Contrasting Paths to Small-School Reform: Results of a 5-year Evaluation of the Bill & Melinda Gates Foundation's National High Schools Initiative

LINDA SHEAR
BARBARA MEANS

SRI International

KAREN MITCHELL

Association of American Medical Colleges

ANN HOUSE
TORIE GORGES
AASHA JOSHI

SRI International

BECKY SMERDON

Education Policy Center, Urban Institute

JAMIE SHKOLNIK

American Institutes for Research

Background/Context: In 2000, the Bill & Melinda Gates Foundation began an ambitious initiative intended to catalyze the fundamental transformation of American high schools. This article summarizes the results of a 5-year national evaluation of the first stage of the foundation's initiative.

Purpose/Objective: This article focuses on two contrasting strategies to small-school reform: starting new small high schools, and converting large schools into smaller learning communities. For each strategy, it reports on the progress of the reform in its first several years, student outcomes, and key implementation factors that shape progress and outcomes for start-up and conversion schools.

Research Design: The evaluation summarized in this article used a mixed-methods design. Data and analyses in this article come from surveys of teachers, students, and school leaders; case study site visits that collected a range of qualitative data; the collection and analysis of teacher assignments and student work; and district records of attendance and achievement.

Findings/Results: Results show that although both strategies have the potential to promote learning environments that are more personalized and that encourage students to work to higher standards, start-up schools in their first several years showed positive results in terms of attendance and some indication of student achievement gains, whereas these outcomes did not emerge for conversion schools during the timeframe of the study.

Conclusions/Recommendations: Results reported by the evaluation must be taken in the context of the typically slow pace of significant educational reform. Despite the many tasks facing start-up schools in their first 3 years and the challenges posed by limited budgets and incompatibilities with district systems, these schools were quickly able to establish strong and supportive school climates without apparent compromise to student academic achievement. Conversion school progress was slowed by an early focus on structural changes and the challenges of equitable reassignment of existing staff and students, with less initial clarity of vision for the learning environment or initial emphasis on instructional change. More evidence is needed concerning the long-term outcomes for students in these schools and the feasibility of creating small schools at scale, and therefore the degree to which either strategy is likely to support the ultimate goal of promoting educational excellence and equity for all high schoolers.

INTRODUCTION

This article describes the findings of a 5-year national study of the Bill & Melinda Gates Foundation's efforts to transform the American high school.

The last decade has seen numerous efforts to improve high schools, but the foundation's initiative is the grandest in scale. With an investment of over $1.5 billion by 2006, the foundation has sought to catalyze fundamental change in a wide variety of schools and districts across the country. The investment was spread among a number of grantees, each with

its own ideas about the paths that school reform should take and its own approaches to supporting school teams. As a result, this initiative offers an opportunity to look at the results that are associated with different paths to small-school reform.

This article focuses on two particular approaches: starting a new small school from scratch, or converting an existing large school into several smaller schools or "learning communities" within the same building. Because these two common small-school reform strategies are often associated with different implementation processes and different results, this article offers a side-by-side comparison of progress, outcomes, and key explanatory factors related to the implementation challenges of each.

The article begins with an introduction to the foundation's initiative and to the policy context in which it was situated, and an introduction to the methods used in the study. Then, for each of the two reform models, we ask the following questions:

For schools in the start-up or conversion sample, how far did the reform progress in the first several years of the initiative? Implementation progress is measured through survey scales that address whole-school aspects of reform, as well as an analysis of the rigor and relevance of teacher assignments collected in classrooms that suggests the depth to which the reform impacted instruction.

What student outcomes resulted from the early years of reform? Student outcomes are described through district records on attendance, progression rates, and scores on standardized tests, as well as a detailed analysis of the quality of student work submitted in response to teacher assignments.

How can key aspects of the start-up and conversion processes help to explain the differences in results? The process of implementation for starting a new school or converting an existing one brings opportunities and challenges that shape the pace and results of reform. Of the many critical factors at play, we focus on organization and staffing models, student recruitment/assignment, instructional change, and the local context of reform.

The article concludes with a comparative summary of progress and outcomes, and implications for future reforms.

BACKGROUND ON THE FOUNDATION'S INITIATIVE

In 2000, the Bill & Melinda Gates Foundation launched its nationwide program to reform American high school education. The foundation sought to catalyze the creation of a new kind of high school—one where every student, including those from low-income and minority backgrounds, would receive a challenging academic program providing sound preparation for college and for family-wage jobs. Tom Vander Ark, then the foundation's education director, summed up the foundation's vision of a new kind of high school in a 2001 article coauthored with Harvard's Tony Wagner: "Even those students most at risk of school failure can perform at very high levels, given the right conditions: much smaller schools, teacher teamwork, a personalized learning environment, and many more opportunities for applied and hands-on learning" (Wagner & Vander Ark, 2001).

To fulfill this vision, the foundation looked not to school districts or state departments of education but to nonprofit intermediary organizations outside the school system that collectively were given the charge to create hundreds of schools with these qualities. These intermediary organizations were awarded grants to establish high schools, ranging from as few as four schools to many as 60. Many of these grants went to organizations that pledged to start new "break-the-mold" schools. Some of these organizations had already established innovative small schools serving diverse student bodies, and grant funds would help replicate them across the country. These model schools—including The Met in Providence, Rhode Island; Minnesota New Country School in Henderson, Minnesota; New Tech High in Napa, California; and High Tech High in San Diego, California—embodied a progressive approach to secondary education, often following the tenets set by Ted Sizer and the Coalition of Essential Schools (see Coalition of Essential Schools, 2001). Other grants went to organizations supporting the restructuring of existing large comprehensive high schools as a set of smaller, more focused units (often called "small learning communities" or "academies") that would, in the foundation's vision, share many of the characteristics of the new start-up schools. The foundation suggested that schools supported under its initiative would be small in size (typically no more than 100 students per grade), not because "small" was an end in itself but because of the conviction that the greater personal attention made possible by a small size promotes implementation of effective schooling.

Specific educational models within the foundation's grantee portfolio varied, but all schools receiving foundation funding were expected to

embody a set of principles that the foundation came to describe as the "new 3 Rs":

- Rigor: A challenging academic program preparing all students for college, work, and citizenship. In support of academic rigor, for example, many initial school designs abolished tracking in an effort to eliminate the possibility of academic programs that came with low expectations for student success.

- Relevance: Studies that students find engaging and meaningful to their current and future lives. In many of these schools, it is common for students to work on projects with a real-world context or on internships to integrate their learning with the community outside the school. In some schools, students develop their own individual academic plans tailored to their interests and goals.

- Relationships: Close and supportive bonds between and among students and faculty. The small size of these schools helps to enable students and teachers to know each other well. School schedules often feature regular whole-school meetings, advisories, and other mechanisms to build school community and to make sure that each student has an adult advocate.

The timing of the foundation's initiative placed school creation and redesign efforts within a challenging regulatory and fiscal context in ways that had particular implications for schools with the above characteristics. The passage of the No Child Left Behind Act of 2001 (NCLB) increased accountability pressures and heightened focus on curriculum standards and teacher subject-area qualifications, making it more difficult for high schools to take a radically different approach to curriculum and instruction or to cover all subject areas with a small teaching staff (AIR/SRI, 2006). Simultaneously, the decline in state revenues and education budgets after the dot-com bust (Lav & Johnson, 2003; Pratt & Walsh-Sarnecki, 2004; Wheeler, 2005) resulted in scaled-back development of new buildings that could house innovative high schools and increased teacher layoffs, particularly for the new teachers whom many of these schools tended to hire (Shear et al., 2005). These contextual challenges would play an important role in shaping the progress of the initiative described in this article.

This article presents and compares the two core strategies of new high school start-up and comprehensive high school conversion. The section titled Evaluating the Initiative summarizes the range of methods and

measures used in the study, and introduces the student populations of the schools in the sample. The Start-up Schools: Starting Small, and Conversion Schools: Starting Broad sections examine for each strategy the markers of implementation and the outcomes that emerged during the timeframe of the evaluation, and seek a partial explanation of those outcomes by unpacking several key factors related to the character and challenges of the implementation process. Finally, the Conclusion section examines the role that each strategy can play in a comprehensive school change initiative and the relative promise of each toward a fundamentally improved high school experience for all students.

EVALUATING THE INITIATIVE

In 2001, the foundation contracted with the American Institutes for Research (AIR) and SRI International to study the processes and outcomes of their grantees' school-creation and redesign activities. The resulting 5-year evaluation began with the first set of high schools created under the initiative and added new samples of schools from each cohort of school openings or conversions between school years 2001-2002 and 2004-2005. Eventually the AIR/SRI data collection encompassed over 100 schools sponsored by 22 different organizations receiving foundation grants.

Throughout this article, school subsamples will be described in terms of the type of school (new "start-up" schools or "conversion" schools undergoing redesign) and in terms of the number of years elapsed since the school's opening or redesign. The number of years that a new school has been operating is particularly salient because most start-up schools opened with a single grade (typically ninth) and then added a grade each year, with the result that they did not have a cohort of graduating seniors until 4 years after their inception. This phasing in of operations at the school level, coupled with the fact that the grantee organizations received their grants in different years and staggered their schools' openings or redesign across multiple years (rather than doing all of them in the first year of their grant term), means that the data collected as part of the evaluation reflect the early stages of small high schools rather than a set of fully mature organizations.

METHODS AND MEASURES

The national evaluation used a mixed-methods design[1] that included four substudies, each with a focus on a particular type of data:

- *School survey substudy:* Surveys of teachers, students, and school leaders, and about school demographic information
- *School case studies:* Site visits that included interviews of school leaders and teachers, classroom observations, and student and parent focus groups, as well as interviews with district and grantee staff
- *Teacher assignment/student work substudy:* Collection of classroom artifacts
- *District substudy:* Extant data, including achievement test scores, attendance rates, and students' grade progression gathered from districts with multiple foundation-supported schools

Data were gathered from five types of high schools: *models* (existing high schools that served as models of some of the school designs that the foundation promulgated); *start-ups* (new high schools that opened with foundation support); *preconversions* (large high schools in which teachers and students were planning for later conversion to small learning communities or other small units); *conversions* (large high schools that were in the process of redesigning into these smaller units); and *comprehensives* (comprehensive high schools not affiliated with the initiative that were selected to serve as comparisons for foundation-supported schools).

Different subsets of the school sample participated in the various data collections that provided the input for the analyses summarized next. Extant data on indicators such as test scores, attendance, and grade-to-grade progression were collected from five districts with multiple foundation-supported schools that could provide student-level achievement test data so that student outcomes prior to high school entry (i.e., Grade 8 achievement and attendance) could be used as covariates.[2] Collections of original data (e.g., student surveys) were conducted in school samples meeting criteria for school type, geographical diversity, range of grantee organizations represented, and years since opening or conversion. The numbers of schools in which data were collected appear in Table 1.

Table 1. School Samples

Type of High School	Number in Survey Sample	Number in Site Visit Sample	Number in Teacher Assignment and Student Work Samples
Model schools	5	5	0
Start-up schools	29	26	12
Preconversion schools	16	11	12
Conversion schools	26	16	30
Comprehensive schools	3	0	8

This article describes a subset of the data that were collected in the overall study. In start-up schools, survey and site visit data were collected each year beginning with the year that the school opened. Some of the start-up school analyses reported here focus on third- and fourth-year start-up high schools, the schools with the longest reform history and for which there are comparison data from comprehensive and other large high schools with similar student populations. In large high schools undergoing conversion, data were collected in each school's preconversion year and again in the second, and in some cases the third, year following conversion.

The design of the evaluation was grounded in the foundation's theory of change (Connell, Kubisch, Schorr, & Weiss, 1995; Shadish, 1991; Weiss, 1995), a logic model that describes the foundation's goals for the initiative; key strategies for funding, grantee selection, support, and advocacy; and assumptions about the best levers for systemic educational change. The foundation's theory of change is described in AIR/SRI (2003). As articulated in 2001, the desired changes were motivated by a vision of schooling described in terms of seven attributes of effective schooling (personalization, a climate of respect and responsibility, high expectations, performance-based decision-making, technology as a tool, common focus, and time to collaborate) and three attributes of powerful teaching and learning (active inquiry, in-depth learning, and performance assessments). As the foundation's vision continued to evolve, these characteristics were later simplified to the 3 Rs: rigor, relevance, and relationships. Both articulations of the foundation's strategies are reflected in the design of this evaluation. Based on the theory of change, survey measures used in this evaluation described the implementation of new schooling models in terms of the attributes of effective schooling, with scales for each attribute[3] that were also summarized into a single measure called the implementation index. Following the foundation's later thinking, the analysis of teacher assignments within this evaluation focused primarily on measures of rigor and relevance.

Profiles of the students who attended these schools serve as an important context for the implementation progress and outcomes measured by this evaluation. The foundation targeted poor and minority students in its reform efforts, especially low-income African American and Latino/Hispanic students. In nine large urban districts in which the foundation worked and where district records were available to the evaluation team in 2004–2005, data for the majority of these districts show that enrollments of poor and minority students in foundation-supported schools exceeded the district average proportion of such students. Additionally, foundation-supported conversion schools tended to be

above district averages in terms of proportional enrollment of students with individualized education plans (IEPs) and English language learner (ELL) students, but proportions of ELL students and those with IEPs were somewhat lower for foundation-supported start-ups (AIR/SRI, 2006).[4] At the time of the evaluation, the supported schools did not appear to be excluding low-income or minority-background students, and many were actively recruiting for these populations.

