The Impact of School Reform on Student Performance Evidence

From the New York Network for School Renewal Project

Amy Ellen Schwartz
Leanna Stiefel
Dae Yeop Kim

ABSTRACT

This paper evaluates the impact of the New York Networks for School Renewal Project, a whole school reform initiated by the Annenberg Foundation as part of a nationwide reform strategy. It uses data on students in randomly chosen control schools to estimate impacts on student achievement, using an intent-to-treat design. After controlling for student demographic, mobility, and school characteristics, the authors find positive impacts for students attending reform schools in the fourth Grade, mixed evidence for fifth Grade, and slight to no evidence for sixth Grade. On average, there is a small positive impact. The paper illustrates how relatively inexpensive administrative data can be used to evaluate education reforms.

I. Introduction

The New York Networks for School Renewal Project (NYNSR), initiated with a $25 million, five-year grant from the Annenberg Foundation, began in New York City in 1995–96 with 80 founding schools (out of approximately 1000 public schools operating in New York City at that time). NYNSR schools represent a model of whole school reform that includes a commitment on the part of the schools

Amy Ellen Schwartz is a professor of public policy, Leanna Stiefel is a professor of economics, and Dae Yeop Kim is a doctoral candidate at the Wagner Graduate School, New York University. *The authors thank Norm Fruchter, Patrice Iatarola, Dana Lockwood, and Dorothy Siegel for their comments and insight and the Institute for Education and Social Policy at New York University for support and assistance. The comments of two reviewers were valuable in revising the paper. All responsibility for errors remains with the authors. The data used in this article can be obtained beginning October 2004 through September 2007 from the authors. Email either amy.Schwartz@nyu.edu* or *leanna.stiefel@nyu.edu.*
[Submitted December 2001; accepted January 2003]
ISSN 022-166X; E-ISSN 1548-8004

to work together voluntarily, to enroll small numbers of students in each school, and to adhere to distinct principles of learning. The NYNSR schools span elementary, middle, and secondary grades in a wide range of grade configurations.

In this paper, we evaluate the effect of this reform on the performance of three cohorts of students—those attending fourth, fifth, or sixth Grade in 1995–96 in founding NYNSR schools. More specifically, we use four years of student level and school level data on student performance, student demographics, school resources, and school characteristics as well as an additional year of pre-reform student test data to estimate the effect of this school reform on student performance on reading and math tests. The students in NYNSR schools are compared to a control group of students attending a set of randomly selected schools in 1995–96. Our cohort studies estimate the impact of attendance in a NYNSR school on the change in a student's performance between 1994–95, the year before the NYNSR initiative began, and subsequent years, through the eighth Grade or their exit from the New York City public schools. Both short-term (one-year) and log-term (three- or four-year) impacts are estimated, and controls are included for various characteristics of students and schools.

In addition to providing useful insight into the impact of this reform on student performance, the paper also demonstrates the feasibility and usefulness of using data obtained from administrative records to evaluate school reforms. The paper is organized as follows. In Section II we discuss whole school reforms in general and the New York Networks for School Renewal in particular. In Section III we provide a brief review of relevant academic studies that have quantitatively estimated the impact of whole school reforms on student achievement.[1] In Section IV we present a model of student performance for estimating impacts of whole school reforms, and in Section V we discuss sampling and data. In Section VI we describe the empirical results, and in Section VII we provide concluding comments.

II. Whole School Reform and the New York Networks for School Renewal Project

Whole School Reforms aim to improve education by changing entire schools; they "create school-wide approaches that integrate curriculum, instruction, assessment, and professional development" (Herman et al. 1999).[2] Some of the better-known whole school reform designs are: James Comer's School Development Program, which emphasizes interpersonal relationships, school climate, and students' psychological welfare; Robert Slavin's Success for All, which emphasizes prevention and early intervention, especially with reading skills; Henry Levin's Accelerated Schools, which involves the entire community in raising expectations and

1. Many studies focus on implementation and on variables other than achievement, but we do not review these because the emphasis in this paper is the effect of NYNSR schools on achievement.

2. According to Northwest Regional Educational Laboratory (1999) entire-school reforms "are based in research, provide schools with a common vision, and deal in some way with the critical areas of professional development, school organization, and curriculum and instruction. A particular strength they bring to comprehensive reform is the increased likelihood that all aspects of the reform process will be coordinated across the school" (page 7).

types of learning experiences for at-risk students;[3] and models developed by New American Schools (NAS) and by the Edison Project.[4]

The reforms initiated by the New York Networks for School Renewal (NYNSR) fit in this group of whole school reforms. NYNSR is a collaborative organization composed of four New York City reform agencies, called sponsors, which organize voluntary networks that normally include three to eight schools based on geographical proximity or instructional level.[5] A centerpiece of the NYNSR initiative is the creation of small learning communities brought about by reducing school size and supporting small schools. Empirical work on the relationship between school size and outcomes has suggested advantages of small schools for student learning. For instance, Fowler and Walberg (1991), Greenwald, Hedges, and Laine (1996), Lee and Loeb (2000), Lee and Smith (1997), and Lindsay (1982) found that students' academic achievement was negatively affected by the size of their schools. A group of studies that used noncognitive outcome measures found that students in smaller schools were less likely to drop out (Pittman and Haughwout 1987; Stiefel et al. 2000) and more likely to participate in social activities at school (Morgan and Alwin 1980; Finn and Voelkl 1993; Lee and Smith 1993 and 1995; Lindsay 1982). In his study of the relationship between school size and inter-district resource allocations, Monk (1984a) found that, among districts with the same total level of enrollment, those with larger average enrollments in their elementary schools spent more of their instructional funds on supervisory services at the expense of classroom instructional services. Thus, NYNSR could lead to improved outcomes, if the reforms effectively reduce school size.[6]

