Center for Research on the Context of Teaching
Stanford University

CRC

Leading School Improvement with Data: A Theory of Action to Extend the Sphere of Student Success

Joan E. Talbert
Stanford University

Nell Scharff
Baruch College

with

Wendy Lin
Stanford University

March 2008

Draft for Review: Please do not distribute without permission

This paper was presented at the annual meeting of the American Educational Research Association, New York City, March 24-28, 2008. It reports findings from the SAM II evaluation, which was supported by New Visions for Public Schools (NVPS) through federal funds allocated for an evaluation it its New York City Leadership Initiative work in Region 1. Ongoing evaluation research on SAM within the NVPS-PSO is funded primarily by the Carnegie Corporation. We are grateful to the SAM facilitators, inquiry teams, principals and teachers for the time and thought they contributed to CRC interviews and surveys. Special thanks to Deirdre DeAngelis for partnering with us on our case study of New Dorp High School.

Leading School Improvement with Data:
A Theory of Action to Extend the Sphere of Student Success

Joan E. Talbert
Stanford University

Nell Scharff
Baruch College

Abstract

This paper describes and evaluates the Scaffolded Apprenticeship Model (SAM), a school improvement and leadership development program that supports a school "Inquiry Team" to close targeted learning gaps and lead their colleagues to do the same. In SAM's theory of change, the school's "sphere of success" continually expands as interventions focused on students' specific skill gaps accelerate their learning. The study tests this theory of change with a range of research methods and data from interviews with SAM facilitators working with the 14 New York City high schools involved in the program, a survey of program participants, two years of teacher surveys in all schools, and a case study of a large restructured high school that showed significant success with this reform model. Results provide evidence of: a) significant growth in team leadership in most of the schools, b) shifts of school culture toward use of student assessment data to focus improvement efforts, and c) improved student outcomes in schools that made expected gains in leadership and data use. The study analyses conditions that support and that inhibit SAM's success and describes processes of change in the school that made greatest progress. Conclusions point to promising strategies for school improvement and hypotheses for further research.

Introduction

Despite policy demands for schools to use assessment data and other sources of information to evaluate and improve their performance (NCLB, 2002) and local education leaders' enthusiasm for this theory of change (Archer, 2005), research on schools' use of evidence to make instructional improvements suggests that few do so. For one, educators and administrators tend to hold different conceptions of valuable data and evidence (Coburn and Talbert, 2006), and therefore district data systems and designs for schools to use them for instructional improvement are often quite out of sync with teachers' needs. Further, teachers typically have little experience or support in using assessment data to detect specific student learning gaps and to design or identify effective instructional interventions (McLaughlin and Mitra, 2003; McLaughlin and Talbert, 2006). In addition, the culture of teaching works against evidence use (Ingram, Louis, and Schroeder, 2004; Supovitz and Klein (2003), and professional and community politics often enforce school practices –such as student and teacher tracking – that undermine learning opportunities of the least successful students (Talbert, 2002).

Organizational, technical, cultural, and political challenges are entailed in developing new practices for using evidence to improve instruction (Talbert, Wood, and Lin, 2007). Challenges of developing teachers' capacity and will for evidence-based school improvement are particularly daunting in high schools because of their size and organizational complexity, subject-specific assessment needs, and professional sub-cultures that resist school-wide instructional interventions (McLaughlin and Talbert, 2007).

The Scaffolded Apprenticeship Model or SAM builds upon research literatures on learning, assessment, and school change to address these challenges. It develops the capacity for a school leadership team, composed of teachers from across high school units and the principal, to use data to identify student learning gaps and to target interventions to expand the school's "sphere of success." Because its theory of change addresses key challenges revealed through prior research – incoherence between administrators' and teachers' conceptions of data for evidence-based practice, difficulty translating knowledge of student learning gaps into instructional interventions, fragmented professional leadership in high schools, and teaching cultures that maintain the status quo – SAM offers promising change strategies to be tried and tested in the field.

This paper evaluates SAM's theory of change for leadership development and inquiry-based school reform. After describing SAM and our research methods, we address three broad questions:

- How do SAM inquiry practices work to expand the sphere of student success?
- To what extent did school leadership and culture change through SAM II?
- How does SAM's program design develop leadership for inquiry-based reform?

The answers to these questions have implications for ongoing SAM work in New York City and beyond, suggesting strengths and weaknesses of the program's theory and design and directions for refinement and further research.

SAM's Theory of Change and Program Design

SAM addresses the problem of leadership and support for evidence-based school improvement in two ways. First it establishes a school team to lead and sustain change. The team is composed of the principal and teachers who broadly represent school units; in high schools the team includes subject areas across the school or within SLCs, thus engaging all professional subcultures. Second, it develops the team's knowledge, legitimacy, and skills through a leadership certification program that includes a curriculum and an on-site facilitator focused on evidence-based school improvement. Unlike typical certification programs, work in SAM is actual school improvement at the school site. The program holds itself accountable to a) improve student outcomes and b) build the school's capacity to sustain improvement.

SAM's stance on the problem for change is that every school has a "sphere of success"—a group of students with whom the school is currently successful, and that the goal of improvement is for a school to learn to continually expand this sphere. This requires developing broad, deep leadership and a change in the beliefs, practices, structures and decision rules that result in current outcomes, a change in school culture and ways of working.

Rather than addressing school culture head on, however, SAM begins with an indepth look at the problem for change. A key idea in SAM's theory of change is "getting small" in order to go big with evidence-based school improvement. A SAM facilitator supports the school team in using data to identify a specific group of students outside the sphere of success; to select one high leverage skill in which the target students are deficient; and to learn, through an action research cycle, how to move these students in the identified skill. The entire program is organized around this core task, and the discipline of assessing interventions according to evidence of accelerated student learning in one measurable skill pushes participants to examine their current practices and the belief systems and structures that underlie these practices, providing a leverage point for individual and organizational change.

Origins

SAM evolved within a larger effort in education to connect leadership preparation as closely as possible to the demands of real world practice. SAM has roots in the Aspiring Leaders Program (ALPs) at Baruch College (Stein & Gewirtzman, 2003), in which practitioners and faculty co-constructed and co-taught problem-based scenarios to teams (Bridges, 1992), and in the New York City Leadership Academy, where team-based curriculum is grounded in complex school simulations. In both ALPS and the Leadership Academy, aspiring leaders work in teams to develop competencies needed to

solve real world problems, and the role of the facilitator is to maintain an environment that simulates the real world. In SAM, however, actual school teams prepare for leadership by enacting it. Leadership is defined in real-world terms—by the collective capacity of a team to move a group of students with whom they were previously unsuccessful and to lead their colleagues to do the same. The curriculum is organized in support of, rather than to prepare participants for, actual improvement.

Currently in its third iteration, SAM was developed through a partnership between Baruch College and New Visions for Public Schools (NV) and piloted with four high schools, two large and two small, in 2003-05. A refined SAM II was then implemented in January 2006 with fourteen high schools in cohorts of two large high schools, five Bronx Region 1 small high schools, and seven small Autonomy Zone schools. The findings reported here are for SAM II. The third iteration involves two cohorts of the 63 schools in New Visions PSO in two versions of SAM: a certification version that replicates SAM II with eleven high schools and a non-certification version involving the other 52 PSO schools spanning all grade levels.[1]

Conceptual grounding, program elements, key strategies

SAM's curriculum for leadership development is organized around core concepts and literature. Central is the concept of shifting the focus of reform leadership from teaching to learning (DuFour, 2002). Research indicates that professional work focused on students and student learning is the centerpiece of effective instructional interventions; that school leaders and teachers learn to improve their practice through such joint work. Further, student assessments must be designed for a grain size that informs instruction (Popham, 2003). As elaborated below, this concept grounds SAM's strategy of starting the change process by focusing on a small group of target students and a specific measurable skill set.

Core SAM concepts also concern conditions for team leadership and ways of thinking about organizations as systems. SAM is grounded in the notion that deep, sustainable reform requires the collective knowledge and concerted action of high performing teams. It is grounded in the belief that a team is greater than the sum of its parts, and therefore capable of greater wisdom and decision making than any individual member. High performing teams embrace cognitive diversity and productively use conflict to improve their practice (Surowiecki, 2004). SAM emphasizes team development a) to support distributed leadership, and b) because the team is a microcosm, thus revealing systemic elements of the larger organization for focused study and change. Also a key element and the focus of a full SAM module, are principles of systems theory, specifically the idea that behavior, structures, decision-rules, and belief systems are linked in organizations and manifest in the individuals and team that represent the whole. Therefore, one small strategic change at the level of the individual or team produces systemic change (Senge, 2000).

[1] In addition, Boston Public Schools is piloting SAM in five schools across all grade levels.

According to systems thinking, current practices result in current outcomes. Another key concept is that in order to expand the sphere of student success, teams must begin at the edge of what is known and learn what is unknown. They must distinguish problems from dilemmas (Cuban, 2001). Problems are responsive to technical solutions, while dilemmas require adaptive leadership (Heifetz & Linsky, 2002). SAM teams continually create and test solutions to contextual, educational dilemmas that have not been solved. A key facilitator competency, then, is the willingness to publicly not know—to be a learner supporting others to learn publicly. The SAM process supports the creation of new knowledge and solutions, rather than providing answers. A core belief as well is that reform at all levels must be linked to and evaluated according to changes in the instructional core (Elmore, 2004).

SAM program elements include a range of seminar, on-site, and off-site activities required of the inquiry teams. Specifically: a four-week summer intensive, weekly seminars, readings of literature tied to curriculum topics and core concepts (per above), in-class activities, activities designed for the school, monthly on-site coaching, apprenticeships overseen by the principal, and monthly inter-visitations.

Finally, two key *inquiry strategies* are central to SAM's theory of change. They are: "getting small in order to get big" and looking objectively at teaching and learning in the classroom. Based on experiences during SAM's piloting years, program leaders learned that the grain size of student assessment data that school teams use makes a difference in their ability to identify and implement effective interventions . More importantly, program leaders learned that keeping the grain size small is a critical element in overturning assumptions that perpetuate current practices and current outcomes. Working at a large grain size (for instance, "writing," instead of "the logical sequencing of ideas") allows participants to stay in the world of generalizations and assumptions. By honing in on their target students' specific skill gaps, teams are able to face student learning gaps head on; evaluate to what extent specific learning needs are being met in the current system; design and evaluate strategic interventions; and see clearly that students can learn. In SAM's theory, successive iterations of this practice significantly accelerate student learning, overturn assumptions about student learning that constrain change, and expand the sphere of student success through a synergy between stronger instructional interventions and growing cultural change. A core purpose of SAM, then, is to reorient a school's gaze to the needs of struggling learners, and for a school to come to hold itself accountable for continually learning to move students with whom they have not previously been successful.