The methods used in this evaluation, as well as the data sets examined for each of the results reported in the sections that follow, are described in greater detail in the Technical Appendix (Appendix B) to this article.

START-UP SCHOOLS: STARTING SMALL

The inventors of the start-up schools supported by this initiative were generally teams of extremely motivated teachers and school leaders who came together with the shared goal of creating a high school environment in which students would feel both academically motivated and personally supported to succeed. Frequently, these schools began with a single grade, allowing the envisioned culture to take root among just 60–100 people before the school grew to its target size, typically about 400 for Grades 9–12. Starting small allowed the daunting task of whole-school design to take place more gradually, with staff designing curriculum for just one new grade each year and delaying detailed requirements for the upper division program (Grades 11 and 12).

Start-up schools within the initiative often had a palpably distinctive character from the first days of their operation. Following the 3 Rs of rigor, relevance, and relationships, many of these schools were designed to embed student learning in projects, internships, and other authentic tasks; foster a strong schoolwide community; and support the academic success of every individual student. These features are illustrated in the descriptions of Oak Park and Franklin[5] schools in Appendix A.

At the time of the evaluation, the start-up schools in the sample were still very young; most had yet to graduate their first class. From this early vantage point, this section of the article looks first at the promise and then at the process of starting up a new student-centered school: What have they accomplished so far, and what does it take? In this section and the one on conversion schools that follows, we first describe the progress of implementation over the first several years of the initiative, then present student outcomes as evidenced over the same time period, and finally unpack several key implementation factors (organization and staffing models, student recruitment/assignment, instructional change, the local context of reform, and the trajectory toward sustainability) that

are particularly important to understanding the outcomes that result from each particular path to reform.

IMPLEMENTATION PROGRESS IN START-UP SCHOOLS

This section describes the early progress of foundation-supported start-up high schools. We begin by using first-year data to describe the early progress of implementation in start-up schools. We will then turn to data collected mainly in the schools' third or fourth year of operation to describe progress and outcomes that may take longer to emerge: the rigor and relevance of instruction; academic engagement as well as the reflection of that engagement in attendance and grade-to-grade promotion rates; and finally, the quality of students' work in class and on standardized achievement tests.

Evaluation findings on implementation and outcomes for start-up schools are summarized in Table 2; each finding is described in more detail in the subsequent text. The table also describes the source of the data on which each analysis is based. Data sets and levels of aggregation vary across these analyses because some use data available only from certain districts, whereas others are based on school surveys or interview samples (AIR/SRI, 2006, n.d.; Shear et al., 2005). Methods and data sets are described more fully in Appendix B.

Table 2. Summary of Start-Up School Findings

	Finding	**Data Source**
School cultures	First-year start-up schools were significantly higher than comprehensive schools on all seven survey-based school culture outcome measures.	Teacher and student survey data from 24 start-up schools in their first year of operation and 15 comprehensive high schools.
Instruction		
Rigor of instruction	The rigor of English/language arts (ELA) assignments given by teachers in start-up high schools was equivalent to that of assignments in nearby comprehensive high schools. The mathematics assignments given by teachers in start-up schools were more likely to be judged as having at least "limited rigor" than were assignments in comprehensive high schools.	Teacher assignments in Grade 10 ELA and mathematics classes in 12 new schools and 8 comparison schools.
Relevance of instruction	Both ELA and mathematics assignments showed significantly higher relevance in start-up than in comprehensive high schools.	Teacher assignments in Grade 10 ELA and mathematics classes in 12 new schools and 8 comparison schools.

Student engagement	Students in third-year start-up high schools reported significantly higher levels of academic interest and educational aspiration than did students in large high schools.	Student surveys collected in 18 start-up schools in their third year of operation compared with 15 large schools planning conversion and 7 large comparison schools drawing students from the same area as selected start-ups.
Attendance and progression rates		
Attendance rates	In three of four districts in the district data substudy that opened multiple new schools with the foundation's help, attendance rates were significantly higher in foundation-supported start-up high schools than in the districts' comprehensive high schools without foundation ties.	Four districts with multiple foundation-supported start-up schools provided attendance data for 2002–2003 and 2003–2004.
Progression rates	In the two districts that opened multiple new schools with the foundation's help and provided grade-to-grade progression data, start-up schools had significantly higher 9th-to-10th-grade progression rates than did the districts' comprehensive high schools.	Two of the four districts with multiple foundation-supported start-up schools provided grade-to-grade progression data.
Academic achievement		
Student work	In start-up high schools, 44% of students' ELA work was judged as being of moderate or substantial quality, compared with 29% of students' language arts work in comprehensive high schools. In mathematics, less than 20% of students' work was judged to be of moderate quality or better in both the start-up and comparison high schools.	Student work samples collected in Grade 10 ELA and mathematics classes in 12 start-up and 8 comprehensive high schools in 2004–2005.
Standardized achievement test data	In three of the four districts that were able to provide achievement test data at the student level, students in foundation-supported start-up high schools and those in comprehensive high schools had equivalent scores on standardized achievement tests in ELA and mathematics. In one of the districts, students in start-up schools had significantly higher scores in both subject areas.	Standardized achievement scores for foundation-supported start-up schools and comprehensive high schools in four districts able to provide student-level achievement data for 2002–2003 and 2003–2004.

Note: Table B1 (in Appendix B) describes the measures, unit of analysis, and analytic methods used.

School Cultures

Over the course of the evaluation, the AIR/SRI evaluation team col-
lected survey data from teachers and students in 24 foundation-sup-
ported start-up high schools that were in their first year of operation. The
surveys asked faculty and students to describe a number of aspects of
school culture: For example, they asked about the extent to which the
relationships between and among teachers and students in the schools
were sustained over time and marked by mutual respect, and the extent
to which all students were held responsible for reaching high academic
standards and given the needed supports to get there. Figure 1 compares
data from start-up schools with survey data gathered from teachers and
students in 15 comprehensive high schools that were planning for later
conversion (Shear et al., 2005).

Figure 1. Implementation Progress in Startup Schools
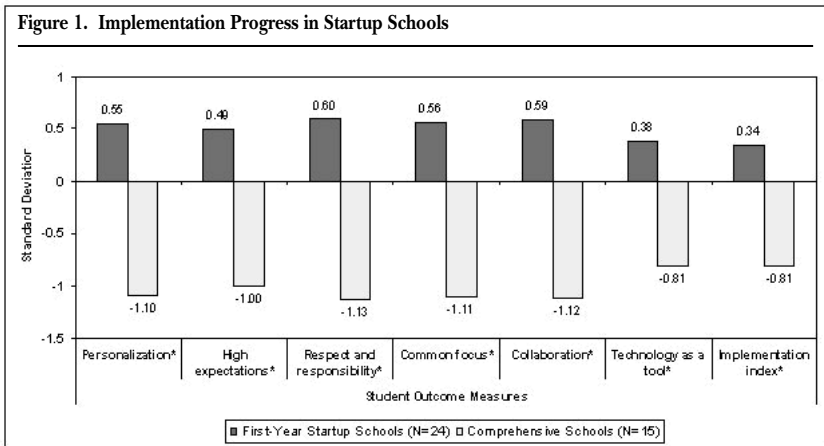
Notes: (1) Number of schools: Data for 24 start-up schools in their first
year of operation and 15 comprehensive high schools planning for con-
version. (2) Data source: Teacher and student survey data from
2001–2002, 2002–2003, and 2003–2004. (3) Analysis method:
Hierarchical linear models (HLM) in which values were computed as
standard deviations around the mean value for all schools in the data set
on each attribute, controlling for the school risk index, a measure of the
percentage of minority students and the percentage of students eligible
for free and reduced-price lunch in each school. (4) In this figure and in
all that follow, an asterisk (*) denotes significance.

In comparison with the comprehensive high schools in the evaluation sample, start-up high schools showed significantly higher levels of personalization, academic expectation, respect and responsibility, common focus, collaboration, and use of technology as a tool. As described earlier, these are fundamental elements of the initial concept of an effective school in the Bill & Melinda Gates Foundation's theory of change, on the assumption that a close, supportive, and respectful school environment is a strong precursor to student success (Bryk & Schneider, 2002; Coalition of Essential Schools, 2000; Lee, Smith, Perry, & Smylie, 1999; Sebring & Bryk, 2000). The data show that most start-up schools in their first year were successful in implementing a culture that students and teachers viewed quite positively compared with perceptions of climate in comprehensive high schools.

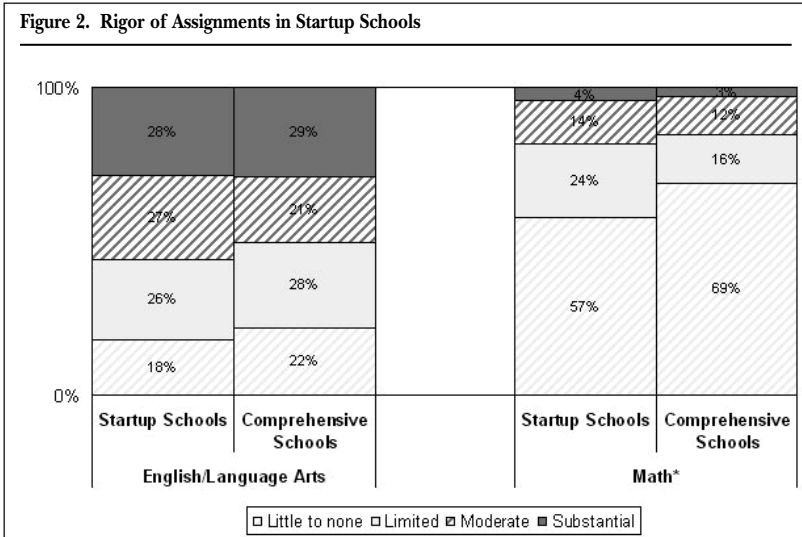*Rigor and Relevance of Instruction*

A second important precursor to better student outcomes is improved instruction. The evaluation team collected copies of the assignments that teachers gave in 12 third- and fourth-year foundation-supported start-up high schools (563 assignments from 71 teachers) and in eight nearby comprehensive schools with similar student populations (262 assignments from 38 teachers). The assignments were gathered from 10th-grade ELA and mathematics teachers.

After the assignments were collected, master teachers from different schools scored them on the extent to which they called on students to do rigorous and relevant work in ELA and mathematics. The scorers examined the extent to which the assignments required students to move beyond the reproduction of information that they had read, heard, or viewed to explore or create ideas that are new to them; to communicate clearly and well; and to use language and mathematics conventions accurately and effectively.

The percentages of ELA assignments that were judged to have substantial, moderate, limited, and little or no rigor are shown in Figure 2. The data on the rigor of teachers' assignments in ELA showed that the assignments that teachers gave in start-up high schools were as rigorous as those given by faculty in nearby comprehensive high schools. In ELA in both sets of schools, about half of the assignments were judged to have either moderate or substantial rigor.

In contrast, relatively few of the mathematics assignments that teachers submitted had moderate or substantial rigor. However, more of the mathematics assignments from start-up schools were judged to have at least limited rigor as compared with assignments from comprehensive schools.

The difference between the rigor of mathematics assignments in start-up and comprehensive high schools was statistically significant (AIR/SRI, 2006).

Figure 2. Rigor of Assignments in Startup Schools

Notes: (1) Number of schools: Data for 12 start-up schools in their third and fourth years of operation and for eight nearby comprehensive schools with comparable student populations. (2) Data source: Teacher assignments in English/language arts and mathematics in 2004–2005. (3) Analysis method: Descriptive statistics. Though the graphed results do not control for background factors, regression analyses (HLM) that controlled for student demographics, prior average classroom achievement, class composition, teacher characteristics, and assignment type (typical or challenging) showed the same results. The difference between the rigor of mathematics assignments in start-up and comprehensive high schools was statistically significant at $p < .05$.

After scoring assignments for rigor, master teachers scored the ELA and mathematics assignments on the extent to which they prompted students to do work with real-world purposes and for authentic audiences. Scorers also examined assignments for the extent to which they gave students choices about what they would learn in ELA and mathematics and how students would demonstrate mastery.

The data showed that ELA and mathematics assignments in these

schools were more likely to have real-world connections and to incorporate elements of student choice, compared with assignments given by teachers in the comprehensive high schools. About 30% of the ELA assignments submitted by faculty in new high schools were judged as having moderate or substantial relevance, compared with only 15% of assignments from faculty in comprehensive schools. These differences were statistically significant.

Though only small percentages of the mathematics assignments were judged to be relevant, more of the mathematics assignments from start-ups than comprehensive high schools had moderate or substantial relevance (20% vs. 4%). The difference between the relevance of teachers' mathematics assignments in start-up and comprehensive high schools was statistically significant (AIR/SRI, 2006).