Alternatively, NYNSR reforms may improve academic performance through increases or changes in the mix of resources. Between the 1995–96 and 1999–2000 school years, NYNSR schools received about 22.5 million dollars for direct support and promotion of networking opportunities. While researchers may disagree about the impact of generically increasing school expenditures, additional resources may have fueled increases in performance in this particular case either because most of these funds were spent on instructional expenditures or because these relatively unconstrained funds were used to leverage existing, more constrained, resources.[7]

NYNSR also may enhance student achievement by encouraging community or parental involvement in student learning activities. In his study of the relationship between educational inputs and resource allocation in classrooms, Monk (1984b) found that parental help on homework substituted for teachers' time spent on small-

3. See King (1994) and Barnett (1996) for more on these models as well as efforts to cost out their implementation.

4. Odden, Archibald, and Tychsen (2000) provide a brief description of these latter two reforms among others.

5. These agencies are: the Association of Community Organizations for Reform Now (ACORN), the Center for Collaborative Education (CCE), the Center for Educational Innovation (CEI) and New Visions for Public Schools. A fuller description of NYNSR principles is available from the authors.

6. Below, we discuss measured differences between NYNSR and other schools in enrollments and school resources.

7. See, for example, Hanushek (1986), Greenwald, Hedges, and Laine (1996), or Sander (1993) for more on the relationship between performance and resources, in general. Lopus (1990) provides evidence that expenditures related directly to instructional services significantly increased students' test scores in economic literacy tests.

group work in math class and increased student achievement. Schneider (1985) found more effective schools had active parental participation through parental councils and volunteer tutors.

Pradl et al. (2001), in a qualitative evaluation of the implementation of NYNSR reforms, reported that NYNSR schools are small; teachers are satisfied; extra funds are used to promote innovative programming, provide additional support to personnel and parents, and acquire new materials (particularly library books and computers); professional development is tailored to teacher needs; and curriculum is somewhat innovative. Iatarola (2002) reported that NYNSR schools had lower pupil-teacher ratios and spent more on professional development. Thus, NYNSR reforms may have led to improved performance, to the extent that these are important determinants of performance and that the NYNSR reforms created these conditions.

III. Literature Review: Estimating the Impact of Whole School Reform with Student Level Data

There is a small but growing literature that evaluates the effects of whole school reforms using student level data and a control group. While there is a much larger literature evaluating a variety of school outcomes, we confine our discussion to the recent studies of the late 1990s that use student level data to evaluate the effects of whole school reforms on performance.[8]

Cook et al. (1999) evaluated program implementation, student climate, and performance of Comer schools in Prince George's County, Maryland, using random assignment for 23 middle schools, and collecting student performance data for two pre and two post program years in both experimental and control schools. They concluded that student learning in Comer schools was not significantly better than in control schools and speculated that this may be due to shortfalls in implementation or to inadequate focus on academics (versus socialization) in the Comer schools. Cook, Hunt, and Murphy (2000) evaluated Comer elementary schools in Chicago, again using random assignment of schools to experimental and treatment groups and student level data, and found some gain in achievement in Comer schools. They noted, however, that "causal inferences about the difference in achievement gains is especially dependent on the design used and the adequacy of the statistical control for selection" (page 589). As an example, a longitudinal analysis based on students remaining in the Comer schools throughout the study period produced different results than a cross section analysis, probably because several schools dropped out of the experiment, and some pilot schools, not originally in the experiment, were added during the study period.

Bifulco (2001) estimated the impact of three whole school reforms implemented in New York City (Slavin, Comer, and More Effective Schools), on reading scores, using a control group of schools selected through a stratified random sampling procedure. The school level analysis yielded some negative estimates of impacts, none statistically significant. Similarly, no effect was found in the analyses using student

8. Note that this review excludes, then, studies based upon school level data such as Berends et al. (2001).

level data on nonmovers (students who stayed in their original schools over the study period).

In a meta-analysis of literature evaluating the effect of various types of whole school reforms on achievement, Herman et al. (1999) stated that in only three out of 24 designs reviewed was there "strong evidence of positive effects on student achievement" and in the same study, in eight of the 24 reforms, there were ". . . no methodologically rigorous studies by which to assess effects of the approach on student achievement" (page 4, Overview).

Finally, Jones, Gottfredson, and Gottfredson (1997) evaluated one Slavin school with one comparison school using three cohorts of 50 to 110 students each and student level data in Charleston, South Carolina. This study found positive effects for reading in kindergarten, mixed results in other early elementary grades, and mostly negative effects for math.

While all of these studies make use of data on students in a group of control schools for comparison with those in the reform schools, there are differences, not always clear or explicit, in their treatment of students exiting or entering the study schools. The appropriate treatment of movers is critical to the research design because of the important role that student mobility plays in determining student performance (see Hanushek, Kain, and Rivkin 2001) and, equally, the composition of the student body. Put simply, studies that exclude movers from their analyses may derive biased estimates of the impact of the reform if "leavers" differ systematically from "stayers." Our empirical strategy is designed to minimize the possibility of this selection bias, as described in detail below.

