

**Improving Teachers' Use of Student Data:  
the influence of America's Choice and Success for All**

by Diane Massell, PhD

University of Michigan  
Consortium for Policy Research in Education

Paper prepared for presentation at the 2009 annual meeting of the American Education Research Association, San Diego, California

This paper would not be possible without the valuable contributions of numerous members of the Study of Instructional Improvement research team, particularly Carol A. Barnes, Kristi Holstrom, Danae de los Rios and Charles Vanover. Research reported in this paper was conducted by the Consortium for Policy Research in Education (CPRE). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of CPRE, the subjects studied or the sponsors of this research study.

## **Introduction**

A central tenet in this era of "new accountability" and standards-based reform is that educators, armed with data on student outcomes, will make better, more informed decisions that will improve instruction and student achievement (Fuhrman 1999; Elmore and Rothman 1999). At the same time, a growing body of research was showing that sound assessment practices could actually be used in the classroom to improve student learning (Black and Wiliam 1998), findings which further kindled interest. And yet the research on educators' actual use of student performance data for instructional improvement suggests a mixed picture, at best, with reformers' aspirations for comprehensive and continuous reflection an often-elusive goal in practice (Simmons and Resnick, 1993; Massell, 2001; Firestone, Monfils, Camilli, Schorr, Hicks, and Mayrowetz, 1999; McNeil, 2000; Ingram, Louis and Shroeder, 2004).

Under this new institutional logic, schools have been saturated with student performance data and an intense pressure to respond. It was not always so. Thirty to forty years ago, performance data was typically collected to satisfy bureaucratic rules and regulations or signal accountability for larger shares of state funding, but then quickly shuffled out of public view (David, 1978). Indeed, even publicly displaying any system-level performance data could launch a firestorm of opposition, at it did when the Reagan Administration decided to publish the "Wall Chart" of state-by-state performance on college admissions tests (Boyd, 1987).

Beginning in the late 1980s, states adopted accountability regimes with strong sanctions and rewards designed to motivate educators to attend to school-level performance outcomes. The interest in data as a central lever for improvement has continued unabated since then, with the No Child Left Behind Act of 2002 increasing testing and amplifying accountability calculations and sanctions beyond what most states had required. By 2006, two-thirds of the states had developed interactive data systems (Education Week 2006), and a growing number of states and districts are providing incentives and resources for schools to use formative assessments (e.g., Gallagher and Worth, 2008).

Educators have scrambled to respond to these new pressures, and have increasingly sought to use data to adjust district, school and classroom practice (Massell 2001). But the "feedback control loop" (Sadler 1989) between student performance data and educational decision-making often remains tenuous, even in schools with a strong reputation for data use (Ingram et al. 2004).

A growing body of literature seeks to explain why educators do or do not become active users of data for instructional improvement. But much of that literature focuses either on general changes in practice expected as a result of standards-based reform, or on the effects of one particular design for data use or one type of student performance data (e.g. Ingram, Louis and Shroeder 2004, Supovitz and Klein 2003; Heritage, Kim,

Vendlinski, Herman 2008; Herman & Gribbons, 2001; Plake & Impara, 1997; Shepard, 2001; Stiggins, 2002).

This paper differs in that it compares the structures and strategies to support data use employed by two comprehensive school reform (CSR) interventions, Success for All and America's Choice. In each of these CSR models, strengthening and deepening school staffs' use of data was an integral part of their larger planned change initiative. While the two designs had many overlapping goals, they established different strategies to help educators learn and act upon data. The focus in this paper is specifically on whether and how those strategies influenced the patterns of data use among elementary teachers in the area of literacy. I rely on qualitative investigations in Spring and Fall of 2002 in six schools selected as part of the larger Study of Instructional Improvement (SII) conducted at the University of Michigan.

After reviewing the literature that framed this investigation of data use practices and discussing the methodology in more detail, I proceed with a descriptive overview of the intended data use strategies in the two designs. This is followed by an analysis of the data use practices as they were enacted in the case study schools

Readers should keep in mind that this paper examines the strategies these CSR models developed and put into play between 1998 and 2002, the years when the SII case study schools adopted their designs. Even in this period, the designs were evolving in response to new research, practical challenges of implementation and scale-up, and other factors. Furthermore, these schools varied in their fidelity and quality of implementation and overall staff commitment to the CSR designs. Nevertheless, these cases do contribute to explanatory theory about the reasons for patterns of data use in schools.

## **Literature Review and Methodology**

### **Literature on Data Use**

Research investigating data use in schools suggests multiple locations where effective use can break down. One is simply a human capital issue—many studies have found that teachers and school leaders simply do not have the knowledge and skills they need to proficiently analyze student results (Supovitz and Klein 2003; Shepard 2005; Heritage & Yeagley, 2005; Herman & Gribbons, 2001; Plake & Impara, 1997; Shepard, 2001; Stiggins, 2002.) Theory and research on knowledge utilization, organizational decision-making, and a literature specifically focused on educators' use of student achievement data suggests that the quality and types of professional learning opportunities influences whether educators acquire the skills to productively act upon this kind of outcome data. Studies show that individuals need extended opportunities to situate and learn new information in their own contexts. “Communities of practice” can help teachers make sense of the knowledge, and foster new organizational norms (see Hood 2002; Lave and Wenger, 1991; see also Cohen and Hill 2001).

But improving the use of student achievement data is more than a human capital problem and question of technical skill, and extends to the organization as a whole.

Schools require district environments and school leaders who provide adequate resources and time, designate data managers, create trust to share and act upon data, model use of data, and employ their power and authority to communicate expectations and motivate attention to data, among other things (Young 2006; Wayman and Stringfield 2006; Young and Kim 2007; Coburn and Talbert 2006). Finally, the characteristics of evidence also affect whether and how it will be used (Farley-Ripple 2008). For example, research has pointed to specific characteristics that influence use, such as whether data is timely, that is, received in time to influence intended decisions (Suppovitz and Klein 2003; Massell 2001); whether it is part of an extremely high volume of data that practitioners must analyze and make sense of (O’Day, 2002); or whether it is appropriately specified to address the particular problems faced by practitioners with different role functions, such as district curriculum administrators, school principals, or teachers (Bond 1995; Bond, Herman and Arter 1993).

Yet little of this literature considers how data use may be affected by the nature of the instructional guidance in play at the school site, the way particular responsibilities for analysis and use are distributed within schools (particularly, in how modifications to instruction are made), and specific leadership routines used to monitor instruction. The two comprehensive school reform models studied here have very different practices in these regards, although broadly similar goals for data use by teachers. Thus they provide us with an opportunity to explore whether and how such factors interact to influence data use.

### **Study Methods**

This research is based on interviews and field visits to a total of six elementary schools in four districts situated in three states—Florida, New York, New Jersey. The research team conducted semi-structured interviews with principals and assistant principals, and school-based coaches or facilitators providing support for the design. In addition, two first and two fourth grade teachers in each school were identified. To the extent possible<sup>1</sup>, each respondent was interviewed twice, once in the Spring of 2002 and again in the Fall of the following academic year, about the influence of the intended designs on their practice. A portion of each protocol explored whether and how respondents used student data in their practices.

We used standard qualitative procedures to triangulate and verify data and to look for repeated patterns and themes in the oral discourse and collected documents (Miles and Huberman 1994). We use pseudonyms to protect the confidentiality of respondents and the schools and districts where they resided. The work here also draws on documents and analyses conducted by other colleagues at the University of Michigan investigating the comprehensive school reform designs as part of the Study of Instructional Improvement.