These results are largely consistent with classroom observations of instruction in start-up high schools and with teacher interviews and student focus groups describing student learning opportunities. A number of the schools, like Oak Park and Franklin profiled in Appendix A, sought to make schooling relevant to students' lives: Topics were chosen to be of interest to students, and projects and internships frequently offered opportunities to interact with authentic challenges or people in the community (e.g., diagnosing a local environmental issue, or conducting interviews with World War II veterans). In math, some schools adopted newer curricula intended to make the subject more relevant to students, but a number of schools reported challenges in integrating standards-based math into cross-curricular projects, finding available curricula that provided both relevance and rigor, finding certified math teachers who were experienced in project-based pedagogies, or using these methods when students came to school underprepared in basic math skills (AIR/SRI, 2003, 2004; Mitchell et al., 2005).

STUDENT OUTCOMES

Outcomes for students are reported here on several measures: engagement as reported in student surveys, attendance and progression rates from district-level data, and academic achievement as reflected in the work that students produce in classes and their scores on achievement tests.

Student Engagement

One reason that many high school reformers stress the importance of relevance is the assumption of a positive relationship between relevant

learning of content and student engagement (Bransford, Brown, & Cocking, 1999; Marks, 1995; Newman & Associates, 1996; Newmann & Wehlage, 1995). In 2004–2005, the evaluation team collected data from students in 18 foundation-supported start-up high schools that were in their third year of operation. In surveys, students were asked to describe their levels of interest in schoolwork and their educational aspirations for the future. These data were compared with survey data from 22 large high schools that were planning for conversion or had similar student populations. The results are shown in Figure 3 (AIR/SRI, 2006).

Figure 3. Student Engagement and Educational Aspirations in Startup Schools
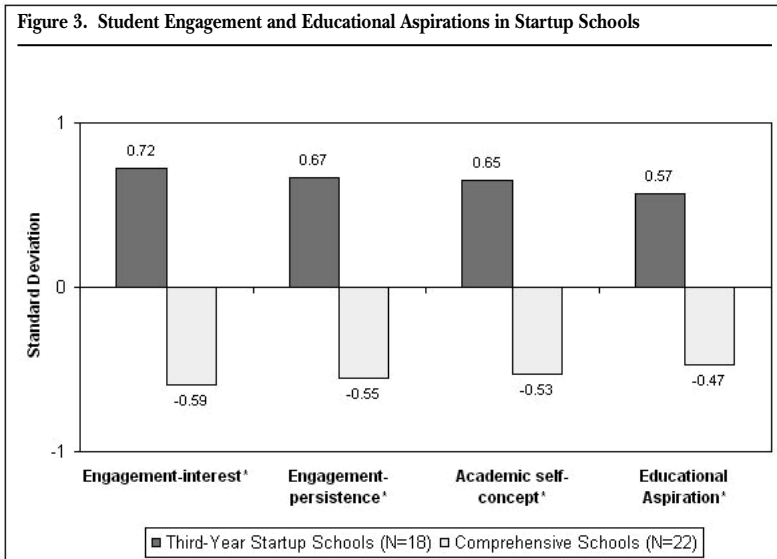
Notes: (1) Number of schools: 18 start-up schools in their third year of operation and 15 large high schools planning for conversion, pooled with data for seven nearby comprehensive high schools with comparable student populations. (2) Data source: Student surveys collected in 2004–2005. (3) Analysis method: HLM in which values were computed as standard deviations around the mean value for all schools in the data set on each variable, controlling for the school risk index, a measure of the percentage of minority students and the percentage of students eligible for free and reduced-price lunch in each school.

The data show that students in third-year start-up high schools reported levels of academic interest and educational aspirations that

were significantly higher than those of students in large high schools, even after adjusting for school demographic characteristics. In focus groups, students often attributed these higher goals to the individual attention and encouragement they received from their teachers. Said one student, "They say, 'I see you've done this and this, but I really think you can go deeper.' . . . If my adviser believes in me, I think I can really do it and strive higher and higher."

Attendance and Progression Rates

Student engagement in foundation-supported start-up schools is also reflected by students' attendance rates and their rates of promotion from Grade 9 to Grade 10. Attendance data were available in 2003–2004 for foundation-supported start-up high schools and comprehensive high schools without selective admissions policies in four of the five districts in the district data substudy. In three of those four districts, attendance rates were substantially higher in foundation-supported start-up high schools than in the districts' comprehensive high schools without foundation ties. These differences were statistically significant (AIR/SRI, n.d.).

The grade-to-grade progression data in foundation-supported start-up schools were also encouraging. In both districts in which the evaluation team had access to the data necessary to compute Grade 9-to-Grade 10 progression rates in start-up high schools and in comprehensive high schools, progression rates in start-up schools were significantly higher than the progression rates in comprehensive high schools.

Discussions with students and teachers rooted students' higher attendance and persistence in their newfound experience of school as a positive place to be and in the commitment of faculty to keeping track of each student with calls home when students were truant. According to one teacher, students often came to school an hour and a half before school started "because this is like their anchor in life, because their outside life for most of our kids is chaos." Said one parent of his child's improved attendance at this school over his previous one, "There's no place to hide here."

Academic Achievement

The evaluation measured academic outcomes through two primary lenses: the quality of student work, and achievement as measured on jurisdiction-sponsored standardized achievement tests.

Student work. Student work samples were gathered from 10th-grade ELA and math classes in the same 12 start-up schools (1,108 work

samples from 624 students) and 8 comprehensive schools (577 work samples from 301 students) in which teacher assignments were collected. Master teachers judged the quality of students' work by examining the extent to which the student products demonstrated the construction of knowledge, deep conceptual understanding of important content, reasoning and problem-solving facility, effective communication, and accurate use of language and mathematics conventions.

The scoring of students' work in ELA classes showed that the quality of student work in start-up high schools was significantly higher than that of students' work in the comprehensive high schools with similar student populations. In start-up high schools, 44% of students did work that was of moderate or substantial quality in ELA. In comprehensive high schools, only 29% were rated as highly.

In mathematics, the data were different. In both the start-up and comparison high schools, less than 20% of students' work was judged to be of moderate quality or better. This result is consistent with the instructional challenges described earlier in math. It is also important to note that this analysis rests on characteristics of student work quality, such as demonstration of knowledge construction and problem-solving skills, which would be difficult for students to exhibit in response to the traditional math assignments that are still prevalent in many schools (AIR/SRI, 2006). This result is also consistent with other small-school research that has suggested that math outcomes are more difficult to achieve and slower to emerge than outcomes in language arts (e.g., Wasley et al., 2000; see Figure 4).

Figure 4. Student Work Quality in Startup Schools
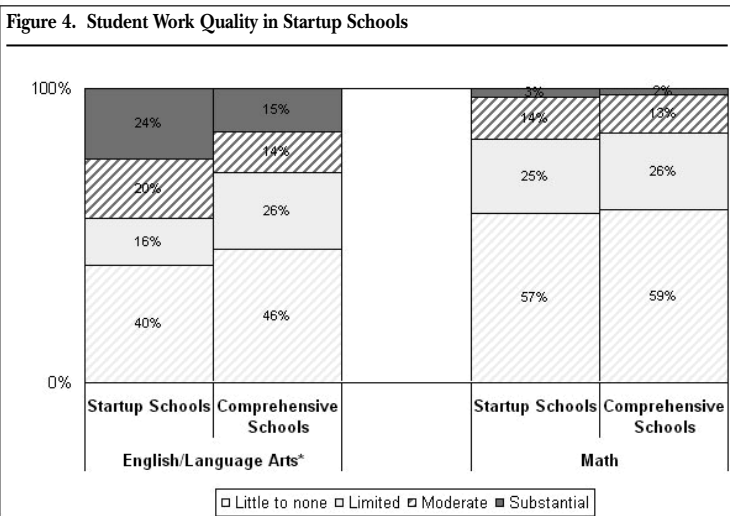
Notes: (1) Number of schools: Data for 12 start-up school in their third and fourth years of operation and for eight nearby comprehensive schools with comparable student populations. (2) Data source: Student work in English/language arts and mathematics in 2004–2005. (3) Analysis method: Descriptive statistics. Though the graphed results do not control for background factors, regression analyses (HLM) that controlled for student demographics, prior average classroom achievement, class composition, teacher characteristics, and assignment type (typical or challenging) showed that students in start-up high schools did higher quality work in the English/language arts than students in comprehensive high schools. The difference between the average scores in the two types of schools was statistically significant ($p < 0.01$). The differences between students' math work in start-up and comparison high schools were not statistically significant.

Standardized achievement test data. In 2003–2004, the evaluation team collected standardized achievement scores for 25 foundation-supported start-up schools in the four districts in the district substudy that had multiple foundation-supported start-up schools. ELA and mathematics scores were collected from jurisdiction-sponsored tests in the grade in which the tests were administered in each district, and data were examined on the score scale that the tests reported. Many of the tests were high school graduation tests administered in the 10th grade.

After controlling for differences between students' prior achievement and background characteristics, when test scores for start-up high schools were compared with those from comprehensive high schools, the data showed that students in start-ups did as well as or better than students in comprehensive high schools in ELA and mathematics. In one of the four districts, students in foundation-supported start-up high schools scored significantly higher than students in comprehensive high schools on standardized achievement tests in ELA and mathematics (AIR/SRI, n.d.). In the three other districts, the two sets of scores were not different statistically.

UNPACKING START-UP IMPLEMENTATION

The previous results show that although most of the start-up schools in the evaluation sample had not reached a steady state of exemplary academic performance by 2005, they had made a great deal of progress toward a supportive culture and engaged student body, with some early indicators of successful results for students and without compromising

student performance on standardized tests. This section describes the implementation process that the schools undertook to bring them to this juncture.

Rather than beginning with an existing school infrastructure and implementing a new vision, the process in most school reform strategies, start-up design teams begin with a vision and create a school around it. This opportunity to design a school from the ground up, though still subject to external constraints, enables school design teams to make decisions based on their concept of what's best for the students rather than on how things have been done in the past. The downside of this start-up strategy is the sheer number of decisions that must be made and implemented before the school doors open, ranging from selection or design of a building, to physical layout and bell schedules, to student and staff recruiting, academic programs, discipline policies, and mechanisms for instituting the desired culture.

Research has widely demonstrated a large number of factors that are of particular importance to the success of whole-school reform. For example, data from this study are consistent with other research that points to leadership capacity as a critical enabler (Daggett, 2004; Harris, 2002; Lambert, 1998; Shear et al., 2005). This study also points to a number of significant challenges, such as the difficulty of finding and funding an appropriately designed building (AIR/SRI, 2004) for new small schools in their early years. In this article, we have chosen to focus on several factors that are particularly helpful in contrasting the outcomes and experiences of start-up and conversion schools, specifically, organization and staffing models, student recruitment/assignment, instructional change, the local context of reform, and school evolution and financial sustainability.

Organization and Staffing Models

Because of funding limitations related to small school size (the number of pupils drives school funding formulae), it was rare for small start-up schools to be able to afford the level of administrative staffing common to larger high schools. As a result, the lines between administrative and teaching staff were often blurred in small schools: School leaders had responsibilities for teaching and/or advising students, and teachers were required to take on administrative tasks such as financial management. In some school models, this flatter organizational structure was a deliberate feature intended to empower teachers, who commonly described a newfound enthusiasm for the teaching profession that was derived, in the words of one teacher, from "creating a system that is making a difference

in these students' lives." However, distributed leadership added to work-loads; teachers had to simultaneously function as school designers, curriculum developers, and student advocates, in addition to their core teaching responsibilities (AIR/SRI, 2003, 2004). After they had been open for several years, some schools that had begun with a very collaborative model of school leadership chose to centralize administrative responsibilities in a more explicit school leader position to allow teachers to focus more of their time and energy on teaching.

Generally, start-ups were staffed by teachers attracted to the vision of serving students more closely, a factor likely to have contributed to the speed with which they succeeded in implementing strongly personalized school cultures. One school leader described his applicant pool as "young, inexperienced, enthusiastic staff who weren't in a rut." This enthusiasm was an essential qualification for teachers, yet staff turnover in the first few years was often considerable as new schools worked to find a team that would gel and to obtain teachers well-suited to the kind of teaching that their school models demanded. Many schools were also challenged to meet the requirements of academic certification, particularly with the advent of NCLB and in the more technical fields of math and science. Given the lean staff demanded by the schools' financial models, a single teacher often had to cover multiple subjects, leading to concerns about the need for multiple credentials and the challenge of obtaining those credentials while working full time.

Student Recruitment

One of the challenges of starting a new school from scratch is that there is no built-in student population. Said one school leader, "We're in the recruiting business, and that's a big difference from normal high schools." Most start-ups in the sample reported that student recruitment was a huge focus early in the life of the school until a reputation was established and that often the first class had to be recruited before there was an identified school building or a defined curriculum. Because school budgets are based on student head count, successful recruitment was an issue of survival, and some schools found themselves compromising school culture to get students, any students, in the door (AIR/SRI, 2004).

