Fleischman, Angela Covert, and Alan Ruby, officers of the foundation, provided us with generous support at various stages of the work. Thanks also are due to the Office of Educational Research and Improvement (OERI) of the U.S. Department of Education, which supported the study as part of its program for national research and development centers. During the course of the study, for example, we received funding from a grant to the University of Pennsylvania for OERI's Center on State and Local Policy (Susan Fuhrman, PI), and from a grant to the University of Washington for OERI's Center for the Study of Teaching and

Policy (Michael Knapp, PI). The National Science Foundation also supported our work through two separate grants from the Interagency Education Research Initiative. Finally, we thank the William and Flora Hewlett Foundation, which gave us funding to complete fieldwork and engage in key dissemination tasks. Mike Smith, a foundation officer, was especially helpful in this regard.

Many other individuals also contributed to A Study of Instructional Improvement. One of the most important was Susan Fuhrman. As Director of CPRE, Susan played a key role in securing funding for the study and was a true intellectual co-traveler throughout the work. Other contributions were made by a talented research team working at the University of Michigan. Lesli Scott and others at the University of Michigan's Survey Research Center were critical to the suc-cessful design, pre-testing, and fielding of the survey component of the study; Sally Atkins-Burnett designed the reading/language arts log and the student assessment procedures used in the study; Carol Barnes directed the qualitative research undertaken as part of the study; and Ruben Carriedo directed our efforts to recruit schools into the study. We also thank Deborah L. Ball and David K. Cohen (co-PIs of the study) for critical contributions at various points in our work. All of these individuals contributed immeasurably to the success of the study.

Two important "behind the scenes" contributors were Steve Raudenbush and Larry Aber. Both colleagues participated in numerous meetings at the Atlantic Philanthropies and helped shape the course of the study as time went on. Steve Schilling contributed to the design and analysis of our measures of teachers' pedagogical content knowledge. And several doctoral students contributed to the work reported here, including Andy Hayes, Chris Johnson, Geoffrey Phelps, and Seneca Rosenberg. Thanks go out to all of these individuals.

Finally, thanks are due to our administrative staff, including Debi Slatkin at CPRE, and Bonita Kothe, Jeanne Kuo, Joseph Hesse, and Jennifer Smith at the University of Michigan.

Introduction

After four decades of education reform aimed at improving the academic achievement of poor and minority students in the United States, No Child Left Behind (NCLB) has once again drawn attention to the problem of America's "failing schools." Forty years ago, when federal policy makers first addressed this problem, a solution seemed easy. All that was needed, it seemed, was to give failing schools more money and (without much additional guidance) the natural capacity of education professionals to improve instruction and schooling would yield gains in student achievement. Within a decade, however, policy makers came to doubt the efficacy of this approach, and as a result, during the 1980's and 1990's, Congress and the states developed additional reform strategies. During this time, agencies at various levels of the education system worked to raise academic standards for students, hold schools accountable for students' test scores, and devise new ways to increase parental choice in schooling. Over this same time period, the tested achievement of American students improved—especially the achievement of poor and minority students. But after more than a decade of higher academic standards, more test-based accountability, and expanded school choice, nearly one-fourth of U.S. schools are still "failing" by the standards of No Child Left Behind (Basken, 2006), and a disproportionate number of these schools serve America's poor and minority students.

Researchers who study school improvement are not surprised by this. Many believe that education policies built around high academic standards, tough accountability, and more school choice are necessary, but not sufficient, conditions for improving instruction and student achievement in schools. Higher standards, increased accountability, and expanded school choice, this argument goes, can motivate educators to work harder at improving schools; but motivation alone cannot produce the magnitude of gains in student learning needed to

turn troubled schools around. Instead, many analysts argue that what low-performing schools in the United States need is a fundamental overhaul of instructional practice. Some reformers would encourage this overhaul through macro-level changes in education policy—for example, by making it easier and more attractive for smarter people to enter (and stay in) the teaching force, or by developing strategies that bring more highly qualified teachers into America's most academically troubled schools. Others call for locally driven approaches to instructional improvement, where teachers and school leaders build strong professional communities—one school at a time—in order to discover and implement more productive approaches to instruction. The evidence on both these approaches, however, is discouraging. For example, teachers with higher academic test scores and better professional training do, in fact, promote increased learning gains in their classrooms, but research shows that these gains are discouragingly small on average (Wayne & Youngs, 2003); and, while locally driven change efforts have produced some dramatic cases of instructional improvement in troubled schools, research shows that locally driven approaches to instructional improvement most often work on a hit-or-miss basis, with only a few (of many) schools trying this approach actually showing real signs of instructional improvement (Berends & King, 1994).

The Concept of School Improvement by Design

By the late 1990's, the uneven progress of education reform was leading many in the education community to become interested in what we call "school improvement by design." As defined here, this approach to school improvement occurs when local schools work with outside agencies to implement new designs for educational practice. As we see it, two elements of this definition warrant attention. First, the concept of "design" suggests a school improvement process guided by a pre-existing blueprint or specification of educational pratices that can be replicated in *many* school settings. Such a design can include core instructional or curricular components, such as new curricular materials and/or sets of teaching routines; but equally important, these designs also frequently include blueprints for organizational practices that allow the core instructional parts of the design to be implemented faithfully and used effectively in schools. Second, our definition of school improvement by design focuses on situations where local schools work with *outside* organizations to stimulate instructional change. This is important because many researchers and reformers want to study approaches to school improvement that can be replicated reliably and faithfully in many settings (not just a single school), and because the outside organizations that support design-based school improvement make such replication possible. In principle, the organizations referred to in our definition can be governmental agencies, but in American education, they are more often for-profit and/or not-for-profit vendors of instructional materials (e.g., textbooks or software), professional development and/or technical assistance services, or increasingly a blend of the two. In fact, situations where local schools enter into contractual relationships with outside organizations to implement new designs for educational practice are ubiquitous in American education (Rowan, 2001).

This report discusses a particular type of design-based assistance to local schools—assistance that is offered by organizations known as

Comprehensive School Reform providers. In discussing such providers, however, it is important to remember that they are just a single example of a larger movement toward design-based, technical assistance that has been present in the U.S. education system for at least 50 years. Indeed, the movement toward design-based school improvement has its origins in the federal government's attempts since the 1950s to build a research and development (R&D) infrastructure in the field of education. This movement originated with the Cooperative Research Act of 1954, which authorized the U.S. Office of Education to conduct research with universities and state departments of education, and then expanded in the 1960s, as first the National Science Foundation, and then the U.S. Office of Education (and its successors), sponsored curriculum development projects, contracted with universities and other non-governmental organizations to build a network of education laboratories and R&D centers, and funded a set of special-purpose organizations to promote the dissemination and utilization of innovative designs for practice. In building this infrastructure, federal policy makers and education researchers hoped to create a "research-based" process of school improvement that would begin with research on practical problems, and then move to the development of new educational programs and practices, which would be disseminated widely to schools, where they would be utilized in practice. This was the well-known "RDDU" (or **r**esearch, **d**evelopment, **d**issemination, **u**tilization) paradigm of school improvement that captured the attention of policy makers and researchers during the expansive period of educational R&D in the 1960s and 1970s.

Although our definition of school improvement by design is not restricted to research-based designs, "research-based" designs for school improvement are especially important today. In 2002, for example, Congress declared that a primary goal of No Child Left Behind was to "ensure the access of children to effective, scientifically based instructional strategies…" and it peppered the bill with over 100 references to "scientifically-based research." Around the same time, Grover Whitehurst, the first Director of the U.S. Department of Education's newly created Institute for Education Sciences, also testified before Congress, arguing that "there is every reason to believe that if we invest in the education sciences and develop mechanisms

to encourage evidence-based practices, we will see progress and transformation [in education]… of the same order and magnitude as we have seen in medicine and agriculture" (Whitehurst, 2002). For these reasons, information about the conditions under which "research-based" designs for school improvement can be successfully implemented in local schools, and the conditions under which such implementation can be expected to improve student learning, is especially important.