---

<sup>1</sup> Some teachers “looped” up to the next grade with their students, so that in the second interview they were teaching in another grade level. A small number of teachers dropped out of the study in Fall 2002, and were replaced with other staff. Finally, one school dropped out of the study entirely. So respondents in one school—Gladrock Elementary, in Sunnyside, New Jersey—were interviewed only once, in Fall 2002.

In this paper, the term “data” refers in general to any systematically collected evidence of student learning. CSR designers and sometimes districts asked school staffs to collect and use a variety of formative or summative assessments in addition to mandated state and district tests. Summative measures ranged from end-of-unit tests to external state or district assessments, and were administered relatively infrequently. We follow Black and Wiliam (1998) in using the phrase formative assessment to encompass all activities undertaken by teachers and/or by students that were intended to modify teaching and learning. Here formative assessments refer both to formal, diagnostic measures, such as the Scholastic Reading Inventory or Running Records, as well as more informal measures, such as teacher questioning techniques, the application of scoring rubrics to written work, editing strategies, and the like.

### **Data and Designs**

Both America’s Choice and Success for All shared similar philosophies about the importance of using data to differentiate instructional practices to target the particular needs of students. Each CSR designed formative classroom assessment activities, and stressed the importance of involving students in evaluation routines. Both recognized and sought to support the use of data in teacher meetings they established to develop collective, sustained learning opportunities, or on-site “communities of practice”. Yet despite these similarities, the ways they structured and carried out these common goals were quite different. The following provides an overview of the data guidance, evidence and support structures in the AC and SFA designs.

#### **America’s Choice**

The America’s Choice design for literacy provides teachers with instructional guidance that among other things includes a set of instructional routines for writers’ and readers’ workshops, core writing assignments, a curriculum guide and scoring rubrics based on AC performance standards. While the guidance is substantial, the sequencing and pacing is not highly prescribed, and AC expects teachers and school leaders to become proficient in using data to focus instruction. They provide specific direction to schools on what evidence to use and how to use it, expecting all school staff to use data from multiple sources, including state and district standardized tests, portfolios, classroom reading assessments that produce detailed scores and analysis, and an AC fourth grade Reference Exam (Glazer, internal AC memo).

In the classroom, teachers are to establish a range of formative assessment practices, including student-teacher conferencing, reading assessments, and student self-assessments and peer-assessments. Each of these components could be quite intensive. For example, teachers are expected to conference one-on-one with students at least weekly or biweekly. In these conferences, teachers ask students to discuss the book s/he has been reading independently, read aloud, or conduct a running reading record. Teachers help students identify goals and tasks to address before their next meeting, and maintain learning logs on these conference interactions. Teachers are also asked to work with students to elaborate the AC rubrics in their own words. This process intends to develop student understanding of the criteria for high quality work and improve their

capacity to self-assess and provide feedback to their peers, which occurs frequently in the course of instruction. In addition, students are to conference with other students in small groups and offer feedback, using “accountable talk” justifying their comments with clear references to the text or agreed-upon criteria.

Teachers are expected to use data to plan their instruction. For example, they are to base their mini-lessons on evidence of student learning problems, or select another recommended strategies for providing additional support to small groups or individuals. In the AC literacy design, students were grouped heterogeneously in literacy classrooms (in contrast to SFA), but in these and other ways teachers are expected to target instruction on the specific needs of individual students.

In addition to helping teachers use and make decisions based upon evidence of student performance, the school-based AC design coach is responsible for facilitating conversations about data in regularly held teacher meetings. AC provides guidance on how to conduct these conversations to stimulate reflection and connections to practice. For example, teachers are supposed to bring samples of their own students’ classroom work to collectively discuss whether and how this work addressed rubrics and standards. In addition, principals are asked to review posted student writing and other work samples in the context of their walk-throughs to monitor instructional implementation; student work was also the focus of leadership team meetings (Barnes, Vanover and Kim 2009)

### **Success for All**

One of the unique, hallmark features of the SFA literacy design is that it calls on schools to place students into multi-age but approximately homogeneous reading groups. Students leave their homerooms to participate in 90 minutes of highly specified reading instruction with these groups every day, using appropriately leveled curriculum materials. School-based SFA facilitators administer diagnostic assessments every eight weeks, and in consultation with teachers use this and other data to make changes in reading group placements (the Joplin method, see Slavin 1987) or identify students for additional tutoring services. In addition to these responsibilities, facilitators are also expected to produce grade- and school-level reports on SFA related tests and state or local assessments to share with other staff.

Although students are grouped homogeneously, SFA recognizes that individual students within each level would still present varying needs and abilities. Thus they embed a range of formative assessments in their instructional materials and routines to help teachers refine and focus instruction, including weekly assessments of story comprehension, vocabulary, and reading fluency and accuracy. In SFA “Treasure Hunts,” students are to respond in writing to questions about their reading texts, and teachers are expected to engage students in active question and answer activities around them using questioning techniques based on Bloom’s taxonomy. Like AC, SFA also incorporates peer- and student self-assessment activities. For example, students are expected to challenge each others’ answers on the “Treasure Hunt” activities, provide feedback to other students on their writing, and conduct “two-minute edits” of their own writing assignments.

Like AC, SFA expects its facilitators to host regular teacher meetings, here called component team meetings, to collectively discuss and learn the design. In 2000-01 SFA introduced new tools to encourage more interaction and dialogue between facilitators and teachers about the eight-week and other assessment data (Peurach, internal communication). In SFA schools, principals' primary responsibilities are to monitor and hold teachers accountable for implementing SFA instructional routines, and create a vision for school improvement based on data (Barnes, Vanover and Kim 2009).



The above descriptions show that while AC and SFA goals were similar, their strategies for data use were situated in very different instructional and organizational structures. The next section of this paper analyzes how these practices played out in these six case study schools, and whether these differences mattered in the extent to which and how teachers used data.

### **Enactment of Data Use in the AC and SFA Schools**

#### **America's Choice**

The three America's Choice case study schools adopted the AC model and components of the literacy design between 1998 and 2000, and were thus in their second to fourth year of implementation when we visited in Spring 2002. Bonds Elementary was located in Sunnyside District, New Jersey, one of the districts included under the state's long-running school finance case, then known as *Abbott v. Burke* (Erlichson and Goertz, 2001). Redmond Elementary is in Coverdale County, Florida, and Westwood Elementary is located in Adderly Community School district, then one of over thirty school districts in New York City (see Appendix A for a summary of the context and way data was used in each school).

Among the teachers we sampled, the design's formative classroom assessments assumed a strikingly dominant place in their discussions about classroom work. These responses were elicited after direct questions about assessment practices as well as questions about what they had changed in their instruction with the adoption of AC. As noted above, AC incorporated a range of assessment strategies intended to help teachers make immediate decisions in the context of their day-to-day instruction. And a majority of interviewed teachers did in fact use these methods, and spoke in often convincing detail about conferencing one-on-one with students, administering running records or other diagnostics of reading fluency and comprehension, having students question their partners and engage in self-assessments, elaborating and applying rubrics and standards to examine products of student work, and more.

Of course, several of these methods were not new for some teachers, and for them AC was more of a reconfirmation or expansion on their extant practices. And a number of teachers complained about administering so many assessments, or felt that assessments consumed too much time, or were quite difficult to do well (see below). But despite such

**Draft Paper: Please do not cite or distribute without the permission of the author**

challenges, teachers nevertheless incorporated these practices broadly into their instruction. One or more teachers in each school reported that the design had brought about a dramatic change in their thinking about assessments and evidence of students' learning. For example, Bako, a fifth grade teacher from Redmond Elementary in Coverdale, said:

[before AC] it was—let's read page 35. I'm gonna throw some questions on the board, you gonna turn them in for a grade, and that's it, you know? I didn't have any time where I actually would conference with the child. I didn't give them any time to read by themselves. For writing, I would...never let them check each other's paper, with writing and making corrections and stuff like that, I mean, that's all a blur to me. I don't want to think about life before. It's horrible. (Bako, Redmond, F02)

Teachers valued the insights into their students that such assessments could provide. For example, several pointed out that conferencing offered a more refined and accurate portrait of students' abilities.