Many of the schools that received funding through the foundation's initiative had the explicit goal of reaching out to underserved students. As discussed earlier, some schools were successful in meeting their targets, but others found that districtwide lottery systems or other targeted

recruitment challenges restricted their ability to select as diverse a population as they had intended. It is still an open question whether this diversity will be maintained. Research on some small-school initiatives (e.g., Sadovnik & Semel, 2006) has found that successful small schools risk becoming boutique options for wealthier families who seek unique educational opportunities for their children. In the AIR/SRI study, staff at two third-year schools expressed some concern that their applicant pool was getting more affluent as the reputation of the school grew (AIR/SRI, 2006), but staff at other schools did not mention such a trend.

Instructional Change

With only a small number of teachers, small start-up schools cannot provide a curriculum that is as comprehensive as that of larger high schools. Instead, most schools in the evaluation sample endeavored to do a smaller number of things well. Some start-up schools were designed around a curricular theme, such as science and technology or social justice. Others stated their focus as "college preparatory," emphasizing the core subjects of math, science, and humanities and offering courses similar to those in the academic track of larger high schools but with fewer choices—for example, a single foreign language option rather than three or four. A third set of schools (like Franklin, profiled in Appendix A) had an entirely student-driven curriculum, so areas of focus might vary according to the interests of the particular student. Other common curriculum characteristics of start-ups included mixed-ability groupings, multidisciplinary courses, project-based learning, or internships to increase the relevance of students' academic programs. These curriculum features contributed to the progress on curriculum relevance reported earlier for start-up schools.

   As with other aspects of start-up for these schools, creating a curriculum from scratch was both an opportunity to develop a unique program guided by a vision and a daunting requirement that continued for the first 4 years (for the schools adding one grade each year). As suggested by the ratings of the quality of teachers' assignments, start-up schools had varying degrees of success in their first several years in implementing a curriculum that met the ideals of intellectual demand and compelling subject matter. Some schools received strong curriculum development support from their grantee and other partner organizations or were required to use a district-mandated curriculum, but in many cases, curriculum development was a responsibility of the teachers. Said AIR and SRI in 2003, "To the extent that we heard about effective and innovative

curriculum modules, it was often attributed to the skill and judgment of the teaching staff, and to their experience as curriculum designers, a demanding expectation for newer teachers" (p. 22).

Local Context Of Reform

Within the initiative, an important aspect of local context for schools was the Gates grantee, the intermediary organization that was charged with selecting and supporting the schools that would receive funding through the initiative. In addition to direct funds, the supports offered by these organizations ranged from process supports and leadership coaching, to curriculum and professional development, to running interference with the district or state; these roles, and schools' perceptions of their value, are described in more detail in AIR/SRI (2004, 2006).

The start-up schools supported by this initiative had varying relationships to their districts. Some were district schools, subject to the same regulations and receiving the same supports as other schools within the jurisdiction. Although some district relationships were supportive, many schools were forced to work within a district context in which policies and curriculum mandates were not developed with these school models in mind. For example, at the time of a spring site visit, one new start-up was facing pending layoffs for four of its five teachers because of seniority-based reassignment policies within the district and, as was common in these schools, a younger-than-average teaching staff (AIR/SRI, 2003). High turnover was particularly threatening to school models based on close multiyear relationships between students and staff.

Some grantees instead chose to operate charter schools outside the purview of the district. These grantees were able to provide supports more tailored to the needs of small start-ups than typical district policies permitted. For example, one school leader credited the grantee organization with providing both the logistical support and the overall vision that took a great deal of the start-up burden off the school staff. However, points of interaction with the districts still existed (e.g., student recruitment and special education services), and schools that were seen as outsiders found that districts could be less than supportive. Whether working within or outside the district, grantees found it necessary to select districts carefully for placement of their schools and to develop and maintain close relationships with district staff.

Another critical aspect of local context for any school is the community it serves, including both families and local businesses. For many start-up schools in the evaluation sample, community partnerships were essential to success on a number of levels, both financial and programmatic. One

school partnered closely with a leading local university that offered strong professional development and a corpus of student teachers; another whose focus was at-risk students with dropout histories worked in partnership with a local social services organization whose staff social workers provided personalized social support for students. Other common models for partnerships were universities that offered courses on their campuses, and local businesses that provided internship sites or mentors as well as specialized equipment and resources for thematic courses.

School Evolution and Financial Sustainability

As described earlier, most new start-up schools followed a staged implementation strategy, starting with just one or two grades and adding another grade each year. Nurtured by very small initial student and teacher populations, school cultures are often successfully implemented in the schools' first year (see Figure 1). However, staff at many second-year start-ups found that these nascent communities could be challenged with the addition of new teachers and students. Individual attention gets more difficult with larger numbers of students, and discipline issues often emerged more strongly in the second year. Some school staff experienced the development of factions, with founding teachers and newcomers sometimes differing on school vision. As a result, many schools experienced a "second-year dip" in the implementation of important cultural attributes (AIR/SRI, 2004; also noted in Borman, 2005; Kahne, Sporte, & de la Torre, 2006). Although there is some indication that cultures, on average, moved toward stability in Year 3 (Shear et al., 2005), ongoing stresses remained, including the overall workload of staff, particularly given the need to develop new curriculum each year for the school's first 4 years. Some schools were still struggling with how to make workloads sustainable at the end of their third year of operation.

A more fundamental challenge for start-ups was school finances. Initially, schools relied on funding from the foundation and on other start-up grants to carry out their operations. As those income sources began to wind down, schools found that the average daily attendance funds (ADA) brought in by their small student populations often were insufficient to maintain unique theme-based course offerings, up-to-date technologies, or even daily operations (AIR/SRI, 2006). A few schools resorted to traditional cost-cutting measures such as increases in class size, which jeopardized the close teacher/student relationships on which these schools' models were founded. Others relied on creative strategies, including the addition of grant-writing resources to school staff,

increased promotion of community partnerships, the addition of middle school grades to increase head count, or the formation of two small schools under one administration to save on overhead. Although the continued existence of the more mature model schools in the evaluation sample suggests that sustainability of these student-centered schooling models is possible, the start-up schools in the sample were just beginning to grapple with life after foundation start-up funding and were still working toward both design and financial equilibrium by the final year of the evaluation.

CONVERSION SCHOOLS: STARTING BROAD

Looking to bring the benefits of a small, rigorous high school to larger numbers of students than can be reached by small start-ups, the foundation funded some intermediary organizations to work with school districts to change the structure of existing comprehensive high schools. The charge of these intermediary organizations was to create a set of small schools from the staff and students comprising the large school. These are variously referred to as "conversion" or "redesigned" schools. Conversions, in contrast to the new schools described earlier, begin with an existing infrastructure, including a physical facility, staff, students, and curriculum. The challenge they face is in restructuring this infrastructure into small, autonomous units. (We will call these smaller units "small learning communities," or SLCs, though some schools use terms such as *houses*, *schools*, or *academies*.)

Reorganizing a large school into multiple SLCs is an avenue through which other changes in teaching and learning are expected to take place. According to one school principal, conversion is "not just physically separating students for the sake of separating them into SLCs," but making changes to the school's "business as usual" ways of educating students. The reason that large high schools plunge into this process (or are plunged into the process by other decision makers) is to reap the benefits of an environment that can provide students with rigorous and relevant learning experiences supported by strong relationships between students and staff. Two schools undergoing conversion with the goal of changing what and how students learn are described in Appendix A.

Typically, schools spend the year or two prior to conversion planning how to redesign the comprehensive school's existing features and resources into SLCs. Design teams usually comprise teachers, school administrators, district staff, and, in some cases, external consultants or

parents. As they work to carve several SLCs out of a large comprehensive school, these design teams need to make numerous decisions. Some of the most basic of these include how many SLCs to create (typically, the goal is to have three to five SLCs of about 400 students each); whether to convert all at once or phase in the reform; how distinct the SLCs will be in terms of thematic emphases and teacher/student assignment; whether to use an existing reform model or design the SLCs from scratch; and how much the large-school context, as reflected in image, sports teams, and common resources, will remain after conversion.

Although design teams begin the planning effort in the preconversion year(s), the conversion schools in the national evaluation sample demonstrated that redesign is a continuing, multiyear process. Important design changes that continued through the evaluation period ranged from efforts to improve the efficiency of SLC operation, such as new bell schedules and physical relocation of SLCs, to more fundamental design issues, such as dropping or changing SLC themes. In one case, a school chose to reduce its SLCs from four to three for reasons of equity and redistributed teachers and students among the remaining SLCs.

Often the schools targeted for conversion were those with existing struggles, such as high dropout rates, low attendance rates, and/or low achievement scores. In addition, most students enrolled in these schools lagged behind their peers in terms of prior measured achievement. Within this challenging context, schools were asked to devise and quickly implement entirely new ways of operating. Initial findings from the AIR/SRI evaluation suggest that conversion schools did not show the same promising early outcomes as start-up schools within the time frame of the evaluation (AIR/ARI, 2006).

CONVERSION PROGRESS IN THE FIRST SEVERAL YEARS

The AIR/SRI evaluation data are more limited for conversion schools than for start-up schools because there was a smaller number of conversion schools in the study sample, and the timing of their conversion was such that they were followed for fewer years in their restructured state. The seven large comprehensive schools surveyed formed 27 SLCs, 26 of which had a sufficient survey response rate for inclusion. Outcomes for conversion schools are summarized in Table 3 and described in the text that follows (AIR/SRI, n.d.; Shear et al., 2005; Shkolnik et al., 2007).

Table 3. Summary of Conversion School Findings

	Finding	Data Source
School cultures	Three of the six effective school attribute scales (personalization, high expectations, and respect and responsibility) showed significant improvement between the year before conversion and the second year of operation as SLCs, but the summary index of attribute scales did not change significantly.	Teacher and student surveys from 7 high schools in the year of conversion planning and 26 small learning communities created from them.
Instruction		
Rigor of instruction	No change was evident in the rigor of either English/language arts (ELA) or mathematics assignments before and after conversion.	Teacher assignments in ELA and mathematics collected from 12 large schools, and 2 years later from 30 of the SLCs that resulted from their redesign.
Relevance of instruction	The relevance of both ELA and mathematics assignments rose significantly after conversion.	Teacher assignments in ELA and mathematics collected from 12 large schools, and 2 years later from 30 of the SLCs that resulted from their redesign.
Student engagement	Students' educational aspirations in the SLCs were unchanged from levels 2 years earlier, before school conversion. Levels of student engagement dropped significantly after conversion.	Teacher and student surveys from 7 high schools in the year of conversion planning and 26 SLCs created from them.
Attendance and progression rates		
Attendance rates	Student attendance rates were equivalent before and after conversion.	Three of the five districts in the district data substudy had conversion schools and provided attendance data for 2002–2003 and 2003–2004.
Progression rates	Grade-to-grade progression rates were equivalent before and after conversion.	Two of the five districts in the district data substudy had conversion schools and provided grade progression data for 2002–2003 and 2003–2004.
Academic achievement		
Student work	The quality of students' work in ELA was significantly higher after school conversion than it had been 2 years earlier. The quality of students' work in mathematics declined significantly over the same period.	Student work in ELA and mathematics collected from 12 large schools, and 2 years later from 30 of the SLCs that resulted from their redesign.
Standardized achievement test data	Students' ELA and mathematics test scores were equivalent in conversion schools and in nonselective comprehensive high schools in the same district.	Standardized achievement scores from three districts with multiple conversion schools able to provide student-level data for 2003–2004.

*Note:* Table B1 (in Appendix B) describes the measures, unit of analysis, sample sizes, and analytic methods used.

School Cultures

Although the summary index of the effective school attributes did not change significantly between the year before conversion and the second year of operation as SLCs, three of the attribute scales (personalization, high expectations, and respect and responsibility) did show significant improvement, suggesting that some aspects of school climate had improved (Shear et al., 2005). Moreover, the qualitative data also pointed to an improvement in school culture, suggesting that grouping staff and students into smaller learning units produced a more positive social environment. As one student put it, "I really like with small schools how my teachers know me on a one-on-one basis, and they know what I'm good at and what I'm not good at." Teachers also mentioned improved relationships with students and improved teacher professional communities. As one teacher described, "The professional community has changed. Teachers are connecting . . . and they're looking to each other to think about what we could do to teach a class together." Closer relationships, efforts to improve teaching and learning, and a more positive climate were strong themes in the qualitative data from SLCs in their second and third years of conversion.
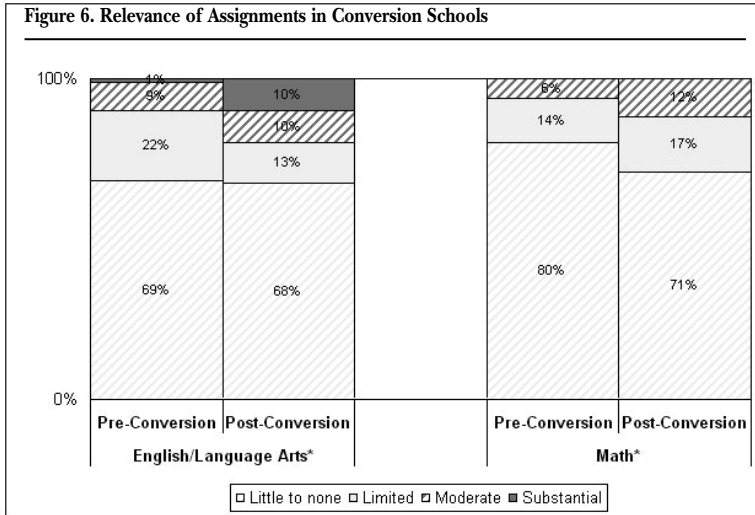
Rigor and Relevance of Instruction

As in start-up schools, the evaluation team collected teacher assignments and student work in 10th-grade English/language arts (ELA) and math classes in conversion schools. Samples came from 30 SLCs created out of 12 large schools (including schools from the national sample and eight preconversion high schools and resulting SLCs from Washington state, where the foundation began its work).