The Emergence of Comprehensive School Reform

This report addresses these questions by examining a particular approach to design-based school improvement that came to be known in the 1990s as comprehensive school reform (CSR). During this time period, the movement toward comprehensive school reform arguably became the "poster child" for scientifically based reform in American education, having been supported initially by business leaders and philanthropists, and then by the Comprehensive School Reform Demonstration Act, and finally by Part F of No Child Left Behind, which gave states funding to award competitive grants to local schools to facilitate adoption of CSR programs locally. Representative David Obey (D-WI), who co-sponsored the first federal bill supporting comprehensive school reform, called this movement "the most important education reform effort since Title I because CSR programs give local schools the tools…they [need to]raise student performance to…high standards" (Congressional Record, 1997).

Interestingly, the CSR movement was not the creation of the federal government. Rather, it was first initiated in 1991 by a private, not-for-profit organization known as the New American Schools Development Corporation (NASDC). Founded as part of President George H.W. Bush's America 2000 initiative, NASDC (later renamed New American Schools [NAS]) provided the kinds of venture philanthropy and political capital that were needed to catapult comprehensive school reform to national prominence. Under the leadership of David Kearns, Chairman emeritus of the Xerox Corporation and a former Deputy Secretary of Education, NAS raised more than $130 million in contributions from the nation's top businesses and foundations with the explicit goal of fostering what it called "a new generation of American schools." Researchers at the RAND Corporation who studied NAS during this key period reported that the organization'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" (Berends, Bodilly, & Kirby, 2002, p. xv).

To make this core idea a reality, NAS funded the development of several new, "break-the-mold" designs for school improvement that it called "whole-school reforms." After selecting 11 organizations from a competitive request for proposals responded to by over 600 applicants, NAS began its work in 1992 with a one-year development phase, during which time the selected organizations (known as "design teams") created new designs for schooling. This was followed by a two-year demonstration phase, during which these organizations worked in a small number of demonstration sites to get their new designs implemented, and then by a five-year scale-up phase in which the organizations worked in a larger set of school districts chosen by NAS to get the new designs implemented even more broadly.

Although the NAS scale-up effort met with uneven success (only seven of the original 11 design teams made it out of the scale-up phase), NAS nevertheless gave a tremendous boost to the idea of school improvement by design. In 1997, for example, when the NAS scale-up phase ended, the surviving NAS design teams were working with over 685 schools around the country. Then, with federal funding from the Comprehensive School Reform Demonstration Act, and later from Part F of NCLB, nearly 7,000 schools across the country adopted CSR designs provided by well over 600 different organizations. By any count, this was a remarkable rate of uptake for an educational innovation. In a few short years, roughly 10% of all public schools in the United States had adopted a CSR design, more than twice the number of schools that were operating as charter schools during the same time period.

The Focus of this Report

This report takes a closer look at design-based school improvement by reporting on the results of a multi-year study of three of America's most widely disseminated comprehensive school reform (CSR) programs. The study described here (known as A Study of Instructional Improvement [SII]) was conducted between 1999 and 2004 and was a large-scale, quasi-experiment involving three well-known CSR programs: the Accelerated Schools Project (ASP), America's Choice (AC), and Success for All (SFA). At the time of this study, these three CSR programs were operating in more than 2,500 elementary schools across the United States. Our approach to studying these programs was to form a sample of 115 elementary schools (31 AC schools, 30 SFA schools, 28 ASP schools, and 26 Comparison schools) located in 17 different states in all geographic regions of the country and to balance the schools in this sample, as much as possible, in terms of geographic location and school demographic characteristics. By design, however, the final sample in the study over-represented schools in the highest quartiles of socioeconomically disadvantaged schools in the United States, since a major goal of the research was to study instructional improvement in high-poverty settings.

In describing the results of this study, we discuss three important issues. First, we look at the strategies each of these CSR programs used to promote instructional change inside of schools. Our aim in this section of the report is to develop some conceptual models to describe the "designs" external agencies use as they go about organizing schools for instructional change. Next, we examine whether or not these designs do, in fact, promote changes in school organization and instructional practice, and if so, in what direction. As we shall see, making changes to school organization and instruction were major goals of each of the CSR programs under study, but each program had a different vision of what this should look like. Finally, we provide

some preliminary data on whether or not each of these CSR programs succeeded in improving student achievement in the schools where they worked.

In addressing each of these questions, we hope to address various puzzles that researchers studying school improvement have confronted over several decades. Among these puzzles are theoretical questions about: (a) how instructional improvement programs can be designed; (b) the relationships between design characteristics and program implementation; and (c) why some externally designed programs work to increase student achievement while others do not. These questions have animated research on school improvement for decades, and they are controversial. For this reason, we turn to a discussion of how these questions have been addressed in previous research on design-based school improvement.

Background

In many ways, research on the design, implementation, and instructional effectiveness of CSR programs echoes familiar themes in research on education reform in the United States. Like many previous reform efforts, the CSR movement began when an influential and dedicated group of reformers (in this case business and government leaders) succeeded in promoting (and, through legislation, institutionalizing) a new template for school improvement. This new template then diffused widely and quickly through the education system, as several thousand schools adopted one or another CSR program. But, while adoption of CSR programs was seemingly quick and easy, implementation at local sites turned out to be difficult (Bodilly, 1996; Berends, Bodilly, & Kirby, 2002; Desimone, 2002; Mirel, 1994), and program evaluations gradually uncovered a pattern of weak effects on the reform's intended goal—to improve the academic achievement of students (Borman, Hewes, Overman, & Brown, 2003). As a result, enthusiasm for the new reform strategy waned, and American education policy veered away from what was once considered a promising approach to school reform in order to find a new magic bullet for school improvement.

There is a problem with this story, however. First of all, while a meta-analysis of CSR program evaluations conducted by Borman et al. (2003) showed that CSR program effects on student achievement were quite small on average (Cohen's $d_{sd} = .12$ in comparison group studies), the analysis demonstrated that there was a great deal of program-to-program variability in effect sizes (with Cohen's d_{sd} varying from $-.13$ to $+.92$ in comparison group studies). Thus, some CSR programs apparently worked much better than others in improving student achievement, a common finding in evaluations of externally-designed school improvement programs dating to the earliest evaluations of Follow Through designs (see, e.g., House, Glass, McLean, & Walker, 1978; Gersten, 1984).

The central objective of this report is to develop an explanation for the variable effects on student achievement that occur when schools embrace design-based instructional improvement programs. Previous research on this issue has tended to focus on three determinants of program success: the nature of the problem being addressed by a social policy or program (e.g., the problem's inherent complexity or uncertainty); the nature of the program itself (e.g., features of the program's design); and the social context in which the intervention or policy change is attempted (e.g., the degree of conflict present over policy or program goals, the coherence of the policy environment in which change is attempted, the motivation and skill of personnel implementing the program or intervention, and the organizational culture, climate, and authority structure under which implementing personnel work). By holding constant the problem being addressed by the CSR programs we studied (i.e., instructional improvement), and by limiting the social context in which these CSR programs operate (to matched samples of elementary schools), our work focuses on an examination of program designs as the key factor explaining program outcomes.

Table 1 shows our assumptions about how a design-based explanation might be used to explain CSR program outcomes. Simply put, this explanation assumes that effective designs resolve two problems of intervention simultaneously. First, organizations providing design-based assistance to schools cannot succeed in raising student achievement unless their designs for instructional practice are different from (and more effective than) existing instructional practices. This statement is akin to the old adage that if you keep on doing the same old things, you cannot expect to get different outcomes. But second, Table 1 also shows that building a CSR program around an effective instructional design does not guarantee improved student learning unless there also exists an effective strategy for getting that instructional design implemented in schools. From this perspective, school improvement by design works under limited circumstance and can go wrong in several ways. An externally-developed program works when it is built around an effective instructional design and a sound implementation strategy. Programs can fail, however, if they are built around an instructional design that is more effective (in principle) than existing practice when it has a poor design for implementation.