Looking objectively at teaching and learning in the classroom – through "low-inference observations" – is another key SAM strategy. Team members are directed to observe classrooms and create transcripts that capture *what* is happening in the classroom, without judging or drawing inferences about *why* particular actions occurred. This practice is meant to promote leaders' deeper and more objective understanding of what needs to change in teaching practice to improve student achievement and to provide a means for determining whether changes have desired effects on student learning. Low-inference transcripts are a vehicle for participants to see their own and other's instruction

more clearly; to help them shift their focus from teaching to learning. They are also are a vehicle leaders can use to engage colleagues in discussions about a classroom event without debating the facts and to identify patterns within and across classrooms to inform targeted professional development.

Facilitator development and learning community

SAM facilitators are pivotal to the success of the program – leading seminars, guiding teams' development of inquiry practices, and supporting the team's work in their schools. Appropriately, the program invests considerable resources in developing facilitators' ability to carry out these roles effectively. Too often in ambitious initiatives aiming to develop school capacity for continuous improvement, coaches receive limited professional development or support for their challenging work with diverse schools.

Individuals recruited into the SAM facilitator role bring diverse perspectives and expertise from their varied backgrounds in positions such as university instructor, professional developer, principal, teacher leader, school support provider – and thus bring diverse perspectives and skills to their new role. Some have more experience using student assessment data to analyze skill sets in content areas and conducting action research than others; some are more versed in the literatures that ground the program than others. SAM is thus challenged to create learning opportunities for the diversity of individuals in this new role that will develop the wide range of knowledge and skills essential to their success in the facilitator role and sustain their professional growth.

SAM's strategy for facilitator development is a weekly Monday session that involves individuals in learning the NYC data systems and analysis skills, discussing key program readings and big ideas, experiencing team-building exercises, bringing examples of successes and struggles for group discussion, role-playing ways to address particular challenges encountered in the work, self-assessing and identifying goals for improving practice, and co-designing work. SAM's lead facilitators guide the sessions around core program principles and practices. The weekly sessions are meant to ground facilitators in SAM's conceptual framework, build a common language and new identities, develop a common repertoire of tools and practices, provide a space for reflection and formative assessment, and ensure that individuals have colleague support in meeting the challenges of their work. In effect, they create a "community of practice" among the facilitators of inquiry-based leadership development and school reform (Wenger, 1998).

Facilitators' work with school teams is organized around a task-based curriculum. At the core is a common set of assignments, all of which scaffold the participants to identify target students and a high leverage skill in which they are deficient, and to provide evidence of moving these students in the identified skill. "The task drives" at all levels of SAM. Monday sessions are designed to provide facilitators with the experience, resources and skills they need to support their participants in successfully completing these SAM assignments. Assignments (participant products) are brought to the Monday sessions and facilitators decide collectively whether participant work meets the program standard according to a collectively designed rubric. The SAM facilitators hold

themselves accountable for the quality of participant work, based on the assumption that the quality of the work reflects the quality of the teaching. Similarly, the program leaders (SAM trainers) view participant products as evidence of the quality of the facilitator trainings, and as a diagnostic about what needs to be "taught" next.

At this point in the evolution of SAM –as theory and design – it is important to draw upon lessons learned through experiences of the fourteen SAM II schools. What do their experiences tell us about the efficacy of the team inquiry strategies and the conditions necessary to develop leadership for school change to expand the sphere of student success?

Evaluation Design and Methods

This study evaluates SAM's theory of action and work with fourteen high schools that participated in SAM II from January 2006 through June 2007.[2] Quantitative and qualitative research methods provide breadth and depth of analysis of SAM implementation and outcomes, as well as conditions that enabled and constrained implementation. A case study of one school's inquiry work over time documents processes through which SAM's inquiry strategies and program design work in practice.

Research methods and data

Quantitative data were developed for all fourteen SAM schools through surveys and student record data. These data collection efforts included:

- a survey of all fourteen schools' SAM inquiry team members (N=76) in February 2007, midway through the second year of their participation. The survey measured the teams' experiences with SAM program features and their self-assessments on intended leadership outcomes.
- baseline and follow-up surveys of all teachers in each school in the Spring of 2006 and 2007.[3] The surveys measured key SAM outcomes: a) school leadership capacities and practices promoted by the program; and b) school culture conditions and practices, such as teachers' use of data with colleagues to identify student learning gaps and target interventions.
- record data on student outcomes – including attendance, credit completion by grade level, scores on Regents exams, and degree completion rates.

Qualitative data were developed to support greater depth of analysis of the teams' experiences with the program design and in leading inquiry-based improvement in their school. Data come from:

---

[2] Of the fourteen SAM schools, two were large restructured high schools located in Jamaica, Queens and in Staten Island; five were in Region 1 of the Bronx, and seven were in the Autonomy Zone (renamed Empowerment Schools in Fall 2007). These school groups are referred to as SAM "cohorts."

[3] All fourteen schools participated in the 2006 survey; teacher respondent N= 570 (81% response rate); Eleven schools participated in the 2007 survey (three Autonomy Zone schools opted out); teacher respondent N=440 (71% response rate).

- principal interviews and SAM team focus groups in six of the fourteen high schools. This subsample represents contrasts in size (large versus small high schools), region and jurisdiction (regular versus Autonomy Zone schools) and all six facilitators involved in the program over the two years[4];
- observations of seminars involving all of the schools; and
- annual interviews with the six facilitators involved in SAM II.

Our case study captures SAM practices and outcomes in the highest-implementation site, a large high school – New Dorp High School in Staten Island – that works with the lead facilitator (study coauthor). Case study data combine quantitative and qualitative data developed through participant observation and external evaluation research. The case study focuses especially on processes through which the SLC teams worked with the facilitator to analyze student data, to design and evaluate interventions with target students, and to involve colleagues to extend the sphere of student success. It also documents ways in which the partnership between New Dorp's principal and the SAM facilitator worked to strategically leverage and support the school change process.

Data analysis and nature of evidence

To address the question of how SAM's inquiry strategies work to expand the sphere of student success, we use evidence from the in-depth case study and from interviews and focus groups in another five high schools to describe how and why the practices bring about change. This analysis provides conceptual grounding and illustration of how inquiry team members' perspectives on struggling students, on teaching and learning, and on inquiry practices shift through the work SAM prescribes.

To assess the extent of school change through SAM II, we use baseline and follow up teacher survey data measuring desired leadership and school culture outcomes and evidence from our New Dorp case study. First we examine overall trends across participating schools. As yardsticks for assessing the extent of change achieved by SAM II schools, we use a) New York state norms from a national survey to gauge the distance moved on a school culture outcome and b) evaluation findings from a California Bay Area initiative to promote inquiry-based reform. Then we test SAM's theory of change by estimating effects of team leadership practices on changes in school culture outcomes, using path analytic techniques. This analysis uses variation among the SAM schools on these outcomes to assess the assumed cause-effect relationships between them. We then describe the ways in which New Dorp's SLC inquiry teams have been leading school culture change.

Finally, to address the question of how SAM's program design develops team leadership for inquiry-based reform we use data from the mid-course SAM team survey and from our New Dorp case study. We describe broad patterns of individuals'

[4] At any one time there was one facilitator in each of the large high schools (2 total), one working with the five Region 1 schools, and two working with the seven Autonomy Zone (AZ). One AZ facilitator left at the end of the first year and a new facilitator joined the program; so ultimately 3 SAM facilitators worked with the AZ schools.

experiences with the program and point to facets that they found to be most and least valuable for their leadership development. The New Dorp case illustrates how the program components work synergistically to advance inquiry-based school reform.

Research Findings

The fourteen high schools involved in SAM from January 2006 through June 2007 varied in their experiences with the program model, implementation of intended leadership practices, expected school culture changes, and student outcomes. Schools where SAM teams developed strong leadership and engaged their colleagues in addressing specific learning gaps of target students were those that realized largest gains in student achievement. In places where SAM's leadership principles and practices developed, teachers began to share accountability for using data to design and evaluate interventions to expand the sphere of their students' success. Although all teams improved outcomes for their target students, those that best navigated challenges of change and made greatest strides in expanding the sphere of student success had the benefit of sustained guidance and support from a SAM facilitator and the principal.

How SAM inquiry practices work to expand the sphere of student success

Inquiry teams in the fourteen SAM schools began their work by examining student assessment data in order to select a group of target students – students in their school or in their SLC in the large restructured schools who were falling below grade-level performance on multiple criteria and whose attendance was not too low for them to benefit from instructional interventions. Teams' criteria for selecting target groups varied widely across the schools and SLCs. One team targeted sophomores who scored 80 or below on the second term of Math A, for example; another focused on lowest-performing students across the grades. Several teams adjusted their target group as they began to develop an intervention and ongoing assessments. For example, one team initially defined its target students as all 6-11th graders who failed both math and English courses or the state exams in both subjects (about 20 students across the grades) and later adjusted their target group to focus on twelve 10th graders who entered 9th grade with very low schools on a literacy assessment.

The next step was to identify specific skill gaps within or across content areas that account for the target students' poor performance on assessments and that therefore represent strategic intervention points for accelerating their achievement. They then designed an intervention that they would all implement in their classrooms and then evaluate. Particular skill gaps identified varied widely across SAM teams as a function of their students and the nature of assessment data they analyzed. The team that targeted sophomores scoring low in math, for instance, identified factoring and LCD (finding the least common denominator) as target skills for accelerating their students' math performance. The team that targeted low-performing 10th graders honed in on the English regents strand of "information analysis" as the skill gap to address in their intervention; while the team whose target students spanned grades focused first on

graphing skills and then shifted to critical thinking, or taking information and both summarizing and interpreting it.

The teams then explored existing curriculum and lessons to see where, if at all, the identified skills are currently taught. They then conducted low-inference observations in each others' classrooms and used transcripts as the basis for insights into, and conversations about, their target students' experiences with instruction specifically in relation to the identified skill gap. Through ongoing inquiry into links between target student learning and curricular and teaching practices, the teams deepened and extended their interventions with target students. In many cases, they discovered that the identified skill gap is simply not something that is being taught. In others, they recognized that a skill was introduced, but not broken down or scaffolded sufficiently. They designed interventions; assessed progress in the identified gap; and either refined or shifted strategies to focus on another identified skill gap. They prompted their colleagues to adopt successful interventions by showing them data on student achievement gains through the strategy.

Each facet of SAM's theory of action challenged the inquiry teams to develop new practices – using data to identify target students and skill gaps, examining curriculum in light of evidence of student needs, using low-inference observations to examine instruction through the lens of target students, designing interventions, and leading evidence-based improvement efforts in their school. How they addressed the challenges determined how well they advanced through the model and the extent to which student achievement improved significantly. Our research points to four key principles for success with inquiry-based reform: keeping the focus small, seeing instruction through the lens of target students, challenging assumptions and surfacing practices that limit student success, and navigating colleague resistance and facilitating their learning to use data to extend the sphere of student success.

Keeping the focus small. SAM teams made progress when they identified a skill that they could measure, teach to, and evaluate change. This is because target students struggle in so many academic areas that the challenge of accelerating their learning can appear overwhelming, and it is easy for teachers to throw up their hands or focus on factors they cannot control. By keeping the focus small, the conversation shifted from the general to the specific—from students' general failure and troubled personal and academic backgrounds to the specific instructional interventions needed. Conversely, small gains translated into new conceptions of what's possible.