Little change was evident in the rigor of assignments before and after conversion. In both ELA and mathematics, the data showed a small but nonsignificant increase in the rigor of assignments after conversion. Overall, more than half the ELA assignments showed substantial or moderate rigor, whereas math assignments showed very little rigor both pre- and postconversion (Shkolnik et al., 2007; see Figure 5).

Figure 5. Rigor of Assignments in Conversion Schools
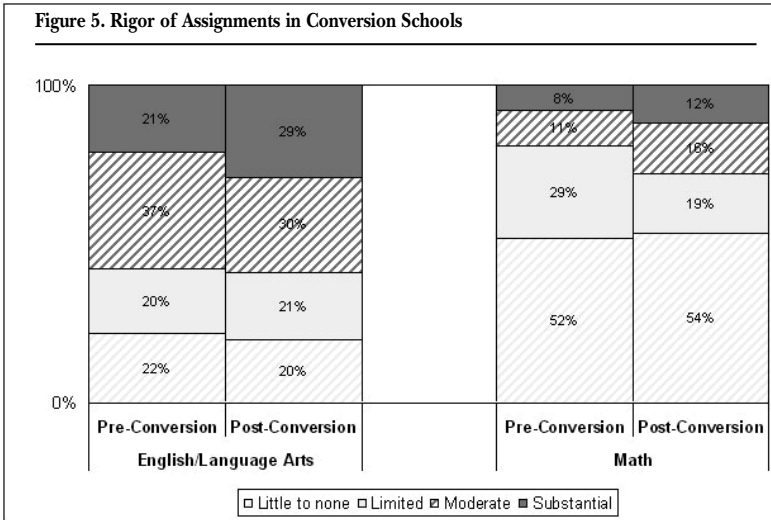
Notes: (1) Number of schools: Data for 12 preconversions and 30 resulting SLCs. (2) Data source: Teacher assignments in English/language arts and mathematics in 2004–2005. (3) Analysis method: Descriptive statistics. Though the graphed results do not control for background factors, regression analyses (HLM) that controlled for student demographics, prior average classroom achievement, class composition, teacher characteristics, and assignment type (typical or challenging) showed the same results.

The findings regarding assignment relevance were more promising at these schools: Both ELA and math assignments showed a statistically significant increase in assignment relevance after conversion ($p < 0.01$). However, assignments with substantial or moderate relevance remained the exception rather than the rule in ELA and math classes, whether before or after redesign (Shkolnik et al., 2007; see Figure 6).

Figure 6. Relevance of Assignments in Conversion Schools

| | Little to none | Limited | Moderate | Substantial |

English/Language Arts*

Pre-Conversion: 69%, 22%, 9%, 1%
Post-Conversion: 68%, 13%, 10%, 10%

Math*

Pre-Conversion: 80%, 14%, 6%
Post-Conversion: 71%, 17%, 12%

Notes: (1) Number of schools: Data for 12 preconversions and for 30 resulting SLCs. (2) Data source: Teacher assignments in English/language arts and mathematics in 2004–2005. (3) Analysis method: Descriptive statistics. Though the graphed results do not control for background factors, regression analyses (HLM) that controlled for student demographics, prior average classroom achievement, class composition, teacher characteristics, and assignment type (typical or challenging) showed the same results. The difference between the relevance of ELA and math assignments between pre- and postconversion was statistically significant ($p < 0.01$).

STUDENT OUTCOMES

Student Engagement

Student ratings of their own educational aspirations in the 26 SLCs in the survey sample, when compared with their original levels before conversion, did not change between pre- and postconversion. Levels of student engagement, however, showed a significant drop (AIR/SRI, 2006). Early in the process of the fundamental reorganization undertaken by these schools, the existing programs that students had come to expect at their high schools were disrupted before improvements in teaching and learning had taken hold. Unlike students in start-up schools, who chose to attend a new school knowing that there would be some growing pains,

students in schools converting to SLCs may focus as much on what is lost in the process of change (e.g., reduced elective choices or sports) as on what is gained (e.g., stronger relationships with teachers).

Attendance and Progression Rates

In three of the five districts in the district data analysis substudy, student attendance data were available for SLCs. Attendance rates from all three of these districts showed no significant changes between comprehensive and SLC schools. In the two districts for which grade-to-grade progression data for both SLCs and comprehensive schools were available, there was no significant difference between rates in the two types of schools (AIR/SRI, n.d.).[6]

Academic Achievement

Student work. Changes in the quality of student work in conversion schools varied by subject. In ELA, there was a statistically significant increase in quality after school conversion ($p < 0.01$). Mathematics student work quality, in contrast, showed a statistically significant decrease postconversion ($p < 0.01$). Students in conversion schools, then, demonstrated improvement in one but not both areas investigated in this evaluation (Shkolnik et al., 2007; see Figure 7).

Figure 7. Student Work Quality in Conversion Schools
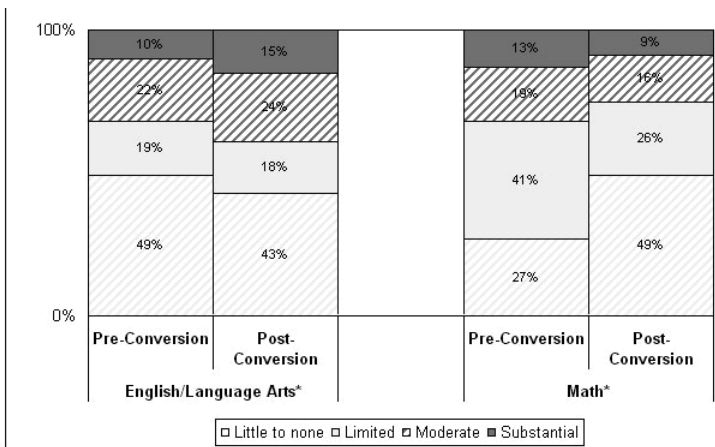
Notes: (1) Number of schools: Data for 12 preconversions and for 30 resulting SLCs. (2) Data source: Student work in English/language arts

and mathematics in 2004–2005. (3) Analysis method: Descriptive statistics. Though the graphed results do not control for background factors, regression analyses (HLM) that controlled for student demographics, prior average classroom achievement, class composition, teacher characteristics, and assignment type (typical or challenging) showed the same results. The increase in the quality of ELA assignments and the decrease in quality of math assignments pre- and postconversion were both statistically significant ($p < 0.01$).

   Qualitative data from school site visits were examined to suggest explanations and provide context for some of the findings of the teacher assignments and student work study—particularly, the finding that student work quality declined in mathematics but not in ELA, despite at least comparable assignment rigor after conversion and increased relevance of assignments in both subjects. The data suggest that there were more staffing challenges in mathematics than in ELA, with more teacher turnover, higher numbers of first-year teachers, and more use of long-term substitutes. In addition, although new curricula introduced in some schools emphasized project-based learning and would be likely to be highly rated for rigor and relevance, this new model of teaching and learning requires a learning curve for teachers to deliver it effectively, as well as for students to express their knowledge in new ways. Interviewees also reported that it was often difficult to integrate math and science into the SLCs' themes. Finally, many of the schools reported that their early curricular initiatives were focused more on ELA and literacy than on math (Shkolnik et al., 2007).

   Standardized achievement test data. After controlling for differences between students' prior achievement and background characteristics, the average ELA and mathematics test scores of students in the SLCs in the three districts providing student-level achievement data showed no significant differences from average test scores in the comprehensive high schools in the same district (AIR/SRI, n.d.). Although some other research has found improvements in student achievement 2, 3, and 5 years after large high schools restructure (Darling-Hammond, Ancess, & Ort, 2002; Lee & Smith, 1995), the results reported here suggest that the first 2 years of the conversion process did not produce a measureable impact on student academic achievement for the schools in this study.

UNPACKING CONVERSION IMPLEMENTATION

Early outcomes from conversion schools show that the rearrangement of staff, students, and classes helped SLCs build a somewhat stronger sense

of school culture but did not produce gains in academic work or district measures of school performance after 2 years. Qualitative data suggest that the very practical issues of space, staff, students, and classes tended to overwhelm concerns of pedagogy and curriculum reform in the SLCs' first 3 years (similar to the findings of Riley & Fouts, 2003). This section discusses these pragmatic issues in high school conversion.

Organization and Staffing Models

The physical organization of multiple SLCs within a facility designed to house one large high school was a struggle for most schools early in the conversion process (AIR/SRI, 2004). Creating separate SLCs within these settings generally meant assigning each SLC to a separate floor or wing of a large building. This proved difficult because specialized rooms like science labs tend to be grouped into one floor or wing and because of the need to share single locations such as the gym or cafeteria. The physical redesign process tended to be a dominant focus early in the conversion process; it was not until the second or third year of conversion that most of the schools reported fewer ongoing structural challenges and therefore a strengthened ability to focus on small-school culture and vision.

Another important aspect of converting a school into SLCs is establishing leadership and administrative structures. Typically each SLC had a designated leader, but the degree of autonomous decision-making power held by the SLC leader varied considerably. In the case of Logan (profiled in Appendix A), a whole-school principal maintained ultimate control of the complex, but each SLC was run by an assistant principal, a lead teacher, and a counselor dedicated to those 500 students. At the other end of the spectrum, one school created SLCs that were housed in separate locations from the original whole school and led by SLC directors. The expectation was that as the large school was phased out, there would no longer be any interconnection between these SLCs.

Many of the challenges faced by SLCs related to the reorganization of their teaching staffs. In distributing teachers across SLCs, schools in the site visit sample generally tried to honor teachers' SLC preferences, which in some cases created imbalances among the SLCs. For example, at Logan (see Appendix A), most of the teachers in the International Baccalaureate (IB) program ended up in one SLC, which in turn led to a concentration of resources and academically advanced students in that small school. Balancing of qualified staff in all subject areas was also an issue: Some schools did not have enough accredited teachers in all needed subjects for every SLC and had to blend students from several SLCs in the same class, or assign a teacher to teach in several SLCs. As

noted in the AIR/SRI 2004 report, "some administrators and staff felt that this 'bleeding' of students and teachers into classes outside their SLCs compromised the 'purity of independence' between small schools that they were trying to achieve" (p. 40). Finally, many teachers were initially frustrated by being reassigned from their academic department structure to collaborate with colleagues from different disciplines. Nevertheless, by 2005, the majority of teachers in the sample described the reorganization as an exciting and positive experience, citing more frequent and deeper collaborations, and more engagement in school governance as reasons for enthusiasm (Shear et al., 2005).

Student Assignment

In contrast to start-up schools, conversion schools did not have to worry about recruiting students for their first class,[7] but their already-existing student populations were not necessarily informed about or involved in the conversion process, and student assignment to SLCs came with its own challenges. Ready and Lee (2006) found that offering students the choice of small-school membership resulted in stratifications of students: Those who were more highly motivated chose the more challenging small schools, and vice versa. Similarly, concerns about small-school equity arose as a result of student distribution in the schools in this sample. Most conversions allowed students some degree of freedom in choosing their SLC while also attempting to balance gender and ethnic composition, special education needs, and low and high achievers across all the SLCs. These efforts were not always successful: In one school, most of the coaches and student athletes concentrated into one SLC as an unintended result of their common choice, and another with a preexisting magnet program was considering redistributing students to equalize the balance of top students and teachers across the SLCs by its second year after conversion. A school that took a different route clustered its English language learners into one SLC, with the rationale that this would better serve this population by concentrating all the resources available to support them.

Instructional Change

School conversions were generally seen as an opportunity for making radical improvements to academic offerings. Unlike start-up schools, which had to develop a curriculum from scratch and generally tried to do so based on an articulated vision, SLCs often had only vaguely articulated curricular goals in the early years of the conversion as planning teams

devoted more focus to structural and organizational changes. Many of the curricular goals expressed by conversions related to equity of academic offerings; strategies included the elimination of tracking, the use of differentiated instruction, and broader access to IB courses. Implementation of curricular changes was typically a growing focus over the first several years of the conversion, and few schools developed answers as early as their planning year. This pace of curricular change is consistent with other school conversion efforts, in which "principals and teachers agreed that the first year emphasis was on launching the schools and they hoped to be able to focus more on teaching and learning in year two" (Kahne et al., 2005, p.14).

Like start-up schools, conversion schools often intended to use a thematic design for their SLCs, but few SLCs accomplished a deep implementation of themes and matching curricula by the end of the evaluation. Some had themes in name only; more typically, the SLC had managed to create a handful of courses specific to their theme but had yet to develop a full, thematic curriculum. The practical issues of conversion, the increase in staff workload as staff took on additional roles within their SLCs, and the ongoing conversion design process left little time available for curriculum development. As academic offerings within an SLC became more focused, either around a theme or simply because of the smaller number of teachers, the range of courses available to students shrank. Both students and parents noted the loss of the honors and advanced courses, arts, and other electives that had been available in the large school. In some cases, a backlash against the perceived reduction in offerings undermined support for the conversion process.

