Charting a Course:
Evaluation Design of the National School District and Network Grants Program

Becky Smerdon
Mike Garet
Jennifer O'Day
Luke Miller
American Institutes for Research

Barbara Means
Karen Mitchell
SRI International

Charting a Course: Evaluation Design of the National School District and Network Grants Program

Over the past decade, a number of reform efforts have aimed at reducing the size of the learning communities in the nation's schools. Most recently, the Bill & Melinda Gates Foundation has committed more than $375 million to improving America's high schools through its National School District and Network Grants Program. Likely to be one of "the most publicly scrutinized educational initiatives" in recent history, the foundation's program is providing a catalyst to educators and policy makers to radically reshape secondary education in the United States. The foundation is working with school districts and network organizations to create education systems that work for all students, including those students who have been characterized in the past as most underserved by the nation's education system. At the secondary level, the foundation is supporting small schools as a corrective to the large, impersonal "shopping mall" high schools that dominate today. To this end, the foundation supports: 1) districts and network organizations in converting large high schools into smaller learning communities, and 2) network organizations' efforts to replicate successful model schools or to implement small, effective schools from the ground up.

Through a competitive award process, the foundation selected our organizations—the American Institutes for Research (AIR) and SRI International—to evaluate its initiative. We are currently beginning our third year of data collection. The purpose of this paper is to describe the framework underlying our evaluation design of a large-scale, multifaceted school reform initiative and to identify some of the methodological challenges inherent in work of this kind. Throughout this paper, we use travel as a metaphor for our evaluation design process. Specifically, we begin by describing the earliest stages of our journey—our starting point. Here we outline the foundation's theory of change—their goals and vision for the initiative. We then turn to a discussion of defining the destination—the research questions we seek to answer—and finding the coordinates—the conceptual framework that will guide our evaluation. Next, we explore the terrain, providing an overview of the grantees and schools in the population, and determine our route (e.g., the design basics of our evaluation). We conclude with a discussion section that we have entitled, Reading the Map While Driving. Here we discuss a few of the challenges of our evaluation and our need for reflection and refinement as the initiative grows and changes and as we learn through our research activities.

Beginning the Journey

We began our evaluation journey by explicating the foundation's theory of change for their reform initiative. Specifically, we sought to understand the goals of the foundation and the organizations they are funding, the strategies they are using to change schooling, and their perspectives on the processes through which their activities will have an impact. Using this approach allowed us to design an evaluation that explicitly tests the fundamental operating assumptions of the program and its primary intended outcomes. That is, this approach provided *a priori* hypotheses about how components of the intervention are to be implemented, and once implemented, how they are expected to produce better outcomes. The theory of change we developed for the foundation initiative guided every aspect of our evaluation design from research questions and instrumentation to analysis and reporting plans.

The Foundation's Theory of Change

The theory of change for the foundation's National School District and Network Grants Program presented here briefly illustrates our understanding of how foundation officials expect the grant program to work. (For a detailed discussion of the foundation's theory of change, see Shear &

Smerdon, 2002.)   The theory of change depicts the goals of the program; strategies for grantee selection, funding, and support; and assumptions suggesting why particular approaches to school change will be successful as described by foundation staff.[1]

Exhibit 1. The Foundation's Theory of Change
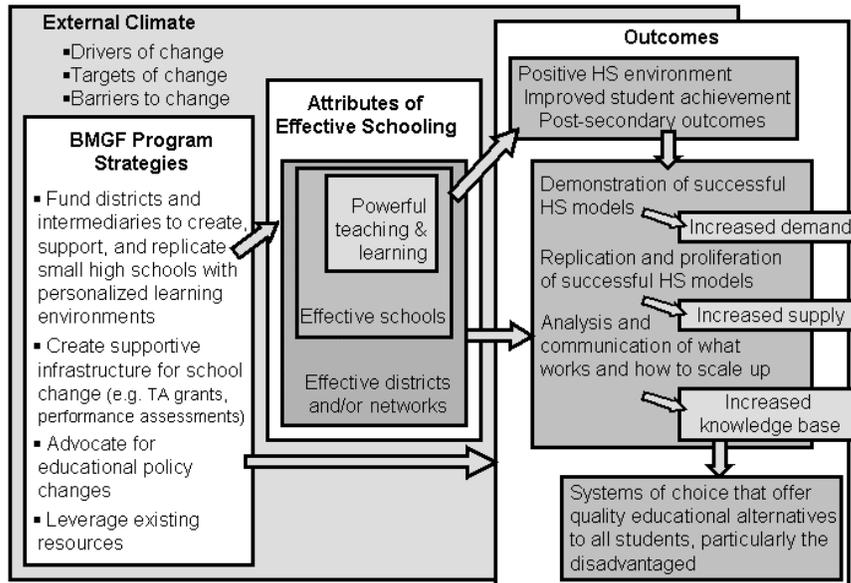
The overall goals and strategies of the National School District and Network Grant Program, as depicted in the schematic diagram above, are as follows:

Drivers of, targets of, and barriers to change. The foundation's efforts are driven by concerns over the current ineffectiveness of large comprehensive high schools, particularly in urban areas. These schools tend to be characterized by a lack of personalism and low expectations; they therefore fail to serve many of their students, who are left without access to the effective educational opportunities that all students deserve.

Foundation program strategies. In this grant program, the foundation is funding initiatives that share the goal of creating small and/or restructured schools, as well as creating a more supportive infrastructure and policy climate in which those schools will operate. Their primary types of grants are as follows:

- *District grants:* grants to districts, in support of district-wide education reform activities
- *Network grants:* grants to organizations whose goal is to demonstrate one or more successful models of schooling or approaches to school change, focusing on small high schools, and to replicate that model or process with the ultimate goal of increasing the supply of effective educational alternatives
- *Urban high school grants:* grants to organizations working specifically with urban school districts to support small high school reform

---

[1] The theory of change described here reflects the foundation's views as of July 31, 2001.

- *Technical assistance grants:* grants to organizations that provide support for change, either to other grantees or to a wider group of schools, districts, and other change agents, or grants intended to support improvements in educational infrastructure or the policy climate (e.g., grants to support the development of new performance assessment models)

There are currently more than 20 grantees (and growing) funded for five years with funding levels between $750,000 to $13,000,000. Foundation grantees, located across the country from Rhode Island to Alaska, are very diverse organizations that share the common goal of creating effective, small learning communities for students.[2] The majority of the foundation's grantees are working directly with schools, supporting the startup of new small schools and/or facilitating the converting or restructuring of existing traditional schools to adopt more personalized and effective schooling models. Ultimately, these grantees intend to offer a widespread and viable alternative to traditional schools that will promote student achievement and well-being, both within and beyond school, as well as increase school choice within the communities they serve. Other grantees are working more broadly to provide assistance to school reform teams or to other grantees, or to advocate for educational policies that are more supportive of small-school reform.