I find that in the whole group setting... a lot of times I just get re-workings of the first answer... a lot of copycat answers, so I don't feel that that always gets you to the core of whether a student is comprehending or not. Now, if I can do a one-on-one with a student, then I can get a more accurate picture (Acta, Westwood, F02).

Similarly, another commented, you could have a reader in your class that could be reading, decoding, every single word, but that person may not understand a thing about what they're reading. And that's why it's very important to have a conference...I can see whether or not that book is matched appropriately to that child (Inge, Westwood, F02).

Perhaps most significantly, teachers reported using data to modify and focus their instructional routines. They used the formative assessments, along with more summative measures such as the New Standards Reference Exams and external test results, to differentiate instruction for small groups or individual students, and to develop mini-lessons to assist the whole class on assessed areas of weakness. Again, teachers noted that one-on-one conferencing led them to shift or refocus their lessons. "Now I'm more specific with what I know they're capable of" (File, Redmond, S02). Another said,

I'm going to document that, and use that, that little piece, as a mini-lesson for the next day. That's how I plan my lessons. I base it on what the students need. Not what I think... they need. I don't theoretically, hypothesize what I think the kids are going to need for writing. I look at what they can do and what they can't do. And I build on their strengths and weaknesses (Inge, Westwood, F02).

In reading, teachers organized students into groups based on assessments and identified appropriately leveled reading materials for the different groups. "I was able to assess my students and organize them into group... not approaching it as a whole class picture but now really trying to key-in on individual students and realizing that it's really not

necessary to go over the same thing with some students when you realize they got it (Acta, Westwood, F02). Observations of this and other sampled teachers in classrooms confirmed that teachers' interactions with different reading groups varied based on their skill.

Teachers in these AC schools also engaged students more centrally in assessing their own and their peers' work. One teacher, who had never used these practices before, commented, "In the past it was just, 'read the books and answer these questions' ... Now it's listening to the kids discuss it with each other, and trying to pull out conversations with them in a group talk... Just having the kids... conferencing together is amazing" (Tait, Bonds, S02). Another, talking about students applying rubrics, said, "They don't know why they're getting the grades they do unless they're involved. I think it's really important to involve children in that part... And, that has changed for me because of AC" (Inge, Westwood, S02.) In their own words, these teachers affirmed research findings indicating that student self-assessment could contribute to improved learning (Black and Wiliam 1998).

The decisions that teachers made based on the data, although scaffolded by AC, nevertheless called for them to do much "inventing" to determine how to adjust their instruction, or, again in Sadler's words, to close the "feedback control loop" in the cycle of improvement (1989). Teachers learned how to invent in the on-going, collaborative meetings between teachers and school leaders required by the design. In AC as in SFA, such meetings were used to review summative assessments, such as state and district tests or New Standards Reference Exam, and aggregated reports of formative classroom assessments.

However, and quite distinctively to this design, staff were also asked to bring authentic products of student work, such as writing samples, and collectively evaluate and discuss them using the rubrics and performance standards. AC design and literacy coaches were diligent about doing so. In Westwood, for example, teachers were not allowed into the teacher meeting unless they brought their standards book and student writing.

In addition to these collaborative work groups, principals looked at this evidence of student learning and scrutinized posted student writing samples as part of AC's administrative routines to monitor teachers' practice. While principals proudly noted their improvements on state test results—and all of the sampled AC schools experienced gains on their state's literacy assessments since entering the design (see Table 1)—they just as often commented about the improvements they perceived in samples of students' writing and other performance. Redmond's principal, for example, said, "[We] see our children being able to express themselves more. I see it in the writing. I see it in just an overall better education that they're receiving through the use of this program [AC] (Principal Lyon, Redmond, F02). Other investigations for this study suggest that these leaders' monitoring routines motivated teachers to focus on such authentic evidence of student learning (Barnes, Vanover and Kim, 2009).

**Table 1. AC Schools' Performance on State Reading and/or Writing Exams (in percentages)**

<b>Bonds Elementary, Sunnyside School District</b>		
New Jersey Elementary School Proficiency Assessment	English-Language Arts Total Proficient (Proficient and Advanced Proficient)	English-Language Arts Advanced Proficient Only
1998-99	24.7	0
1999-00	25.0	2.0
2000-01	49.4	2.4
2001-02	70.6	4.9
2002-03*	71.4	2.2
<b>Westwood Elementary, Adderly Community School District, New York City</b>		
New York State exam	Reading Total Levels 3 and 4	Reading Level 4 Only
1998-99	23.5	n/a
1999-00	27	n/a
2000-01	26.7	n/a
2001-02	26	n/a
2002-03	30.7	n/a
<b>Redmond Elementary, Coverdale County</b>		
Florida Comprehensive Assessment Test	Reading Total Proficient	Writing Total Proficient
1997-98	18	92
1998-99	29	49
1999-00	37	86
2000-01	44	97
2001-02	45	85
2002-03	50	94

\*In 2002, ESPA changed to NJ-ASK.

AC design and literacy coaches also believed that the guidance AC provided for teacher meetings had enabled instructional conversations based on evidence of student learning. The Bonds Elementary coach argued that although they had always cared about student work before, AC standards, rubrics and use of student writing gave them a

framework for systematically looking at and reflecting upon student work as evidence of educational practice, and had significantly changed the content and focus of their teacher meetings. The design coach in Redmond described how teachers brought samples of student writing at different performance levels to foster reflective conversations about why certain practices are not working, and discuss alternative strategies for improving these outcomes.

These instructional leaders noticed changes in teachers' attitudes as a result of such activities. Westwood's design coach noted that these activities had prompted one very "vociferous" teacher—the kind "you don't want at the meetings"—to reflect on her instruction, and led other teachers to think about what they needed to look for in their students' work. While not all teachers found these meetings to be sufficient to provide them with the answers they needed to change their instruction, they were motivated to pursue new ideas from their colleagues. For instance, while one Redmond teacher said these meetings tended to focus on weaker, less experienced staff, she turned to a colleague for answers—an indication that she had assumed responsibility for finding answers, and accepted the importance of doing so.

These illustrative comments suggest that for these teachers and many school leaders we interviewed, AC's data use structures and strategies contributed to changes in the professional culture in these buildings, in what is "taken for granted" as best practice. Research and theory suggest that when teachers have a deep understanding of the pedagogical principles of a reform, they are better able to adapt it to new demands and changing contexts—in other words, to sustain and scale up the fundamental ideas (McLaughlin and Mitra 2001; Coburn and Meyer 1998; Coburn 2003). While some respondents in each school predicted that they would sustain AC practices even if the program were eliminated, actual evidence of sustainability came in Westwood, a school located in Adderly, New York. In the summer of 2002, this district changed course and mandated that its low-performing schools introduce a new, relatively more scripted literacy program called Voyager. Voyager came with a prescribed set of tests. A computer program analyzed the results and identified three reading groups. Classroom teachers were to follow a specified practices for each group. Acta, a first grade teacher, lamented that this design had robbed her of the ability to determine the skills of her students and follow what she believed they needed. She said she refused to give up some of the AC formative assessment practices, such as running records, and sustained other elements of the AC literacy design. She also requested permission to depart from Voyager to serve her gifted students the way she had been able to do under AC.