Alternatively, a program can fail if it has a very strong design for program implementation but is built around a weak and ineffective instructional design. Finally, in the worst case scenario, an external program might be built around poor ideas about both instruction and implementation. From this perspective, building an effective design is difficult and requires attention to both instructional design and implementation support.

Table 1. Relation of Design to Instructional Improvement

Effective Implementation Strategy?	Effective Instructional Design?	
	Yes	No
Yes	Changes in instruction and learning	Changes in instruction without effects on learning
No	No change in learning or instruction	No change in instruction or learning

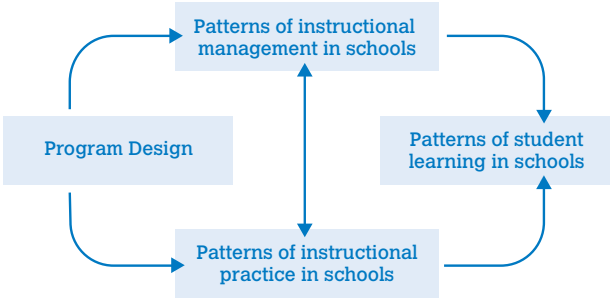
It is worth noting that while this basic idea seems obvious, much prior research on design-based instructional improvement has failed to gather data on these twin issues of instructional design and implementation simultaneously. As a case in point, consider the large body of research on curriculum development projects supported by the National Science Foundation (NSF) in the 1960s—arguably America's first attempt at large-scale school improvement by design. Research on this educational reform effort often took for granted that these NSF-supported curricular designs were more effective than existing materials (especially since the innovative curricula were developed by universities and prestigious not-for-profit organizations). One result of this assumption was that a great deal of research on this reform effort focused on problems of implementation. Indeed, a major finding from this body of this research was that few NSF curricula were implemented with any fidelity at all (for reviews of this literature, see Welch, 1969; Darling-Hammond & Snyder, 1992; Elmore, 1996). Obviously, this is an important finding, but it does not tell us whether the new curricula—if implemented—would have improved student outcomes, and so we, in fact, gained only partial information about the process of school improvement by design from this research.

A somewhat different problem plagued the next generation of research on innovative programs. Consider, for example, the so-called "planned variation" experiment designed to evaluate alternative Follow Through designs. Here, researchers focused on measuring student outcomes, but as a result of funding problems, failed to collect measures of program implementation (House et al., 1978). A major finding in this research was that there was great variability in effects on student outcomes across different Follow Through designs. But explanations for this finding were the subject of a huge debate, largely because researchers could never tell if the variability in program effects was due to differences in the effectiveness of the instructional designs across programs themselves, differences in the ability of program designers to get their instructional designs faithfully implemented in the multiple school settings where they operated, or both (for a review of this literature, see the essays in Rivlin & Timpane, 1975).

Approach

Fortunately, much has been learned in succeeding decades about how to study design-based intervention programs. The general idea has been to build a "logic model" that describes the "theory of action" underlying a particular reform effort, and to use that model to lay out both the intermediate and final outcomes that reformers hope to achieve as a result of their reform efforts. Our effort to formulate such a model for the process of school improvement by design is shown in Figure 1. That figure begins on the left-hand side with the assumption that *any* provider of design-based assistance has a program design, which we earlier defined as a blueprint for change laid out along two dimensions: (a) an instructional design, and (b) a design for school organizational practices that encourage faithful implementation and productive use of that instructional design. Moving to the right in the figure, we have included a set of arrows describing our assumption that these designs influence the ways schools are organized to manage and support instruction and to encourage the use of particular instructional practices in schools. Finally, the arrows in Figure 1 suggest that organizational and instructional practices in schools are reciprocally related and affect student outcomes.

Figure 1. Logic Model of Design-based Instructional Improvement



In the field of education policy research, much effort has gone into building highly general conceptual frameworks to describe each step of this process. For example, Berman (1978) developed an influential conceptual framework that described intervention designs as either "programmed" or "adaptive" in order to capture the fact that programs can be more (or less) explicit and directive about the kinds of organizational and instructional practices they want to implement in schools. Others have developed conceptual frameworks to describe organizational practices for managing instruction, for example, the contrast between "mechanistic" or "organic" forms of management that Miller and Rowan (2006) used to signal the extent to which patterns of instructional supervision, monitoring, and decision making in schools are either centralized, standardized, and routinized, or decentralized and flexible. Finally, attempts have been made to characterize instructional practices in schools as oriented either to "basic" or to "higher order" instructional goals (e.g., Cohen, McLaughlin, & Talbert, 1993), or to "reform" versus "traditional" practices (Stecher et al., 2006).

In what follows, we depart from these familiar categories to describe how the CSR programs we studied actually worked. That is not because we have a quarrel with these more generalized conceptual frameworks. In fact, we find them useful. However, in our own work, we have found that such general categories do not provide the kinds of nuanced descriptions of CSR program designs, intended organizational practices, or intended instructional practices that are needed to explain program-specific outcomes in a logically compelling fashion, either at the intermediate stage of our model (where we are looking at organizational and instructional change resulting from CSR participation) or at the final step of our model (where we are looking at student achievement outcomes that result from working with a specific CSR program).

The remainder of this report is divided into four sections that mirror Figure 1 and sketch out our logic model for studying the process of school improvement by design. The section immediately following this one provides an overview of this logic model by briefly reviewing the main findings from A Study of Instructional Improvement. In this section of the report, we provide brief "portraits" of the three CSR designs we studied, and discuss what was found about how they were implemented in schools, and to what effects on student learning. The

following three sections then report in more detail on these portraits. Here, we follow Figure 1 by discussing in turn: (a) how the different CSR models organized to promote instructional change in classrooms; (b) whether the CSR programs succeeded in moving schools toward their preferred instructional designs; and (c) whether, once implemented, the CSR programs succeeded in improving student achievement at the schools under study. Considering these topics as a whole, we seek to shed light on the specific mechanisms through which the CSR designs under study influenced student achievement, particularly reading achievement, a major emphasis of improvement in all of the schools we studied.

Brief Portraits of the CSR Programs Under Study

Over a four year period, SII researchers used program documents, field research, and survey data to examine the design,[1] implementation, and instructional effectiveness of the three CSR programs examined in A Study of Instructional Improvement. Not surprisingly, these analyses showed that each of the CSR programs studied had different "designs"—that is, ideas about the kinds of instructional practices they wanted to implement in schools and the ways they wanted to organize schools to support the process of instructional change. More importantly, we have found that these designs had important consequences for the schools working with these programs. For example, we have found that schools working with the different programs varied in: (a) how they were organized to manage instruction, (b) the kinds of literacy instruction they provided to students, and (c) the patterns of student achievement growth that occurred at different grade levels. This section lays out our findings in these areas and provides a preview of results that will be discussed in more detail at later points in the report.

The Accelerated Schools Project (ASP)

In previous studies (Rowan, Camburn, & Barnes, 2004; Correnti & Rowan, 2007; Rowan & Miller, 2007), we have argued that at the time of our study, ASP was using a pattern of "cultural controls" to bring about instructional change in schools. We use that description because ASP's approach to working with schools revolved around promoting a normative commitment among school leaders and faculty to the program's abstract vision or ideal of "powerful learning" for all students. Indeed, from the onset of working with schools, ASP

[1]It is important for the readers to keep in mind that our analysis of the design of the CSR programs reflects the designs as they existed at the time of our data collection (from 2000 to 2004). The designs evolved during the course of the study, and continue to evolve in response to research, current demands of the reform environment, and their own intuitions about the school improvement process.

facilitators used the staff development process to emphasize the program's commitment to this abstract construct, and to define powerful learning as constructivist in nature, with an emphasis on authentic, learner-centered, and interactive forms of instruction. However, ASP was not prescriptive in nature. For example, it did not target particular school subjects for improvement, nor did it provide teachers with a great deal of explicit guidance about curriculum objectives or teaching strategies. Instead, ASP facilitators helped schools use a systematic process of organizational development to design a unique path toward powerful learning and to adopt locally-appropriate forms of instrutional practice consistent with this approach. In this sense, ASP had a design best labeled as "adaptive" in form.