IV. A Model of Student Performance

The centerpiece of our empirical work is a value-added specification of an education production function in which student performance (measured by test scores) is produced using a combination of school and student inputs. This formulation is relatively standard although the specification does reflect, in part, the availability and definition of variables provided in the administrative data. Todd and Wolpin (forthcoming) provide a thorough discussion and up-to-date review of education production functions as well as models of cognitive development. We begin our analyses with a baseline model as follows:

$$(1)\quad TEST_{i,s,t} = \alpha + \beta X_{i,s,t} + \delta NYNSR_i + \gamma_1 TEST_{i,s,t-1} + \gamma_2 TEST_{i,s,t-2} + e_{i,s,t}$$

where $TEST_{i,s,t}$ refers to the performance of student i in school s in year t, and $TEST_{i,s,t-1}$ and $TEST_{i,s,t-2}$ are defined similarly for tests in the previous two years. $X_{i,s,t}$ refers to a set of student characteristics, some of which are time invariant (for example, sex and race), some of which vary across years (for example, attendance rate and retention), and some of which are slow to vary (for example, free lunch status). Table 1 provides a list of test scores and student characteristics with descriptive statistics. $NYNSR_i$ is a dummy variable that takes a value of one if student i attended a NYNSR school in 1995–96, the year the NYNSR program began, and $e_{i,s,t}$ is an error term with the usual properties. In this model, the coefficient on the NYNSR dummy, δ, provides an estimate of the short-term (specifically, the one year) impact

Table 1
Descriptive Statistics by Sample Group; Grade Four Cohort, 1995–96

	NYNSR		Random	
	Mean	Standard Deviation	Mean	Standard Deviation
Standardized test scores				
Reading: 1994–95 pretest (for 1995–96 year)	−0.23	0.95	0.05	0.99
1995–96	−0.23	0.91	0.01	1.00
Math: 1994–95 pretest (for 1995–96 year)	−0.30	0.91	0.06	1.00
1995–96	−0.42	0.91	0.33	1.01
Student characteristics				
Attendance rate (percent)	91.34	8.13	92.24	7.09
Female	0.48	0.50	0.51	0.50
Black	0.30	0.46	0.33	0.47
Hispanic	0.55	0.50	0.42	0.49
Asian	0.06	0.25	0.11	0.31
Free-lunch eligible	0.78	0.42	0.79	0.41
Reduced-price-lunch eligible	0.06	0.25	0.07	0.25

Table 1 (*continued*)

	NYNSR		Random	
	Mean	Standard Deviation	Mean	Standard Deviation
Recent immigrant	0.02	0.13	0.03	0.18
Resource room	0.13	0.34	0.09	0.28
Exposure to language other than English	0.47	0.50	0.48	0.50
Took LAB test	0.17	0.37	0.18	0.39
LAB test percentile	15.29	18.57	17.25	20.04
Retained in previous year's grade	0.05	0.21	0.02	0.14
Advanced to higher grade than the typical	0.002	0.04	0.001	0.03
School resources and characteristics				
Number of teachers per 100 students	6.40	1.32	5.937	1.82
Percent teachers fully licensed and permanently assigned	75.03	9.35	84.65	11.05
Percent teachers with more than five years experience	55.89	12.66	73.66	12.41
Percent teachers with master's degree	65.99	7.87	75.41	11.39
Percent teachers more than two years in the same school	73.46	23.32	82.86	11.86
Teachers' average number of days absent	7.75	0.13	7.26	1.16
Total enrollment	789.54	636.40	942.84	403.12

Notes:
i) Data on student eligibility for free or reduced price lunch, participation in resource room, retention in previous year's grade, advancement to higher grades than the typical, and number of teachers per 100 students are not available for 1995–96. Values are for 1996–97.
ii) The means of school resources and characteristics are not pupil weighted.

of NYNSR participation on the performance gains of students, controlling for the measured student characteristics.[9] This baseline model is estimated using data on a cohort of students that includes both students who attended NYNSR schools and those who did not. As described in more detail below, we use data on three cohorts of students—students in Grades four, five, and six in 1995–96—termed the Grade four, Grade five, and Grade six cohorts, respectively.

Three specific attributes of this baseline model are worthy of further discussion. First, it includes no other school level variables that describe the characteristics of the schools, such as school size, grade configuration, resources, or demographic descriptors of the school community. Thus, to the extent that these differ between NYNSR and non-NYNSR schools, the coefficient on the NYNSR variable will capture the impact of all of the differences across schools. This is purposeful. The NYNSR strategy is to effect whole school change and, thus, our intention is to capture the full impact of NYNSR—including the impact that might be accomplished through changing school resources or characteristics. Put differently, if these differences are viewed as features of the NYNSR intervention then these baseline estimates may be most appropriate estimate—NYNSR gets credit for these improvements. As an example, an important feature of the NYNSR program is to create small learning communities. The impact of NYNSR participation, then, is in part due to school size. This analysis will yield NYNSR impact estimates that give credit to NYNSR participation for performance gains due to changes in school size. Subsequent analyses isolate the impact of the various NYNSR components (including school size) and identify the impact of the NYNSR participation net of the changes in school resources. (As described below we also estimate models that include school variables and/or school fixed effects to control for variations in school characteristics.)

Second, the model in Equation 1 includes two prior test performance variables. While the use of a single prior performance variable has become a standard formulation of a value-added model of student performance, a single test score provides a noisy measure of previous educational achievement. Two provide a more robust measure of prior academic achievement and control for prior growth in test scores.[10]

Finally, this specification characterizes a student as participating in the NYNSR "treatment" based on the school of her attendance in 1995–96. Thus, students are considered members of the treatment group (the NYNSR sample) if they attended a NYNSR school in 1995–96. Alternatively, they are members of the control group if they did not attend a NYNSR school in 1995–96 but, as described below, attended one of a set of randomly selected schools. In our analyses, student groups are fixed—that is, their membership in the NYNSR or random sample groups is unchanging over the course of their education—even if they later move into or out of the schools

9. We include a dummy variable indicating whether or not the student was retained in grade in order to control for the impact of retention on student performance. In general, if NYNSR schools differ systematically from other schools in retention, or other such policies, then the NYNSR coefficient will capture the impact of these policies. In the specific case of retention, it is likely that the decision to retain the student was made prior to the current academic year and is, therefore, best viewed as capturing the student's prior academic experience.