Although the evidence on these AC schools' use of data is largely positive, teachers and school leaders identified significant challenges in using the assessment practices associated with the design. Perhaps the most commonly expressed issue pertained to the time and paperwork pressures associated with all the formative assessments to be administered in the classroom. Said one first grade teacher from Bonds, when asked if she had changed the way she assessed students with AC:

Yes, it's changed. . . . The positive things are the rubric, the writing rubric is wonderful. I love that, it's really good. But the paperwork, keeping records of the assessment, and constantly assessing, it's too much. It's burying us. I mean this is going to be the death of teaching, the death of inner-city teaching anyway.. it's too much, we spend more time..showing what we do than doing what we show. . . . every second of the day we're assessing it seems. . . (Alma, Bonds, S02)

While Alma is referring to more than the AC record keeping, it alone was quite demanding, and when placed in the context of other mandates was often overwhelming. However, as the earlier discussion makes clear, many AC teachers nevertheless embraced the importance of doing the formative assessments, even while complaining about the burdens it imposed. When asked about the most challenging part of implementing AC, a first grade teacher exclaimed,

All of the paperwork! Every time I turn around they want something else! There's a lot of work involved...But in the long run, yeah, it works, because that way [my principal] knows not only what I know, but she knows that I'm following through with what I'm learning in the assessments... (File, Redmond, S02)

Finally, carrying through on the formative assessments that AC planned requires not only the kinds of bureaucratic organizational skills that teachers identified, but also classroom management skills and other resources. Teachers were often beset with the challenges of collecting data and tailoring instruction based on student needs, in the context of large or even moderately sized classrooms. Student discipline issues also contributed to teachers' reported difficulty in conducting the kind of one-on-one conferences or small group instruction that AC envisioned.

### **Success for All**

The three case study schools adopted the Success for All model and the literacy design between 1998 and 1999, and were thus in their third or fourth year of implementation when we first visited in Spring 2002. Gladrock Elementary, like the AC school Bonds in our sample, was located in Sunnyside District, New Jersey. Doris Elementary was situated in Adderly, just as Westwood was. The third SFA school, Nightingale, was located in another New Jersey district also covered by the *Abbott v. Burke* court case, called Freightville (see Appendix A for more complete school descriptions).

The pattern of data use in this sample of SFA schools was distinctively different from what we found in America's Choice. Although SFA literacy teachers similarly administered the range of formative classroom assessments prescribed by SFA, and school leaders focused on helping teachers use this data to tailor and modify their instruction, teachers varied greatly in the extent to which they reported using this data in their instruction. They were hesitant to change the instructional routines they had come to know. Student achievement data did not emerge as a major theme in their discourse on how SFA changed their practices. This was the case even though these schools were in

their fourth or fifth year of SFA, and two were considered exemplary SFA schools with leaders who were proficient and motivated to use data.

Rather in these sites, the school-based SFA facilitators were the most prominent and consistent users of student achievement data. As intended by the design, facilitators analyzed the results of a commercial reading assessment every eight weeks, along with summary reports of the SFA classroom assessments as well as state and local test results, to determine student placement into leveled reading groups for the 90 minutes of SFA literacy instruction. In practice, making such decisions based on this composite data was not a “paint-by-numbers” task—it was technically demanding, uncertain and logically complex. In the Wings component for grades 2-5, for example, although SFA provided a recommended list, schools had to select which literacy assessment to use to determine student placements. Facilitators from all three schools discussed the challenges they faced in sorting through issues of these assessments’ reliability and validity, questioning whether the various commercial assessments appropriately identified students’ reading abilities, and were well-aligned to the pedagogical goals of SFA or other tests used in the school. Their choices could have significant consequences for how the program functioned. In Nightingale, facilitator Fina noted that the original assessment used for grouping ended up placing students in levels that were too high for them and which had long-lasting repercussions for students and teachers. Some facilitators modified the assessments to better align them to other goals and tests. In making placement decisions, facilitators also had to weigh other issues, such as their teachers’ skills and weaknesses, or potential class sizes.

SFA Foundation staff recognized this complexity, and believed that facilitators’ able analysis and use of data to make student placements was extremely critical to a well-functioning and successful program. Intriguingly, one external SFA liaison to Gladrock Elementary commented that this was even more important than the quality of teachers in a school:

I have got to say I think a lot of elements of Success for All are working well in this school. I think the number one best piece of this puzzle is the facilitator... The fact that it looks like all but 9 or 10 1<sup>st</sup> graders are going to be out of the Reading Roots program (and) well into a 2<sup>nd</sup> grade reader by the end of this school year is amazing. And that does not always happen when you don’t have a strong facilitator. You can have the best teachers in the world, but when you have a strong facilitator that’s keeping everybody chugging along and looking at the data and really focusing in on it and using what she knows about the data, about the teachers, and about the students to make some placement decisions, you see that kind of fantastic growth. (Moor, SFA, S02)

SFA guided facilitators to consult with teachers about their placement decisions. In our schools, some teachers mentioned their participation in this activity. In Doris Elementary, a fourth grade teacher commented, “We more or less tell [the facilitator] where we think the child is and then we look at the score to see if the actual work in the class is matching the score” (Camp, Doris, F02). But several other teachers from Doris

and Gladrock indicated that they had no engagement, and no understanding, of the quarterly assessment data. For example, another Doris teacher acknowledged not even knowing what the test items were, or how the grouping and tutoring decisions were made. Some teachers viewed the quarterly assessment data and grouping decisions as out of their province, as something done for them rather than with them. A Gladrock teacher noted that she had never received any support for understanding or using these SFA assessments, “No. We were just given..them at the end because they [the facilitators] test the children” (Fumo, Gladrock, F02).

Whether they were involved or not, many interviewed staff expressed confidence that using data to create relatively homogeneous groups of readers had helped the school as a whole better differentiate instruction to meet students’ learning needs. In fact, all of the schools showed growth on their state literacy assessments since adopting SFA (see Table 2), and increasing numbers of students were reading on grade level. These results could be motivating to teachers. First grade teacher Goff commented, “I started from blank, and at the end of this year they will be reading. That keeps you going” (Goff, Nightingale, F02). Instructional leaders in both Doris and Nightingale noted that leveled reading grouping had “saved the top” by providing their strongest students with more demanding literature and reading assignments, and maintaining these students’ interest and motivation—an essential aspect of a productive learning process (Cohen and Ball 1999). Facilitator Zann in Doris observed that the collection of student performance data every eight weeks had enabled them to keep better track of students and to provide more timely attention and tutoring resources, just as SFA had planned. Teachers also valued when the 90-minute SFA period functioned as intended, e.g. when they were assigned with smaller classes of students at similar reading levels—an outcome that simplified their instructional work. Commented one fourth grade teacher: “Basically..we have so much less in the classroom. You aren’t dealing with a group of 20 students. You are dealing with a concentrated group of 14 students that are all at the same level. I mean that’s a huge help” (Dunna, Gladrock, F02). Echoing this sentiment, other teachers longed to replicate homogeneous grouping across other subject areas.

**Table 2. SFA Schools’ Performance on State Reading and/or Writing Exams (in percentages)**

<b>Nightingale Elementary, Freightville School District, NJ</b>		
<b>New Jersey Elementary School Proficiency Assessment</b>	<b>English-Language Arts Total Proficient (Proficient and Advanced Proficient)</b>	<b>English-Language Arts (Advanced Proficient Only)</b>
1998-99	5.3	0
1999-00	14.9	0
2000-01	33.9	1.7
2001-02	43.2	0
2002-03*	56.1	0

<b>Gladrock Elementary, Sunnyside School District, NJ</b>		
New Jersey Elementary School Proficiency Assessment	English-Language Arts Total Proficient (Proficient and Advanced Proficient)	English-Language Arts (Advanced Proficient Only)
1998-99	8.7	0
1999-00	38.0	0
2000-01	59.6	1.9
2001-02	69.5	0
2002-03*	50.9	0
<b>Doris Elementary, Adderly Community School District, New York City, NY</b>		
New York State	Reading Levels 3 and 4 (High)	Reading Level 4 Only
1998-99	17.6	n/a
1999-00	22.6	n/a
2000-01	19.9	n/a
2001-02	20.5	n/a
2002-03	30.6	n/a

\*In 2002, ESPA changed to NJ-ASK.