Outcomes. Target outcomes include supportive high school environments and improved achievement for students, both in high school and beyond, as well as systemic outcomes that lead ultimately to more widely available systems of choice for all students, particularly the most underserved. Through the demonstration of successful alternatives and attention to issues of scale, the foundation intends to increase the demand for and supply of high-quality high schools, while increasing the knowledge base that can support successful long-term, effective, and systemic changes to education.

Attributes of effective schooling. Central to the foundation's theory of change and grant program is their definition of effective schools and classrooms. In their effort to fund projects that successfully reform high schools, the foundation has identified seven attributes of effective high schools:

- Common focus on a few key, research-based goals

- High expectations, with all students completing a rigorous course of study

- Small, personalized learning environments

- Respect and responsibility among students and among teachers and between these groups

- Time to collaborate and the inclusion of parents and the community in an education partnership

- Emphasis on performance, with student promotions based on demonstrated competency

- Technology tools for designing and delivering engaging and imaginative curricula

According to the foundation, *effective classrooms* are characterized by:

- Active inquiry
- Indepth instruction, and
- Performance assessment.

---

[2] For some of the grantees, the Bill & Melinda Gates Foundation has partnered with other foundations to support the reform initiative.

Foundation officials have asserted, success is not predicated on being able to check off the individual attributes; rather, it is a function of the coherence and comprehensiveness of the overall vision that includes each of the attributes and makes them part and parcel of the school.

Defining the Destination

Our charge is to explore—and to the extent possible, to test—the foundation's theory of change. Therefore, we developed three central research questions that mirror the foundation's assumptions.

1. To what extent do the projects funded (wholly or partially) by the foundation initiative lead to secondary schools and classrooms with the desired attributes and to better, more equitable outcomes for students?

2. What factors influence the success of the foundation-supported projects, including:
   - Grantee vision and strategies?
   - Grantees' capacities?
   - Environmental conditions?
   - Implementation depth?

3. To what extent have grantees developed mechanisms to scale up and sustain their efforts when foundation funding end?

The first (main effects) question asks about the extent to which the projects receiving foundation grants contribute to the creation or transformation of schools and classrooms with the desired attributes and to better and more equitable outcomes for students—that is, is the initiative producing the schools, classrooms, and student outcomes envisioned by the foundation? This question is by far our most complex and analytically challenging. For example, there are multiple units of analysis implied in this question—school, classroom, individual. As discussed later, we focus on the schools and students as units of analysis. In addition, this question implies a relationship between the foundation initiative and change in schools, classrooms, and students, although we know that causally linking any one reform activity to particular outcomes is very difficult. Finally, this question implies a comparative framework—better compared to what or whom?

Research Question 2 focuses on variation in the results of the initiative across grantees, schools, and contexts. This question is one of process and context. Our goal here is to understand and explain what strategies and forms of assistance are proving the most effective in what sorts of circumstances (differential effects). Finally, Research Question 3 considers whether and how grantees are putting in place the conditions to scale up and sustain successful practices over time and which of those long-term strategies appear to be the most productive. The overarching purpose of these questions and this evaluation is to better understand the foundation's initiative and its potential impact on schools and students. Therefore, these questions focus on general aspects of the initiative and the grantees' capacity and activities and do not imply an evaluation of particular grantees or schools.

Finding the Coordinates

Once we explicated the theory of change and developed the research questions, we turned to the research literature to develop a conceptual framework to guide the evaluation or to "find our

coordinates" (see American Institutes for Research & SRI International, 2001a for our review of the literature). We turned to the research literature to add specificity to the foundation's theory of change and to generalize our work by locating it in the larger literature of school change.
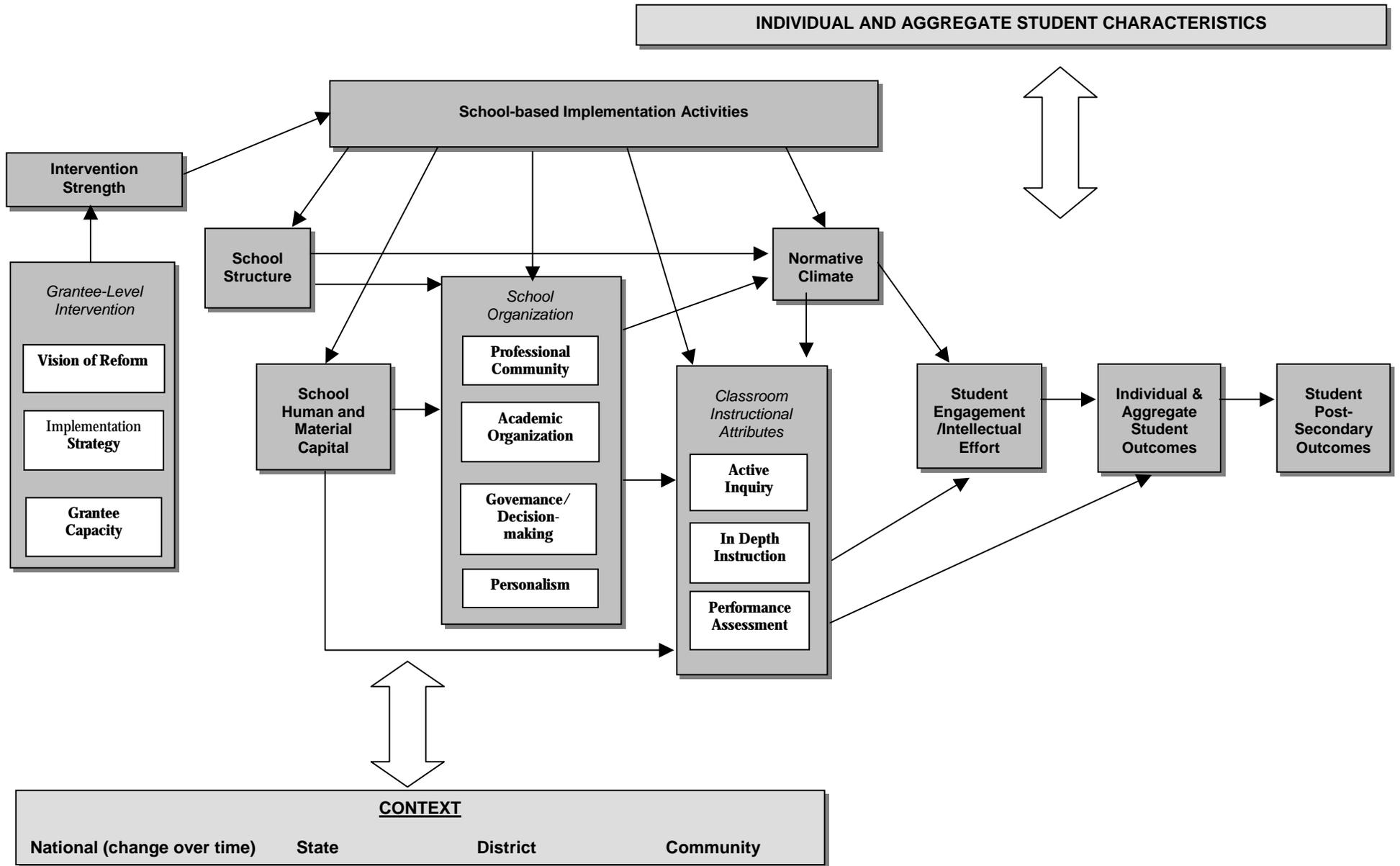
In this section, we outline briefly the main constructs and relationships that make up the conceptual framework, which is depicted graphically in Exhibit 2. The framework begins with the foundation's theory of change as its starting point, concentrating on the predicted impact of grantees on the schools and students with which they work. Thus, reflecting the general structure of the theory of change, the framework depicts student outcomes (on the far right-hand side of Exhibit 2) to be the ultimate dependent variable of interest, and the foundation-supported activities (funded through the grantees; far left hand side of the model) to be the main independent variable or intervention of interest. Also, as in the theory of change, the impact of the intervention is predicted to occur through the grantees' influences on specified characteristics of schools and classrooms. More specifically, the conceptual framework portrays grantee interventions as influencing the presence and strength of characteristics associated with high-performing schools (in the literature on school organization) and of in-depth, inquiry-based instruction (in the literature on teaching for understanding). These in turn are posited to positively influence student learning.