Local Context of Reform

Intermediary organizations receiving Bill & Melinda Gates Foundation grants to promote high school conversions were less likely to provide a whole-school model for design and curriculum than grantee organizations creating start-ups—an understandable difference given that converting schools tended to preserve many existing elements of their school rather than starting fresh with a blank slate.[8] In addition, among conversion schools in the AIR/SRI sample, the motivation for change tended to be from districts, either as part of a large-scale district reform effort or in response to very low individual school performance. As a result, grantee/school relationships tended to be somewhat more distant for conversions than for start-ups, but grantee work within districts tended to be more active. Of the 12 intermediary organizations in the evaluation sample that supported conversion schools, 10 provided professional

development or coaching to districts in how to organize and support conversion efforts within their jurisdiction. In two cases, grantees chose to move someone on their own staff into the district organization as a more direct means of providing support for converting schools.

The schools' close ties to their districts led, at times, to implementation challenges. For example, some districts required the group of SLCs to provide one whole-school leader to interface with the district even though the whole school had ostensibly been replaced by the set of smaller units. Others required budget and hiring to be managed as a whole school, and some were bound to information and other data systems that did not have the ability to identify and organize student records by SLC. In these ways, district procedures and policies sometimes undermined the SLC effort, even in cases in which the district's top leaders expressed support.

Another contrast to start-up schools is in the role of the families of students. Start-up schools typically worked with a self-selected group of families drawn to the school's planned design, whereas the comprehensive schools undergoing conversion served an existing group of families who often did not choose this change. Decomposition of a large, often longstanding high school into unknown and unproven SLCs is a dramatic change, and it is not surprising that conversion met with mixed feelings among parents and other community members. For parents of students who had been involved in advanced classes and athletic, art, and other specialized programs that were dropped, conversion was seen as removing opportunities for their children. Moreover, within the community, the conversion of a high school into separate SLCs was, for some, the loss of a community institution. When a student would no longer graduate from the school that her father had attended, or when community members could no longer gather for local football games, conversion represented the end of an institution that had long provided social connection.

Sustainability of Conversion

Because the conversion process begins with operating schools that already have full complements of staff, students, and physical facilities, there is little risk that the new schools formed through conversion will disappear entirely. For conversion schools, the sustainability issue relates more to the ability of the SLCs to maintain distinct identities, improvement processes, and autonomy. Over the course of the study, one converted school changed back into a large school structure when its district chose to stop supporting the reform work of the SLCs, ostensibly because

of a lack of academic improvement. Other schools in the sample cited district funding cutbacks and a loss of district-level support for reform as reasons for concern about SLC sustainability. One school leader facing this type of challenge said, "I have this huge fear . . . that they'll change our bell schedule to make us be more economically efficient. . . . The question of staffing is huge. The easiest way to staff for the district is to change the bell schedule and allow crossover," essentially ending the design of distinct small schools.

Whereas some other high school reinvention work has shown success 3 and 5 years after redesign into smaller schools (Darling-Hammond et al., 2002), the reform trajectory for the schools in this conversion sample exhibits a shallower slope than for the start-up schools described earlier in this article. The structural changes that moved staff and students into smaller units produced some noticeable differences in the relationships among staff and students, but the extent to which this increased personalization will impact teaching, learning, and ultimately student achievement remains to be seen, testing the patience of funders and stakeholders for whom these deeper changes are of primary concern.

CONCLUSION

Critics of America's education system have long lamented the system's seeming imperviousness to change, and the high school has been widely cited as the least malleable layer of K–12 education (Hess, 2004; McLaughlin & Talbert, 2001). The Bill & Melinda Gates Foundation's high school initiative tested the prevailing pessimism, applying substantial resources and catalyzing the efforts of scores of education reformers and social entrepreneurs. Although most of the schools created or converted with support from this initiative were still in their early years at the conclusion of this study, it is nevertheless appropriate to ask what has been learned thus far about the initiative's two basic strategies: the creation of brand-new small schools and the conversion of existing schools into smaller entities.

The two strategies of school creation and school redesign share the goal of providing a rigorous, relevant educational program in a setting marked by strong, supportive relationships, but their trajectories toward achieving the desired end state are clearly divergent. Despite the many tasks facing start-up schools and the challenges posed by limited budgets and incompatibilities with district systems, these new schools have been able to provide a very different and demonstrably more supportive climate for their students. This climate has been accompanied by stronger student engagement and by markedly better attendance.

Nevertheless, as a whole, these schools have not produced the hoped-for significant improvements in achievement results for students to date. Controlling for prior student achievement and demographic variables, the AIR/SRI evaluation found better achievement test performance in both ELA and mathematics in one district for students in the new small schools, but in three other districts, scores for students in start-up schools and in conventional high schools were essentially equivalent. The study did, however, find indications of promise. An examination of the work produced by 10th graders, for example, found higher quality work in ELA in start-up schools than in the large schools that their students likely would have attended if the new school had not been established.

The advantages of starting fresh are highlighted when the start-up school experience is compared with that of school conversion efforts. Although improvement in some of the indicators of school climate (personalization, high expectations, and respect and responsibility) was found between the year before school conversion and the second year after conversion, these improvements did not translate into a more positive climate across the board. Student engagement actually declined during the 3 years that the conversion schools were followed. The study found no improvement after conversion in student attendance or grade-to-grade progression. In terms of achievement test scores, there was no change in the three districts where appropriate achievement measures and data on prior achievement and student demographics were available for analysis. Findings from the study of student work collected before and after school conversion were mixed: The quality of students' ELA work went up, but their mathematics work became weaker.

The qualitative data emerging from the conversion schools, moreover, paint a picture of extensive effort going into structural change and assignment of students and staff to different smaller units with less emphasis, in these years, on the design and quality of curriculum and instruction. Although qualitative data captured enthusiasm for the more positive schooling environment among some students and teachers in the SLCs, a mindset of loss was evident among others.

Prior research on school reform efforts suggests that school improvements typically take more than 3 years to appear (Borman, 2005), and it is possible that the school conversions will simply require more time to achieve deeper differences in school climate, student engagement, and achievement. But the early data from this study are also compatible with the interpretation that the conversion change model is insufficiently powerful. Although this strategy can reorganize large numbers of teachers and students into smaller learning communities in a relatively short amount of time, reorganization per se does not appear to have all the

accompanying benefits, in terms of rigor and relationships, that the foundation had hoped for.

Start-up schools enjoy the tremendous advantage of being able to start with a relatively clean slate, giving them the opportunity to develop coherent school visions, structures, and practices. Start-up schools were more likely than SLCs to have a dynamic leader who could attract others to the school vision. By recruiting staff and students who are attracted to the new school's distinctive character, a start-up school leader can create community cohesion to a degree that is difficult to achieve among teachers and students who are comparing a new SLC to the offerings and facilities available in the large school from which the SLCs were created.

In summary, we judge the experiences of this sample of start-up schools as grounds for optimism concerning the feasibility of establishing public secondary schools that can engage and challenge students from a broad range of backgrounds. Clearly these small, mission-driven high schools create a very different social environment, and there are promising signs of their potential to make good on higher academic expectations as well. What is less clear from these data is the feasibility of creating enough such schools to substantially improve the educational experiences of historically underserved groups and to sustain both the existence of these schools and their unique schooling models at scale.[9]

The Bill & Melinda Gates Foundation initiative has provided existence proofs for the premise that it is possible to "break the mold" of the large comprehensive high school. More evidence is needed concerning the long-term outcomes for students in these schools and the feasibility of creating them at scale, and therefore the degree to which this strategy is likely to support the ultimate goal of academic success for students of all backgrounds. The 5-year study described in this article documents an important, but formative, period in this ongoing pursuit.

Notes

1.   More extended presentation of the analytic findings and details of the samples, methods, and analyses can be found in the series of reports produced by the AIR/SRI evaluation team (AIR/SRI, 2003, 2004, 2006; Mitchell et al., 2005; Rhodes et al., 2005; Shear et al., 2005) available at http://www.gatesfoundation.org/UnitedStates/Education/ResearchAndEvaluation/Evaluation/NHSDGEvaluation.htm.

2.   District data were requested from four other districts with multiple foundation-supported schools. Three districts were not used in the analysis because they did not have any nonselective comprehensive high schools during the years for which data were being analyzed. One district was excluded because it could provide data for only a single year.

3.   We report scales for six of the attributes. Performance-based promotion is not

included because the survey measure for this attribute was unreliable and poorly correlated with the other attributes and with the implementation index.

4. Because they offered a highly personalized education environment, some start-up schools were actually attracting large numbers of students with IEPs (AIR/SRI, 2003). Their small staff sizes, however, often meant that they had limited ability to provide specialized supports to students with learning disabilities or with limited English, and a few schools even chose to refer these students to schools that would be able to serve them better, which may explain the lower numbers of special needs students than would be expected based on districts' overall enrollment patterns.

5. Throughout this article, the names of schools are pseudonyms.

6. It should be noted that for schools that phased in conversion, beginning with lower grades or limited SLCs, these data come from very initial stages of conversion. The perceived improvement in school culture and in relationships between teachers and students does not yet seem to have impacted rates of student attendance or grade-to-grade progression in the first several years of conversion.

7. Roughly half of the conversion schools in the AIR/SRI sample reported engaging in student recruitment activities with eighth graders in their district. Various recruitment strategies included a small school fair where SLCs provided informational booths on their design and theme, as well as sending SLC leaders and students to local middle schools for promotion presentations. Because SLCs are made up of existing students from a district, however, the student recruitment efforts were much more limited than those of start-up schools.

8. There were several notable exceptions among the conversion schools in the evaluation sample. In particular, one school chose to convert into several fully autonomous small schools that were each designed according to a national school model, including those of organizations funded under this initiative, such as the Big Picture Company. In this case, the school was working with multiple grantees: one that funded the overall conversion process, and others that each supported the design of a particular resulting small school.

9. Districtwide effects must also be considered in any analysis of feasibility at scale. Other research on wide-scale reforms, for example, has reported negative effects on conventional large high schools that are not involved in the school opening or restructuring efforts (Saulny, Moynihan, & Fisher, 2005).

References

AIR/SRI. (2003). *High time for high school reform: Early findings from the evaluation of the National School District and Network Grants Program.* Washington, DC: Author.

AIR/SRI. (2004). *The National School District and Network Grants Program: Year 2 evaluation report.* Menlo Park, CA: SRI International.

AIR/SRI. (2006). *Evaluation of the Bill & Melinda Gates Foundation's High School Grants Initiative: 2001–2005 final report.* Washington, DC: American Institutes for Research.

AIR/SRI. (n.d.). *Evaluation of the Bill and Melinda Gates Foundation's High School Grants: Achievement studies: 2003-2004 demographic and student outcome information in six districts.* Unpublished manuscript.

Borman, G. D. (2005). National efforts to bring reform to scale in high-poverty schools: Outcomes and implications. *Review of Research in Education, 29,* 1–27.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (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*. Chicago: Consortium on Chicago School Research.

Bryk, A. S., & Schneider, B. L. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.

Coalition of Essential Schools. (2000). *CES School Benchmarks*. Oakland, CA: Author.

Coalition of Essential Schools. (2001). *Principles at work*. Oakland, CA: Author.

Connell, J. P., Kubisch, A. C., Schorr, L. B., & Weiss, C. H. (Eds.). (1995). *New approaches to evaluating community initiatives: Concepts, methods and contexts*. Washington, DC: Aspen Institute.

Daggett, W. R. (2004). *Reforming American high schools: Why, what, and how*. New York: International Center for Leadership in Education.

Darling-Hammond, L., Ancess, J., & Ort, S. W. (2002). Reinventing high school: Outcomes of the Coalition Campus Schools Project. *American Educational Research, 39*, 639–673.

Harris, A. (2002). Effective leadership in schools facing challenging contexts. *School Leadership and Management, 22*, 15–26.

Hess, G. A. (2004). Districtwide strategies of high school reform: The case of Chicago. In *Using rigorous evidence to improve policy and practice: Colloquium report* (pp. 55–78). New York: MDRC.

Kahne, J. E., Sporte, S. E., & de la Torre, M. (2006). *Small high schools on a larger scale: The first three years of the Chicago High School Redesign Initiative*. Chicago: Consortium on Chicago School Research.

Kahne, J. E., Sporte, S., & Easton, J. Q. (2005). Creating small schools in Chicago: An early look at implementation and impact. *Improving Schools, 8*, 7–22.

Lambert, L. (1998). Building leadership capacity in schools. Alexandria, VA: ASCD.

Lav, I. J., & Johnson, N. (2003). *State budget deficits for fiscal year 2004 are huge and growing*. Retrieved March 11, 2005, from Center on Budget and Policy Priorities Web site: http://www.cbpp.org/12-23-02sfp.htm

Lee, V. E., & Smith, J. B. (1995). Effects of high school restructuring and size on early gains in achievement and engagement. *Sociology of Education, 68*, 241–270.