This description, it is worth noting, suggests that ASP's approach to producing instructional change is at least partially at odds with many of the design features that previous research on educational innovation has identified as promoting implementation success. For example, at the time we studied ASP, the program's goals for change were generic in form—aiming at broad changes across the board rather than targeting specific areas of the curriculum for change. By contrast, prior research on design-based assistance suggests that successful programs often focus on changing specific, curriculum-embedded elements of instructional practice as opposed to more diffuse elements that cut across curricular areas or represent generic forms of teaching (Cohen & Hill, 2001; Desimone, Porter, Garet, Yoon, & Birman, 2002; Fennema et al., 1996). Moreover, the kinds of changes teachers were expected to make as a result of participating in ASP were not formally specified, and instead, each school (and each teacher within a school) was asked to "discover" the most appropriate means to producing powerful learning within his or her own particular context. Again, this differs from the stance taken by more successful programs, which tend to have clearly defined goals for change in particular curricular areas, that is, a clear specification of what features of curriculum and instruction will be changed, and of the steps to be taken to achieve these changes (McLaughlin & Marsh, 1978; Elmore & Burney, 1997; Nunnery, 1998). For these reasons, schools and teachers had a great deal of autonomy in the ASP system, with the result that there was little real focus on implementation fidelity, either from external

program facilitators, or from internal leaders. In fact, in previous research, we have found that ASP schools had the lowest reported levels of instructional leadership of all the schools in our study sample (Camburn, Rowan, & Taylor, 2003). This too contrasts with previous research, which suggests that successful programs are those in which program designers, local program facilitators, and local administrative leaders demand fidelity to planned changes in instructional practice (Loucks, Cox, Miles, & Huberman, 1982; Huberman & Miles, 1984; Stringfield & Datnow, 1998).

In light of these patterns, it is not surprising that a particular pattern of instructional improvement emerged in ASP schools. First, as we report below, staff in ASP schools reported that school improvement plans were weakly (not highly) specified; that is, they saw plans as centered more on shared values and group investigation than on specific curricular goals and objectives. In this environment, teachers in ASP schools were trusted to make innovations in their classrooms as they saw fit, so long as these innovations were consistent with the normative ideal of powerful learning. The autonomy afforded by this system of instructional control promoted high levels of motivation for improvement among faculty. But, as we demonstrate below, the approach did not promote implementation of distinctive forms of literacy instruction in schools; instead, we will show that literacy instruction in ASP schools looked very much like literacy instruction in our sample of comparison schools. Moreover, because instructional practices were so similar, patterns of student achievement were also quite similar across ASP and comparison schools.

America's Choice (AC)

The America's Choice program took a contrasting approach to instructional change at the time of our study, stimulating instructional improvement using what we have called a pattern of "professional controls" (Rowan, Camburn & Barnes, 2004; Correnti & Rowan, 2007; Rowan & Correnti, 2006; Rowan & Miller, 2007). The AC program had its origins in the standards-based reform movement, and, as a result, the program was built around some definite ideas about the curricular content and methods of teaching it wanted to occur 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). Moreover, 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 expected schools that adopted the program to create two new leadership positions—a design coach and a literacy coordinator. Design coaches were expected to help principals implement the program, while AC literacy coordinators were expected to work with classroom teachers. Previous research showed that levels of instructional leadership were highest in the AC schools in our study sample (Camburn, Rowan, & Taylor, 2003).

Our findings (discussed below) show that the "professional" control strategy used by AC had strong effects in the schools where they worked. First, AC schools tended to organize for instructional improvement in ways that were quite different from schools in the control group or in ASP schools. For example, given all the instructional guidance that AC teachers and leaders received, it is not surprising to find that staff in AC schools viewed their school improvement plans as clear and well-specified. Moreover, as a result of extensive coaching, AC schools tended to be characterized by strong instructional leadership. But this pattern of strong leadership and clear instructional guidance came at a price—with teachers in AC schools reporting lower levels of autonomy of practice than did teachers in comparison group schools. Still, as we shall see, the presence of strong instructional leadership—coupled with a well-specified instructional design—produced distinctive instructional practices in AC schools. As our findings demonstrate, AC teachers were much more likely than other teachers to emphasize writing instruction as part of their literacy program, and to place more emphasis on students' production of extended written text. This approach to literacy instruction, in turn, seemed to produce accelerated growth in students' literacy achievement at the upper elementary grades.

Success for All (SFA)

Success for All illustrates yet a third model for promoting instructional change in schools: what we elsewhere have called "procedural controls" (Rowan et al., 2004; Correnti & Rowan, 2007; Rowan & Miller, 2007). Of the three programs under study, SFA gave schools the clearest and most highly-specified plan for instructional improvement by producing a set of highly-specified instructional routines for the teaching of reading. In particular, the SFA program was built around two core principals: (a) offering students cooperative learning opportunities during class instruction, and (b) a clear and well-defined reading curriculum that provided teachers with a weekly lesson sequence, where each lesson in this sequence was designed around a "script" intended to guide teaching activities through a 90-minute reading period. In kindergarten and first grades, moreover, these scripts were accompanied by program-provided curricular materials for use throughout the school. For grades two through six these scripts were designed to be used in conjunction with existing reading curricula in the school.

SFA schools also were more centrally managed than other schools in our study. For example, the SFA design sought to intervene on the students (e.g., through tutoring and family support) in order to increase their attendance in school and to better prepare them to benefit from classroom instruction. Additionally, schools implementing SFA were expected to appoint a full-time literacy coordinator, and this staff member was given substantial responsibility for school-wide coordination of the reading program, including the task of constituting reading groups and making teaching assignments to these groups on a school-wide basis every eight weeks. In addition, instructional leaders in SFA schools and SFA linking agents were asked to supervise implementation of SFA instructional routines. In prior research, levels of instructional leadership in SFA schools were not statistically different than AC schools, but were higher than levels of instructional leadership found in ASP schools (Camburn, et al., 2003).

Given SFA's approach to promoting instructional change, it is not surprising that the SFA schools in our study were characterized not only by distinctive patterns of organizing for school improvement, but also by distinctive patterns of literacy instruction and student achievement. For example, as we demonstrate below, staff in SFA

schools saw school improvement plans as highly specified and as focused squarely on a particular instructional target (reading). In addition, school leaders worked mainly to monitor instruction, reducing teacher autonomy and standardizing instruction. This emphasis on faithful implementation of instructional routines, we found, produced a distinctive pattern of teaching practices that was generally faithful to the program's instructional design. In particular, more than any group of teachers in our sample, SFA teachers enacted what we have called a "skills-based" reading program focused on direct/explicit teaching of reading comprehension strategies and an emphasis on having students demonstrate their reading comprehension through simple, direct responses to oral questions and/or short written work. As we shall see, this instructional approach appeared to accelerate students' reading achievement in the early elementary grades.

Findings on Instructional Improvement Processes in CSR Schools

Having presented initial portraits of the CSR schools under study, we turn now to several detailed discussions of the data analyses from which these portraits were constructed. The first analysis to be discussed in this report examines the organizational processes for school improvement found in the four groups of schools under study. A detailed discussion of these findings has been presented in Rowan and Miller (2007). That paper used measures of organizational processes constructed from surveys administered annually over the four-year course of the study to approximately 5,500 teachers and about 800 school leaders. Specifically, measures of organizational processes for school improvement were developed from teacher and school leader survey responses in four analytic domains: (a) the degree of *instructional guidance and standardization* in schools; (b) the degree of intensive *instructional leadership* present in schools; (c) schools' emphasis on faculty *innovation, discretion, and autonomy*; and (d) the strength of faculty professional *communities* in schools.