Although sharing these central commonalities, the conceptual framework expands on the theory of change in several ways. For example, on the basis of our review of the relevant literature, we have unpacked the attributes of high-achieving schools, placing them into two inter-related dimensions of schooling: school organization and normative climate. We have also added variables left out of the theory of change, such as the quality and the stability of instructional staff (which the framework posits as part of the human and material capital available to the school). Exhibit 2 also recognizes that student learning is co-produced by teachers and students together. Thus, the impact of instruction occurs primarily through its effect on the active engagement and effort on the part of students. We explain each of these constructs and relationships in greater detail below.[3]

Overall, the conceptual model depicted in Exhibit 2 posits that the impact of foundation-supported activities will be influenced by a combination of **grantee-level factors**, including the grantees' *vision, strategy,* and *capacity* and the ways these interact to influence the overall *strength of the grantee's intervention.* These grantee-level factors will then influence various aspects of **the schools** with which the grantees work (including the *school's structure, organization, human and material capital, and normative climate*) and of the *classroom instruction* within those schools. Some of these school characteristics are also posited to influence one another, and all are expected to be affected by the **environmental context** in which the school resides. The school characteristics are then expected to have an impact on **students**, first through *student engagement and intellectual effort* and ultimately through specified *student outcomes* in high school and beyond. As with environmental context, **individual and aggregate student characteristics** are posited as important influences on the success of grantees' work that are extrinsic to the intervention.

---

[3] Although we believe that the conceptual framework depicted in Exhibit 2 provides a useful guide for data collection and analysis, two limitations of the diagram are important to note. One concerns the direction of the predicted relationships. In the diagram, all arrows are uni-directional, suggesting a clear causal linkage from one variable to another. In fact, we assume that the relationships are much more complex and interactive. This brings us to the second limitation: the diagram is currently limited in its ability to depict changes over time.

Exhibit 2. Conceptual Framework for Network and Urban Grants

INDIVIDUAL AND AGGREGATE STUDENT CHARACTERISTICS

School-based Implementation Activities

Intervention Strength

Grantee-Level Intervention

Vision of Reform

Implementation **Strategy**

Grantee Capacity

School Structure

School Organization

Professional Community

Academic Organization

Governance/ Decision-making

Personalism

Normative Climate

School Human and Material Capital

Classroom Instructional Attributes

Active Inquiry

In Depth Instruction

Performance Assessment

Student Engagement /Intellectual Effort

Individual & Aggregate Student Outcomes

Student Post-Secondary Outcomes

CONTEXT

National (change over time)   State   District   Community

Exploring the Terrain

To state the obvious, our research questions, conceptual framework, and evaluation design must all be both informed by and adapted to the initiative as it unfolds. Our understanding of the "terrain" requires that we know as much as we can about the grantees funded under the initiative and their school partners. This section briefly describes our understanding of the grantees and schools.

<u>Overview of population of grantees and schools</u>
Foundation officials intend that the grant-making portion of their strategy will demonstrate and replicate successful high school models and will support the organizations that are enacting change, either by direct technical assistance or by the development of a more supportive infrastructure. Exhibit 3 provides a thumbnail sketch of the first 12 grantees included in our evaluation.

Exhibit 3
Grantee Models and Approaches

Aspire Public Schools
Aspire Public Schools is establishing a network of California charter schools that emphasize authentic learning and a balance between constructivism and basic skills. Its strategy is to grow in clusters, with elementary schools of grades K-5 and secondary schools of grades 6-12. Bill & Melinda Gates Foundation funding will be used to cover design and start-up costs for five secondary schools over the next 5 years.

Bay Area Coalition for Equitable Schools (BayCES)
Working with schools in Oakland, California, and surrounding communities, BayCES supports school design teams striving to create new small autonomous schools and to transform large high schools into small learning communities. In both cases, the goal is equitable learning opportunities and high achievement for all students. BayCES works with the Oakland Community Organization to build support for small equitable schools.

Big Picture Company
The Big Picture Company is creating new small high schools in the next 5 years, based on the model of the Metropolitan Regional Technical and Career Center ("The Met") in Providence, RI. Big Picture's model school embodies its philosophy of "one student at a time" through individualized curricula based on student interests and passions, authentic learning through internships and mentors, and public exhibitions of student work.

Center for Collaborative Education (CCE)
The Center for Collaborative Education supports networks of progressive schools doing whole-school reform. CCE coordinates five networks, one of which is the New England Small Schools Network (NESSN), which is funded by the Bill & Melinda Gates Foundation. NESSN intends to further the development of the small-schools movement by converting large schools, starting new small schools, and functioning as a clearinghouse and advocacy organization for the movement.

Center for School Change (CSC)
Established 12 years ago within the Hubert H. Humphrey Institute of Public Affairs at the University of Minnesota, the Center for School Change is working with urban districts to support school design teams working to restructure large high schools and convert them into small, personalized schools of choice. With a foundation grant to support work with Cincinnati, St. Paul, and West Clermont (OH), CSC takes a community-based approach to designing and implementing programs.