But SFA also wanted and planned for teachers to use data to modify their instruction to address particular student needs, recognizing that within each leveled reading group students would still have varying skills and needs. And, once schools confidently implemented instructional practices with fidelity and pedagogical principles were deeply understood, SFA encouraged and supported significant, mindful elaborations of instructional routines (Peurach, internal communications). Thus facilitators also used component team meetings or other collective work time to discuss the implications of the quarterly and classroom assessment data to help teachers reflect on and modify their instructional practice. Facilitators also provided summary data charts for teachers' reading groups, extra skills packets for teachers to use, and other tools to facilitate instructional change based on data. These conversations and tools focused on improving implementation of SFA instruction, or differentiating and varying instruction to assist particular individuals. For example, the Doris facilitator looked at the eight-week assessments with teachers to “see where the weaknesses are...comprehension, word out loud. So I would do a component meeting on those weaknesses. And a lot of time the teachers would talk about what worked for them in meaningful sentences, for example” (Zann, Doris S02). A fourth grade teacher from Gladrock described using the facilitator’s data charts:

she writes down the scores and then ... we can line them up, we can see if all of our students were weak in let's say meaningful sentences, we know we have to concentrate on how to write a sentence with the vocabulary word and what it means... We look at it per student but we also look at it as a whole group because if the whole group is missing something, obviously it's the teacher who didn't teach it. So we use the tests and every week we go over the scores of the children and we go over the strengths of the group and then the strengths of the individual and the weakness as well (Dunna, Gladrock F02).

Teacher Goff from Nightingale noted that they looked at the data to see what “we might be overlooking in our lessons that we could concentrate a little bit more on. Put a little more time into” (Goff, Nightingale S02).

School leaders in Nightingale Elementary were especially devoted to the importance of using data to identify problems and seek solutions. In fact, both the SFA literacy facilitator, Fina, and Espy, the new school principal appointed in Fall 2002, viewed the use of data as *the* critical lever for school improvement, and both identified developing teachers' use and understanding of data a major school priority. Fina attributed her interest and skill in using data to SFA. Before they adopted the model, she said,

I never knew anything about data. I was actually one of those people, I was afraid to share data because it never looked really good and I was like, “I don't want to share this!” But I've learned. I've learned how to look at it, and I've learned how to, you know, go through and how to say, “Well, this is why this isn't working.” (Fina, Nightingale, F02)

Fina and Espy's commitment to data was reflected in numerous actions. They continued to bring SFA in for more data training in the fifth year of implementation—a time when other schools had begun to scale back their interactions with the foundation. Whereas all schools, including Nightingale, lost some or all component team meeting time for teacher discussions about SFA in 2002, Fina and the principal arranged coverage and set aside time for data conversations. In the fall, she introduced new SFA procedures that involved teachers more directly in the analysis of assessment data, for example, by having teachers analyze and highlight their own data on computers. Such engagement began to improve teacher ownership and understanding “last year we looked at our data but we didn't understand it as much and we didn't go through and highlight it...it kind of helps you realize which students where and what you need to do better” (Goff, Nightingale, F02). One fourth grade teacher described how this led them to reflect on students' learning and their own practices:

We look at our class and then like the last time we highlighted [the eight week assessment in] three different colors. And then if you had a certain color, [we asked] is it your technique? Then looked at the student. If it is a mainstream student, they may be lower in writing because of this and that. [But] it's not like don't look at yourself as the reason....For instance, now I moved from a third

grade book to a fourth grade book and some of my kids are having more difficulty, and we discuss why. Well, because in the fourth grade book, the questions are much more higher order thinking. And then we sort of share all that with each other” (Web, Nightingale, F02).

Armed with such information, and including state and district assessment results, Fina organized or encouraged numerous types of instructional modifications and mindful elaborations to the SFA literacy program. For a group of special needs students not making progress on the state ESPA assessment, for example, she worked with SFA to introduce a letter-recognition and phonics program and make it SFA-like (e.g. a 5-day schedule, storybooks to incorporate reading and comprehension) “so that if kids ever do go into the regular [SFA] reading groups...it’s more consistent for them.” (Fina F02). The facilitator also temporally resequenced skills objectives to emphasize areas where students performed poorly on the state assessments.

Most importantly, Fina encouraged teachers to stretch the SFA pacing or SFA instructional cycle beyond what strict fidelity to the program would have required. When one teacher found students not doing well on the Treasure Hunt classroom assessments, for example, Fina came into the classroom and modeled taking significantly more time for review than allowed in the usual SFA pacing schedule. She directed two of the Nightingale teachers we interviewed to add another day to the usual five-day weekly cycle, because student results on formative assessments indicated that they were not grasping the content.

Such a change surprised teachers, who were reluctant to alter the SFA pacing and instructional routines. For example, although Fina modeled and encouraged major instructional changes for Web, the latter continued to say that under SFA she was not able to “make an assignment and pull out that small group, like that lower group...and work a little with them one-on-one.” Another Nightingale teacher, Mr. Gray, perceived that although they discussed identifying students different learning styles in component team meetings, he could not really alter his teaching much to accommodate such differences because of the structured instructional design. Similarly, sampled teachers from Doris and Gladrock frequently commented about the pressure they felt to strictly adhere to the instructional timetable in the design, pressures which they felt prevented them from adjusting the routines to address student needs. Said one Gladrock teacher, “With SFA, we don’t always get time to check the “Treasure Hunts” that they’re working on. You see? So it’s kind of like, ‘Okay, rush, rush. Day one, you do this. Day two, you do this here.’ And it’s not often I get back and see if they’re really understanding what they’re reading. Whereas the old-fashioned way, they’d come the next day in their groups and we’d go over it together. SFA, you go day one, day two, and you just move on.” (Fumo, Gladrock, S02). A teacher in Nightingale perceived the message in SFA mathematics to be, “If the kids don’t get it, oh well, move on...you have to be on this lesson, on this day, and you have to test on a certain day, whether they get it or not” (Web, Nightingale, F02).

As this suggests, the early SFA message about fidelity was difficult for staff to

“unlearn.” The message to adhere to pacing and other routines was reinforced by external SFA site visitors who oversee implementation and make recommendations. As one of the Doris facilitators, Lane, described, “the teachers get all in a panic when they[external SFA site visitors] come, and the teachers feel “they’re in and out of my room. They don’t know what I do most of the 90 minutes. Maybe they caught me at a bad time. Maybe I wasn’t on the right schedule.” And that’s another thing as far as scheduling and consistency. Everyone tries. They know when they come walking in a room, they know at this point you should be at this, and this point this. So they really know where you should be and if you’re not” (Lane, Doris, S02).

Similarly, principals monitored classrooms with an SFA checklist that led them to look for surface features of teachers’ implementation of instructional routines, and who in general were not expected to acquire a deep understanding of the SFA instruction. This reinforced the idea that program pacing and instructional routines should be strictly maintained (Barnes, Vanover and Kim 2009). So, for example, Nightingale’s Fina noted that the principal would talk to a teacher who had not used SFA’s Think-Pair-Share peer-related routine. But, that “is just a tiny thing of SFA where I talk to my partner before I answer a question. But, that’s not getting us to where we need to get. That’s just a little thing. We need to address it, but that shouldn’t be the most important thing you see wrong with somebody’s lesson.” Without the principal’s deeper understanding of the pedagogical principles of SFA, even Fina—a figure well-supported by her principal and among teachers—felt that some teachers simply did not listen to or feel comfortable with her messages to alter and modify their practices because she does not have the formal authority to evaluate them.