One team, for example, initially resisted developing a finely grained assessment to identify target students' skill gaps. A participant said repeatedly: "I know what the problem is. The students aren't motivated. They never do their homework." The facilitator explained that the SAM process can be used to test assumptions, to see—for example, whether completing the assigned homework would have provided target students practice in the identified skill. Being forced to stay small and evidence-based helped reorient the team members, who came to see that a target student missing a foundational skill might never be exposed to what he or she needs to learn to be

successful. In this team, and others the conversation shifted from student motivation or ability to when (if at all),how and how well students are taught what they need. This inquiry work raised the question of how teachers can meet curriculum mandates *and* address student skill gaps, teaching students from where they are—a leadership dilemma that many SAM teams came to face head on. SAM's theory is that if specific skill deficiencies are targeted and addressed efficiently and strategically, student learning can be accelerated. -

Seeing instruction through the lens of target students. A key tool for helping participants identify and revise assumptions is low inference observations. In SAM, participants first observe each other and then branch out to observe teachers across departments or the school. Ideally, they observe classes that contain their target students. They create verbatim, low inference typed scripts of everything they see and hear, and then analyze these scripts for patterns in relation to identified skill gaps. The observation process has many purposes, including helping participants see instruction more deeply, become knowledgeable about patterns in practice schoolwide, and forge a culture of collaboration. Most notably, however, the tool helps participants shift their focus from teaching to learning. One participant said: "When I used to observe, I would just look at what the teacher was doing. Now I look at the students, at whether or not there is any evidence that they understand whatever it is they are supposed to be learning."

Another SAM participant described the shift as follows:

> You're no longer looking at a class as a whole, you're looking at a classroom as a group of individual people. So it's the difference between looking at the class as a group versus a series of individuals and realizing that each individual has separate wants, needs, desire, skill, ability. So you cannot make one shoe to fit everybody. You have to make shoes to fit the feet.

The principal of New Dorp echoes this sentiment, seeing a direct connection between participants' shifting focus and their ability to meet student learning needs, as demonstrated in improved student performance. A SAM participant at New Dorp experienced a striking success this past January, when every student in her two repeater math classes—each of whom entered high school level 1 or level 2 in math and failed math the prior year—took and passed the Math A regent 1 semester early—all but one with a 65 or above. The principal explained the change she saw in the SAM participants' teaching as follows:

> Before SAM, she was a strong teacher, but she was all about coverage. If you asked her how she knew if the students understood something, she'd say: "Because I taught it." Now she can tell you exactly what each student knows and doesn't know. She stops all the time to see where students are, if they're not all with her, she doesn't move on.

Although, many inquiry teams struggled initially with low-inference observations and some resisted doing them at all, most came to see their value over time. In the mid-course SAM team survey, a little over three-fourths (79%) reported feeling well prepared to conduct these observations and only about half of the team members (55%) rated them as useful in guiding instructional improvement in their school. By the end of the year, however, team members reported in focus groups that they were an important vehicle for their learning through SAM. At New Dorp, the principal and SAM II participants have found the process so powerful that they are spreading the process schoolwide. Every teacher will conduct two low inference observations of colleagues this school year.

Challenging assumptions and surfacing practices that limit student success. SAM enters the problem of culture change via identifying and shifting outcomes for specific students, not by addressing culture or belief systems head on. In SAM, the process of challenging assumptions happens in layers, deepening over time. First, participants experience "aha" moments about their own teaching and/or schoolwide practices. At New Dorp, for instance, one veteran English teacher said she thought that she and other members of her department were teaching writing. When she analyzed transcripts in module I, however, she was shocked to learn, in her words, that: "We're not really teaching writing, myself included! We're doing a lot of do-now's, but we're not breaking down or really teaching the specific skills students need to write well."

Typically, however, the deeper assumptions, about the inability of target students to learn, for example, become challenged later in the process when team members hit the wall in their own learning, and have to break through that wall.

For instance, one team at New Dorp identified "the logical development of ideas" as their target skill and was able to boost performance for all but five of their target students through use of an instructional tool (a linear graphic organizer) that was already in their knowledge bank, but that they could implement more broadly. In module III they realized that this strategy was effective with all but their five ELL (English Language Learner) target students, and that they simply did not know how to move these students, who had not even scored at the first level of their self-designed rubric.

At first, the team wanted to drop these five students, since they felt they did not know how to move them and that their needs would be better served by the ELL teacher. The principal, assistant principal leading that team, and the facilitator aligned in agreement that these target students could not be dropped (that this was a teachable moment in SAM), and pushed the team to design an intervention and demonstrate progress for these five students by June. The team, when pushed, finally decided upon an intervention—to give the same assessment to these five students, but in their native language. When the team presented their results in June, the entire SAM cohort was rapt, sitting on the edge of their seats. All five students had scored a 5 (the highest score) on the rubric. The team still needed to learn how to help these students with writing proficiency in English. Challenged however was their belief that somehow these students couldn't learn, that they "couldn't even score a 1" on the attached rubric, as opposed to the fact that the rubric was an insufficient tool to capture student knowledge. Knowledge

of their literacy strengths in their native language led the SAM participants to design entirely different interventions to move these students forward.

At the end of the third module, after identifying and moving almost all target students in the identified skill, SAM II participants at New Dorp reported that at first they believed "the problem was the kids," but that they did not feel that way anymore.

Navigating colleague resistance and facilitating adult learning. A current challenge in SAM is ensuring that the process of continually expanding the sphere of student success is sustainable within a school. While it seems clear that all SAM II schools experienced some degree of success moving target students in an identified skill, and while many describe having to navigate colleagues' resistance to do so, it is less clear that the processes for continuing to create new learning and practices, and to evaluate their impact according to evidence of improved outcomes for students, will continue or spread. The two large high schools, both of which have inquiry teams in every SLC, have made the most progress in scaling up the program and embedding it in the workings of the school. New Dorp in particular has made notable strides in this area, with every teacher in the school now participating in a SAM-based inquiry team.

The challenge here though, interestingly, involves participants cycling again to confront assumptions, and to challenge these assumptions in leading learning for their colleagues. The first round involved a realization that they didn't believe students could really learn. The second round at New Dorp has involved surfacing a belief that their colleagues cannot learn. It makes sense systemically, in terms of the deprofessionalizing of teaching as a profession, that a core assumption that needs to surface and be challenged is belief in the limited capacities of colleagues. Before entering Module 4, where the focus is sustainability, each SAM team at New Dorp analyzed transcripts of their leadership. When asked to look at what underlying belief systems might make sense of patterns in their behavior, they quickly decided upon "We don't trust the teachers." All the SLC's are currently working on this in behavioral and measurable terms. In one SLC, for instance, team members conduct monthly transcripts of their leadership of common time, looking for increasing evidence of more teachers' talking over time.

This raises questions about how to scale up the process since realizations seem to have to be learned over and over in increasing depth in an iterative process. What is the ongoing need for an external facilitator in this regard? At what point can we feel confident that continual learning is sustainable?

How school leadership and culture changed through SAM II

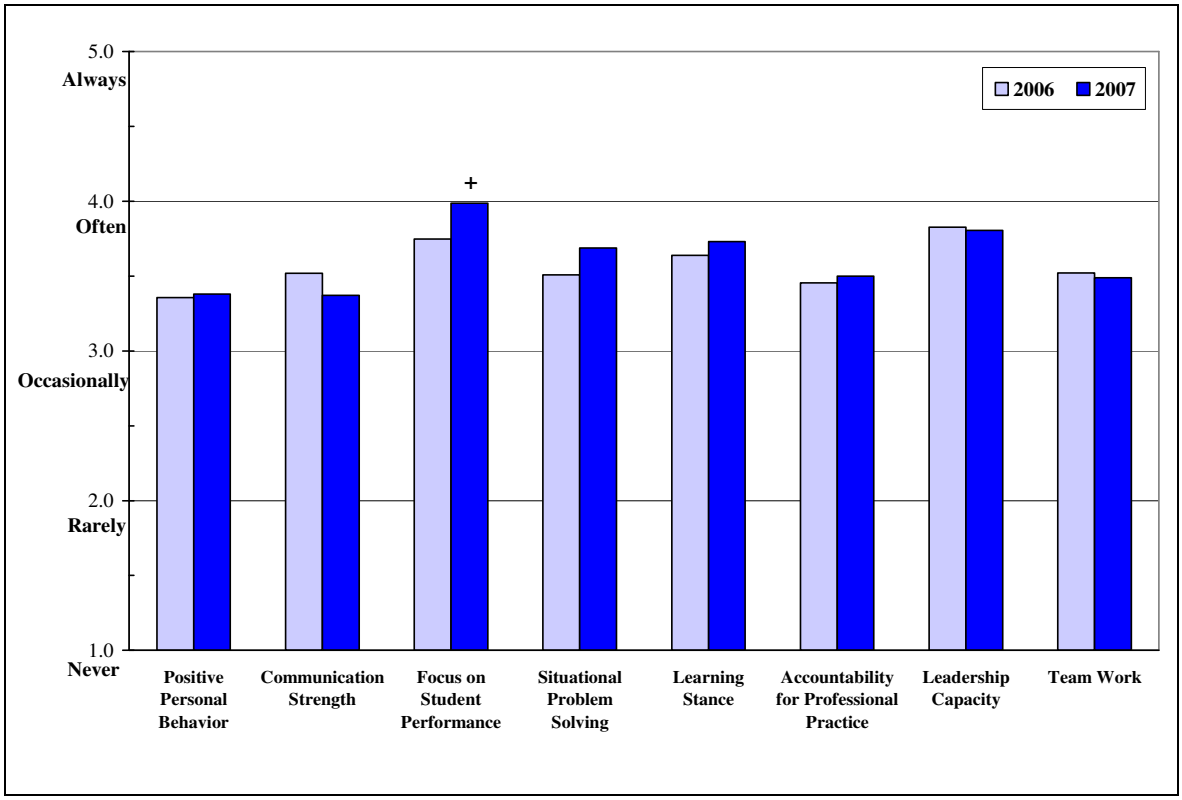
SAM's key concept of starting small to get big anticipates that school teams, through their scaffolded inquiry practices, will develop the capacity to lead inquiry-based reform in their school. Their leadership will be grounded in assessment data analysis skills, evidence of how focused interventions boost student performance, team collaboration, team building skills, and a vision of how to expand the sphere of student

success through targeted interventions with students who fail under existing school practices.

After little more than a year of SAM teams' work, we found promising evidence in support of SAM's theory of change. This assessment relies mainly on teacher survey reports of professional practices in their school. Using baseline data from Spring 2006 and follow-up data from Spring 2007, we analyze school scores on survey scales that measure SAM school leadership and professional culture outcomes (see Appendix for scale definitions). To what extent did the SAM teams' inquiry practices described earlier make a difference in the broader school culture? In particular, did school leadership focus more on student performance; did teachers increase their assessment of individual student skills and collaboration with colleagues on using data to improve instruction.