Exhibit 3
Grantee Models and Approaches (Cont'd)

Colorado Children's Campaign (CCC)
Colorado Children's Campaign (CCC), a statewide nonprofit organization, was formed in 1985 to advocate for children. CCC was funded to work with partners across the state of Colorado to: (1) convert several existing large high schools into smaller schools, (2) establish several new small high-tech high schools focused on mathematics and science; and (3) develop a Charter School Network to promote instructional improvement and assessment development among small charter schools.

EdVisions Incorporated
EdVisions provides startup funding and professional development to efforts to replicate its model school, Minnesota New Country. The model's design includes a commitment to high-quality, personalized, project-based learning; active partnerships with parents and the community; and systems that are decentralized but stress accountability. EdVisions schools operate on a teacher-owner governance model with a teachers' cooperative offering performance-based, non-tenured teacher contracts, professional development, payroll management, and instructional materials.

High Tech High Foundation
High Tech High Foundation is funded to establish nine small charter high schools based on the design principles of the High Tech High (HTH) school in San Diego. The new schools are intended to provide rigorous college preparatory courses with an emphasis on building skills needed in an information-age economy. Instructional design focuses on authentic problem-based inquiry and real-world application, with required internships and roughly 1/3 of the teachers hired from industry.

Model Secondary Schools Project (MSSP)
MSSP was established in August 2000 with the mission of creating new small secondary schools in eight urban districts across the country. Working in Detroit, Compton (CA), Boston, Cincinnati, Cleveland, East St. Louis (IL), Rochester, and Las Vegas, MSSP supports local school teams involved in designing small schools. MSSP does not impose a school design but rather encourages the planning teams to develop their own designs incorporating the principles of small size; distributed leadership; personalized learning; project-based learning; performance accountability; extensive use of technology; room for talking, movement, and visitors; community internships; and trust and respect.

National Council of La Raza (NCLR)
Established in 1968 with the mission of improving life opportunities for Hispanic Americans, the National Council of La Raza is seeking to develop a large network of charter schools to better serve Latino populations. With foundation funding for establishing 15 small schools, NCLR provides funding and technical assistance to community organizations that meet prescribed eligibility requirements and that seek to start-up or improve these schools. NCLR emphasizes the humanities, Latino culture, bilingual competence, and training in conflict resolution, as well as high parental involvement, small school and class sizes, technological literacy, and high academic performance.

New Technology Foundation
New Technology Foundation was established to provide management and fundraising support for New Tech High School, established in Napa, California, in 1996. With funding from the foundation, New Tech seeks to create a network of nine new Northern California schools with the essential features of the Napa model. These include project-based learning; a culture of respect and responsibility; community internships; and the use of technology to support learning.

The foundation envisions the grantees as catalysts for change that will shepherd both new small and existing large schools through the complex process of becoming flourishing small learning communities. No two grantees perform this responsibility in identical ways. They differ in their replication target and geographic reach. Specifically, grantees and schools that are part of this reform initiative are clustered in specific areas (i.e. Oakland, St. Paul); however, those areas are spread out across the country (i.e., Arizona and California). Consequently, there is considerable variation in the communities within which the schools reside and the student bodies that these schools serve. Grantees also vary in the number of schools with which they work and the timeline for opening or converting those schools (see Table 1). As illustrated in Exhibit 3, the number of foundation-supported schools is expected to grow from 13 to almost 200 over the five years of our evaluation.

Generally speaking, there are three general categories of schools that receive support through this reform initiative: models, startups, and conversions. Model schools are small high schools that existed before this initiative and display the key attributes detailed in the theory of change. Start-up schools are either replicated versions of the model schools or new schools that are designed based on a set of specific principles or a process of reform. Finally, conversion schools are the small schools carved out of pre-existing large comprehensive schools. For example, a school of 1,600 students would be converted to four schools of 400 students each.

While these differences present a challenge to applying a common design that addresses this diversity, the differences do provide an excellent opportunity to examine various approaches to this reform initiative, to see what works for whom under what circumstances. However, like any well planned road trip, we must select a route that takes us to all the major tourist attractions while also allowing us enough time to go beyond them to soak in the local culture of the people who live, work, and play there.

Table 1. Brief description of grantee and school populations.

Grantee	Replication Target	Geographic Reach	Types of Schools Served	Number of Affiliated High Schools (Number of Affiliated Spin-off High Schools)			
				2000-2001	2001-2002	2002-2003	2003-2004 [a]
Aspire Public Schools	Design principles	Local/ regional	Start-up	0	3	4	4
Bay Area Coalition of Equitable Schools	Process for reform	Local/ regional	Model, Start-up, Conversion	5	7	11 (2)	17 (3)
Big Picture Company	Explicit school model	Multistate	Model, Start-up	3	3	10	16
Center for Collaborative Education	Design principles	Multistate	Start-up, Conversion	0	7 (5)	16 (12)	16 (17)
Center for School Change	Design principles	Multistate	Start-up, Conversion	0	3	9 (24)	11 (24)
Colorado Children's Campaign	Design principles	State	Start-up, Conversion	0	1 (3)	2 (6)	9 (9)
EdVisions	Design principles	Multistate	Model, Start-up	3	7	10	13
High Tech High Foundation	Explicit school model	Multistate	Model, Start-up	1	1	4	5
Model Secondary School Program	Design principles	Multistate	Start-up	0	2	8	8
National Council of La Raza	Process for reform	Multistate	Model, Start-up	2	6	9	9
New Technology Foundation	Explicit school model	Local/ regional	Model, Start-up	0	1	3	3
New Visions for Public Schools	Process for reform	Local/ regional	Start-up	0	0	18	27

[a] 2003-2004 numbers are projected.

Determining the Route

As we developed our coordinates and explored the terrain, we also began sketching out the design of our evaluation. Generally speaking, the route we determined for our evaluation consists of three inter-related studies: (1) Core Study; (2) Student Work Study, and (3) Longitudinal Study. Each is described below.

Core Study
The Core Study, for which we first received funding, is a cross-section, school-level study that examines change over time in school characteristics and aggregate student outcomes. The evaluation will use a mixed methodology design that comprises inter-related research strands with integrated sampling, data-collection, and analysis plans. This strategy will enable us to answer each question with the most relevant and powerful sources of data, to triangulate the data and explain the results obtained from different sources of data, and to understand both the processes and the outcomes of project implementation.