Thus, particular aspects of the SFA instructional routines and leadership practices, as well as its strategy of emphasizing fidelity before scaling up to mindful elaborations, appeared to unintentionally dampen teachers’ motivation to use the formative classroom assessments or other data to modify instructional routines within the 90 minutes of SFA literacy instruction. This may help explain the contradictory finding that while the homogeneous grouping in SFA developed staff confidence that as a school they were better differentiating reading instruction and meeting students’ needs, some teachers felt disconnected from deeply knowing the learning needs of their own individual students, or that it was especially challenging to identify in the eight week time frame. “It’s kind of a challenge when you have a whole new group of students ... and then you have to assess them quickly and see what level they’re at and change it for each one of those... that’s just a challenge to assess a whole new group every eight weeks and then find out their strengths and weaknesses quickly and help them with that” (Dunna, Gladrock F02). Another parallel reiterates this point. Like Westwood, the AC school described above, Doris Elementary was located in Adderly, New York and was similarly mandated to begin phasing in the new Voyager program in fall 2002. Recall that a first grade Westwood teacher perceived that the assessment strategy used in Voyager did not provide her with the same level of knowledge of her students’ learning needs than AC did, and insisted on maintaining some of the practices she had acquired through AC. By contrast, Datz, a first grade Doris teacher, believed that Voyager strategies *improved* her knowledge of her students’ abilities. Whereas she admitted knowing very little about the

SFA eight-week assessments or grouping decisions, she felt more confident that she understood students' learning needs because Voyager allowed her to administer the assessments that identified reading levels, and these small groups were located within her own homeroom; furthermore, she would maintain these same students throughout the year.

## **Conclusion**

As the foregoing discussions illustrate, teachers and school leaders in both America's Choice and Success for All attributed an increased attention to student learning and performance data to their participation in the models. In most schools, we saw a growing commitment to the value of using evidence to generate improvements in student performance. They felt validated by growth on state assessments and other internal evidence of success, such as more students reading at grade level in SFA schools, and a higher caliber of student writing in AC schools.

Although not discussed in the analysis above, contextual factors, such as district and school leaders' commitment to the CSR design and leaders' skillful management of data practices, certainly influenced the extent of teachers' use of this evidence. In schools like Doris, for example, frequent changes in principal leadership and lack of steady district commitment to any CSR design or school improvement strategy undoubtedly had a negative impact on teachers' willingness to engage with the SFA facilitators. Their efforts to improve teachers' understanding of data and use of assessments were often ignored.

But the designs' planned guidance, structures and strategies for using data, as well as the nature of their literacy instruction, seemed to matter more in the overall patterns of use that emerged across these six schools—*regardless of the quality of implementation in any specific site*. Specifically, these patterns suggest that even though AC teachers faced obstacles in managing and finding time for classroom assessments, they were more likely to be engaged in using them in their instruction. They also perceived themselves to be more centrally responsible for, and able to, plan and differentiate instruction based on that evidence. These teachers felt keenly aware of and connected to their students learning problems and needs. While some of the SFA teachers were also attentive to data and used it to adjust instruction, this was a more variable phenomenon. Rather in SFA schools the literacy facilitator stood out as the prime mover and more exclusive user of both the summative and formative data.

One key difference was the basic organizational structure and role distribution for data use (Young 2006) employed by the two designs. As originally conceived by SFA and first used in these schools, the quarterly placement task was framed in a way that distanced teachers from one of the most significant ways that SFA schools differentiated instruction—the assignment of students into level reading groups. Although facilitators were supposed to consult with teachers and use their teachers' classroom assessment summaries in the placement decision, in practice the extent to which individual teachers were involved ranged considerably among our sample. Thus, this decision was often out

of the proximal zone of teachers' direct duties, leaving many teachers with little to no understanding of this data. In fall 2002, the Nightingale facilitator introduced a more active role for teachers in analyzing this evidence, a process that seemed to be bridging this gap and developing teacher knowledge, ownership and commitment to the data.

A second key difference between the designs was in the nature of their literacy instruction, and in their expectations about the scope of teachers' discretion to modify or target that instruction based on evidence. SFA encouraged fidelity to its more prescribed and specific literacy practices, especially when schools first adopted the model or exhibited signs of weak implementation over time. In these circumstances teachers were expected use formative and summative data to intensify focus on individual or class weaknesses, but maintain overall routines. However, once SFA instruction was learned, broadly implemented, and deeply understood, they encouraged more extensive elaborations of practice. The Nightingale facilitator recognized this and directed more radical adjustments. But our evidence shows that teachers were reluctant to undertake significant change, even when data suggested that their students were not making progress. Their hesitance is explained partly by the specified nature of the SFA instruction, which emphasized rapid pacing and sequencing. A number of teachers perceived that the instructional design left them with little room to invent or time to target student needs. Some teachers also felt that within the eight-week span it was difficult to gain firm knowledge of students' competencies and challenges in order to make such determinations. In AC, by contrast, teachers maintained the same children for one or even two years if looping occurred. AC built in substantial professional discretion for teachers from the beginning, and created intentional space for teachers to make decisions based on the range of evidence. As others have noted, SFA's approach reflected a more centralized theory of change, while AC's theory of change emphasized more extensive professional control with professional support (Rowan, Camburn and Barnes 2004). Their strategies and expectations for teachers to use data inside their literacy instruction reflect these different theories.

Finally, the patterns of data use were also influenced by the monitoring and teacher support practices scaffolded by the two designs. AC principals monitored implementation by examining artifacts of student writing and other authentic evidence of student learning, whereas SFA principals visited classrooms in short intervals and focused on the appearance of SFA instructional routines and pacing. The AC strategy motivated teachers to emphasize performance outcomes by backing this focus with the authority and power of the top administrator. In addition, AC teacher meetings were dominated by a focus on student work products (For further discussion on this point, see Barnes, Vanover and Kim 2009.)

This comparison between the America's Choice and Success for All schools suggests that prior research and theory on data use has not paid sufficient attention to three critical elements:

- (1) the nature of the instructional design in play in schools. Instructional programs build in more or less flexibility and time for teachers to invent or elaborate. This

- aspect of instruction can influence whether teachers feel able to use the evidence they have to target students' needs.
- (2) the way that specific data-relevant decisions are distributed between school leaders and teachers. While prior research has indicated the importance of identifying key roles of managing data in schools, it has not attended to what decisions are made by whom using this evidence. Our findings from the SFA schools suggests that where key decisions for differentiating instruction reside affected the extent to which teachers felt responsible for taking further action.
  - (3) the way instructional monitoring and professional support strategies motivate teachers to focus on student performance data.

One question not answered in the foregoing analysis is whether or to what extent the different patterns of data use that emerged deeply impact student learning and tested achievement. Understanding that connection is critical, for it may well be that these patterns do not matter to performance outcomes at all, particularly on the performance of the school as a whole—clearly, a hypothesis that requires future study.

## References

Anhalt Erlichson, B. and Goertz, M. E. (2001). Implementing whole school reform in New Jersey: Year two. New Brunswick, NJ, Rutgers University: Center for Government Services

Barnes, C.A., Vanover, C., Kim, J. (2009) Evidence based leadership by design. Paper prepared for the 2009 annual meeting of the American Education Research Association. San Diego, CA.