School leaders increase their focus on student performance. In each school, teachers who were not part of the SAM teams were asked to rate their school's or SLC's leadership on items measuring eight SAM standards (see Appendix for measures). Data summarized in Figure 1 show average teacher ratings for 2006 and 2007 across all SAM schools.[5]

Figure 1. Change in Teacher Ratings of School Leadership: 2006 and 2007



+: The overall scale mean in 2007 was higher than the mean in 2006 by over 1 standard deviation.

[5] Data comparing 2006 and 2007 are based on survey results for eleven of the fourteen SAM II schools, since three of the Autonomy Zone schools opted out of the follow-up survey.

A significant overall increase is shown for "focus on student performance". This measure combines teacher responses to the question of how often leaders in their school:

- Align every action with improvement of student outcomes
- Use data to identify patterns to inform decision making
- Demonstrate high expectations for all students

This finding is consistent with SAM's theory of action, which expects the inquiry team to lead school change through a focus on student performance – engaging their colleagues in implementing focused interventions to address identified skill gaps.

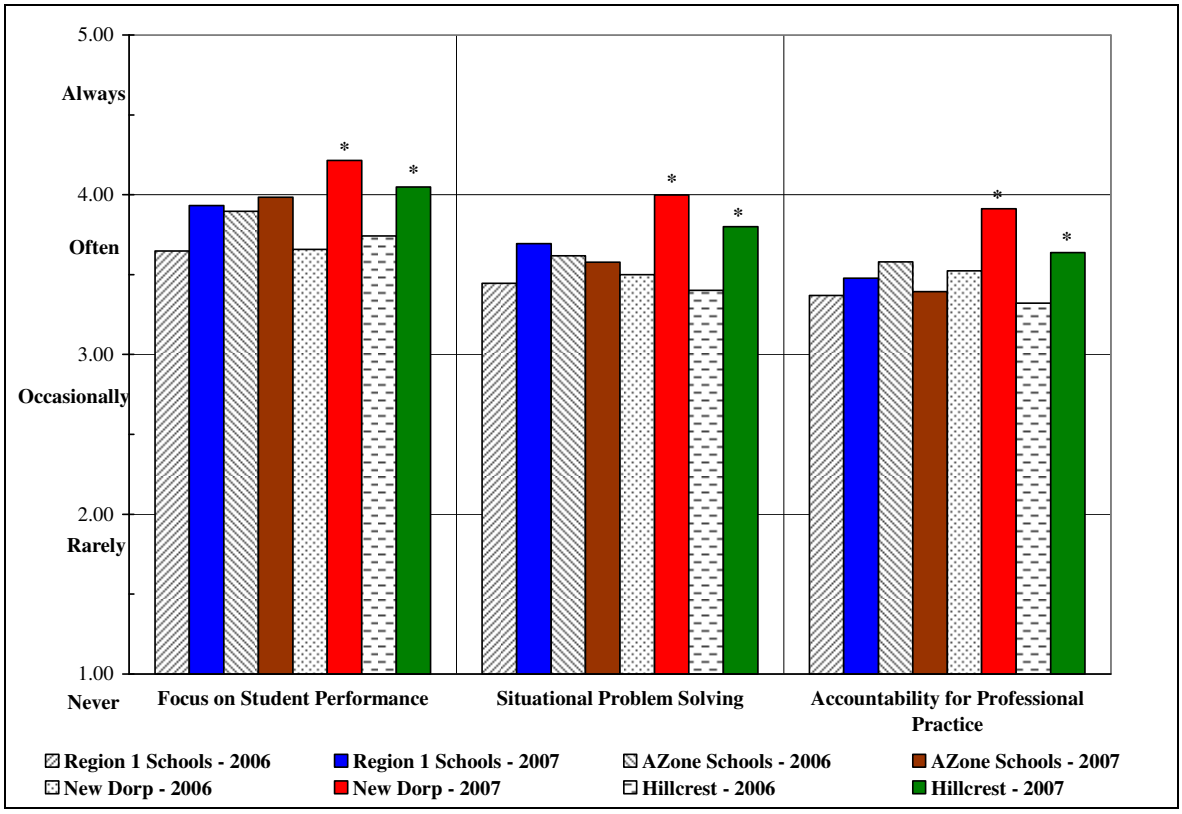
Disaggregated data show that this pattern results mainly from changes in teacher ratings in the two large high schools. The large schools also evidence change on two other SAM leadership outcomes:

- "situational problem solving" (help others solve problems, use objective evidence to identify and solve problems, use data to evaluate the effectiveness of decisions) and
- "accountability for professional practice" (work actively to support improvement of instruction, work with individual teachers effectively to improve practice.

Notably, the large schools had ratings in 2006 that were essentially the same as those for the Region 1 and AZ cohorts, but their 2007 ratings were substantially higher (see Figure 2).

Cohort differences perhaps reflect the more pervasive presence of SAM team leadership across the large high schools – eight SLC inquiry teams in New Dorp and nine SLC inquiry teams in Hillcrest – and their coherent focus on data-based improvement efforts. Further, these principals were relatively active in SAM work and leveraged school resources to support it; while principal support of SAM team leadership varied across schools in the other cohorts. Teachers outside SAM teams in the large schools clearly were more aware of the SAM inquiry teams' work and more influenced by it.

Figure 2. Change in Teacher Ratings of School Leadership by SAM Cohort



School cultures became more assessment-centered and collaborative. Teachers also were asked in the 2006 and 2007 surveys to provide ratings that describe the ways in which they relate to their colleagues in the school.

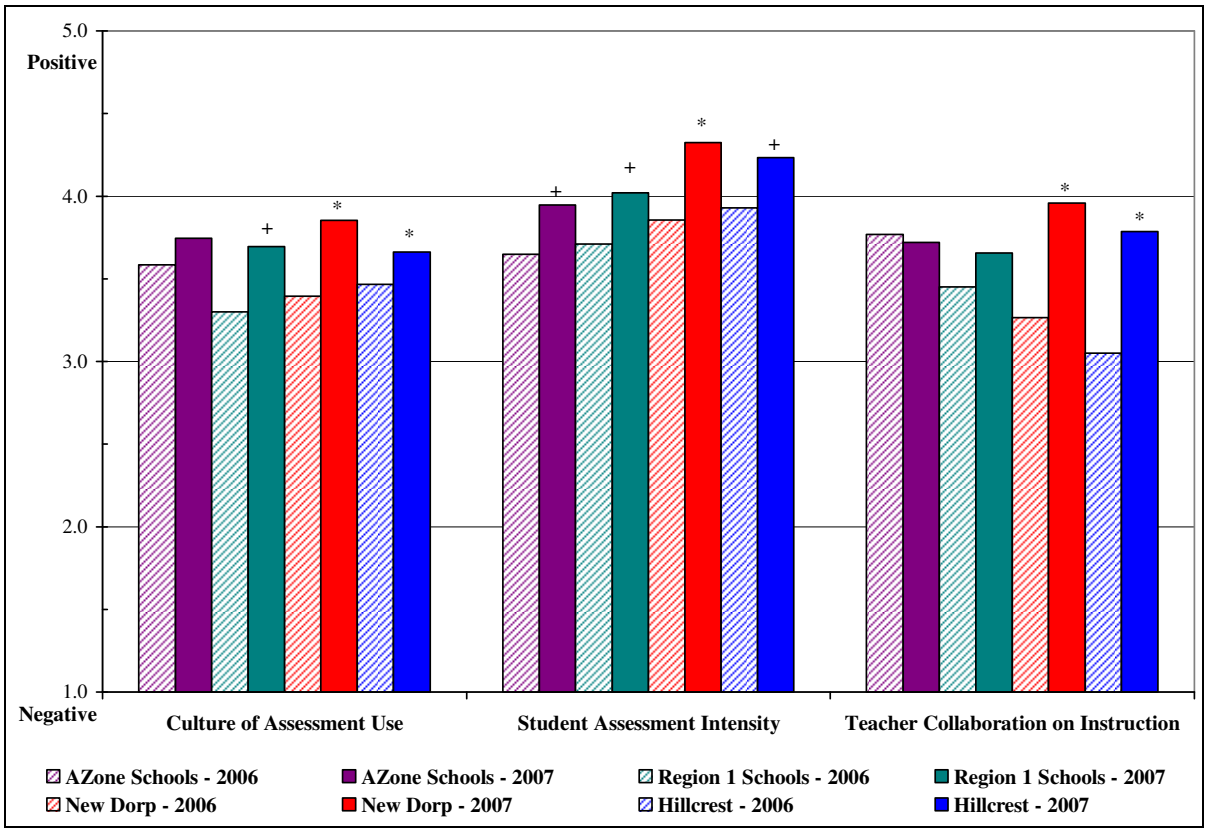
Comparisons of their responses for the two years show significant changes on measures of key professional practice outcomes promoted by SAM, specifically:

- *Culture of assessment use*:
  - We use a variety of assessment strategies to measure student progress
  - This school uses assessment data to evaluate teachers' instructional practices
- *Student assessment intensity*:
  - I use benchmarks to assess student achievement
  - I closely follow the progress of individual students performing at different levels of academic achievement
  - My lesson plans include specific instructional strategies for students who differ in their academic skills
- *Collaboration on instruction:*
  - Teachers meet regularly to review student performance in order to adjust their practices

- Teachers discuss particular lessons that were not very successful
- Staff work together to improve instruction
- I receive meaningful feedback on my performance from colleagues in this school
- I share and discuss student work with other teachers at my school

These results are summarized in Figure 3 by SAM cohort. The fact that all cohorts evidenced significant gains on the student assessment intensity measure lends support to SAM's theory of change. Inquiry teams across the school types appear to have increasingly engaged their colleagues in assessing individual student performance and gearing instruction to particular skill gaps.

Figure 3. Change in Teacher Ratings of School Professional Culture:



*: Survey scale mean in 2007 was statistically higher than that in 2006, $p < .05$ (one-tailed)
+: Survey scale mean in 2007 was over 1 standard deviation higher than that in 2006.

Trends for the other two outcomes differ across the cohorts. School culture change evidenced for the AZ school cohort is limited to teachers' assessment intensity[6],

[6] Across the four AZ schools with trend data, only one shows another significant change: an increase on "culture of assessment use."

and changes for Region 1 schools are modest and statistically significant for only one of these outcomes.[7] Consistent with data for leadership outcomes, the large high schools evidence the greatest change: each have statistically significant gains on all three school culture outcomes. Especially strong are increases for teacher collaboration on instruction, practices that were relatively uncommon in the large schools when SAM began in 2006.