Overall, a total of 12 separate measures of these organizational processes were constructed in three steps. First, person-level scale scores were created by applying the Rasch model (a one parameter item response theory model) to the survey responses of teachers and school leaders (see Bond & Fox, 2001 for an accessible discussion of the Rasch model). Second, these scores were modeled as outcome measures in three-level hierarchical linear models (HLM) which nested individuals' annual scores on a particular scale within individuals, who were then nested within schools. In a third step, school-level scores on the outcome measures were derived from these HLM models. Specifically, the variables reported in this section are school-level empirical Bayes residuals from the models and indicate a school's average score on a measure, aggregated across all teachers (or leaders) and all four years; after controlling for staff characteristics, student characteristics, and school size. In all instances, these empirical Bayes

(EB) residuals are standardized scores with mean = 0 and standard deviation = 1. The results reported below are based on comparisons of the overall means of the four quasi-experimental groups on the 12 outcome measures, where group means were "bracketed" by their standard errors. By comparing the ranges defined by the standard errors of the means, we were able to get a sound understanding of both the magnitude and degree of uncertainty of group mean differences in these organizational measures.

Findings

As expected, the four quasi-experimental groups differed in important ways with respect to the amount of instructional guidance and standardization that was reported to exist in schools. Here, three separate measures of instructional guidance were examined: (a) school leaders' reports of the extent to which there was a press for a standardization of practice in the school, (b) teachers' reports of how closely their improvement efforts were monitored, and (c) teachers' reports of instructional guidance they received.

Rowan and Miller (2007) showed that schools participating in the AC and SFA programs exhibited much higher levels of instructional guidance and standardization than schools participating in the ASP program and comparison schools. For example, we have found that among the four quasi-experimental groups in SII, the press for standardization was reported to be much greater in schools implementing AC and SFA than in comparison or ASP schools. In addition, teachers in AC and SFA schools felt that their improvement efforts were monitored more closely than did teachers in ASP and comparison schools. Teachers in schools implementing AC and SFA also reported receiving greater levels of instructional guidance than teachers in ASP and comparison schools; however, mean differences on the instructional guidance variable were not statistically significant.

The second domain of organizational processes examined was instructional leadership. Here, school leaders reported on three dimensions of instructional leadership—their involvement in staff development, their advising of teachers on matters of instruction, and their efforts at setting a vision for teaching and learning in the school. The data indicate that schools implementing the AC design were higher than schools in the other three groups on all three dimensions

of instructional leadership. The differences between AC schools and schools in the other three groups are statistically significant in nearly every case. These findings on leadership processes illustrate a key difference in implementation strategy between SFA and AC. Both programs provided teachers with strong instructional guidance. However, SFA's strategy of procedural controls relied primarily on scripted lesson routines to secure faithful implementation, complemented by school leaders working closely with teachers to monitor and reinforce the use of scripted lesson routines. By contrast, AC relied more on having school leaders work closely with teachers to help them develop the knowledge and capabilities to use the design effectively.

The third organizational process we examined was schools' emphases on teacher autonomy and innovation. Here, we examined a measure of leaders' reports of teacher autonomy, leaders' reports of the prevalence of values-based decision making in the school, and teachers' reports of support for innovation in the school. Given ASP's strategy of cultural control, we anticipated that ASP schools would have higher average scores than AC and SFA schools on these three measures. The results confirmed this prediction. The means for schools implementing the Accelerated Schools Project design were higher than the means of the three other quasi-experimental groups on all three measures of teacher autonomy and innovation, and with only a few exceptions, these mean differences were statistically significant.

The fourth and final organizational process examined was the strength of professional community in the schools under study, which was assessed through three measures: teachers' reports of trust and respect among the faculty, the prevalence of collaboration on instruction, and the prevalence of critical discourse among school staff. Recall that the ASP strategy of cultural control relied strongly on teachers to generate locally-proposed instructional improvements. For this approach to be successful, however, schools would seem to need strong professional communities in which high levels of trust, critical discourse, and collaboration were evident. We found evidence that this was indeed the case in ASP schools. As with the indicators of teacher autonomy and innovation, ASP schools were higher than the other three quasi-experimental groups on all three measures of professional community, though the differences on these indicators were not as large as differences

found in the areas of autonomy and innovation. In particular, ASP teachers reported substantially greater levels of trust among faculty members than did teachers in AC schools; further, ASP teachers also reported a greater prevalence of critical discourse among colleagues than did teachers in AC schools.

Summary

Our analyses of data on organizational processes in schools illustrate how externally-designed and operated instructional improvement programs can pursue very different strategies to produce instructional change in schools. In particular, the survey data just discussed show that ASP's strategy of cultural control led to the development of school environments that were quite strong in professional community (i.e., trust among faculty, critical discourse, and teacher collaboration) and led to a great deal of teacher autonomy in pursuit of classroom instructional innovations. In contrast, AC and SFA used very different approaches to instructional reform. SFA, for example, pursued a strategy of "procedural controls" to promote instructional change, and as our data show, this approach led to relatively high levels of instructional guidance and to an associated press for instructional standardization. But, as we saw, SFA relied heavily on procedures and routines, with school leaders monitoring and reinforcing the enactment of routines. However, this approach failed to stimulate a sense of strong professional community. Finally, America's Choice also emphasized a significant amount of guidance and press for instructional standardization as part of its instructional improvement strategy, but it did so not by emphasizing scripted instructional routines, but rather by encouraging development of strong instructional leadership in schools. As with SFA, this emphasis on standardization and leadership worked against the formation of strong professional communities and also decreased the press for innovation and autonomy in AC schools.

Patterns of Literacy Instruction in CSR Schools

An interesting question is whether these different approaches to promoting instructional change were consequential, especially in promoting distinctive instructional practices in the schools under study. Results from SII provide evidence on this issue. As we are about to see, both AC and SFA managed to get their preferred instructional practices implemented faithfully in schools, whereas instruction in ASP schools was indistinguishable from that observed in comparison schools.

To examine instructional practices in the schools under study, SII researchers analyzed data from 75,689 instructional logs that were collected from 1,945 classroom teachers in grades one through five over the course of the study. In general, log data have been analyzed by SII researchers using three-level hierarchical linear models that nest multiple log reports within teachers within schools (see, e.g., Rowan, Camburn, & Correnti, 2004). The point of the analyses reported here was to test for mean differences in instructional practices across schools in the different quasi-experimental groups after adjusting, through propensity score stratification for many different school-level, pre-treatment covariates, as well as important lesson and teacher level characteristics that might differ across quasi-experimental groups.

Results

The results of these analyses have been discussed in considerable detail in Correnti and Rowan (2007). In particular, these researchers reported very distinctive patterns of instruction for schools in the AC and SFA quasi-experimental groups, but little distinctiveness of instruction for the ASP schools in the sample. In the analyses presented here, we discuss the findings from this report in three main areas of literacy instruction: word analysis, reading comprehension, and writing.

We begin by noting that we observed no significant differences in literacy teaching practices between ASP schools in the study and the comparison schools. That means that, on average, students in ASP schools would have experienced instructional opportunities that were virtually the same as students in comparison schools. This result is not surprising, especially in light of the school improvement strategy pursued by the ASP program. ASP's strategy of "cultural controls" did not prescribe specific instructional practices in the area of literacy but rather left it up to individual schools and their teachers to determine which instructional practices to implement. When left largely to their own devices, teachers in ASP schools apparently implemented the same patterns of instruction that were common in comparison schools. This indicates that ASP's approach to reform, similar to at least some other unsuccessful reform efforts, is not well-suited for creating large-scale instructional changes.