Berends, M., Kirby, S., Naftel, S. and McKelvey, C. (2000). Implementation and performance in new American schools: Three years after scale. Santa Monica, CA: Rand Corporation, MD-1145-EDU.

Black, P. and Wiliam, D. (1998). Assessment and classroom learning, *Assessment Education*, 5(1)

Black, P. (1993) Formative and summative assessment by teachers, *Studies in Science Education*, 21, 49-97.

Bond, L. (1995). *Norm-referenced testing and criterion-referenced testing: The differences in purpose, content, and interpretation of results*. Oak Brook, IL: North Central Regional Educational Laboratory.

Bond, L., Herman, J., & Arter, J. (1993). Does American educational assessment pass the test? In N. McGinn & W. Cummings (Eds.), *Handbook of development education: Past and future*. New York: Garland Publishing Co.

Boyd, W.L. (1987) President Reagan's School-Reform Agenda", Education Week accessed at <http://www.edweek.org/ew/articles/1987>

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.

Coburn, C. E. and Meyer, E.R. (1998) *Shaping context to support and sustain reform*. Paper presented at the American Education Research Association Conference, San Diego, CA.

Coburn, C.E. and Talbert, J. (2006) Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, Vol. 112, No. 4: pp. 469-495

Cohen, D. K. and Ball, D. L. (1999). Instruction, capacity, and improvement. (CPRE Research Report No. RR-043). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education (CPRE)

Cohen, D. K. and Hill, H. (2001). *Learning Policy: When State Education Reform Works*. New Haven, CT: Yale University Press.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students, *Review of Educational Research*, 58, 438-481.

David, J. L. (1978). Local uses of Title I evaluations. Research report EPRC 21. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, Department of Health, Education and Welfare.

Elmore, R. F. and Rothman, R., eds. (1999). *Testing, teaching, and learning: A guide for states and school districts*. Washington, DC: National Academy Press.

Farley-Ripple, E. N. (2008) Accountability, evidence and school district decision-making. PhD dissertation, Philadelphia, PA: University of Pennsylvania.

Firestone, W., Monfils, L, Camilli, G., Schorr, R., Hicks, J., & Mayrowetz, D. (2002). The ambiguity of test preparation: A multimethod analysis in one state. Teachers College Record, 104(7), 1485-1523.

Fuhrman, S. (1999) "The New Accountability," RB27, Philadelphia, PA: Consortium for Policy Research in Education

Gallagher, C. and P. Worth (2008) Formative assessment policies, programs and practices in the Southwest region. (Issues & Answers Report, REL 2008–No. 041). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

Glazer, J. (nd) Internal memo for the Study of Instructional Improvement on the America's Choice design.

Gwynne, J. and Easton, J. Q. (2001, April). Probation, organizational capacity, and student achievement in Chicago elementary schools. Paper presentation. Seattle, WA: American Education Research Association.

Heritage, M., Kim, J., Vendlinski, T.P, and Herman, J.L. (2008) From evidence to action: A seamless process in formative assessment? CRESST report 741. Los Angeles, CA: UCLA.

Herman, J., & Gribbons, B. (2001). Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation. Los Angeles: University of California, Center for the Study of Evaluation.

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

Ingram, D., Louis, K.S., and Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. Teachers College Record, 106(6), 1258-1287.

Council of Chief State School Officers (1992) Model standards for beginning teacher licensing, assessment and development: a resource for state dialogue. Interstate New Teacher Assessment Consortium, CCSSO, Washington, DC.

Lave and Wenger, 1991

Massell, D. (2001). The theory and practice of using data to build capacity: State and local strategies and their effects. In S. H. Fuhrman (Ed.), From the Capitol to the classroom: Standards-based reform in the states. Chicago: National Society for the Study of Education.

McLaughlin, M. (1987) "Learning for experience: Lessons from policy implementation," *Educational Evaluation and Policy Analysis*, Vol. 9, No. 22, Summer 1987, pp. 171–178.

McLaughlin, M. W. & Mitra, D. (2002). Theory-based change and change-based theory: Going deeper, going broader. *Journal of Educational Change*, 2(4), 301-323.

McNeil, L. M. (2000). Contradictions of school reform: Educational costs of standardized testing. New York: Routledge

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

National Board for Professional Teaching Standards (1989) What teachers should know and be able to do.

Oakes, J. (1985). *Keeping Track: How Schools Structure Inequality*. Birmingham, N.Y.: Vail-Ballou Press.

O'Day, J. (2002) Complexity, accountability, and school improvement. *Harvard Educational Review* 72(3).

Peurach, D. (2006) *Designing and Managing Comprehensive School Reform: The Case of Success for All*, Ph.D dissertation, Ann Arbor, Michigan: School of Education.

Peurach, D. (nd) Internal communication on the Success for All design.

Plake, B. S., and Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom assessment* (pp. 53-68). San Diego, CA: Academic Press

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.

Sadler, R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-144.

Shepard, L. A. (2001). *The Role of Classroom Assessment in Teaching and Learning. Handbook of Research on Teaching*. V. Richardson. Washington, D.C., American Educational Research Association.

Simmons, W. and Resnick, L. (1993) Assessment as the catalyst of school reform. *Educational Leadership* (50)5, 11-15.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83, 758-765

Supovitz, J. and Klein, V. (2003) Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement. Philadelphia, PA: Consortium for Policy Research in Education.

Wayman, J.C. and Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education*, Vol. 112, No. 4: pp. 463-468

Young, V. M. (2006) Teachers use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education* Vol. 112, No. 4: pp. 521-548

Young, V.M. and Kim, D. (2007, draft) Using assessments for instructional improvement: A literature review. Philadelphia, PA: Consortium for Policy Research in Education.

## **Appendix A: Background of the Case Study Schools<sup>2</sup>**

### **America's Choice Schools**

**Bonds Elementary** is located in Sunnyside District, New Jersey, one of the 30 urban districts in New Jersey included in the 1997 Abbott Consent Decree, and thus required to select a comprehensive school reform. Bonds adopted AC in 2000, beginning with writers' workshop the first year and rolling out readers' workshop the second and mathematics the third year. Thus, when we visited in Spring 2002, writers' workshop had been in place for two years, but readers' workshop had begun only about half-way through that academic year.

All teachers and instructional leaders interviewed were supportive of the program and most were very supportive, and attributed their substantial growth on state assessments to America's Choice. The principal had been in the building since adoption, and was reported to be an active participant in instructional meetings related to the design. In addition, she reported spending an average of two hours a day in classrooms observing, modeling, and providing feedback

Site visitors from the SII research team noted that teachers and school leaders spent a great deal of time and energy attending to a variety of student assessment data. Teachers and leaders also noted that they worked together in grade level meetings to study results on the district tests and to use these to make decisions in the school about changes in instructional programs. Finally, teachers noted that in grade level meetings they shared and discussed student work and also shared and discussed classroom assessment tools such as running records. However, all teachers noted that they felt overwhelmed by the record keeping requirements of America's Choice and in general feel overextended. However, this sentiment did not lead teachers to express overall dissatisfaction with the program.

**Redmond Elementary** is located in Coverdale County, Florida. Coverdale's superintendent was deeply influenced in his philosophy and approach to improvement and school management by staff from America's Choice, and considered this design to be a centerpiece of district policy. Redmond's principal volunteered to be one of the pilot schools for the 1999-2000 school year, because she felt that teachers in the building were working hard but the school's performance on state assessments remained very low.