The four linked research strands follow:

1. *School Survey Strand* to assess both intermediate and longer term outcomes and the relationships between project interventions and outcomes

2. *School Qualitative Strand* to obtain a dynamic picture of the processes of school change and its relationship to school and student outcomes

3. *Case Studies of Foundation Grantees* to describe and assess the several intervention strategies the foundation is supporting

4. *Student Achievement Strand* designed specifically to assess student learning and the relationship between student learning and school and instructional attributes[4]

Our Core Study design focuses on key constructs and measures, used across grantees, schools, and studies, that indicate the degree to which the attributes of effective schools characterize the new or converting schools and the extent to which individual schools and networks of schools have fostered the student outcomes emphasized by the foundation. We developed a set of indicators for assessing grantees' scale-up and sustainability plans. As indicated in the next section, our design follows schools and grantees over time.

The evaluation design includes baseline data collections for schools that are converting or beginning operations. Implementation and progress data will be collected over the course of the evaluation. Again, we will use these data to provide formative evaluation information to the foundation and grantees. Comparison data for schools not supported by the foundation will be collected in years 4 and 5 of the evaluation. These and other program outcome data will be used to provide summative evaluation information.

Data collection. Table 2 illustrates the data collection schedule for grantees and two cohorts of schools for the Core Study.[5] Grantees will be site visited each of five years. Converting schools will be site visited each of three years. Survey data from converting schools will be collected twice over a three year time period.[6] The first year of survey (and site visit) data will be collected in the year prior to conversion and serve as a baseline for comparative purposes. The start-up schools will be surveyed and site visited each of three consecutive years with comparison schools surveyed and site visited during the final year. Our design calls for a total of 10 converting schools, 20 start-ups and 20 comparison schools for the survey sample. It also calls for 10 converting schools, 10 start-ups, and 5 comparison schools for the site visit sample. (American Institutes for Research & SRI International, 2001b.)

---

[4] The design of this strand is currently in process and will not be described here.

[5] Schools were assigned to cohorts depending on when they opened their doors. Those that were open in 2001-2002 are cohort 1 and those that opened in 2002-2003 are cohort 2.

[6] We originally planned to survey in converting schools every year, but learned that during the first year of conversion it is much more difficult to get high enough response rates to report the data given the amount of change occurring in these big organizations.

Table 2. Data Collection Schedule: Core Study

	Year 1 2001-02	Year 2 2002-03	Year 3 2003-04	Year 4 2004-05	Year 5 2004-05
GRANTEES	o	o	o	o	o
SCHOOLS					
Cohort 1					
Conversion		x, o	o	x, o	
Start-up		x, o	x, o	x, o	
Comparison				x, o	
Cohort 2					
Conversion			x, o	o	x, o
Start-up			x, o	x, o	x, o
Comparison					x, o

x = Core Study survey

o = Core Study site visit

<u>Data analysis.</u> In this section, we describe some of the key analytic tasks we will undertake as part of the Core Study and a brief description of our methods.

Change over time in foundation-supported schools. Survey data collected from students, teachers, and administrators in foundation-supported schools will be used to examine change over the course of the evaluation on such dimensions as the presence and penetration of desired school and instructional attributes and various student outcomes. In other words, we will investigate whether foundation-supported schools are developing the characteristics of effective schools and classrooms and exhibiting changes in student outcomes over time and changes in the outcomes of *certain populations* of students (e.g., high-poverty students). We will undertake these analyses by examining changes on the attributes and the teacher-, student-, and school-level outcomes in foundation-supported schools at the two time points during which we will collect survey data. We will conduct these analyses by using Hierarchical Linear Modeling (HLM) to examine change on key constructs within foundation-supported schools and then looking at the variation in change across these schools.

Comparisons of foundation-supported and "other" schools. Data collected from foundation-supported and comparison schools will be pooled to examine intermediate outcomes comparable across schools (e.g., students' intellectual effort and the attributes of effective schools as measured by the survey). Here, we will use a two-level HLM, with comparable outcomes and one independent variable (foundation-supported versus other high schools). The results of these analyses will tell us whether significant differences exist between foundation-supported and comparison schools on the presence and penetration of the foundation attributes and the level of student outcomes (non-achievement). In addition, we plan to examine the relationships between certain intermediate outcomes (e.g., do students in schools with higher levels of personalism exhibit more intellectual effort compared with their counterparts in schools with lower levels of personalism?). We will again pool the sample, use a two-level HLM, and designate students' intellectual effort as the outcome (for example).

Description of change process. We will use the qualitative data to describe grantee and school visions and strategies, grantee capacities, environmental conditions, and school implementation activities. We will

examine the narrative and content analysis data for relationships between the factors, (e.g., grantee capacity or school context) and implementation levels. We will use the coding data and the narrative information to portray these relationships. We will also look for the relationships between implementation and student outcomes; for example, we will study the relationships between depth of implementation and both judgments about student learning and achievement data (where available from extant sources). We will continue to examine these associations over time.

Description of grantee development. The foundation initiative is predicated on the idea that grantees will be able to scale up their efforts over the next few years and sustain their work when foundation funding ends. The case studies of foundation grantees will examine the plans that grantees develop to market their programs, price their services, protect their intellectual property, and support substantial numbers of schools. The data for these analyses will come primarily from annual interviews with grantees and semi-annual interviews with principals. We will follow grantees and their school partners over time, looking for increased capacity, service, and support. Information on the numbers and success of replication schools will come from the analyses of the survey data and the extant data collected from those schools.

Student Work Study

As we considered an evaluation of this initiative, it become clear to us that we would want to supplement our Core Study design with a study that addresses two key challenges that we faced: (1) because the schools are spread throughout the country, we don't have a single measure of achievement, and; (2) short-term classroom observations and close-ended survey items provide only a limited picture of the dynamic process of teaching and learning. Collecting and analyzing teacher assignments and student work allows us to examine firsthand the nature of instruction in the classrooms of small foundation-supported schools and the link between learning opportunities (what teachers are delivering) and student learning. We can look for evidence of students' reasoning, understanding, and complex problem solving. The student work can show the extent to which students understand the concepts, principles, and procedures of inquiry and the extent to which assignments actively engage them in broadening their understanding and extracting underlying themes and principles. Students' work can be examined to study the extent to which the written feedback from their teachers provides information to support learning and the frequency and types of opportunities that students have to revise their work in light of feedback (Bransford, Brown, & Cocking, 1999; Vye 1998a, 1998b).

Research questions. We revised our original research questions to guide this study:

1. To what extent do projects funded by the foundation initiative lead to secondary schools and classrooms with a higher level of academic challenge (e.g., active inquiry, indepth instruction, performance assessment)?

2. Do students who are exposed to more challenging learning opportunities have better student performance as reflected in the quality of student work and scores on standardized achievement tests?