Given evidence that SAM appears to be especially powerful in large schools, we further investigated professional practices by examining patterns at the SLC level in each school. Importantly, the disaggregated data showed no significant differences across the eight SLCs in New Dorp or the nine SLCs in Hillcrest. In each school, two or three SLCs appear relatively strong in their inquiry practices, but none stands out as especially weak. This finding is important because research on professional cultures in large high schools consistently finds wide variation in the strength and practices across subject departments or other units. Our data suggest that SAM teams in each of the SLCs successfully engaged their colleagues in inquiry and collaboration to improve their students' learning.

SAM outcomes on school culture stack up well against available norms. The school culture changes observed for the SAM schools after less than two years in the program are impressive against two kinds of yardsticks.

First, we compared change data for SAM schools with data for high schools that were involved in a different initiative that targeted inquiry-based school reform. This analysis used a survey measure of school culture that was replicated across the initiatives: "culture of assessment use." We compared changes made by the 11 SAM schools over one year (2006 to 2007) with changes made by 9 high schools that participated in the Bay Area School Reform Collaborative (BASRC) over three years or two years.[8] We found that 6 of 11 SAM schools (55%) had statistically significant gains on this measure over one year, compared to 4 of the 9 BASRC schools (44%) that were involved in the reform for three years. Stronger results for SAM schools in much less time suggest that the program design is considerably more powerful than BASRC, or other similar initiatives promoting evidence-based school reform, and that results after several years of SAM work are likely to be quite impressive. Further encouragement comes from evidence that BASRC schools with strong inquiry practices made significant gains in student outcomes (CRC, 2002; Talbert, Wood and Lin, 2007).

Second, we used norms from the national Schools and Staffing Patterns Survey of 2003-04, disaggregated for New York state high schools, as a yardstick against which to measure the extent of change in SAM school cultures. In order to do this we replicated a set of SASS survey items measuring school culture in the SAM teacher survey.

---

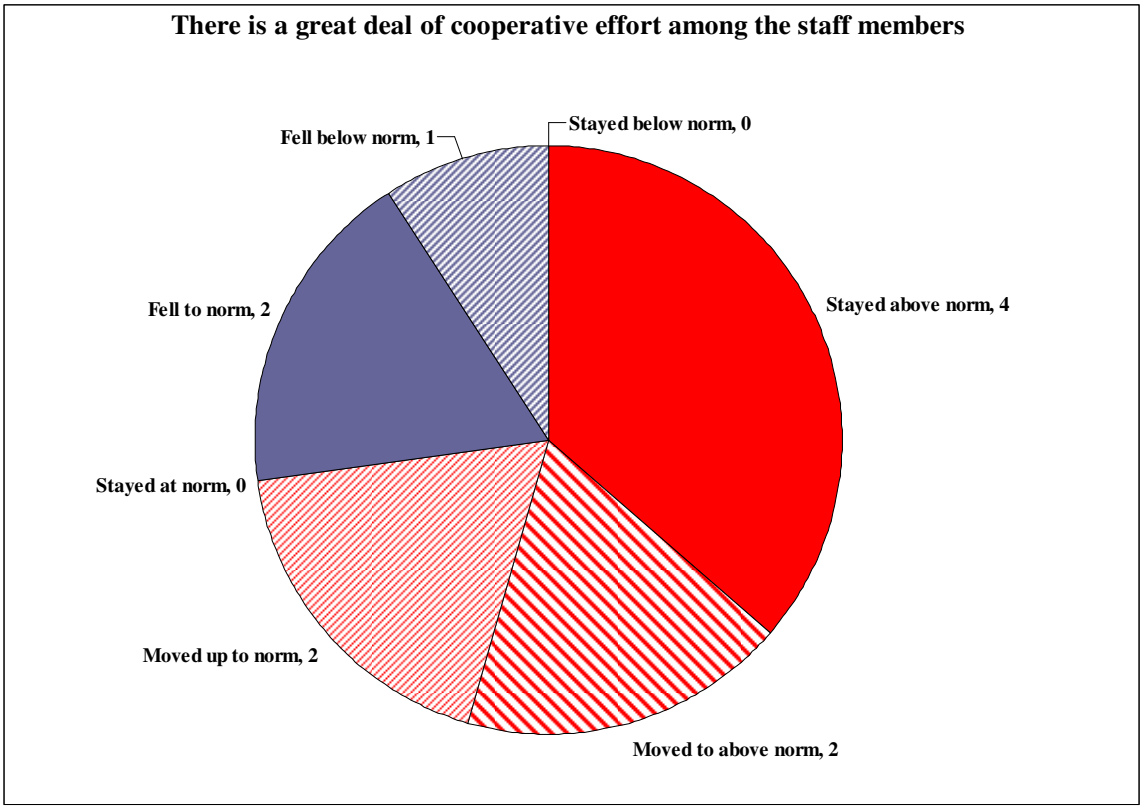
[7] However, results are uneven across the five Region 1 schools. Three schools have significant increases on "culture of assessment use," and two show increases on "collaboration on instruction."

[8] BASRC was an intermediary organization, funded by the Hewlett Foundation from 1996-2006, that provided grants, coaching support, and tools (including its Cycle of Inquiry model) to schools and districts to use data to improve student outcomes. Survey data available for BASRC high schools are for 7 schools from 1998 to 2001 (3 years) and 2 schools for 2003 to 2005 (2 years).

Unfortunately, none of the national survey items is an indicator of teachers' use of data to target instruction to student skill gaps. The closest is a measure of teacher collaboration on instruction – "cooperative effort among staff members" – which, though a weak indicator of school culture changes that SAM seeks, correlates quite highly with our survey scales Culture of Assessment Use (r=.70) and Teacher Collaboration on Instruction (r =.60).

Figure 4 shows that eight (73%) of the eleven SAM schools surveyed in both years have scores on staff cooperation that represent significant advances against New York norms. Four of the schools stayed significantly above the state high school norm, and two moved from the norm to above norm and two moved from below norm to the norm on this measure.

Figure 4. Change in School Professional Culture against Norms for New York High Schools, 2006 and 2007



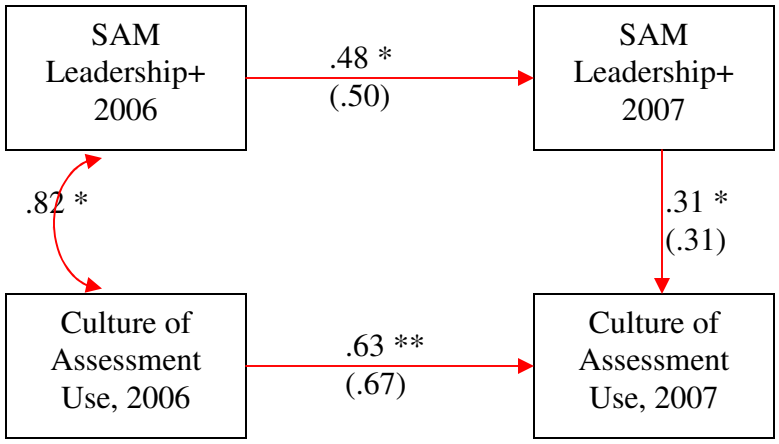
Note: For each survey year, a school's score on the survey item was determined as statistically "at", "above", or "below" the state high school average, using means and standard errors for a state-representative sample of New York high schools. The pie chart shows numbers of SAM schools that stayed the same or shifted their status against these norms between the two years. Areas shaded red indicate positive change; blue areas show negative change.

Effects of SAM leadership on school culture lend support to the theory of change. SAM's theory of change begins with the development of a school inquiry team's capacity

to lead inquiry into student performance, design interventions to address skill gaps, and hold one another accountable for results. In turn, the team's leadership across the school will engage teachers in these practices and result in shifts in professional culture that continually expand the sphere of student success.

In order to assess this logic for school reform, we asked the question: do schools that are most advanced on SAM leadership show the greatest progress on desired school culture outcomes? We analyzed variation in SAM II schools' leadership scores in 2007 in relation to change in the school's culture of assessment use between 2006 and 2007. Results of our path analysis are shown in Figure 5. We find an independent effect of SAM Leadership in 2007 on school assessment culture in 2007, after controlling for baseline school culture measures in 2006. This finding lends support to SAM's theory of change and its investment in developing a school team's leadership of inquiry-based reform.

Figure 5. Statistical Effects of SAM Leadership on School Assessment Culture



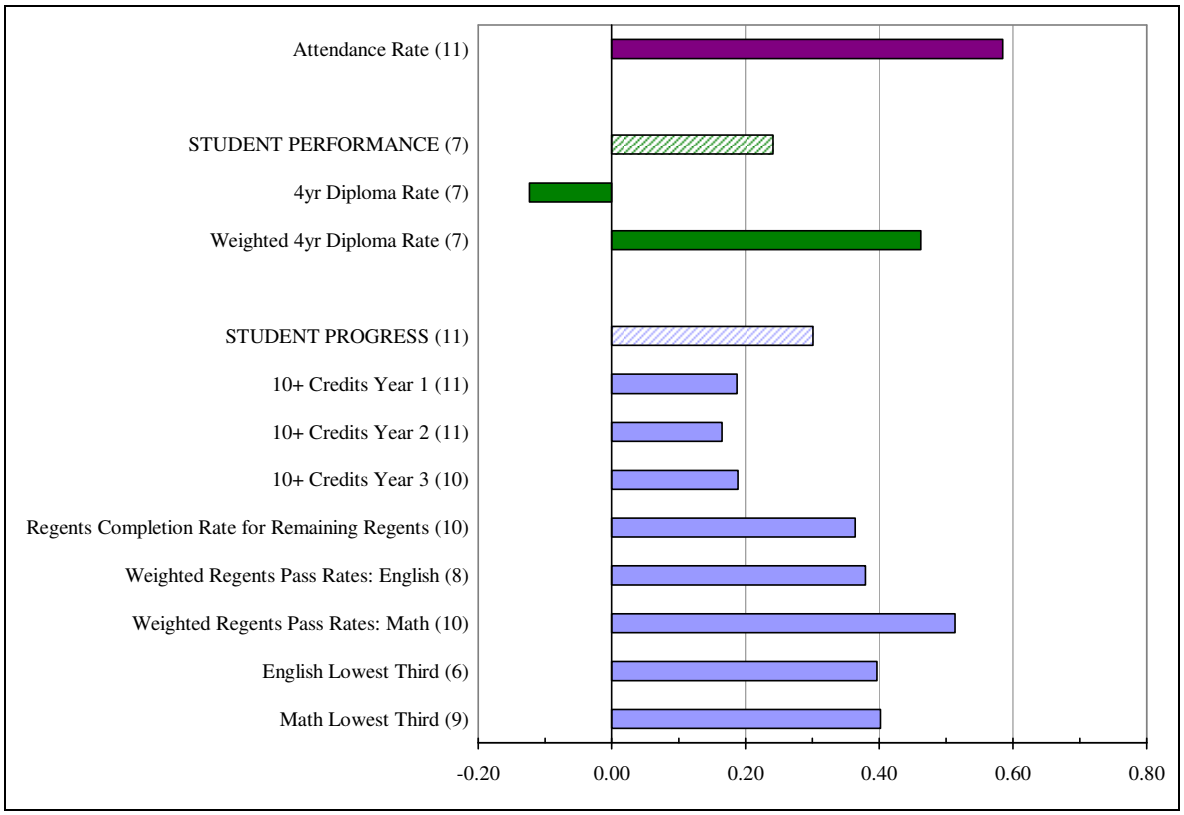
+ SAM Leadership is the mean of a school's score on three teacher survey scales: Focus on Student Performance, Situational Problem Solving, and Accountability for Professional Practice
Note: Model chi-square = 0.05, df = 2, n.s.; N = 11.
Non-standardized and standardized (in parentheses) coefficients are shown.
** p < .01;
* p < .05 (one-tailed)

School assessment culture predicts student outcomes. We further tested SAM's theory of change by examining how well desired school culture outcomes predict student performance on a range of outcomes. Lacking longitudinal data on the outcomes for all of the SAM schools, we examine correlations between school scores on our survey measure of Culture of Assessment Use and how schools were rated on student outcomes relative to their Peer Horizon in 2007.[9] We ask: do the SAM II schools that became most advanced on SAM reform practices have the strongest student outcomes?