By contrast, Correnti and Rowan (2007) found substantial differences in literacy instruction between teachers in AC and comparison schools. Moreover, these differences occurred precisely where the AC instructional design was most prescriptive—in the area of writing and in the production of written text by students. In fact, the magnitude of these differences was quite large by social science standards. Controlling for lesson, teacher, and school characteristics, for example, Correnti and Rowan (2007) showed that AC teachers focused on writing in 54% of all lessons, whereas comparison teachers focused on writing in just 38% of all lessons. AC teachers also differed in the instructional practices and curricular content they covered when they taught writing. For example, on days when writing was taught, AC teachers were more likely than comparison teachers to have engaged in six of the 10 writing-related instructional practices measured by SII researchers. In particular, when they taught writing, AC teachers were more likely than comparison teachers to also have the lesson focus on reading comprehension and to directly integrate work in reading comprehension with their work in writing. They also were more likely to explicitly teach the writing process, more likely than comparison teachers to provide instruction on literary techniques or different writing genres, and more likely to have students share their writing and do substantive revisions to their writing. Additionally, AC teachers were more likely than comparison teachers to have their

students write multiple connected paragraphs as they taught writing. Sensitivity analyses revealed that these findings were not likely due to omitted variable bias (Correnti & Rowan, 2007). Finally, we have recently examined measures of variability of these measures of writing instruction. This evidence is compelling because a measure of variability—the confidence interval for the coefficient of dispersion (Bonnet & Seier, 2006)—reveals that variability in writing instruction among AC teachers was less than it was for comparison teachers, or indeed, for teachers in each of the other CSR designs. Thus, not only did AC teachers have higher means, on average, they also were less variable in their use of these instructional strategies. This is further evidence of the design's effect on literacy instruction. Moreover, this reduction in variation was largely due to a reduction in variance among teachers within schools and less to a reduction in variation across schools.

Correnti and Rowan (2007) also found large differences in instruction between SFA and comparison schools. In SFA schools, teachers were more likely to teach reading comprehension on a daily basis and they also taught comprehension differently from comparison teachers when it was taught. Here, for example, the average SFA teacher taught reading comprehension in 65% of all lessons, while the average comparison school teacher taught comprehension in 50% of all lessons. Moreover, when reading comprehension was taught, SFA teachers were more likely than comparison group teachers to use teacher directed instruction, to focus on literal comprehension strategies, to check students' comprehension by eliciting brief answers from students, and (due to extensive use of cooperative grouping arrangements) to have students discuss text with one another. It is noteworthy also that teachers in SFA schools did *not* compromise any other aspect of comprehension instruction in order to obtain these significant differences. That is, in lessons where comprehension was taught, teachers in SFA schools were no less likely than comparison school teachers to focus on more advanced reading strategies or write extended text about what they read. They did, however, more frequently provide direct instruction on reading strategies with more frequent checks for student understanding requiring brief oral or written answers from students. And, as was the case with AC schools, teachers in SFA showed less variability in their reading comprehension

instruction than did teachers in the comparison schools or teachers in schools participating in the other CSR designs.

Summary

The analysis of literacy instruction practices in CSR schools is important for two reasons. First, it suggests that the ways in which CSR programs organized schools for instructional improvement was consequential, not only for the kinds of organizational processes that emerged to support instructional change within schools, but also for the kinds of instructional practices that ended up being implemented. The evidence presented thus far suggests that although ASP's use of cultural controls promoted a strong professional community of teachers working hard on instructional innovation, the lack of a clear instructional design or strong instructional guidance for teachers, coupled with weak instructional leadership, tended to produce quite ordinary instruction that was not different from what was observed in comparison schools. By contrast, AC and SFA were far more prescriptive in their instructional designs. Both used different, but apparently quite effective strategies of "professional" and "procedural" controls to stimulate instructional change, and in both cases, SII researchers observed very distinctive forms of instructional practice in program schools.

A second point is that although both AC and SFA were "prescriptive" in instructional design and developed organizational processes in schools that emphasized faithful implementation of their preferred instructional designs, the instructional designs implemented in AC and SFA schools were quite different. Literacy instruction in AC schools was "literature-based" in emphasis. As a result, students were far more likely to be exposed to direct instruction in writing and to work on extended writing assignments than were students in other schools. By contrast, SFA's instructional design placed more of an emphasis on what might be called "skills-based" reading instruction, that is, explicit instruction in reading comprehension tasks, coupled with a tendency to have students work on providing brief written and oral answers to check for basic comprehension. As we demonstrate in the next section, these differences in instructional practices provide at least one explanation for the patterns of reading achievement found in SII schools.

The findings presented here are based on cross-sectional comparisons of *teachers'* instruction in treatment and comparison schools. But it is important to consider these results from the students' perspective, since the instructional differences observed in a single year add up over time for students who remain in the treated schools. For example, across grades three to five in our study, students in SFA schools experienced about 28% more reading comprehension instruction (341 days vs. 265) than did students in comparison schools. Similarly, students in AC schools experienced about 36% more writing instruction (264 days vs. 194) than students in the comparison schools. These differences are quite substantial for that portion of students who did not move out of the treated schools. Unfortunately, however, rates of student mobility in high poverty schools can be very high. In the SII sample, for example, only 46% of the students originally sampled in third grade remained in their same school by the end of fifth grade.[2] Given our logic model (that differences in achievement growth are likely to be caused by differences in instruction), the accumulated instructional histories of students provide two working hypotheses. One hypothesis is that students in SFA and AC schools are more likely to show differences in achievement growth due to the large differences in accumulated instruction in target areas. A second hypothesis is that such gains will be especially true for non-mobile students who remain in treated schools for multiple years and thus benefit from a greater increase in instructional opportunities than do mobile students in those schools.

[2]Rates of attrition did not vary substantially by treatment or comparison group.

Patterns of Literacy Achievement in CSR Schools

The final step in our work has been to study patterns of literacy achievement in the schools under study. To do so, SII researchers followed two cohorts of students as they passed through the schools under study, examining differences in achievement growth among students in the four quasi-experimental groups. One cohort of students in the study was followed as it passed from kindergarten to second grade over the course of the study, with SII researchers assessing these students' achievement in spring of kindergarten, the fall and spring of first grade, and the fall and spring of second grade. A second cohort of students was followed as it passed from third to fifth grade over the course of the study, and this group was assessed during the fall and spring of third, fourth, and fifth grades. The kindergarten cohort included approximately 3,600 s tudents, while the third grade cohort included approximately 4,000 students. Students in the sample analyzed here were enrolled in a total of 114 schools: 28 ASP schools, 31 AC schools, 29 SFA schools, and 26 comparison schools.

SII researchers used the *TerraNova* assessment published by CTB McGraw-Hill to measure students' growth in literacy achievement. This assessment produced two literacy scale scores for students at each administration—a *Reading Comprehension* score and a *Language* score. The Reading Comprehension scale charted students' academic growth as they moved beyond basic oral and reading comprehension to a level at which they are able to analyze and evaluate more extended text by employing various reading comprehension strategies. The Language scale charted the degree to which students moved from a basic understanding of sound/symbol relationships to a more complex understanding of the structure of the English language. In both the lower and upper grades the findings from the language scale score closely mirror findings from the reading comprehension scale score. As a result, in this paper, we discuss only the findings from the reading comprehension scale score.

The sample of students assessed during the course of the study was more disadvantaged than a representative sample of U.S. students, reflecting the fact that schools in the SII sample were disproportionately drawn from high and medium poverty neighborhoods in order to study the effects of specific instructional interventions on student achievement in high poverty settings. For example, over half of the students in the SII sample were African American, while another nineteen percent were Hispanic. Moreover, over half the parents of SII students had only a high school education or less, and over 40% of the students' mothers were single parents. Finally, substantial percentages of students in the SII sample were at risk of school failure. For example, 20% of SII students received services for learning difficulties, 18% received special education services, and 13% repeated a grade early in their elementary school career. Although schools in the SII sample varied in their degree of disadvantage, on average the schools served a highly disadvantaged student population.

Achievement Growth Models

To examine general patterns in reading achievement in the schools under study, we have been fitting a series of three-level hierarchical linear models (HLM) in which multiple test scores per student are nested within students, who are nested within schools. In these analyses, we have been especially interested in examining differences in rates of literacy achievement growth between each set of CSR schools and the set of comparison schools. In the evaluation literature, such analyses are commonly referred to as "intent-to-treat" models since they examine achievement patterns for any schools nominally involved with a CSR program regardless of its level of implementation. Since schools were not randomly assigned to treatments, we also have been using propensity score stratification to statistically equate schools on 34 observed pre-treatment characteristics and then match the schools (using optimal matching) based on their propensity to have received treatment. Under the assumption of strongly ignorable treatment assignment, the average treatment effect in our statistical models is determined by pooling within-stratum treatment effects—the difference in mean rates of achievement growth between treated and untreated schools with similar pre-treatment characteristics. We caution the reader that the results about to be presented are *preliminary* and have yet to be peer-reviewed.