10. Estimates suggest the two-year prior test score is a significant determinant of achievement. Models estimated using only one prior test score for comparison purposes yield substantially similar results, as described below.

they originally attended. This is similar in spirit to the "intent to treat" methodology utilized in medical research—students are considered to be members of the treatment group if they were originally exposed to a NYNSR school at the outset.

The implication is that our methodology provides an estimate of the impact of NYNSR participation on students in the NYNSR treatment group, for years beyond 1995–96, which reflects the impact on both students who remain in NYNSR schools over the study period and those who move to other schools. If students move out of NYNSR schools during our analysis period, they receive less than a full dose of NYNSR education, but our methodology continues to characterize them as NYNSR students. As an example, if a student attends a NYNSR school in 1995–96 and moves to a non-NYNSR school in 1996–97 she has received only one year of NYNSR treatment but is still considered a NYNSR student. The most important reason for doing this is to reduce the potential for selection bias induced by differential exit or entry into schools across samples. In this formulation, the NYNSR reform gets no "credit" for attracting (or repelling) higher (or lower) performing students in years subsequent to the reform. The impact is estimated based only upon the students enrolled at the time the reform was implemented.

As noted, while the baseline model includes only student characteristics, excluding characteristics of the school, so that the NYNSR impact estimates may reflect differences in school resources or other characteristics between NYNSR and other schools, we might also be interested in estimating the impact of NYNSR above and beyond these differences. For example, students attending the NYNSR founding schools in 1995–96 stayed in the same school for a greater number of years than those in the control sample throughout all three cohorts, suggesting the NYNSR impact might be due to differences in student mobility. Thus, we might be interested in an alternative model that controls for school characteristics and student mobility to identify an impact of the NYNSR program independently of these characteristics.

We explore this alternative in two ways, first, by augmenting the models with student mobility and school variables, and, second, by substituting school fixed effects for the school variables. More specifically, we augment Equation 1 with socioeconomic characteristics of the student body (for example, race and poverty), characteristics of the school (for example, school size and resources), and student mobility:

$$(2)\quad TEST_{i,s,t} = \alpha + \beta X_{i,s,t} + \delta NYNSR_i + \gamma_1 TEST_{i,s,t-1} + \gamma_2 TEST_{i,s,t-2} + \phi S_{i,t} + \eta M_{i,s,t} + e_{i,s,t}$$

where $S_{i,t}$ represents a vector of characteristics of the school attended by student i in year t, and $M_{i,s,t}$ is a student mobility variable measured as the number of consecutive years in the same school. In this model, then, the estimate of the NYNSR impact should be interpreted as the impact of the reform, controlling for student mobility and differences in the characteristics of NYNSR and non-NYNSR schools. See Table 1 for brief descriptions and statistics on these additional school level variables.

We also estimate a model that substitutes school fixed effects for the school level variables:

$$(3)\quad TEST_{i,s,t} = \alpha + \beta X_{i,s,t} + \delta NYNSR_i + \gamma_1 TEST_{i,s,t-1} + \gamma_2 TEST_{i,s,t-2} + \eta M_{i,s,t} + \theta_s + e_{i,s,t}$$

where θ_s represents a school fixed effect. Note that the school fixed effect is identified (that is, it can be differentiated from the NYNSR dummy) because of the movement of students between schools over time.[11] Some of this mobility is due to institutional features such as the end of an elementary school at a certain grade rather than due to individual student characteristics. The intent to treat design means that the NYNSR dummy remains the same for each student regardless of her transitions to different schools. Thus, the NYNSR dummy captures the difference between the performance of students in the NYNSR treatment group and the performance of students in the control group, controlling for the average performance of each school.[12]

Notice the coefficient on the NYNSR dummy will decline if the (positive) impact of NYNSR is due to differences between NYNSR and non-NYNSR schools in the school characteristics or student mobility variables introduced in Equation 2 and 3. Put differently, the impact estimate δ will decline with the addition of school and student mobility characteristics, if these variables capture differences between NYNSR and non-NYNSR schools that are correlated with the NYNSR intervention.

We supplement the short-term analyses with longer-term analyses that yield estimates of the impact of NYNSR over a three-year period, between 1995–96 and 1998–99. In fact, few will argue that whole school reform will lead to large immediate improvements in student performance. Instead, the impact is expected to be felt over the longer run as students benefit from several years in the NYNSR educational environment.

Our long-term impact estimates are based on a set of regression models following the short-term regressions. Corresponding to Equation 1, we estimate:

$$(4) \quad TEST_{i,s,1999} = \alpha + \beta X_{i,s,t} + \delta NYNSR_i + \gamma_1 TEST_{i,s,1996} + \gamma_2 TEST_{i,s,1995} + e_{i,s,t}$$

where all variables are as defined previously. For the Grade six cohort, 1998–99 test scores were unavailable,[13] so we estimate:

$$(4') \quad TEST_{i,s,1998} = \alpha + \beta X_{i,s,t} + \delta NYNSR_i + \gamma_1 TEST_{i,s,1996} + \gamma_2 TEST_{i,s,1995} + e_{i,s,t}$$

Robust standard errors are used to account for the clustering of students in schools.