[9] Given system changes in measures of student performance in recent years, we lack sufficient longitudinal data for assessing change in student outcomes for the SAM II schools. Future evaluation reports for these

Results in Figure 6 show the expected pattern. SAM schools where teachers have become relatively advanced in using assessment data to guide their instruction also have strong student outcomes relative to similar schools. These results validate SAM's outcomes and investment but do not provide evidence that its design and resulting changes in SAM II schools improve student outcomes. This assessment awaits longitudinal data on student outcomes for SAM II and NVPSO/SAM III schools.

Figure 6. Student Outcomes[10] in Relation to School Culture of Assessment Use: Correlations for 2007



* Notes: 1) School N used to compute each correlation is shown in parenthesis; Ns vary according to availability of student outcome data for SAM schools.
2) Student outcome measures are school scores relative to their Peer Horizon, as reported in New York City's 2006-07 High School Progress Reports.

---

schools and New Visions PSO/SAM III schools will assess changes in student outcomes across schools and as predicted by variation in their progress in the program.

[10] Summary measures included are: 1. "Student Performance" is a composite of : a) 4-year diploma (or graduation) rate, b) weighted 4-year diploma rate, c) 6-year diploma rate, and d) weighted 6-year diploma rate. (Since many SAMII schools are missing data for the 6-year diploma rates, we report scores separately only for 4-year diploma rates.) 2. "Student Progress" is a composite of: a) Credit accumulation: Percentage earning 10+ credits in first year; Percentage earning 10+ credits in second year; Percentage earning 10+ credits in third year; Average completion rate for remaining Regents b) Weighted Regents pass rates: English, Math, Science, US History, and Global History; c) Weighted Regents pass rates in school's lowest 3rd: English, Math, Science, US History, and Global History

At this juncture of the evaluation, evidence that SAM-supported changes in school leadership and professional culture are the engine for improved student outcomes depends especially upon case studies. Evaluating whether and how SAM engenders expanding spheres of student success requires process analysis situated in particular schools. Given evidence from multiple indicators that New Dorp has made major strides on SAM's leadership development and school change model, we highlight both the ways in which inquiry practices have become systemic in the school and evidence of student benefits.

Leading system change in a school: the case of New Dorp. Our case study of this large restructured high school shows how SAM works to leverage and support student outcomes within units and how they come together to create systemic change in a school (Scharff, Talbert, and DeAngelis, 2007). New Dorp's eight SLCs used data for their own students to identify target students and analyzed fine-grained assessment data to identify their learning gaps and focus interventions. Following SAM's key principle, they started small in order to get big. Over the course of a year, each team created one or more interventions that developed their target students' skills and performance on assessments. And they engaged their colleagues in the interventions and in looking at data to assess the students' skills and evaluate interventions.

Through their work with student data and interventions and their low-inference observations of classroom instruction, the teams became focused on individual student learning needs. This shift was guided by their SAM facilitator who kept them on task, guided their creation and use of data, pushed them to question their assumptions, and supported their struggles with change and leadership in the school. Cross-team seminars created a place where common struggles could be discussed and teams could learn from one another, and they created the space for stepping back from practice and learning from the literature that grounds SAM.

> One team member captured the shift in these terms:
> When we shift our thinking from 'how we teach' to 'how the students learn' it's not just the teacher teaching the whole, but when you look at the individual students who are in front of you and how *they* learn, it's really created this entire shift in the way as a school community we're approaching what we're doing. I can say very honestly that this is [now] something that the entire school community is looking at. Guidance counselors, non-instructional assistant principals, AP organization, AP guidance, we *all* are part of the process in really looking at the way our kids are learning.

The systemic changes described here are outcomes of a strategically guided change process that was co-lead by New Dorp's principal and the SAM facilitator. The facilitator brought deep understanding of the program principles, guided and supported the work of eight SLC teams, used SAM readings and assignments to scaffold team leadership development, helped the teams become skilled in using a wide range of data,

pushed them to have hard conversations, and celebrated their successes. Critically, the work was designed with the principal, and adjustments were strategic in response to problems that arose and assessments of opportunities to move the work in particular teams and across them in seminars.

The SLCs that moved forward used data to identify student learning gaps and design interventions, examined their practices in ways that changed the teaching culture, and developing shared accountability for expanding the sphere of student success. Their focus on moving students in one small skill lead them to identify concrete skill gaps, confront curricular gaps relative to student need, intervene to provide students' instruction targeted to need. They saw that understanding students' specific needs would allow them to accelerate student learning.

As the SLC team inquiry skills and leadership developed, the teams involved their colleagues in a new phase of school reform. At the end of two years in SAM, the New Dorp participants decided that they wanted to bring the program school wide, that they would teach a streamlined version of SAM in their SLC-based common time the following year. Currently each teacher in the school participates in an inquiry team, identifying and moving target students in target skills. Teams will share the results of this years' work on a professional development day in June.

New Dorp made substantial improvements on a range of student outcomes beyond those of the target group between end of years 2006 and 2007. The school saw changes on the following student outcomes:

- Advance in 9th graders' credit accumulation for promotion from 57% to 71%
- Improved scores on Regents exams in most subjects: percent scoring 65-100 increasing 10% for ELA, 4% Global Studies, 2% US History, 12% Math A, 12% Math B, 27% Physics, 4% Living Environment. (Declines were shown for Chemistry (-11%) and Earth Sciences (-15%).)
- Improved course pass rates in academic subjects: English (6%), Social Studies (5%), Mathematics (4%), Science (33%), Foreign Language (11%).

They see sustained trajectories of student improvement in the current year and expect even greater gains on these key institutional measures of student outcomes.

School leaders, SAM facilitator, and evaluator regard these signs of an expanding sphere of student success as outcomes of their SAM involvement and the school leadership for inquiry-based change that it has grown. Coupled with the school's restructuring effort, SAM has had a powerful impact on teaching and learning in New Dorp.

How the SAM program develops leadership for inquiry-based reform

School inquiry teams participating in SAM II were positive about their experiences in the program, as evidenced in their survey ratings and comments during team focus groups and interviews. Those that participated in the credentialing program run through Baruch College (all of the fourteen schools, except New Dorp) were

especially positive that the course work focused on their own school's improvement efforts and involved school teams.

Nonetheless, the teams gave varying ratings to particular elements of the SAM model and benefited unequally from the program, as documented earlier. Given the strong progress on SAM outcomes noted for the two large high schools, it will not come as a surprise that they gave the most positive ratings of all program elements. They appear to have been better able to take advantage of the SAM curriculum and resources to develop leadership and school cultures focused on using data to improve teaching and learning. In these schools, SAM teams from 8 and 9 SLCs worked to assess and improve their own students' skills, came together for cross-team seminars on site, and had a facilitator dedicated to the school. In contrast, the five Region 1 schools and seven Autonomy Zone schools had one SAM team in each school, shared a facilitator who worked in the school every 2-4 weeks, and came together at an off-site location for seminars. These design differences may well account for differences in the teams' experiences with facets of the program and their schools' benefits from SAM.

Credentials offer incentives and "gravitas." The link between school reform and a leadership credentialing program is a distinctive feature of SAM II, and –with the notable exception of New Dorp's stunning success without it –the opportunity for teachers to obtain credits and credentials through SAM appears to have been strength on the whole.

In some instances, team members said that they never would have taken on the considerable extra work and responsibilities had it not been for the opportunity to earn an administrator credential. Others said that it made no difference at all: "Personally, for me, I have at this point zero desire to be an administrator. I did it because I thought the work would be beneficial to the school –and that, in the long run…it *may* end up being something I want to pursue. And, obviously, not having to pay for it if in ten years I want to be an administrator [is a plus]…" Although we lack data to address the question of how important the credential incentive was in attracting team members, we do have evidence that large portions of them likely or definitely will pursue administrative positions in the future. In our 2007 SAM team survey, over half (53%) said that they will pursue a position of assistant principal and 42% aspire to a principalship (not mutually exclusive statistics). Interestingly, 60% indicated that they would prefer a leadership position in their own school. In fact, several SAM II graduates are currently in positions of principal or assistant principal in their school, indicating that the program contributed to both the NYC administrator "pipeline" and sustainable leadership of SAM school reform work.

Other SAM team members pointed to the way in which participation in a credentialing program sustained their commitment to the hard work and professional challenges SAM entails. When asked in a focus group about the value, or not, of SAM's credentialing role, one team member commented:

> I think that the administrative credential part of it has really sort of leant it gravitas and created a lot more buy-in for us in terms of taking an insane amount of work and dedicating a *lot* of time in a way that an administrator does to a school, and sort of an ownership for the easy the school is running in the way an administrator does. And I think that while teachers are definitely invested in developing that, in developing the school and helping kids, I think that one of the most valuable things about this program has been its leadership development for me. And I'm not sure what the program would look like without that aspect. And I'm not sure if it would have moved me forward as much if it hadn't had that [credentialing] aspect… While there have been frustrating moments, some of the coursework has been absolutely invaluable in the way that we were thinking about our school, the systems work this summer [especially], and really mapping out the way our school is run… It has helped me think about school differently. And I don't think I would have been able to without the coursework.

Another team member who teaches 6th grade said:

> It's about motivating myself to stay up until whatever hour, or it's giving up the last prep that I had in 2006 to do this work… And particularly looking schoolwide, it would be difficult to image teachers really working with … I meant the 10th graders [the team's target students] are so far from the kids that *I* teach. And its so hard for me to imagine a teacher in my position giving as much time and thinking as widely across the school if they weren't thinking of themselves as a possible leader in the future of a school.