We also have been examining differences between each set of CSR schools and all other schools in our sample (e.g., SFA schools vs. comparison, ASP and AC schools). Once again, we are using propensity score stratification to statistically equate schools on 34 pre-treatment covariates and to match them using an optimal matching program. In many ways this comparison is superior to the one between each set of CSR schools and the set of comparison schools, not only because the larger sample of schools provides for better matches between treated and untreated schools, but also because some of the schools in our so-called "comparison" group participated in a variety of whole-school reform programs (e.g., Expeditionary Earning/Outward Bound, Direct Instruction).

To date, the findings of all of these analyses seem to follow logically from our discussion to this point. For example, thus far, we have found that the school improvement strategy followed in ASP schools failed to produce instructional practices that were different from instructional practices in comparison schools. As a result, it is not surprising that patterns of achievement in ASP schools were also indistinguishable from patterns of achievement in comparison schools. For both cohorts of students (kindergarten through second grade and third through fifth), our preliminary analyses failed to find any significant differences in students' rates of achievement *growth* over time across ASP and comparison schools—or across ASP and all other schools. Again, given the lack of differences in instructional practices across ASP and control schools, this finding is not surprising.

In contrast, we *have* found differences in patterns of achievement between SFA and comparison schools (under certain model conditions) and between SFA and all other schools. In our preliminary analyses, these differences are most apparent for the lower grades cohort, but the magnitude of these differences varies depending on the statistical adjustments used in the models. In a model with no controls other than the propensity score, for example, SFA students gain about six points more than comparison students over the two-year interval from the spring of kindergarten to the spring of second grade. After adjusting for patterns of student mobility in the schools under study, however, the SFA advantage over comparison schools during the same interval increases to more than 10 points. A similar 10-point advantage holds when SFA students are compared with all other students in AC, ASP,

and comparison schools. This SFA advantage is especially impressive considering that students in the SII study gained in percentile rank over this interval relative to the norming population. An average student beginning our study at the 30th percentile in a comparison school, for example, finished the end of second grade at about the 40th percentile. The SFA effect moved a comparable student in the average SFA school from the 30th percentile to the 50th percentile.

In a similar vein, our preliminary analyses have found statistically significant differences in patterns of achievement growth for students in AC schools in the upper grades. Students in AC schools grew at a significantly faster rate than students in comparison schools and faster than students in all other schools. From the beginning of third grade to the end of fifth grade, for example, our analyses suggest that students in AC schools, on average, scored an additional nine to 12 points on the reading comprehension outcome, depending on the model adjustments. The size and interpretation of the AC effect on reading comprehension is similar to the one found for SFA schools in early grades reading, except that in the upper grades cohort, students in the SII were losing ground relative to the norming population. For example, our statistical models suggest that the average student in a comparison school who began third grade at about the 40th percentile nationally ended the study at about the 30th percentile; by contrast, our models suggest that the equivalent student in the average AC school who began third grade at the 40th percentile would achieve at or above the 40th percentile nationally at the end of fifth grade.

We should note that the effects on reading comprehension for SFA schools in the lower grades and for AC schools in the upper grades represent the average "intent-to-treat" effects of those interventions. We have also have been examining whether exposure to the treatment has influenced student growth in our schools. In preliminary work we have been able to demonstrate effects of exposure to the treatment in a number of ways. For example, we examined the degree of implementation at each school by examining the organizational, instructional, and staff development profiles of schools and the degree to which they reflected the aims of the SFA and AC designs, respectively. Using this strategy, we have found that, in both interventions, schools with better implementation scores show higher rates of achievement growth. Additionally, we also observed an implementation effect

for SFA schools in the upper grades. Within our sample, a number of SFA schools indicated that they participated in Roots and Wings, while others indicated they did not. Thus, while all of the schools considered themselves SFA schools, Roots and Wings schools were more resource intensive, especially in the upper grades, where the Wings curriculum (grades two to six) is enacted immediately upon completion of reading Roots (grade one). Indeed, in the upper grade models, SFA schools that participated in Roots and Wings had students with greater achievement growth than students in other SFA schools that was also significantly better than the growth rates for students in comparison schools.

Additionally, it has become apparent in our work that student mobility plays a large role in moderating effects of the CSR programs on student achievement. For example, we have shown that the overall effects of both SFA in the lower grades and AC in the upper grades increase when the statistical models adjust for student mobility, thus demonstrating that students who stay in treated schools for a longer period of time make greater gains in achievement. These same effects can also be demonstrated when a variable indicating students' entry into the study at the initial starting point is entered into the model. Students remaining longer in an SFA school (in the lower grades) or an AC school (in the upper grades) show significantly higher achievement growth. Third, we have run statistical models adjusting for student demographic characteristics and for school-level propensity strata (but without the CSR program variables entered into the analysis). We then examined student-level residuals, calculating the model predicted gains for each student. When we examined bar plots with 95% confidence intervals for SFA versus other students in the lower grades and for AC versus other students in the upper grades, we found that students making the greatest gains were less mobile and located in schools with less student mobility. More importantly, the *magnitude of the CSR program effect* was greatest for those same students, presumably because they received greater exposure to the treatment.

These results are important for two reasons. First, the fact that the CSR program effect varies with exposure to the treatment strengthens our causal arguments about the effects of program designs on achievement gains. Since students varied in their treatment dosages and since higher dosages tended to predict higher achievement gains, we can

infer that the treatment is more likely than not the causal agent producing the achievement gains. Second, the results also show that the potential for achievement growth after the adoption of one of these interventions is greater than the average "intent-to-treat" effect *if schools have a faithful implementation of the design*. In addition, our results show that the "intent-to-treat" effect on students is a lower bound *if students are present to receive the treatment in successive years*. Consider, for example, that the SFA effect in the lower grades is averaged over all SFA students in the model. But, of these students, only 29.6% received SFA-like instruction in both first and second grade. Another 15.4% were present both years and had one year of SFA-like instruction, while only 4.1% were present both years and received instruction very similar to students in comparison schools in both years of the study. About half of the SFA students were only present for one year of the study.[3] Thus, 37.2% of the SFA students in our achievement models received SFA-like instruction in the one year they were present, while the remaining 13.2% received instruction that was very similar to the instruction received by students in the comparison schools in the one year they were present. Thus, even for an intervention with high rates of implementation fidelity, the transfer of the treatment to individual students over successive years reaches less than one-third of all SFA students in our achievement models. When we examine "intent-to-treat" effects, we must keep in mind that not only is implementation fidelity incomplete (a condition that is potentially manipulatable by CSR design), but also that student mobility severely limits the treatment dosage received by students (a condition that is less easily manipulated by CSR programs or, indeed, by social policies).

[3]Daily log data submitted by teachers was used in a discriminant analysis to identify teachers as being either AC-like, SFA-like, ASP-like or comparison-like in their instruction (see Rowan & Correnti, 2007). Once teachers' instruction was identified as being in one of these groups, the instructional-type was written back to individual students. We then constructed instructional profiles for students across the years of the study. The discriminant analysis correctly identified 76% of SFA teachers as being such, 62% of AC teachers as such, only 44% of ASP teachers as such, and only 36% of comparison school teachers as such. These data show that even when intervention programs produce strong effects on instruction (as did SFA and AC), many teachers do not implement the intended instructional regimes faithfully. The analysis also shows that teachers in both comparison and ASP schools implemented patterns of instruction that was close to the preferred regimes of SFA and AC.