3. To what extent does the level of academic challenge experienced by traditionally underserved students differ between foundation-supported schools and other schools in the district?

Through a systematic collection and assessment of teacher assignments and student work, we will be able to examine the relationships between the intellectual quality of teacher assignments and feedback and student performance. We will measure student performance by using assessments of student artifacts related to the assignments and student scores on standardized achievement tests. By examining the link between the intellectual quality of actual teacher assignments and the student work generated in response to these assignments, we will be able to supplement the overall evaluation with a more authentic analysis of the link between effective classroom attributes and student performance. By examining how the relationship between

assignment challenge and student performance differs in foundation-supported and non-supported schools, we will be able to determine whether the totality of other changes that occur in foundation-supported schools have a beneficial effect on that relationship.

Data collection activities. For the start-up schools, our design calls for an overlap in the Core Study and Student Work Samples (see Table 2). We were unable to consider a similar overlap for conversion schools because our Student Work Study design calls for collecting data in such schools the year prior to conversion. Therefore, the Student Work Study includes data collection activities in cohort three schools (i.e., those large school that convert to small schools in 2005-2006, which are not part of the Core Study.) Following the approach used by Newmann, Bryk, and Nagoaka (2001), we will ask teachers at the schools participating in the student work option to submit both typical and challenging assignments in mathematics and writing. Teachers will be asked to provide copies of the student work generated by the challenging assignments only. The first year, 2002-2003 will be used to develop and pilot scoring rubrics and the data collection and scoring procedures. We propose to start collecting data in 2003–2004 and 2004–2005 for 12 foundation-supported start-up schools and 4 comparison schools in their districts and in 2003–2004 and 2004-2005 for 6 large schools undergoing conversion into smaller learning environments. Thus, we will be able to make comparisons between foundation-supported and demographically comparable non-foundation-supported classrooms and to conduct longitudinal analyses of teachers in foundation-supported classrooms.

Table 2. Data Collection Schedule: Core Study and Student Work Study

	Year 2	Year 3	Year 4	Year 5	Year 6
	2001-02	2002-03	2003-04	2004-05	2005-06
GRANTEES	o	o	o	o	
SCHOOLS					
Cohort 1					
Conversion	x, o	o	x, o		
Start-up	x, o	x, o	x, o, **A**	**A**	
Comparison			x, o	**A**	
Cohort 2					
Conversion		x, o	o	x, o	
Start-up		x, o	x, o, **A**	x, o, **A**	
Comparison				x, o, **A**	
Cohort 3					
Conversion				**A**	**A**

x = Core Study survey

o = Core Study site visit

A = Teacher assignments and student work collected

Data analysis. Several analytic techniques will be used to examine the data collected in the student work component. The outcome variable consisting of the teacher assignment measures will be generated from many-facet Rasch measurement (MFRM). We propose to estimate the mean intellectual demands of assignments in English and math. Since students are naturally clustered within classrooms, and classrooms within schools, we intend to use hierarchical linear modeling (HLM) in some of our analyses to properly estimate standard errors. We will use HLM (1) to compare the foundation-supported and non-foundation-

supported *classrooms* on the level of academic challenge; (2) to examine the relationship of teacher assignment quality with student performances; and (3) to compare the level of academic challenge in classrooms within schools undergoing conversion using the pre-conversion data as the baseline comparison. Our analyses will be modeled on those described in Bryk, Nagaoka, and Newmann (2000), Newmann, Bryk, and Nagaoka (2001), and Newmann, Lopez, and Bryk (1998).

Longitudinal Study

As described earlier, the Core Study is a longitudinal study of schools and aggregate student outcomes. As such, the Core Study design has a number of important strengths (e.g., it looks at change in school attributes and outcomes over time, and it allows a comparison of foundation-supported and traditional schools). However, the Core Study design does not allow us to examine students' academic growth in an unambiguous way. Measuring growth requires that one follow individual students over time. Furthermore, although the Core Study's cross-sectional design allows us to examine important outcomes at the *aggregate* level (e.g., dropouts, detentions, graduation), it does not allow us to link presumed *student-level* precursors of these important events accurately (e.g., course-taking behaviors) to outcomes for students. Only by following students longitudinally can we draw robust conclusions about student-level growth. This longitudinal study provides an important complement to the work of the Core Study while the students are in high school, and also permits the assessment of outcomes in the years after high school, when students make crucial decisions such as those regarding postsecondary education and employment options. Moreover, a longitudinal study makes it possible to control for 8th-grade test scores and other background characteristics to determine and adjust for differences in student selection.

Research questions. The longitudinal data we collect will be used to address three basic research questions:[7]

1. Do students attending foundation-supported high schools exhibit better outcomes (e.g., engagement, grade advancement, graduation, college attendance) than students in traditional high schools?

2. Are particular types of students (e.g., traditionally underserved) more likely to benefit from attending foundation-supported schools than others?

3. What features of high schools (e.g., personalism, normative climate, instruction) are associated with student outcomes?

Data collection activities. Unlike the Core Study, our focus for the Longitudinal Study is on students. We propose to study one group of students from start-up schools from cohorts 1 and 2 for six years starting in the 2003-2004 academic year (see Table 3).[8] We have chosen 2003-2004 to begin data collection for the longitudinal study because it strikes a good balance between several considerations. First, by this time start-up schools will have had time to implement their reform programs. By 2003-2004, cohort 1 start-up schools will be in their third year of operation, and by 2004-2005, cohort 2 start-up schools will be in their third year as well. In each of 40 schools (20 foundation-supported schools and 20 comparison schools) we sample for the longitudinal study, we will select 70 students from the 9th grade. We will track these students and collect data over three phases of the study, ending in 2008-2009, with the option of following students for two additional years to 2010-2011. During the high school years, we will survey students; once students leave high school—either graduate or drop out—we will conduct telephone interviews with them.

---

[7] The research questions are still under refinement, a final version will be included in the study's analysis plan to be completed at the end of April.
[8] Our goal is to have an overlap in study samples such that all of the Student Work study schools are also in the Longitudinal and the Core studies.

Table 3: Tracking One Cohort Over Time In Startup Schools

| | | | | | | | OPTIONAL | |
| | | End of Phase I | | End of Phase II | | End of Phase III | | End of Phase IV |
Class/Grade	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11
Frosh (9)	Cohort 1	Cohort 2						
Soph (10)		Cohort 1	Cohort 2					
Junior (11)			Cohort 1	Cohort 2				
Senior (12)				Cohort 1	Cohort 2			
Post-HS 1 (13)					Cohort 1	Cohort 2		
Post-HS 2 (14)						Cohort 1	Cohort 2	
Post-HS 3 (15)							Cohort 1	Cohort 2
Post-HS 4 (16)								Cohort 1

Data analysis. A discussion of all of the analytic methods we plan to employ for the longitudinal study are too numerous to describe in this paper and some of the methods overlap with those previously discussed. Therefore, we briefly describe one approach that we will undertake that is unique to the longitudinal study—comparisons of foundation-supported schools and traditional schools on the *growth trajectories* of student outcomes.