As in the short-term equations, we also estimate a specification that includes school level variables for each of the schools attended in each of the years of the study period as well as the student mobility variable.

Finally, note that, although statistically and intuitively appealing, the available data did not allow for the estimation of a model using student fixed effects, since there was insufficient information about students and schools prior to the implementation of the NYNSR reform. Specifically, while information on test scores prior to

11. See Table 2 for statistics on student mobility.

12. To be clear, we do not estimate the impact of the reform as the average performance of NYNSR schools compared to non-NYNSR schools. Instead, we compare the average performance of students attending the NYNSR schools in 1995–96 to students attending a control school, even as the students move from one school to another over the years. The school effects capture the average performance in each of the schools in the sample, identified by both the students who move in and those attending from the outset. The NYNSR dummy captures the relative performance of NYNSR treatment group students, then, relative to the average performance in the school they attend.

13. Test score data were available through the eighth Grade only. In 1999, Grade six cohort students were primarily in ninth Grade.

implementation was available, no other student information (including school attended) was known for prior years.

V. Samples and Data

A. *Samples*

The empirical analysis makes use of two samples of students—a sample that includes all of the students in Grades four, five, or six who attended any founding NYNSR elementary or middle school in 1995–96 (the NYNSR sample) and a sample that includes all students in Grade four, five, or six who attended one of a set of randomly selected schools in that year (the random sample). Although not used in the analyses reported here, an additional sample of students, attending a set of purposively chosen schools (the comparison sample), was used in auxiliary analyses.[14] To be more specific, random sample schools were chosen by lot from all elementary and middle schools in New York City, excluding schools implementing NYNSR, comparison group, and Staten Island schools. (No NYNSR schools are located on Staten Island.)

Notice that the NYNSR schools were not randomly assigned to the NYNSR reform. In contrast, such assignment would negate a part of the NYNSR philosophy, which is that schools either choose or agree to be sponsored for the reform. The implication is that NYNSR schools may be systematically different from other schools and the results cannot be interpreted as estimates of what would happen to the academic performance of a student randomly assigned to a school that had been itself randomly assigned the NYNSR reform. Instead, these are estimates of the impact on students attending a school that had chosen this reform, perhaps, from alternative reforms available.[15] While there are circumstances under which whole school reforms are imposed on schools, rather than chosen with their consent and involvement, this is relatively rare in New York City. Further, it is undoubtedly true that some of the schools in the random sample were themselves implementing some other sort of reform. Thus, the estimates can be viewed as measuring the impact of the NYNSR reforms on student academic performance, relative to other schools in New York.

B. *Data*

This study uses student level and school level data routinely collected and stored by the New York City Board of Education (BOE) and is similar to data available

14. The comparison sample consisted of schools sharing characteristics of NYNSR founding schools—location of the school in the same community school district, enrollment, student race/ethnicity, poverty, attendance and test scores, and school application procedures. Unfortunately, the proximity of the schools to NYNSR schools raises the concern that they may have been affected by the NYNSR reforms themselves (that is, adopting some or all of the NYNSR reforms), which may mean that impacts estimated using this sample would be biased. Results obtained using the comparison sample are qualitatively similar to those reported in this paper. Data on the universe of New York City public schools and their 1.1 million students were not available for this study, due, largely, to the difficulty and expense of assembling these and small expected benefits of an increased sample size.

15. Students or their parents are unlikely to have known about this reform before enrolling (or even after enrolling).

in other medium and large size school districts. The student level data provide rich and detailed information on reading and math test scores, attendance, school and grade codes, and other student demographic and socioeconomic characteristics. We obtained school resource and student body characteristic data from the BOE Annual School Reports (ASR) and linked them with the student level data.

Records were obtained for each student in our sample groups for 1995–96 and each of the subsequent years through 1998–99, eighth Grade, the highest grade tested citywide, or their exit from the BOE schools. In addition we obtained test scores for these students for 1994–95, the year before the NYNSR reforms were implemented. Thus, we have five years of test data for the two cohorts beginning in Grades four and five in 1995–96, (the Grade four and Grade five cohorts, respectively). Only four years of data are available for the cohort of students who were sixth Graders in 1995–96, the Grade six cohort, since they reached eighth Grade in 1997–98 and exited from the sample in 1998–99. Notice that since NYNSR schools cover different grade configurations, including middle school and high school grades, only a subset of NYNSR schools enrolled students in each of these cohorts. For example, only 15 NYNSR schools enrolled fourth Grade students. The implication is that these cohort studies will provide impact estimates for only a subset of the schools—those enrolling students in fourth, fifth or sixth Grade in 1995–96. These schools are largely elementary schools but include some middle schools.[16]

Table 2 shows the number of schools attended by fourth Grade cohort students, in each of the study years, distinguishing the number of founding NYNSR schools, the number of random sample of schools, as well as the number of other schools. Table 2 also shows the distribution of students across the schools by sample group for each year. Over time, the number of original schools attended by students in the study samples declines and the number of students attending their original schools declines because of student mobility. The decline is particularly large in the years corresponding to the end of elementary school (typically fifth or sixth Grade) and the beginning of middle school (typically sixth or seventh Grade).