Lee, V. E., Smith, J. B., Perry, T. E., & Smylie, M. A. (1999). *Social support, academic press, and student achievement: A view from the middle grades in Chicago*. Chicago: Consortium on Chicago School Research.

Marks, H. M. (1995). *Student engagement in the classrooms of restructuring schools*. Washington, DC: U.S. Department of Education.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago Press.

Mitchell, K., Shkolnik, J., Song, M., Uekawa, K., Murphy, R., Garet, M., et al. (2005). *Rigor, relevance, and results: The quality of teacher assignments and student work in new and conventional high schools*. Washington, DC: American Institutes for Research.

Newmann, F. M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass.

Newmann, F., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report*. Chicago: Consortium on Chicago Schools Research.

Newmann, F. M., & Wehlage, G. G. (1995). *Successful school restructuring: A report to the public and educators by the Center on Organization and Restructuring of Schools*. Madison, WI: The Center on Organization and Restructuring of Schools.

Pratt, C., & Walsh-Sarnecki, P. (2004). Detroit schools' shortfall worsens. *Detroit Free Press*. Retrieved March 11, 2005, from http://freep.com/

Ready, D. D., & Lee, V. E. (2006, April). *Choice, stratification, and schools-within-schools.* Paper prepared for the annual meeting of the American Educational Research Association, San Francisco, CA.

Rhodes, D., Smerdon, B., Burt, W., Evan, A., Martinez, B., & Means, B. (2005). *Getting to results: Early outcomes in new and redesigned high schools.* Washington, DC: American Institutes for Research.

Riley, S. C., & Fouts, J. T. (2003). *The Bill & Melinda Gates Foundation's Model School Initiative, Cohort 1: Year 3 summary report.* Mill Creek, WA: Fouts and Associates.

Sadovnik, A. R., & Semel, S. F. (2006, April). *Curriculum and pedagogic practices, leadership, and sustainability in small schools: A comparison of Central Park East Secondary School and North Star Academy Charter School.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Saulny, S., Moynihan, C., & Fisher, J. (2005, January 4). City cites gains in combating strife in schools. *The New York Times.* Retrieved January 4, 2005, from http://select.nytimes.com/gst/abstract.html?res=FA0916F7395D0C778CDDA80894DD 404482&showabstract=1

Sebring, P. B., & Bryk, A. S. (2000). School leadership and the bottom line in Chicago. *Phi Delta Kappan, 81,* 440–443.

Shadish, W. (1991). *Foundations of program evaluation: Theories and practices.* Newbury Park, CA: Sage.

Shear, L., Song, M., House, A., Martinez, B., Means, B., & Smerdon, B. (2005). *Creating cultures for learning: Supportive relationships in new and redesigned high schools.* Washington, DC: American Institutes for Research.

Shkolnik, J., Song, M., Mitchell, K., Uekawa, K., Knudson, J., & Murphy, R. (2007). *Changes in rigor, relevance, and student learning in redesigned high schools: An evaluation for the Bill & Melinda Gates Foundation.* Washington, DC: American Institutes for Research.

Wagner, T., & Vander Ark, T. (2001). A critical fork in the road. *Education Week.* Retrieved March 5, 2005, from http://www.edweek.com/ew/ewstory.cfm?slug=30wagner.h20

Wasley, P., Fine, M., Gladden, M., Holland, N., King, S., Mosak, E., et al. (2000). *Small schools: Great strides. A study of new small schools in Chicago.* New York: The Bank Street College of Education.

Weiss, C. H. (1995). Nothing as practical as a good theory: Exploring theory-based evaluations for comprehensive community initiatives for children and families. In J. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65–92). Washington, DC: Aspen Institute.

Wheeler, K. (2005). *Cleveland school board continues process of cutting budget tonight.* Retrieved March 11, 2005, from WKYC-TV Web site: http://www.wkyc.com/news/news_fullstory.asp?id=28318

APPENDIX A: SCHOOL SNAPSHOTS

These vignettes describe the design, culture, and implementation challenges of two start-up schools (Oak Park and Franklin) and two conversions (Logan and Western). Although the names are pseudonyms, the descriptions are of actual schools from the study's site visit sample. They are offered here as a brief illustration of the variety of schools represented in this study and some of the ways that the design goals of the "3 Rs" are instantiated in practice.

START-UP 1: OAK PARK HIGH SCHOOL

Oak Park is a start-up school in a large West Coast city that opened in 2004–2005 to serve a diverse mix of students, primarily African American and Hispanic. Its school leader is experienced in small-school start-up and is adamant about taking the process step by step, beginning with a thoroughly entrenched school culture. "It takes five years to build a culture," he says. "The real culture of a school shows up when [the current class] becomes seniors." The strategy, he says, is to invest in this first class of students as role models so that the values of safety, responsible behavior, and academic excellence are transmitted from peer to peer.

During the spring 2005 site visit, the 100+ students of the future first graduating class were all freshmen, and the school had had about 8 months to build a culture. People and policies at the school consistently encouraged students to take collective responsibility and to support each other. Students and teachers described a recent community meeting (a twice-weekly gathering of all faculty and students for open discussion of any issues that affect the school community) at which two girls stood up in front of their peers to take responsibility for the graffiti in the bathroom and commit to repainting the stall the next day; the time when $100 was stolen from a teacher and each student paid $1 to return the money to her; and the way pending fights were nearly always diffused by another student before they began. "That's what I like," said one student, "it's like a safe community."

With regard to academics as well as to personal integrity, expectations at Oak Park were high. "Here they don't just give you the grade," said a student, "you need to earn it." In their 90-minute blocks for each core subject, students worked on projects within extended units that were integrated across several subjects, like the apartheid unit that spanned literature, biology, and European and African history, though school staff acknowledged that integration across subjects was not yet consistent nor inclusive of all subjects. Curriculum was primarily homegrown, devel-

oped by teachers in collaborative sessions that mapped standards in each subject to possible cross-curricular themes shaped by student-generated questions about the topic. Each unit culminated in presentations to teachers, parents, and peers at an exhibition night event, typically attended by all students: Students told us that these were group presentations, and even those who hadn't been coming to classes regularly did not want to let their project teams down.

Despite these accomplishments, Oak Park faculty were struggling with a number of difficult school design issues. (In Oak Park's distributed leadership model, one teacher said, "the majority of the decisions are made by the teachers together.") At the beginning of the year, workloads were unmanageable, leading teachers to devise a modified school schedule with fewer electives to allow teachers more planning time—an improvement, interviewees agreed, but workload remained a significant challenge. In teacher meetings, faculty grappled with whether to expel students whose behavior problems detracted from the community and the learning environment, and whether low-performing students would be better served in a targeted math class despite the essential school design principle of mixed-ability classes.

Despite these challenges, early signs were promising. After just 1 year, the school already had 200 students on its waiting list. Although a first graduating class was still 3 years away, families were attracted to Oak Park's growing reputation as a safe place to learn.

START-UP 2: FRANKLIN HIGH SCHOOL

Franklin is a start-up school located in the downtown section of a large West Coast city. The school serves a diverse mix of students reflective of the area (White, African American, Asian, and Latino). Roughly 75% of students are ethnic minorities, and 44% qualify for free or reduced-price lunch.

The school opened in fall 2002 with only a ninth-grade class and then added a new grade in each of the next 3 years. In its first year, Franklin enrolled 32 ninth graders. The entire school was housed in two classrooms and an office in a community college, a site that was confirmed only 2 weeks before the school's start date. The school's small staff (which in its first year included a principal who also served as adviser to some of the students, one teacher, and a full-time office administrator) ran the school from its one office, the administrative heart of the school. The office was the school's staff meeting room, its library, its breezeway for student and staff traffic, its lunchroom, and its student work display case. To establish more space for a 10th-grade class and beyond, the

school later moved into a district building. By the time of the spring 2005 site visit, the school had three grades, 100 students, and nine teachers and administrators.

The primary instructional feature of Franklin, patterned after a national grantee model, is a student-directed approach to learning in which students select their own internships and projects rather than participating in a common course curriculum. Internships and projects are fueled by student curiosity and interests. They are negotiated with advisers and selected on the basis of academic rigor and complexity, as well as on the basis of relevance to the student's particular learning goals. For example, work that students presented to family, school staff, mentors, and peers at one end-of-semester exhibition night included the creation of a local community map as a precursor to an analysis of the effects of local family migration patterns on school enrollment, and lessons learned about curriculum development by a student in her internship teaching sign language to 3–5-year-olds at a childhood development center. An internship coordinator position that was created in the school's second year proved essential to placing students in appropriate internships that would allow them to fill this academic requirement, and the school is growing its list of strong community partners and other volunteers who offer internships and provide mentoring.

To scaffold academic experiences, the school groups students into the same 16- or 17-student advisories from the ninth grade until graduation. As advisers, the teachers and the principal help students construct individual high school learning plans and navigate their individual academic paths. Advisories are designed to create a personal learning environment in which trusting relationships are built and strengthened between advisers and students and among the students themselves. Advisories meet as a full group 3 days a week for 90 minutes in the morning, and students meet individually with their advisers in structured conferences throughout the week.

Although the school has the second highest pass rate in the district on standardized exams, Franklin's principal worries about getting enough academic rigor into students' programs: "The danger of this model is to become seduced by personal growth and not by academic growth." The school's mission is to graduate 100% of its students ready for college, a goal made more challenging by the fact that the primary university system in the state does not recognize internships and projects as valid alternatives to traditional coursework. In response, in its first several years Franklin added a formal literacy program, options for students to enroll in community college courses while in high school, and some structured courses.

CONVERSION 1: LOGAN HIGH SCHOOL

In spring of 2003, Logan High School, located in a large Midwestern city, had a population of approximately 1,400 students. Of these, 55% were eligible for free or reduced-price lunch. The student body was approximately equally divided between White and non-White students. Logan's conversion is part of a strategy to convert all the district's comprehensive high schools into SLCs, with the goal of increasing relevancy and rigor for all students.

Logan's conversion process was intentionally slow: In 2002, the principal noted that it would be 2006 before teaching and learning would have significantly changed at the school. The school originally planned to open six SLCs, but in 2003–2004, four theme-based communities were opened. These four were limited to 10th and 11th graders; that year's graduating class was not part of the conversion process, and that year's ninth graders were in a separate program to introduce them to high school and to all the SLCs. Curriculum specific to each SLC was to be developed during the opening year rather than the year prior to conversion.

Logan experienced some challenges during its first year as a converted school. There was a great deal of crossover in which students and staff from one SLC participated in or taught classes in another SLC. This was particularly true for International Baccalaureate (IB) and other specialized courses because there were not enough teachers in each SLC to offer these subjects in all SLCs. Also, because of a lack of staff now that they were divided into four SLCs, some classes were as large as 50 students. During Logan's first year operating as SLCs, course offerings, curriculum, and classroom activities had not yet changed, and students were not entirely aware that a conversion had taken place. Staff and physical location had not yet been designated for each SLC. The advisory system (in which students have a period with an adviser, somewhat similar to homeroom), however, was put into place in each SLC during the first year.

In its second year as a converted school, Logan appeared to have smoothed out most of its processes. Each of the SLCs had a separate hallway, with different wall colors and banners. The school leader noted that "having teachers closer together is better" because teachers could communicate with their SLC teams more easily. Teachers were also pleased with the change; one said, "It feels more like a home." The developing teacher community was further assisted by grant funding, which provided pay for teachers to collaborate and plan together over the summer. One SLC instituted a successful team teaching program in English and social

studies, which the other SLCs looked to as a model. Also helping to strengthen the SLCs was the expansion to include all grades, as opposed to just 10th and 11th grades.

CONVERSION 2: WESTERN HIGH SCHOOL

In the spring of 2003, Western High School was serving approximately 1,200 students in a midsize New England city. Just over half the students were considered low income, and about half were non-White. Western's conversion efforts built on its experience with a successful health sciences magnet program that had been in operation for 15 years prior to the conversion, and with ninth-grade teams that had been used to create a more personal environment for incoming students. The school wished to expand on these successes, reaching the entire student body. Western converted the magnet program into an SLC in 2002–2003 and opened two additional SLCs in 2003–2004 (though they had originally planned to open all three in the same year). All the SLCs were theme-based: health sciences, technology and business, and social systems and justice.

By the spring of their first conversion year, although not all the SLCs had been formed, Western had put into place several critical components of its conversion. Teaching and learning were becoming more relevant and personalized for students thanks to senior-year internships and a looping practice that kept students with the same teacher(s) for multiple years. By its second year of conversion, all three SLCs had been formed, and the school's advisory program was beginning to take off, with teachers using the AVID (Advancement Via Individual Determination) program to "know the person as a kid, not just as a student," according to one SLC leader. Although seniors did not seem to like advisory much, ninth graders were more enthusiastic.

Western administrators never intended for the three SLCs to be autonomous, as is the goal for many conversion schools. Staff told site visitors that Western "as a whole will retain its school identity." This posed a challenge for the staff as they worked to create and strengthen SLC identities. Large amounts of student crossover between SLCs further tested staff's ability to develop a team atmosphere within each SLC.