By Spring 2002, Redmond was implementing all major components of the AC literacy design. For the first two years, the school administered the AC Reference Exam

---

<sup>2</sup> These descriptions are based on write-ups conducted by various members of the SII research team.

in language arts and math to its students, but the district cut funding for the third year, and Redmond did not administer the exam that year or in any successive year. While the school had a design and literacy coach up through spring 2002, the district eliminated the literacy coach position from its first cohort of AC schools in an effort to share resources with newly implementing schools. However, staff remained committed to the design, particularly pointing to their significant growth on state assessments and their alignment with policies in the external environment.

For the first two years, Redmond made extensive use of the data generated by the AC Reference Exam, which they administered at the end of third and fourth grades. They instituted changes in their instruction in light of the results from the Reference Exam. The school's relationship with its external AC liaison was positive; in fact, this person had previously served as its design coach and was well-regarded. The new design coach focused much attention on using data and supporting teachers in its use.

**Westwood Elementary** is located in Adderly Community School District, New York City, New York. Westwood was performing poorly on state assessments and was encouraged by the prior superintendent to select a CSR design. They adopted AC in 1998, but lost the original principal who spearheaded its adoption the following year. The new principal arrived with little knowledge of AC, and felt that she did not get the support that she needed from AC her first year. While the design and literacy coaches thought their AC training was extremely good, in general this school did not have the kind of strong, on-going relationship with AC staff experienced by faculty in our other two case study sites.

Nevertheless, the principal reported that her confidence in the program grew over time because she began to see improvements in student work, especially writing, and design implementation did improve. By the time we visited in Spring 2002, accounts suggested that teachers' use of writers' workshop had moved forward. Though some pockets of resistance remained, the early grade teachers were reputedly committed to the design, as were fourth grade teachers who had seen their students make gains on state assessments. While the teachers we interviewed and observed used reading practices consistent with AC readers' workshop, including adjusting instruction on the basis of formative assessments, school leaders thought that this component had made much less headway in the building.

While the Adderly school district incorporated several elements of AC into district policy and professional development, such as leveled libraries and professional development on AC curriculum units (core assignments), it was the classic "christmas tree" district that seized on many reform ideas. Thus, while staff saw some of their efforts as aligned with district policy, they also had to fend off other intrusions. Finally, despite notable progress on state assessments the school remained in corrective action status, and in Fall 2002, the district required Westwood to begin phasing in a new literacy program called Voyager. Yet our interviews suggest that school leaders and teachers had embraced the AC philosophy and wanted to sustain it, despite the new mandate and despite the fact that they no longer had CSR funds.

Accounts from our respondents suggest that AC influenced schoolwide data collection and use, as well as teachers' use of data from formative classroom assessments. For example, school leaders reported administering a new summative assessment as part of AC, and found that it helped better prepare their third grade students for the fourth grade state assessments. They used the results to create a booklet for fourth grade teachers to help them integrate particular test skills into the literacy block. The teachers we interviewed were conversant in many of the AC classroom formative assessments, and talked about making instructional decisions based on this information. Student work had become a source of discussion at teacher meetings.

The city and state mandated a high volume of assessments, which were largely seen as aligned to the model. School leaders and teachers also received professional development on how to interpret data. However, although some of these assessments were seen as useful for instruction (especially the early literacy assessment), the principal did not think most teachers were actually using it. Furthermore, the principal repeatedly complained about being overwhelmed by too much data.

### **Success for All Schools**

**Doris Elementary** is located in the Adderly Community School District in New York City. As a school identified for corrective action under the state accountability program, it was pressed to select a comprehensive school reform model. In 1999 Doris adopted SFA, and brought in its Roots and Wings literacy program. However, the principal who brought in the model retired, and the school subsequently hired a series of administrators with little to no knowledge of or commitment to SFA, and who did not use their authority to support implementation or monitor instruction and thus little incentive for resisters to move forward with the program. By the time of our second visit in Fall 2002, the school finally received a new, more experienced leader who had been in a successful SFA school. But that previous summer the district required that Doris, like Westwood, begin phasing in another literacy program called Voyager.

The school did have stability in its facilitator position, which was shared here by two individuals. They used data to place students into groups for SFA instruction, and developed skills packets and other strategies for teachers to address student weaknesses. However, most sampled teachers indicated that they did not actively use this information. And the time set aside in the design for collective professional learning—component team meetings—were voted out by Doris staff for the 2002-03 academic year, and were only sporadically attended in earlier years. For these and other reasons, the defining theme of implementation in this building was inconsistency. While the school had made some improvement on external state tests, reading progress was less than in the other two schools and SFA was not uniformly viewed as the main contributing factor in this growth. They also had a group of students who did not make progress in reading, and they struggled to handle this challenge.

**Gladrock Elementary** was located in Sunnyside, New Jersey. This school adopted SFA in 1998, and by 2002 they were implementing the SFA literacy program and the World Lab program for science and social studies. SFA had a strong presence in this school, which was designated a “visitation site” by SFA to show others considering the model. Although the school received a new principal in Fall 2002 who was unfamiliar with the design, he quickly became an advocate for the model, perceiving it to provide the kind of specified instructional guidance that their largely young, inexperienced teachers needed.

Thus in general, teachers and school leaders were supportive of the SFA program, and credited the model with raising student achievement in reading. Evidence of its implementation in classrooms was widespread, and any obstacles and challenges were viewed as minor. Respondents noted many supports to implementation, including a stable and experienced SFA facilitator, ample materials, sufficient SFA and district professional development, and a general sense that program was well aligned with state assessments and district requirements. However, while SFA was clearly at the center of Gladrock’s instructional program, in 2002 the Sunnyside district began pressing all of its schools to strengthen writing using America’s Choice methods. Consequently, component team meetings no longer centered on SFA, and these on-site professional learning opportunities for the program diminished.

The external SFA liaison assigned to the school described the Gladrock facilitator as extremely proficient in using multiple sources of SFA data to make well-reasoned decisions about student groupings. Although external data and SFA quarterly assessments were used as intended to group students and chart growth over time, only one teacher discussed modifying instructional effort based on student achievement data.

While school leaders reported close examination and attention to the New Jersey state tests and to SFA data for regrouping, data was not a strong focus for most interviewed teachers here. Component team meetings were focused on the recently introduced AC writing program in 2002, and so professional learning time focused on SFA data was infrequent.

**Nightingale Elementary** adopted SFA in the 1998-99 school year, and unlike the other two case study schools here, they selected both the SFA literacy and the SFA mathematics components. The principal at the time of adoption continued to lead the school throughout this period, thus providing a high level of stability. To try to ensure a staff committed to the model, she counseled teachers to leave if they disagreed with the design and was in other ways known to be a harsh enforcer of SFA. Then, as in other well-functioning CSR schools, she signaled her support by attending SFA meetings regularly and monitoring SFA instruction as intended. Although the original staff member selected as SFA’s literacy facilitator did change, the individual who assumed her role was widely perceived to be a dynamic, knowledgeable and trusted instructional leader. The school made substantial gains on external assessments in reading during this period, and staff attributed the gains to SFA and an adept use of data to modify school and teacher practices.

The literacy facilitator maintained close and on-going interactions with the SFA Foundation, and became an skilled advocate for data use. In their fourth year of implementation, when the other schools were winding down their staff time for SFA component team meetings, she and the principal arranged coverage and scheduled meetings to provide consistent time for Nightingale teachers to discuss SFA to consider instructional strategies. In fact, teachers' understanding and use of data use was identified as a top school priority.

As in other schools, the facilitator used the quarterly assessment data to regroup students as intended in the model. But she also strongly encouraged teachers to make noticeable adaptations to the pacing and sequencing of SFA instruction to address particular learning needs identified in the data, and/or to better align instruction to the assessment expectations in the external assessments. However, as in the other SFA schools, teachers were reluctant to make significant changes to the specified routines.