Despite student mobility, we continue to track students as long as they remain enrolled in some New York City public school. Students are lost only if they exit the New York City public schools or if they graduate to ninth Grade.[17] The students in schools other than founding NYNSR or original control schools consequently rise over time.[18]

Student performance on standardized tests is reported for a test in reading (CTB) and mathematics (CAT) for each year, except 1994–95 and 1998–99. Students were given the Degrees of Reading Power test (DRP) in 1994–95, and we have its percentile and raw scores. In addition, in 1998–99, fourth and eighth Graders took new statewide reading and mathematics tests and the BOE reports the scaled scores of the new state tests. To make these different scores comparable over years and to each other, we convert them to z-scores, which are standardized to have mean zero

16. A list of NYNSR schools included in our evaluation is available from the authors.

17. Even if we are missing some data for right-hand side variables for a student, we continue to include that student in our analyses. We use a set of missing dummy variables to account for the variables with missing values.

18. Data analogous to Table 2 for fifth and sixth Grade cohorts are available from the authors.

Table 2
Number of fourth Grade Cohort Students and Schools Attended by Year, by Sample Group

Year (Typical grade)	Number of Schools				Number of Students						
					NYNSR Sample Group			Random Sample Group			
	Original NYNSR Schools	Original Random Schools	Other NYC Schools	Total	Original Sample Schools	Other NYC Schools	Exit from Database	Original Sample Schools	Other NYC Schools	Exit from Database	Total
1995–96 (Grade four)	15	57	0	72	668 (100)	0	0	5,837 (100)	0	0	7,701
1996–97 (Grade five)	15	56	333	404	562 (84.1)	57 (8.5)	49 (7.3)	4,964 (85.0)	506 (8.7)	367 (6.3)	7,701
1997–98 (Grade six)	13	37	387	437	276 (41.3)	309 (46.3)	83 (12.4)	1,673 (28.7)	3,408 (58.4)	756 (13.0)	7,701
1998–99 (Grade seven)	8	19	330	357	124 (18.6)	439 (65.7)	105 (15.7)	127 (2.2)	4,684 (80.2)	1,026 (17.6)	7,701

Note: Numbers in parentheses are the percentage of students in the 1995–96 original sample schools who, in subsequent years, attended the original sample schools, transferred to other NYC public schools, or exited the NYC public school system.

and standard deviation of one.[19] Table 1 presents descriptive statistics for the fourth grade cohort, by sample, for 1995–96. As shown, NYNSR students' average reading and math test scores were lower than those of random sample students in 1994–95 and 1995–96.

The student level data include a variety of demographic and socioeconomic characteristics: attendance, sex, eligibility for free or reduced price lunch, race (black, Hispanic, Asian, and white),[20] Language Assessment Battery (LAB) scores, as well as a group of indicators for recent immigrants who arrived in the United States within the past three years, resource room participation (that is, part-time special education services), and exposure to languages other than English at home. Since data on free and reduced price lunch eligibility, and on resource room variables and variables indicating whether a student was retained or advanced to a higher grade than the typical one, were unavailable in 1995–96, descriptive statistics are reported for 1996–97.

As shown in Table 1, the two sample groups look similar in most categories, however, there are some disparities. The proportion of female students in the NYNSR sample (48 percent) is lower than the random sample (51 percent). The NYNSR sample includes a higher percentage of Hispanic students (55 percent) than the random sample (42 percent), while the representation of Asian and black students is greater in the random sample. Proportions of resource room participants and students retained in their previous year's grade are slightly higher in the NYNSR sample.

The school level data include information on teacher resources such as teacher student ratio, licensure for assignment, years of experience, stability, average number of days absent, and enrollment. Note that the mean values of these variables reported in Table 1 are not pupil weighted and, thus, represent the average characteristics of schools attended by the NYNSR or random sample students. As shown in Table 1, schools attended by the NYNSR sample students average a slightly higher number of teachers per 100 students, and lower enrollments. On the other hand, NYNSR sample students attend schools with less experienced teachers, lower proportions of teachers with license and master's degree, and higher percents of new teachers in 1995–96. NYNSR schools seem to have made some tradeoffs in spending their resources, hiring more teachers, with less experience and education, than the random schools.

VI. Results

The estimated production function models of student test scores produce coefficients and R-squared values that are consistent with those in the literature. For example, in the short-term baseline Model 1 for the Grade four cohort for 1997–

19. Specifically, z-scores are computed as the deviation of the test score from its mean divided by the standard deviation. For CTB, CAT, and the new state tests, we use citywide averages and standard deviations. Since citywide averages of 1994–95 CAT and DRP tests are not available, we use the sample means obtained from all our available samples (NYNSR, Random, and comparison samples) for this calculation.
20. A very small number of students characterized as Native American or race/ethnicity unknown were included in the white category.

98, the model explains more than two-thirds of the variation in test scores. Both prior test scores have positive and significant coefficients. Hispanic students do a little worse and Asian students and recent immigrants a little better than white and black students. Poor students score lower, as do students receiving resource room services (part-time special education) and students who score poorly on tests of English language proficiency. (Estimates of all equations are available from the authors.)

Table 3 shows the impact estimates from three alternative models estimated for each of the three cohorts (Grades four, five, and six), for each of three academic years (1996–97, 1997–98, and 1998–99). The top panel shows results for reading tests; the bottom panel for math tests. (Again, parameter estimates for other coefficients are available from the authors.) In the baseline model, estimated impacts of NYNSR participation are mixed. Impact estimates for the 1996–97 school year (the first followup year) are insignificant for all cohorts for both reading and math.

The 1997–98 results (the second followup year) for the baseline model are statistically significant for two cohorts. A positive impact, estimated for the Grade four cohort for reading in 1997–98 (0.150 standard deviations), is matched by a similar impact in mathematics in that year (0.144 standard deviation), and the Grade five cohort also shows positive significant impacts. No significant impact is estimated for the Grade six cohort. In the final year of our followup, 1998–99, both Grades four and five cohorts show positive significant impacts in reading, and the Grade four cohort shows positive significant impacts in math. Thus, two observations summarize the results for the baseline model. First, it may take a couple of years for NYNSR participation to have an impact on test scores for students who attended NYNSR schools in Grades four or five, but positive impacts were estimated in both reading and math. Second, the results provide no evidence that NYNSR participation had any impact on the sixth Grade cohort. We discuss these results more fully in the conclusions.