Western also faced challenges with equity across SLCs. Because the magnet program had more years of experience and a reputation that allowed it to draw strong students, Western's staff had to consider how to equitably distribute students across the SLCs. They planned to consider racial composition, special education needs, and achievement levels, while also taking into account whether a student was already enrolled in the magnet program. The two new SLCs struggled to provide internships

for seniors, which the former magnet program had done thanks to an existing partnership with a medical center. By 2006, one of the new SLCs had developed a partnership with a local business, but the other still had not yet secured this type of affiliation.

Western also faced a major facilities challenge. The 1,200 students are housed in what was originally intended to be a junior high for 800 students. The building was in poor condition; one parent called it "an abomination." As of 2004, the school had been approved for a new building, but they were on a waiting list and remained unsure when the new building would be built. Despite these challenges, the administration and staff tried to create a "home" for each of the SLCs within the building.

APPENDIX B: TECHNICAL APPENDIX

AIR and SRI collected data for the evaluation of the High School Grants Initiative from fall 2001 through spring 2005. The evaluation used mixed methods to study the foundation-supported schools' progress, outcomes, and challenges, relying on four major sources: school surveys, site visits, teacher assignments and student work, and district measures, each explained in the following text. Table B1 displays the specific data sources used in this report.

Table B1. Measures and Analyses

Measures	School Types	Academic Years	Comparisons Made	Analysis Methods
Implementation of school-based supports for personalization, high expectations, respect and responsibility, common focus, collaboration, and technology as a tool	- Start-up schools in their first year of operation - Conversion schools in their second year as SLCs	2001–2002 2002–2003 2003–2004	- 24 start-up vs. 15 comprehensive high schools planning for conversion - 26 small learning communities (SLCs) in conversion high schools vs. 7 original high schools in their planning year	HLM with adjustment for school demographics
Teachers' assignments in English/language arts and mathematics classes	- Start-up schools in their third and fourth years of operation - Converting schools the year before and 2 years after conversion	2002–2003 2003–2004 2004–2005 2005–2006	-12 start-up vs. 8 nearby comprehensive schools with comparable student populations -12 preconversion schools vs. 30 SLCs created from them	HLM with adjustment for prior student achievement and student and school demographics

Student engagement in schoolwork and educational aspirations	- Start-up schools in their third year of operation - Conversion schools in their second year as SLCs	2004–2005	- 18 start-up vs. 15 large high schools planning for conversion pooled with 7 nearby comprehensive schools with comparable student populations - 26 small learning communities vs. 7 original high schools in their planning year	HLM with adjustment for school demographics
Average daily attendance and 9th-to-10th-grade progression	- Start-up schools in their first, second, and third years of operation - Conversion schools in their first and second years as SLCs	2002–2003 2003–2004	- Start-up vs. comprehensive high schools in the three districts in the district data substudy with both types of schools and the capability to provide grade progression data - Conversion vs. comprehensive high schools in the districts in the district data substudy with both types of schools and the capability to provide attendance data (three districts) or grade progression data (two districts).	Significance tests of mean differences in school-level data
Student work in English/language arts and mathematics classes	- Start-up schools in their third and fourth years of operation - Converting schools the year before and two years after conversion	2002–2003 2003–2004 2004–2005 2005–2006	-12 start-up vs. 8 nearby comprehensive schools with comparable student populations -12 preconversion schools vs. 30 SLCs created from them	HLM with adjustment for prior student achievement and student and school demographics
Achievement data in English/ language arts and mathematics from jurisdiction-sponsored tests	- Start-up schools in their first, second, and third years of operation - Conversion schools in their first and second years as SLCs	2002–2003 2003–2004	- Start-up vs. comprehensive high schools in the four districts in the district data substudy with start-up schools - Conversion vs. comprehensive high schools in the three districts with conversion schools	HLM with adjustment for prior student achievement and student and school demographics

SCHOOL SURVEYS

The AIR/SRI team used as a framework the attributes identified by the foundation as critical for schools to effectively serve low-income and other historically underserved students. School surveys were administered to students, teachers, and principals in spring of each year of the study. Schools were added to the sample on a rolling basis, with several exceptions. Model schools were only surveyed once, in the first year of the evaluation. Conversion schools were surveyed in the year before the conversion (in their large-school state) and then in each year after their conversion year, skipping their first year of conversion (see Table B2; AIR/SRI, 2006, p. 91).

For analysis, only schools with adequate teacher and student survey response rates were considered. Adequate response rates for schools in the selected sample were defined as at or above 50%, with at least one of the two response rates (teacher or student) at or above 60%. Adequate response rates for comprehensive schools were defined as at or above 40%, with at least one of the two response rates (teacher or student) at or above 50%. For a comprehensive description of the school surveys and their corresponding analyses, see AIR/SRI, 2006.

Table B2. Survey Sample

	2002			2003			2004			2005		
	Response rate criteria			Response rate criteria			Response rate criteria			Response rate criteria		
School type	Met	Not Met	Total	Met	Not Met	Total	Met	Not Met	Total	Met	Not Met	Total
Model	5	0	5	0	0	0	0	0	0	0	0	0
Start-up	7	2	9	21	0	21	27	1	28	22	1	23
Comparison	0	0	0	0	0	0	3	2	5	4	1	5
Preconversion	8	1	9	2	1	3	4	1	5	2	1	3
Conversion	0	3	3	0	0	0	26	1	27	1	0	1
Total	20	6	26	23	1	24	60	5	65*	29	3	32

*Note: The large increase in the number of schools surveyed from 2003 to 2004 was due largely to the fact that each preconversion school in 2002 had broken into multiple redesigned schools by 2004.

SITE VISITS

Site visits were conducted across 4 years of the study at selected surveyed schools, as well as in several of the grantee organizations and districts with which schools were affiliated. Table B3 shows number of schools and type (model, start-up, preconversion, or conversion), as well as grantee organizations and districts, visited in each year of data collection

(AIR/SRI, 2003, 2004, 2006; Mitchell et al., 2005; Rhodes et al., 2005; Shear et al., 2005).

Table B3. Site Visit Sample

Site	2001–02	2002–03	2003–04	2004–05
Model Schools	5	0	0	0
New Schools	8	15	22	23
Preconversion Schools	6	2	4	4
Conversion Schools	0	4	4	8
Total Schools	19	21	30	35
Districts	6	12	21	17
Grantee Organizations	12	14	30	22

Site visits were generally conducted by two-person teams and lasted 2–4 days. School visits included interviews with school leaders, student focus groups, parent focus groups (in selected schools), interviews with teachers, and classroom observations. Interviews focused on such topics as mission, vision, governance, learning environment, relationships between teachers and students, teacher community, and ability to serve students. Classroom observations were conducted using structured observation forms to code instructional activities and teacher and student actions, as well as a narrative description of activities. Site visitors also completed two structured inventories of the school that described the school's physical environment and school design components. At grantee organizations, site visitors typically interviewed 3–5 staff members, including the principal investigator for the grant. Interviews included topics such as vision, supports for schools, organizational capacity, plans for activities with schools, and barriers encountered. For districts, site visitors typically interviewed the superintendent or deputy superintendent, as well as the reform manager or high school director. Areas of focus included community context for schools, districtwide reforms, requirements for schools, and district support for and perceptions of foundation-supported schools (AIR/SRI, 2004).

Following site visits, site visitors wrote up data capture forms, which were entered in an ATLAS.ti database. A team of trained coders coded the text to sort it by topic, using a coding manual created on the basis of the foundation's theory of change and the conceptual framework of the study. Coder agreement rates on commonly used codes ranged from 74% to 82% across the 4 years of data collection. Analysts queried the ATLAS.ti database by relevant code to review data on particular topics. These topics were chosen to explain, confirm, or deny findings in the survey data or to find and substantiate prevalent themes in the qualitative

data. Larger groups met weekly to evaluate the consistency of findings across the site visit and survey data and to determine areas for further review.

TEACHER ASSIGNMENTS AND STUDENT WORK

For start-ups, teacher assignments and student work were collected throughout the second, third, and fourth years of the study from a total of 12 new schools, for comparison, and 19 comprehensive schools. Four of these were preconversion high schools in the national sample, seven were comparison high schools for the national sample, one of the pre-conversion high schools served as a comparison, and eight were preconversion high schools from Washington state, where the foundation began its work.

Student work was collected from the same 12 new schools across the 3 years; most of these schools were surveyed as well. The selection of the 19 comprehensive schools used as comparisons ranged in the different years. The 11 comprehensive schools in the first and second years of collecting teacher assignments and student work were foundation-supported large high schools planning to convert into small autonomous schools and did not adequately match the surveyed new schools in terms of location or demographics. The eight comprehensive schools in the third year of this data collection were large high schools that were geographically and demographically similar to the 12 surveyed new schools and were not receiving foundation support.

For conversion schools, teacher assignments were collected from 12 large schools in the planning year prior to their redesign, and 2 years later from a total of 30 SLCs that resulted in their conversion. Teacher assignments and student work in ELA were provided in 26 of these SLCs, and teacher assignments and student work in math were provided in 27.

In all, eight assignments were collected from each participating 10th-grade ELA teacher and mathematics teacher. Student work in response to three of these assignments was collected from a random sample of 6–12 tenth graders in each teacher's class.

The teacher assignments and student work were scored by 10th-grade master teachers during the summers of each year of their collection. On the basis of the "Authentic Intellectual" framework, as conceived by Newmann and Bryk (Bryk, Nagaoka, & Newmann, 2000; Newmann, Bryk, & Nagaoka, 2001; Newmann, Lopez, & Bryk, 1998), two scoring rubrics—one for teacher assignments and another for student work—were created. The criterion across the scoring rubrics mapped onto three foundation-identified standards of assignments and work: rigor, rele-

vance, and quality. The measures for the standards were determined using the psychometric model, many-facet Rasch measurement. The measures indicated the overall scale of rigor and relevance of the teacher assignments and the quality of student work, ranging from none to substantial. For a comprehensive description of the teacher assignments and student work collection and their corresponding scoring, see AIR/SRI, 2006.

DISTRICT DATA

Extant data from districts include attendance rates, 9th- and 10th-grade progression rates, and achievement scores from jurisdiction-sponsored standardized tests. The AIR/SRI evaluation team examined school- and student-level data in five large urban districts, comparing data for schools that had received foundation support with districtwide data or with data for schools without foundation ties. The districts used different measures and metrics, so the analysis team did not compare or aggregate data across districts. Not all the districts provided data for all the measures reported in this article. Most of the analyses of district data controlled for prior student academic achievement (eighth-grade test scores) and student demographic characteristics such as race/ethnicity, gender, and socioeconomic status.

LINDA SHEAR is a senior learning consultant at SRI's Center for Technology in Learning, helping public and private sector clients to understand and maximize the classroom effectiveness and learning outcomes of the programs they deliver. She is currently directing research and evaluation projects in the areas of international education reform, youth development/college readiness, and knowledge management systems for teachers. She has an MA in education in math, science and technology from the University of California, Berkeley.

BARBARA MEANS (PhD, educational psychology, University of California, Berkeley) is the director of SRI's Center for Technology in Learning and served as co–principal investigator for the evaluation of the Bill & Melinda Gates Foundation's High School Grants Program. Her research focuses on evaluation of the implementation and effectiveness of innovative educational approaches. Her published works include the jointly authored volumes *Using Technology Evaluation to Advance Student Learning* and *The Connected School*, as well as the edited volumes *Evaluating Educational Technology* and *Teaching Advanced Skills to At-Risk Students*.

KAREN MITCHELL directs the Medical College Admission Testing program at the Association of American Medical Colleges. She was among the SRI International authors of the 2006 reports, *Changes in Rigor, Relevance, and Student Learning in Redesigned High Schools* and *Evaluation of the Bill & Melinda Gates Foundation's High School Grants Initiative*.

ANN HOUSE is an education researcher at SRI International's Center for Technology in Learning. Her professional interest centers on the use of innovative opportunities for teaching and learning, in a variety of settings ranging from informal learning environments to the process of school reform. Her current work includes case studies of technology use in the home, analyzing collaboration challenges across academic disciplines, and professional development programs bringing technology skills to teachers.

TORIE GORGES is an education researcher at SRI International's Center for Technology in Learning. Her recent work has included evaluations of several school reform projects, knowledge management in education initiatives, and educational technology programs. She earned her MA in international education administration and policy analysis at Stanford University.

AASHA JOSHI is a researcher at SRI International's Center for Technology in Learning. She is currently enrolled in a PhD program in the Faculty of Social and Political Sciences at the University of Cambridge (UK). Her recent work includes studies of organizational change within schools.

BECKY SMERDON is currently senior research associate at the Urban Institute's Education Policy Center, where she is leading three projects examining the implementation and impact of high school reform. She earned her PhD in education foundations and policy from the University of Michigan.

JAMIE SHKOLNIK received her PhD in economics from the University of California at San Diego, where she researched school resource allocation and its effects on education. Dr. Shkolnik is a senior research scientist in the Education group at AIR. She was the project director for an evaluation of the quality of teacher assignments and student classwork in small high schools, funded by the Bill & Melinda Gates Foundation, and serves as the quantitative analysis lead for the National Evaluation of the Early College High School Initiative.