Despite the benefits of a credentialing version of SAM's inquiry-based school reform model, it also carries considerable challenges to program leaders and facilitators. Three that emerged during the course of SAM II center on the unorthodox nature of the course work and instructors, quality standards for the individual and group assignments, and tensions between facilitators' roles of evaluating and supporting participants' work. In one instance, a team complained that as students of Baruch College they should be able to have the benefit of multiple instructors who specialize in program topics, criticizing SAM for its departure from a traditional credentialing program. Grading standards surfaced in instances where a team was rated lower than other teams, and facilitators were challenged to come up with common standards and to take account of the wide variety of SAM work in assigning grades that participants deemed fair.[11] Further, facilitators walked a tight rope created by their dual roles of evaluating individuals' course performance and coaching the team in their school improvement

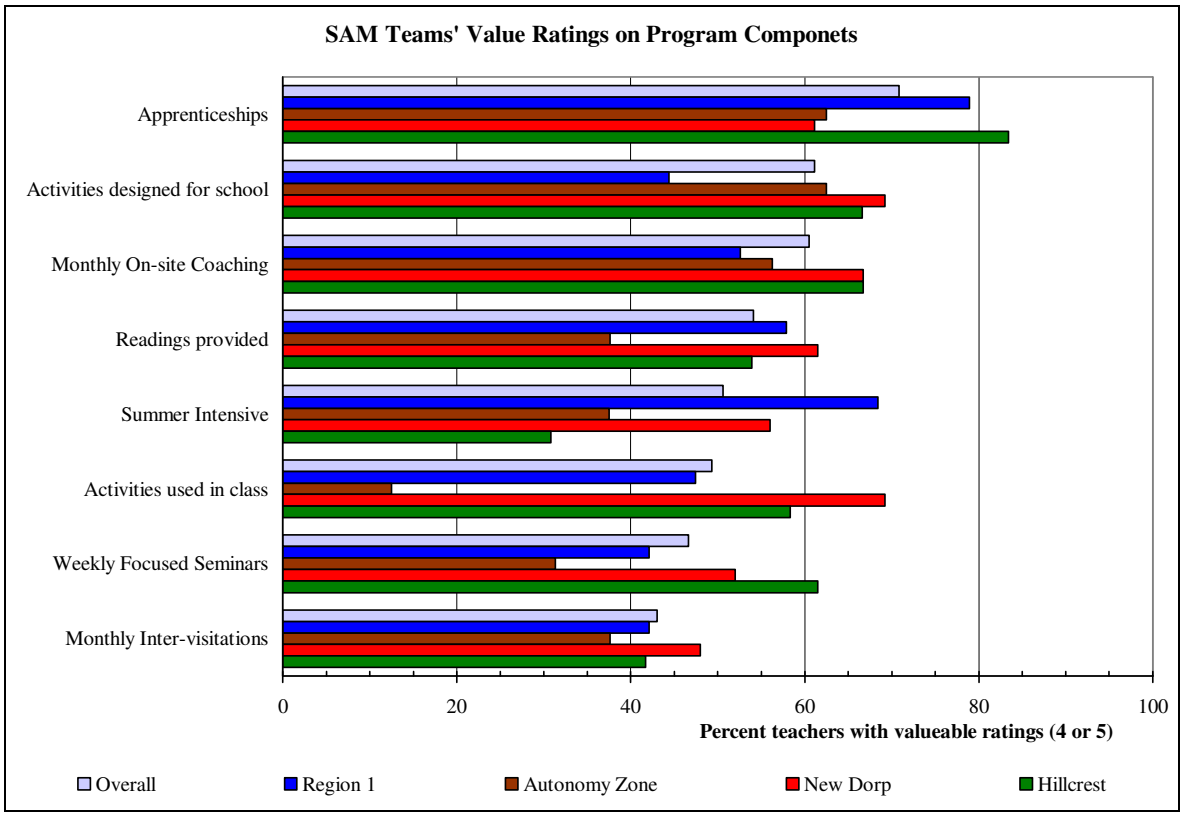
[11] The current SAM-cert program is using norming products to address "quality control" and set standards for credit. The facilitators are bringing participant work and determining together what "meeting standard" looks like. They are developing rubrics for each product based on mock versions of the assignments that facilitators create first, with pretend data, to work out any problems/confusions in the assignment.

efforts. Although not incompatible roles, managing the tension requires considerable facilitator skill and a great deal of clarity on the program side.

SAM program activities that happen at the school are most valued. SAM teams gave quite positive ratings to most program components. Differences in ratings – between components and between SAM cohorts on the same component – suggest that teams value most the program supports that happen on their site.

Figure 7 shows SAM program components ordered from highest to lowest overall ratings. The bars show how each cohort rated a particular program facet. Two patterns show up. First, components rated highest are those that happen on the school site – apprenticeships (tasks guided by school administrators), activities designed for school teams, and facilitator coaching. Second, two of the lowest rated components – in-class activities and weekly seminars – receive high ratings from the two large high schools, where the work happens on site. Ratings on these program features are particularly low for the Autonomy Zone schools, reflecting several factors – the large distance many of the teams traveled to meet at New Visions, the challenge facilitators faced in grounding the seminars in particulars of the work of seven different schools, and turnover of facilitators between school years.

Figure 7. School Teams' value ratings of SAM program components



Although there are efficiencies in bringing schools together for seminars, costs are that large numbers of teams have trouble making connections between the class work and their school reform work, and facilitators are stretched to design ways for that to happen. Our data do not suggest an ideal number of schools to involve in cross-school seminars, but seven appears to be too many; SAM I involved four schools and did not encounter the challenges documented for the AZ cohort in SAM II.

The teams' quite low ratings of SAM's monthly inter-visitation component reflects their experience of uneven quality of the visits, in terms of sufficient scaffolding and norms for giving and receiving feedback, and their sense that the visits were not customized to their work. Some felt that this activity took away valuable time from the team's work in its school. Based on experience, facilitators refined the design for inter-visitations over the course of the year, and ratings most likely would have been higher had the survey been conducted later. However, it may be that this program component is least well suited to supporting the teams' situated, inquiry-based leadership development and school reform. The data raise questions of what kind of cross-school learning opportunities advance SAM inquiry work, e.g., to observe data briefings or targeted interventions, and under what conditions and strategic timing.

Program elements work in synergy. When SAM team members were asked in the survey how well the program components work together to support their ability to lead school reform, just over half (56%) indicated that they work "well" or "extremely well" together. Yet three-fourths (77%) of New Dorp's team members reported that the work was synergistic. What made the difference in New Dorp? Part of the answer is captured by our earlier description of how the teams and facilitator worked around data from student assessments and classroom observations to examine assumptions about teaching and learning. Beyond the daily, micro processes of facilitated change in teachers' thinking about students and teaching, what organizational conditions mattered?

Notably, New Dorp is a large high school and the only SAM II school that did not participate in the credentialing facet of the program – two conditions that some might argue would inhibit its progress. In fact, our data indicate that both of the large SAM schools made greater strides on leadership development and culture change than small schools. There are several possible explanations for this pattern. One is that multiple teams on one campus create a critical mass of reform leadership and mutually reinforcing efforts across the SLCs. Another is that the simultaneous creation of a new organization structure in both large high schools presented both need for new leadership and a press and opportunity for teachers to become more student-focused.[12] A third account is that each of the large schools had a dedicated SAM facilitator, who therefore spent more time on campus than the facilitators who were spread across several schools and were in a position to work more closely and strategically with the principal to push the SAM work. Evidence at hand lends support to all three hypotheses.

[12] New Dorp's principal sees SAM and restructuring as interlocking forces in the school's rapid change. The partnership was strategic in that she managed the structural change while the SAM facilitator lead the SLC teams' inquiry into student achievement. It was a powerful partnering strategy.

That New Dorp stands out as a site where facets of SAM worked synergistically during a relatively short period of time, when it lacked incentives and authority brought by the Baruch credentialing program, begs further explanation. Our observations, interviews, and focus groups suggest that the synergy of SAM work was created through an exceptionally effective partnership between the principal (and other administrators) and the SAM facilitator. This partnership developed through mutual, evolving understanding of SAM and shared commitment to using them consistently as the vehicle for improving student achievement across the school. Partners collaborated in adapting the SAM program to the school and in supporting all SLC inquiry teams' involvement and leadership. After becoming deeply engaged and skilled in using data and low-inference observations to move their students, the first SAM cohort of SLC leaders proposed spreading the work to a new SAM cohort in the school, and ultimately to the whole school. The principal and facilitator backed this bottom-up initiative and continue to collaborate strategically to help deepen and spread the work across the school.

New Dorp's effort this year to broaden SAM to include all teachers has encountered resistance in the faculty. Although teachers respect the SLC team leaders and have bought into interventions they designed on the basis of data, many resist the idea that they should be responsible for doing the research. Some point to the current mandate from the DOE to create inquiry teams and use standardized tests to judge schools as bureaucratic strategies that undermine their best work. They jump into a compliance mentality. Such reactions in a school that has made greatest strides on using SAM to build teacher leadership and expand student success raise questions for the future. How can SAM's vision and practice for broadening teacher leadership for inquiry-based reform be sustained by New Visions and its schools in the context of bureaucratic mandates for inquiry?

Lessons and Research Issues

Lessons from this study can help guide the implementation of SAM's core ideas in NYC schools being asked to use student data to improve achievement and inform initiatives for evidence-based reform across the nation. Certainly, mandating inquiry teams and making data available will not be enough to transform leadership and school culture in the ways that SAM envisions. Absent significant investments in developing school team leadership and external facilitator's capacity to support this movement, the gains in student achievement that we documented for New Dorp and several other SAM II schools will not be realized.

Lessons and issues that we draw from experiences of the SAM II schools center on: synergy of conditions for change, adapting strategies to diverse school cultures, and navigating system conditions that can inhibit change.

School change depends upon mutually reinforcing system elements. When SAM worked best, the change process was:

- spearheaded by school administrator(s) committed to data-based reform;

- skillfully guided by an external facilitator who brought tools, organization savvy, and teaching skills to shepherd data use for change; who enforced the SAM principle of staying small to get big;
- lead by school leaders who represented key school units and expertise;
- scaffolded by readings and seminars that provided school leaders with new cognitive frames and knowledge essential to leading change;
- legitimized and supported by external partners.

Evidence from SAM suggests that these conditions worked synergistically to promote change. In the school most successful in using data to improve student achievement, all of the elements interacted at a high level of quality. Some were established by the SAM program design, but the success of coaching and seminars and assignments depended upon the SAM facilitator's skilled enactment of them. Other elements depended on school administrator leadership, including the principal's strategic moves to select and nurture leaders and learning communities across school units. Where and when these conditions for change came together, teacher leaders became engaged in data-based improvement efforts and in leading their colleagues to accelerate the learning of target students.

The fact that many SAM schools did not experience synergy of program and leadership conditions points to a fundamental issue for practice and research: how are these multiple capacities developed in tandem? Since the facilitator is a primary change agent in SAM: how do individuals in this role develop deep understanding of SAM's conceptual underpinnings and principles to anchor their judgments and actions in working with schools? What investments in facilitator training are necessary, at a minimum, to meet this goal? Despite SAM II's substantial training investment, most individuals in this role said they had a steep learning curve and "got" the core principles only after they had been working with their schools for some time. They pointed to weekly facilitator sessions as a setting in which their learning took shape around shared readings and sharing of particular struggles and successes with their schools.