Turning to the models estimated with school characteristics, we find that including the additional variables generally has the effect of decreasing the magnitudes of the impact estimates, suggesting NYNSR schools systematically differ from control schools in ways that enhance student gains. However, the change in the estimates is relatively small and, for many, the difference is not statistically significant.[21] As before, the Grade four cohort shows positive and significant impacts in both reading and math in the second followup year, and in math for the third followup year. The Grade five cohort shows positive impacts in reading in the second and third followup years, but a negative impact in math in the first followup year. Again, there is no evidence of an impact in the Grade six cohort.

Substituting school fixed effects for the school level characteristics variables yields slightly different results. Positive impacts are estimated for NYNSR participation for the Grade four cohort for reading and math for both the first two years of followup and in math in the third year. The Grade five cohort shows little evidence

21. School characteristics are statistically significant as a group. Interpreting any individual coefficient is confounded by the interaction with the other school characteristics. Since the main purpose of including the school characteristics is to yield unbiased estimates of the reform impact these coefficients are not presented, however, it is worth noting that, where patterns emerge, they are consistent with previous studies.

Table 3
Annual Short-term Impact of NYNSR Participation on Standardized Reading and Math Scores by Cohort

	Fourth Grade Cohort			Fifth Grade Cohort			Sixth Grade Cohort		
	1996–97 (Grade 5)	1997–98 (Grade 6)	1998–99 (Grade 7)	1996–97 (Grade 6)	1997–98 (Grade 7)	1998–99 (Grade 8)	1996–97 (Grade 7)	1997–98 (Grade 8)	1998–99 (Grade 9)
Baseline reading regressions									
NYNSR	−0.003	0.150***	0.070**	0.015	0.071*	0.095**	0.013	0.061	—
	(0.032)	(0.052)	(0.036)	(0.067)	(0.042)	(0.040)	(0.035)	(0.037)	
R^2	0.692	0.700	0.677	0.678	0.702	0.679	0.671	0.687	—
Including school characteristics[a]									
NYNSR	−0.017	0.105**	0.062	−0.016	0.082**	0.064*	0.003	0.033	—
	(0.045)	(0.043)	(0.042)	(0.044)	(0.039)	(0.034)	(0.036)	(0.037)	
R^2	0.695	0.711	0.682	0.691	0.707	0.689	0.679	0.691	—
Including school fixed effects									
NYNSR	0.381**	0.074*	−0.011	−0.037	0.089**	0.026	0.053	−0.001	—
	(0.154)	(0.043)	(0.039)	(0.70)	(0.044)	(0.036)	(0.059)	(0.045)	
R^2	0.720	0.745	0.717	0.722	0.729	0.720	0.697	0.708	—
Number of observations	5,580	5,360	4,947	5,596	5,254	4,842	6,176	5,981	—
Baseline math regressions									
NYNSR	0.012	0.144***	0.140***	0.007	0.073*	0.001	0.016	0.054	—
	(0.048)	(0.038)	(0.035)	(0.060)	(0.040)	(0.038)	(0.038)	(0.035)	
R^2	0.722	0.724	0.726	0.722	0.719	0.735	0.725	0.688	—

Table 3 (*continued*)

	Fourth Grade Cohort			Fifth Grade Cohort			Sixth Grade Cohort		
	1996–97 (Grade 5)	1997–98 (Grade 6)	1998–99 (Grade 7)	1996–97 (Grade 6)	1997–98 (Grade 7)	1998–99 (Grade 8)	1996–97 (Grade 7)	1997–98 (Grade 8)	1998–99 (Grade 9)
Including school characteristics[a]									
NYNSR	−0.036	0.136***	0.171***	−0.084**	0.040	0.026	−0.024	0.043	—
	(0.046)	(0.038)	(0.033)	(0.042)	(0.032)	(0.030)	(0.038)	(0.053)	
R^2	0.727	0.741	0.733	0.744	0.725	0.746	0.733	0.696	—
Including school fixed effects									
NYNSR	0.299**	0.158***	0.104**	−0.063	0.039	0.050	0.070	0.059	—
	(0.143)	(0.036)	(0.044)	(0.041)	(0.041)	(0.037)	(0.055)	(0.061)	
R^2	0.758	0.776	0.762	0.769	0.753	0.771	0.755	0.715	—
Number of observations	5,762	5,388	5,024	5,787	5,367	4,977	6,469	6,153	—

Notes: i) Test scores are measured in z-scores transformed from NCE's for the CTB or CAT exams, except for the DRP reading test scores in 1994–95 and state reading (ELA) and math test scores for eighth Grade in 1998–99.
ii) There are no consistent ninth Grade test data available, hence, no sixth Grade regression for 1998–99.
iii) * indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** indicates significance at the 1 percent level.
iv) Huber's robust standard errors are reported in parentheses.
v) All regressions include one and two year lagged test scores; attendance rate; LAB percentile; dummies for students who are female, free or reduced price lunch eligible, exposed to a language other than English, Asian, Hispanic, black, recent immigrant, retained in previous year's grade, and advanced to a grade higher than the typical one; and a set of missing value indicators. Regressions with school characteristics and school fixed effects include an additional variable capturing student mobility measured as the number of consecutive years in the same school.
vi) "Recent immigrant" and "advanced to higher grade than the typical one" are dropped from the 1998–99 regressions. As of 1998–99, no student considered a recent immigrant in 1995–96 retained that status. None of fourth and fifth Grade cohort students who advanced to a higher grade than the typical one in 1998–99 had valid reading or math test scores for that year.
a. School level controls are total enrollment, number of teachers per 100 students, teachers' average number of days absent, the percentages of students who are black, Hispanic, Asian, free lunch eligible, LEP, recent immigrant, special education, resource room participant, and the percentages of teachers fully licensed and permanently assigned, with master's degrees, with more than 5 years of experience, working more than two years in the same school.