As explained earlier, one important advantage of a longitudinal research design is that by collecting data in multiple points in time, we can examine not only the status of student outcomes at a given point in time, but also the growth trajectories of certain student outcomes over time. Many of the hypotheses underlying our evaluation imply growth over time. For example, we expect that as a result of the personalized learning environment and meaningful school work promoted by the foundation's initiative, students attending foundation-supported schools will become increasingly engaged with schooling and exert increasingly higher levels of intellectual effort on their school work. On the other hand, students in the traditional large impersonal high schools may become progressively bored and disengaged with schooling as their high school careers progress. This suggests that not only might student engagement increase at a faster pace in foundation-supported schools, the direction of change may differ—engagement increases in foundation-supported schools and decreases in traditional schools.

In later phases of the study, after we have collected the same outcome data over multiple years, we will be able to examine multiple measures of a particular outcome (e.g., engagement) over time and model a growth trajectory for each student. Using a three-level HLM model (i.e., time periods in students in schools), we will be able to examine these trajectories and relate them to both student and school characteristics. That is, we can examine if students with particular characteristics (e.g., high SES) exhibit greater increases in engagement than their counterparts. We can also examine if students in foundation-supported schools exhibit a greater increase in engagement than students in comparison schools—or, more dramatically, if students in the former

school type exhibit increased engagement, whereas, student in the latter school type exhibits decreased engagement over time.

General Issues

There are a number of general issues that we continue to consider as we develop and refine our evaluation design. Among these are:

Diversity of interventions and contexts. A great deal of diversity exists among the grantees and the schools they are supporting and creating. This diversity includes wide variation in interventions, school locations, and target populations. For example, some interventions are very prescriptive in their reform strategies and plans; others are based on broad guiding principles. Some intervention strategies include creating new schools, breaking up existing schools, or restructuring existing schools. (These variations exist among grantees, but also within grantees; a grantee may use different interventions in different locations.) Additionally, schools are located in diverse communities, districts, and states all over the country. The target populations for schools vary; some interventions target the most underserved students, whereas others do not.

To address the challenge of diversity in interventions and contexts, we plan to capture and study it. We will articulate specific theories of change, activities, and progress for each grantee and examine how these evolve over time. We will classify grantees in terms of key features of their intervention designs. We will also examine implementation activities in schools, the various contexts within which schools reside (e.g., community, district, state), and the types of students and teachers they attract. This approach should help us examine our data to detect relationships among intervention design, implementation, and context variables. These analyses will also help us understand the impact and processes of school reform as they play out in different ways and in different contexts.

Schools change more slowly than expected. Evaluating schools in the early stages of development or reform is difficult at best. Experience tells us that educational reform is a slow, incremental process and that it takes time for its effects on school organization, instruction, engagement, and achievement to unfold. To address this challenge, we will chart the progress of foundation-supported schools over a minimum of three years and adapt the timing of our analysis of student outcomes to fit the progress in school capacity improvements. Although we will prepare for student outcome analyses from the beginning, our initial quantitative data collection will not be used to test for differences in student outcomes in foundation-supported schools. Rather, in the early stages of the evaluation, we will use quantitative and qualitative data to understand and describe the reform process. We will conduct comparative analyses on intermediate and later outcomes as reform proceeds and the evaluation that focuses more on the outcomes of the grant effort.

Small and changing school populations. The population of grantees and schools is both diverse and changing. These schools are, by design, small. The small number of grantees and schools and the "smallness" of the schools will limit the statistical power for estimating relationships. We will address this challenge in two ways. First, we will limit the number and types of analyses we will conduct using statistical procedures. We have conducted power analyses to ensure that we will have the statistical power necessary to test the relationships we plan to investigate. (See American Institutes for Research & SRI International, 2001b.) Second, our design relies heavily on qualitative methods to address all three research questions in greater depth. We will be able to collect qualitative data from the grantee population and a large proportion of the population of schools (approximately 50%).

Matching comparison schools. Identifying and recruiting comparison schools will be challenging. Foundation-supported and comparison schools must be fairly carefully matched on a range of potentially confounding factors to maximize the chances of detecting program-related effects. Ideally, as reflected in the

research questions described earlier, the Core Study and Longitudinal Study compare schools and students in foundation-supported schools with large, comprehensive high schools and similar students in those schools. Specifically, the longitudinal study will focus on a comparison of students who did attend foundation-supported schools and students like them who attended the types of schools the "foundation students" would have attended. The key to this approach is finding both similar types of students (described in the next section) and similar schools that foundation students would have attended. We describe the latter process below.

The comparison school selection process involves two basic steps:

(1) **Determine which high schools the foundation students would have attended had they not attended the foundation-supported schools.** We started by identifying all of the middle schools that had been attended by 9th graders in the foundation-supported schools during the 2002-03 academic year. We then determined the high schools to which these middle schools "feed" students—that is, where students from these middle schools go for high school. From this information, we are able to develop an initial pool of possible comparison schools.

(2) **Collect demographic and other descriptive data from the foundation-supported and comparison schools.** We will collect demographic and descriptive data, including such variables as geographic location, number of students by grade, number of teachers, number of students by grade and ethnicity, type of school, and number of students qualified for free or reduced-price lunches.

We use this information to identify a first- and second-choice comparison school for each of the foundation-supported schools. Three criteria were considered on a case-by-case basis in selecting the most appropriate first- and second-choice comparison schools: (1) schools with comparable demographic characteristics; (2) schools where the majority of the foundation students would have gone if they had not gone to the foundation-supported school; (3) schools that are located in the same district as the foundation-supported school.

Because patterns associated with "feeder" middle schools and the targeted comprehensive high schools attended by their students differ markedly for the foundation-supported schools, the strategy for selecting comparison schools will be tailored to each of their individual contexts. The goal is to select one or more schools that will yield students who mirror foundation students closely on key dimensions and that provide an education typical of comprehensive public high schools. Thus, private, charter, magnet, and alternative schools were not included in the pool of eligible comparison schools.

Reading the Map While Driving

With this evaluation, we have a unique opportunity to examine various reform efforts aimed at creating small, effective learning communities, and our work allows us to learn about a large number of high schools from students, teachers, and principals perspectives over time. We are also able to examine the work being assigned by teachers and completed by students as well as follow individual students for many years beyond high school to determine if there are long-term effects of the initiative. Our study is comparative in nature and our goal is to conduct high-quality quasi-experimental studies of reforming and traditional high schools. However, our work is also very complex and requires a great deal of flexibility and problem solving as we undertake this work. In this section we describe some of our general challenges and what we mean by "reading the map while driving."