SAM's theory of change is notable in taking on all of these challenges simultaneously. It is deeply grounded in learning theory and research on system change, and through its current instantiation in New Visions' PSO will help to advance understandings of principles and practices for data-based school improvement.

Schools' origins and reform pathways condition their readiness for data use. In SAM and in other initiatives that promote evidence-based practice in schools, documentation research finds that schools differ in their initial capacities to respond to designs for using data to improve teaching and learning. Some have prior experiences to build upon, while others' histories may pit them against this theory of change. For example, schools involved in the Bay Area School Reform Collaborative – a ten-year initiative to promote and support inquiry-based school reform – made significantly greater progress if they had a reform history of inquiry into student learning (Talbert, Wood, and Lin, 2007). Teachers in such schools were used to looking at student work and using formative assessment data to focus instruction on student learning needs.

Further, schools where strong professional communities had developed were best able to take advantage of resources for improvement and bring them into their practice, since teachers were comfortable sharing their students' data and opening their practice to colleagues.

SAM's limited success with Autonomy Zone schools suggests a revision of conclusions from prior research. Strong nontraditional teaching cultures are not always conducive to reform. It appears that, in some instances, they resist the use of data as a strategy to improve student achievement. The AZ schools had developed a student-centered school culture that prided itself on addressing the whole child; they did not readily embrace SAM's model of using fine-grained student data to design instructional interventions. These established school cultures pushed back on the SAM model and, for the most part, made little progress in improving student outcomes beyond their target students.

This finding lends support to a view that a school's reform pathway frames problems and opportunities for change. What are the entry points for change in each school? How can the facilitator build upon a schools' prior experience with using data on student performance to leverage change? This challenge frames several practical and research questions. Is this easier or more difficult when partners come from the same reform history? What developmental process moves a school culture from its origins – as traditional teaching-centered cultures or reformed student-centered models – to one that focuses on strategic interventions to expand the sphere of student success? How do the routes differ for different starting places? The literature on partnering for change is scant, and documentation research on change efforts will be key to building it.

Schools are challenged to navigate system mandates. In New York City and elsewhere, school systems have embraced the vision and taken up the challenge of evidence-based school improvement. Driven by federal policy under No Child Left Behind (NCLB), local systems are pressed to use student achievement data to assess their progress. Districts across the country are now mandating professional learning communities (PLCs) and data-driven decision making as vehicles for improving student achievement. On one hand, the resources that systems devote to these efforts can improve school's capacity for change. On the other hand, educators' resistance to the top-down initiatives can squelch change.

Schools undertaking SAM and related initiatives to develop teacher leadership for inquiry-based reform are challenge to navigate the bureaucratic frames and resources that system initiatives bring (Talbert, in press). Current NYC education policy creates conditions that can undermine or support school progress on SAM. For one, teachers may take issue with central office directives for inquiry teams to lead change and resist the model. This presents the challenge for facilitators to make the work authentic and co-lead change with principals. On the other hand, centralized student assessments and data summaries have never been better. Yet, the data are at times a grain size that SAM teams cannot act on, which creates a need for school teams to produce data and to press their systems to create finer breakdowns of their students' performance. School leaders and

facilitators are challenged to create the capacity and culture for evidence-based improvement, and to keep the focus on the purpose of bureaucratic mandates (improved student outcomes).

References

Archer, J. (2005). Guiding hand. *Education Week*, September 14: S5-S10.

Bridges, E. M. (1992). *Problem Based Learning for Administrators*. Eugene: ERIC Clearinghouse.

Center for Research on the Context of Teaching, Stanford University. (2002). *Bay Area School Reform Collaborative: Phase one (1996-2001) evaluation.* Stanford, CA: Author.

Coburn, C. and J. Talbert (2006). Conceptions of data use in school districts: Mapping the terrain. *American Journal of Education*, 112: 469-495.

Cuban, L. (2001). *How Can I Fix It? Finding Solutions and Managing Dilemmas*. New York: Teacher's College Press.

DuFour, R. (2002). The learning Centered Principal, *Educational Leadership*, January 1: 59(8), 12.

Elmore, R. F. (2004). *School Reform from the Inside Out: Policy, Practice, and Performance*. Cambridge, MA: Harvard Education Press.

Heifetz, R. A. and Linsky, M. (2002). *Leadership on the Line: Staying Alive through the Dangers of Leading*. Boston: Harvard Business School Press.

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

Lieberman, A. and Wood, D. R. (2003). *Inside the National Writing Project: Connecting Network Learning and Classroom Teaching.* New York: Teachers College Press.

McLaughlin, M. and D. Mitra (2003). The Cycle of Inquiry as the Engine of School Reform: Lessons from the Bay Area School Reform Collaborative. Stanford University: Center for Research on the Context of Teaching.

McLaughlin and Talbert. (2007). Building professional learning communities in high schools: Challenges and promising practices. In Stoll, L. and Seashore Louis, K. (Eds), *Professional Learning Communities: Divergence, Depth and Dilemmas.* Berkshire, England: Open University Press.

McLaughlin, M. W. & Talbert, J. E. (2006). *Building School-based Teacher Learning Communities: Professional Strategies to Improve Student Achievement.* New York: Teachers College Press.

No Child Left Behind (2002). Public Law 107-110.

Popham, J. W. (2003). *Test Better, Teach Better: The Instructional Role of Assessment*. Alexandria, Virginia: ASCD.

Scharff, H., Talbert, J., and DeAngelis, D. (2007). Staying small to make a big difference: A successful strategy for large high school improvement. New York: Baruch College.

Senge, P., Cambron, N., Lucas, T. et al (2000). *Schools that Learn: A Fifth Discipline Fieldbook for Educators, Parents, and Everyone Who Cares About Education*. New York: Doubleday.

Supovitz, J. and V. Klein (2003). *Mapping a course for improved learning: How Innovative Schools Systematically Use Student Performance Data to Guide Improvement.* Philadelphia: Consortium for Policy Research in Education.

Stein, S. and Gewirtzman, L. (2003). *Principal Training on the Ground*. Portsmouth, NH: Heinemann.

Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Doubleday.

Talbert, J. E. (2002). Professionalism and politics in high school teaching reform. *The Journal of Educational Change.* 3: 339-363.

Talbert, J. E. (in press). Professional learning communities at the crossroads: How systems hinder or engender change. In Lieberman, A. (Ed.). *International Handbook of Educational Change.* Dordrecht, The Netherlands: Springer.

Talbert, J. E., Wood, A. and Lin, W. (2007). Evaluation of BASRC Phase II: Evidence-based system reform: Outcomes, challenges, promising practices. Stanford, CA: Center for Research on the Context of Teaching.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity.* Cambridge, England: Cambridge University Press.

APPENDIX
Teacher Survey Scale Definitions

Survey scales combine teacher responses to a number of related survey items in order to obtain reliable measures of key research variables. These scales were developed with data from the teacher survey administered in 2006 in 14 SAM schools (teacher respondent N = 570). Principal components analysis was used to identify survey items that loaded on a common factor. Alpha coefficients were computed to determine the internal consistency of the scales, or how highly correlated the items are with one another. Here we report Alpha coefficients obtained for 2006 survey data and for 2007 follow-up survey data (teacher N=440). Scale scores analyzed in the paper are the unweighted mean of an individual's rating for each item in the scale, with school scores computed as the mean of individuals'scores.

SCHOOL LEADERSHIP

These scales measure **leadership conditions in the school or SLC** that are aligned to SAM Leadership Standards.

> *Now consider **leadership in your school or SLC**. Please indicate the extent to which leader(s) do each of the following.* **School/SLC leaders …**
>
> [5-point frequency scale, ranging from 1 ("Never"), 2 ("Rarely"), 3 ("Occasionally"), 4 ("Often") to 5 ("Always")]

Positive Personal Behavior **(2 items. Alphas = .71 and .79)**

	2006	**2007**
Actively seek and make use of diverse and controversial views	6a	6a
Are willing to admit and learn from mistakes	6d	6c

Communication Strength **(1 item)**

	2006	**2007**
Negotiate successfully between opposing points of view	6c	6b

Focus on Student Performance **(2 items. Alphas = .67 and .76)**

	2006	**2007**
Demonstrate high expectations for all students	6e	6d
Use data to identify patterns to inform decision making	6g	6e

Situational Problem Solving **(4 items. Alphas = .82 and .91)**

	2006	**2007**
Help others to solve their problems themselves	6h	6f
Use objective evidence to identify, frame and solve problems	6i	6g
Use data to evaluate the effectiveness of decisions	6j	6h
Create structures for dealing with recurring problems	16b	6i

Learning Stance **(2 items. Alphas = .53 and .78)**

	2006	**2007**
Actively pursue their own learning	6k	6j
Communicate clear expectations that everyone in the community is responsible for the learning of their colleagues	16h	6k

Accountability for Professional Practice **(3 items. Alphas = .77 and .91)**

	2006	**2007**
Work with individual teachers effectively to improve practice	6m	6m
Differentiate professional learning opportunities based upon identified patterns in faculty proficiencies and needs	16k	6n
Create system of on-going feedback and evaluation to improve practice	16l	6o

Leadership Capacity **(2 items. Alphas = .60 and .83)**

	2006	**2007**
Cultivate a shared vision and common purpose among staff	6n	6p
Demonstrate the capacity to delegate and trust others with real leadership tasks	16p	6q

Team Work **(1 item)**

	2006	**2007**
Take responsibility for others' learning	6l	6l

SCHOOL PROFESSIONAL CULTURE

Culture of Assessment Use (2 items. Alphas = .58 and .81)

5-point Likert scale, ranging from 1 ("Strongly disagree") to 5 ("Strongly agree")

*How well does each of these statements describe **teachers' practices** in your school?*

	2006	2007
This school uses assessment data to evaluate teachers' instructional practices	3b	3d
We use a variety of assessment strategies to measure student progress	3e	3c

Student Assessment Intensity (3 items. Alphas = .76 and .79)

5-point Likert scale, ranging from 1 ("Strongly disagree") to 5 ("Strongly agree")

*Now indicate the extent to which you agree or disagree that each of the following statements describes **your instructional practices**.*

	2006	2007
I use benchmarks to assess student achievement	4a	4b
I closely follow the progress of individual students performing at different levels of academic achievement	4b	4c
My lesson plans include specific instructional strategies for students who differ in their academic skills	4c	4d

Collaboration on Instruction (5 items. Alphas = .85 and .88)

5-point Likert scale, ranging from 1 ("Strongly disagree") to 5 ("Strongly agree")

*How well does each of these statements describe how **teachers work together in your school or SLC?***

	2006	2007
Teachers meet regularly to review student performance in order to adjust their practices	3c	3b
Teachers discuss particular lessons that were not very successful	3d	3e
Staff work together to improve instruction	3f	3f
I receive meaningful feedback on my performance from colleagues in this school	3g	2e
I share and discuss student work with other teachers at my school	3h	3a