Conclusions and Recommendations for Future Research

What do the specific findings from A Study of Instructional Improvement tell us about the process of school improvement by design and how to study this process in the future? First, our findings suggest that design-based instructional improvement can come in many and varied forms. Intervention providers, for example, can vary in terms of how they organize schools to manage the process of instructional improvement *and* in the kinds of instructional practices they seek to put in place in schools. As we argued at the beginning of this report, and as we demonstrated through a discussion of SII findings, both of these design elements must be explicitly studied if we are to understand the conditions under which design-based instructional improvement efforts actually succeed in improving student learning. Indeed, SII illustrates this lesson in a telling way. As we have seen in this report, one program studied by SII researchers (ASP) managed to produce a very distinctive pattern of instructional management in schools, but that pattern did *not* produce any real changes in instruction, and as a consequence, the program had very little effect on students' reading achievement in the schools where it worked. By contrast, two other programs that were studied by SII researchers appear to have changed schools in ways that *did* promote instructional change. Here, however, the patterns of instruction produced were differentially effective. The AC program, for example, produced what we called a "literacy-based" pattern of reading instruction in schools, but that type of instruction appeared to promote reading gains only at the later grades—not in earlier grades. By contrast, the SFA program produced a pattern of "skills-based" reading instruction in schools that—in some analyses at least—was only effective at the early grades. The obvious implication for future studies is the need to closely examine not only how schools organize to make instructional change, but also to examine explicitly the kinds of instructional changes being made in these schools. Only then can researchers understand if null effects on student achievement

that emerge in many studies are due to implementation failures or to implementation of ineffective instructional practices.

Our findings also show the benefits of moving beyond highly generalized conceptual frameworks for describing program designs, patterns of instructional management, and specific instructional practices in schools. For example, we argued earlier in this report that program designs are often discussed as being either "programmed" or "adaptive" in form (Berman, 1978). Yet SII demonstrates why this dichotomy is simplistic. In this report, we advanced an alternative (and more nuanced) conceptual framework. To be sure, the design features emphasized by ASP appear to be consistent with Berman's discussion of "adaptive" designs, and to a large degree, SFA appears to have been built around a design that can easily be labeled as "programmed" in nature. But the AC program, which we also studied, is not easily classified within this dichotomy, incorporating as it does at least some features of programmed designs and some features of adaptive designs. A similar point can be made about the patterns of instructional management that emerged in AC schools. While organization theorists in education often contrast mechanistic with organic forms of instructional management (e.g., Miller & Rowan, 2006), AC's organizational design once again incorporates features of both. The implication for future research is to continue to look in more fine-grained ways at programs designs, for it is not at all clear that the patterns that we labeled as cultural control, procedural control, and professional control exhaust the "types" of designs that might be encountered as the design-based school reform movement progresses.

Third, we would argue that A Study of Instructional Improvement shows the benefits of looking closely at instruction when studying the process of instructional improvement. While that point seems obvious, it is nevertheless amazing how much research on school reform avoids directly measuring instructional practice when trying to explain student achievement outcomes. Indeed, in SII, the pattern of achievement results that were found make sense only if we appeal to our findings about instructional practices in schools. Why, for example, would two well-implemented instructional programs produce very different patterns of reading growth in the schools where the worked?

After looking at instruction in the schools under study, the answer is obvious—because instruction differed across schools implementing these programs.

Finally, we would argue that the results of SII confirm the larger "logic model" of school improvement by design that we described earlier in this report. That logic model points to the importance of two dimensions of design—the way designs organize schools to produce instructional change, and the kinds of instructional changes the design envisions. Obviously, the results of SII show just how fragile school improvement by design is: A design can fail because of a poor approach to implementation, it can fail because of an ineffective approach to instruction, or it can fail on both counts. That is an important lesson, and one that should guide all future research and development in the area.

References

143 Congressional Record S97-100.H6849-H6890. (September 4, 1997). Statement of David Obey. Retrieved July 29, 2008 from http://frwebgate3.access.gpo.gov/cgi-bin/waisgate.cgi?WAIs-clocID=3583079271+2+0+0&WAISaction=retrieve

Basken, P. (2006, March 29). States have more schools failing behind. *Washington Post*. Retrieved July 28, 2008, from http://washington-post.com/wp-dyn/content/article/2006/03/28/AR2006032801794.html

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

Berends, M., & King. B. (1994). A description of restructuring in nationally nominated schools: Legacy of the iron cage? *Educational Policy*, *8*(1), 28-50

Berman, P. (1978). *Designing implementation to match policy situation: A contingency analysis of programmed and adaptive implementation*. Santa Monica, CA: RAND.

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

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bonett, D., & Seier, E. (2006). Confidence interval for a co-efficient of dispersion in nonnormal distributions. *Biometrical Journal, 48*(1), 144-148.

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*, 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), 347-343.

Cohen, D., & Hill, H. (2001). *Learning Policy: When state educational reform works*. New Haven, CT: Yale University Press.

Cohen, D., McLaughlin, M. W., & Talbert, J. (Eds.). (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey Bass.

Cooperative Research Act of 1954, Pub. L. No. 83-531. 83rd Congressional 2d. Session (1954).

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal, 44*, 298-338.

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: MacMillan.

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

Desimone, L., & Porter, A., & Garet, M., Yoon, K., & Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Elmore, R. (1996). Getting in scale with good educational practice. *Harvard Educational Review, 66*(1), 1-26.

Fennema, E., & Carpenter, T., Franke, M., & Levi, L., & Jacobs, V., & Empson, S. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403-434.

Gersen, R. (1984). Follow through revisited: Reflections on the site variability issue. *Educational Evaluation and Policy Analysis, 6*, 411-423.

House, E., Glass, G., McLean, L., & Walker, D. (1978). No simple answer: Critique of the follow-through evaluation. *Harvard Educational Review, 48*(2), 128-160.

Huberman, A. & Miles, M. (1984). *Innovation up close: How school improvement works*. New York: Plenum Press.

Loucks, S., Cox, P., Miles, M., & Huberman, M. (1982). *People, policies, and practice: Examining the chain of school improvement, Vol. 2: Portraits of the changes, the players, and the contexts, A study of the dissemination efforts supporting school improvement*. Andover, MA: Network of Innovation Schools.

McLaughlin, M., & Marsh, D. (1978). Staff development and school change. *Teachers College Record, 80*(1), 69-94.

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

Nunnery, J. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society, 30*(3), 277-295.

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

Rowan, B. (2001). The ecology of school improvement. *Journal of Educational Change, 3*, 283-314.

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, D.C.: National Clearinghouse for Comprehensive School Reform.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement. *Elementary School Journal, 105*, 75-102.

Rowan, B., & Correnti, R. (2006, April). Interventions to improve instruction: How implementation strategies affect instructional change. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Rowan, B., & Miller, R. J. (2007). Organizing strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal, 44*, 252-297.

Stecher, B., Le, V. N., Hamilton, L., Ryan, G., Robyn, A., & Lockwood, J. R. (2006). Using structured classroom vignettes to measure instructional practices in mathematics. *Educational Evaluation and Policy Analysis, 28*(2), 1010-130.

Stringfield, S., & Datnow, A. (1998). Scaling up school restructuring designs in urban schools. *Education and Urban Society, 30*(3), 269-276.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89-122.

Welch, W. W. (1969). Curriculum evaluation. *Review of Educational Research, 39*(4), 428-443.

Whitehurst, G. (2002, June 25). Statement before the Senate Committee on Health, Education, Labor and Pensions. Retrieved July 28, 2008, from http://ies.ed.gov/directors/speeches2002/06_25/2002_06_25.asp.

Nondiscrimination Statement

The University of Pennsylvania values diversity and seeks talented students, faculty, and staff from diverse backgrounds. The University of Pennsylvania does not discriminate on the basis of race, sex, sexual orientation, religion, color, national, or ethnic origin, age, disability, or status as a Vietnam Era Veteran or disabled veteran in the administration of educational policies, programs or activities; admissions policies, scholarships or loan awards; athletic, or University administered programs or employment.

Questions or complaints regarding this policy should be directed to Executive Director, Office of Affirmative Action, 1133 Blockley Hall, Philadelphia, PA 19104-6021 or 215.898.6993 (Voice) or 215.898.7803 (TDD).

Consortium for Policy Research in Education