<u>Merging lanes.</u> One of our most critical and daunting challenges in this evaluation is to tell a coherent story about a very complex and dynamic initiative. Our evaluation design incorporates three interrelated studies, each with their own research questions, designs, sampling and data collection plans, and methodological approaches. And while the results of each study will provide a rich picture of the implementation and impact of the foundation's reform initiative, together these studies are much stronger than the sum of the parts. The combination of these studies, with overlapping questions, sampling and data collection plans, will allow us to draw a multi-dimensional picture of the initiative, using multiple sources of evidence and triangulating results.

As we undertake this work, however, it is necessary for us to balance attention to the micro details of the project and the larger landscape of reform, research, and policy. We are addressing this challenge through a number of organizational and managerial structures. For example, we have staffed the project with a diverse group of researchers who specialize in certain aspects of reform (e.g., district change, implementation, professional development, student learning, etc.) and have regularly scheduled discussion on the substance of our work. In addition, we have a team of researchers who work across the studies in various roles (e.g., site visitor, survey developer, etc.), including key leadership positions. We schedule regular staff meetings to discuss the details of the projects, cross-cutting issues, and early findings. Moreover, we have a strong leadership team that facilitates the articulation across projects and keeps the work on track.

<u>The road(s) not taken</u>. Even though this is a large-scale evaluation with a great deal of financial resources behind it, we still find ourselves making decisions based on what we can reasonably do with the resources that we have. It takes a great deal of work to understand the diverse interventions (all with varying degrees of specificity), the complex organizations in which they are being introduced (i.e., schools), the diverse social and political contexts within which these schools reside, and how the dynamics of schooling and reform play out over time. In many ways, design can be a zero-sum game. Even with many resources, it is often necessary to make difficult decisions about the scope of the project and often when you gain something, you also have to give something up. For example, there are a many different roads we could have taken to evaluate this initiative. We could have directed all of our energy and resources toward one, narrow question (e.g., Is pedagogy changing in foundation-supported schools?) and addressed this one question very deeply. It's possible that we could have designed an experimental study examining the impact of the interventions of student outcomes, rather than using quasi-experimental designs.

Our understanding of the foundation's needs, however, led us to pursue three broad research questions. With this approach, we are able to learn about many important aspects of the reform, including pedagogy, for example, but we are not able to learn about any one aspect to the same degree we would have had we devoted all of our resources to just one topic. Moreover, although an experimental would have put us in a position to make causal inferences, as evaluators, we are unable to control the conditions under which the treatment is administered and it is our understanding that the foundation is most interested in examining the landscape for positive change, rather than necessarily causally attributing their efforts to improved outcomes. We find, however, that it is challenging to find and maintain a balance between the numbers of questions we can pursue and the level of detail we can attain. The more we learn, the more questions we have. Restraint and careful monitoring of the integrity of our evaluation design has been and will remain critical for the success of our work.

<u>Retracing our steps</u>. As described earlier, this is a very complex initiative and a complex evaluation design. One of the critical steps that we undertake throughout the evaluation as it moves forward is backward mapping. That is, we continuously reflect on our activities and findings and retrace our steps back to our research questions, conceptual framework, and the foundation's theory of change. These reflections serve as important benchmarking activities that allow us to consider both the extent to which we are indeed addressing our research questions with our evaluation activities and the degree to which we are testing the foundation's underlying theory for their initiative. This activity also allows us to reflect on the conceptual

framework that we developed prior to data collection to determine if it remains relevant and truly reflects what we are learning. We consider this latter activity to a form of theory building that guides our continued work on this evaluation.

This activity of first backward and then forward refining is critical for the success of our evaluation. The school reform initiative that we are evaluating is very generally defined and the "treatments" are very diverse as are the grantee organizations, the schools, and the school contexts. The foundation's initiative is fluid and evolving over time, as is its funding strategy. New grantees are being funded and new and different types of schools are being created and converted. As the initiative changes and grows and we learn and adapt, it is important that we reexamine our coordinates, re-acquaint ourselves with the shifting terrain, and reevaluate our course of action. These steps will allow us to be thoughtful and flexible with our evaluation design.

<u>Miles to go before we sleep</u>. This spring marks our second year of large-scale data collection and findings from our first data collection have just been released. As the previous sections illustrate, we are currently in the early stages of our evaluation and will be pursuing the research questions described in this paper for the next nine years or so. Throughout this time period, we will be reporting findings to the foundation, grantees, schools, and general public. As the next year of the initiative unfolds, we will continue to observe the processes and results and refine our evaluation design. Answers to these questions will guide the next part of our journey and we anticipate a very interesting road ahead.

REFERENCES

American Institutes for Research & SRI International. (2001a). Evaluating the National School District and Network Grants Program: Draft of the Literature Review. Washington, DC: AIR.

American Institutes for Research & SRI International. (2001b). Evaluating the National School District and Network Grants Program: Draft 1 of the Evaluation Design. Washington, DC: AIR.

Bransford, J., Brown, A., & Cocking, R. (1999). *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

Bryk, A. S., Nagaoka, J. K., & Newmann, F. M. (2000). *Chicago Classroom Demands for Authentic Intellectual Work: Trends from 1997-1999.* Chicago, IL: Consortium on Chicago School Research.

Newmann, F., Bryk, A., & Nagaoka, J. (2001). *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence.* Chicago, IL: Consortium on Chicago School Research.

Newmann, F., Lopez, G., & Bryk, A. (1998). *The Intellectual Work in Chicago Schools: A Baseline Report.* Chicago, IL: Consortium on Chicago School Research.

Shear, L. & Smerdon, B. (2003). Mapping the Terrain: Year 1 of the Evaluation of the Bill & Melinda Gates Foundation's National School District and Network Grants Program. Paper Presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL, April 2003.

Vye, N., Goldman, S., Voss, J., Hmelo, C., Williams, S., & Cognition and Technology Group at Vanderbilt. (1998a). Complex Mathematical Problem Solving by Individuals and Dyads. *Cognition and Instruction, 15*(4).

Vye, N., Schwartz, D., Bransford, J., Barron, B., Zech, L., & Cognition and Technology Group at Vanderbilt. (1998b). SMART Environments that Support Monitoring, Reflection, and Revision. In D. Hacker, J. Dunlosky, and A. Graesser (Eds), *Metacognition in Educational Theory and Practice*. Mahwah, NJ: Erlbaum.