of any impact (a positive and significant impact estimate in reading in the second followup only) and the Grade six cohort shows no evidence of any impact in any year.[22] Taken together, the short-term impact estimates suggest that attending a NYNSR school had a positive impact for students in Grade four, no impact by Grade six, and an impact in Grade five that falls somewhere in between.

We next turn to estimating impacts over a longer-term period. For our Grades four and five cohorts, these are three-year impacts. For our Grade six cohort, followup data only allow two-year impact estimates. Estimated baseline models of long-term performance growth perform well, explaining more than 60 percent of the variation in reading test scores for the Grade four cohort, for example. The coefficients are generally consistent with estimates from the short-term regressions and with previous literature.[23]

Table 4 presents the long-term impact estimates for the different cohorts, providing the same mixed evidence about the NYNSR reforms as the short-term models. In the baseline models, NYNSR participation seems to have been good for Grade four cohort students—both reading and math tests are significantly higher for NYNSR students. There seems to be no impact for the Grade six cohort students and, again, the experience of the Grade five cohort is in between, showing gains in reading, but no impact in mathematics.

As before, we proceed by introducing additional variables to these regressions, to control for the characteristics of the student body and schools attended by NYNSR and random sample students in each of the study years.[24] While results for the Grade four and Grade six cohorts are essentially unchanged, the estimates here suggest that the long-term impact for the Grade five cohort is insignificant for reading and for math negative. Alternatively, it may be that, in the higher grade cohort, the NYNSR reforms "work" through changes captured by the additional included variables describing school and student characteristics. Further work, including, potentially, qualitative analyses and survey research, would be required to disentangle these effects.

VII. Discussion and Conclusions

This paper reports the results of an evaluation of the impact of a particular whole school reform model adopted by 42 New York City public schools in 1995–

22. The fixed effects are jointly significant at the one percent level. Statistical tests suggest there is little difference in the coefficients on the student characteristics between the two models.

23. Coefficients are available from the authors. We note selected interesting findings for the Grade four reading regressions. Only the current year's attendance rate shows a significant positive effect on student performance in reading in 1998–99. Long-term analyses show a marginally significant positive coefficient on the 1998–99 resource room participation. Limited English proficient students earn lower scores, but scores increase with English language skills. Interestingly, students held back in the previous year's grade in 1996–97 performed significantly better on the 1998–99 reading test, although the same pattern does not obtain in subsequent years. Students who advanced to higher grades than the typical ones in 1996–97 and 1997–98 scored significantly greater gains than others.

24. Although estimating a long-term model including a school fixed effect for the school each student attended in each year might seem attractive, the large number of schools attended by cohort students over the three years, as shown in Table 2, is prohibitive. A different school fixed effect would be required for each school for each year, thus introducing more than 500 additional dummies into the model.

Table 4
Long-term Impact Analysis of NYNSR Participation on Standardized Reading and Math Scores by Cohort

	Fourth Grade Cohort	Fifth Grade Cohort	Sixth Grade Cohort
	1998–99 (Grade 7)	1998–99 (Grade 8)	1997–98 (Grade 8)
Baseline reading regressions			
NYNSR	0.161***	0.165***	0.064
	(0.036)	(0.063)	(0.044)
R^2	0.627	0.634	0.646
Including school characteristics[a]			
NYNSR	0.155**	0.029	0.062
	(0.063)	(0.065)	(0.043)
R^2	0.636	0.655	0.658
Number of observations in all models	4,947	4,842	5,981
Baseline math regressions			
NYNSR	0.251***	0.039	0.047
	(0.045)	(0.048)	(0.040)
R^2	0.666	0.678	0.645

Including school characteristics[a]			
NYNSR	0.229***	−0.113*	0.001
	(0.056)	(0.062)	(0.077)
R^2	0.680	0.699	0.667
Number of observations in all models	5,024	4,977	6,153

Notes:

i) Test scores are measured in z-scores transformed from NCE's for the CTB or CAT exams, except for the DRP reading test scores in 1994–95 and state reading (ELA) and math test scores for eight Grade in 1998–99.

(ii) * indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** indicates significance at the 1 percent level.

iii) Huber's robust standard errors are reported in parentheses.

iv) All regressions include 1994–95 and 1995–96 test scores; dummies for students who are female, exposed to a language other than English, Asian, Hispanic, black, and recent immigrant; and, for each year, attendance rates, LAB percentiles, and dummies for free or reduced price lunch eligibility, resource room participation, retention in previous year's grade, and advancement to a grade higher than the typical one; and a set of missing value indicators. Regressions with school characteristics include an additional variable capturing student mobility measured as the number of consecutive years in the same school.

v) "Recent immigrant" and "advanced to higher grade than the typical one" are dropped from the 1998–99 regressions. As of 1998–99, no student considered a recent immigrant in 1995–96 retained that status. None of fourth and fifth Grade cohort students advanced to a higher grade than the typical one in 1998–99 had valid reading or math test scores for that year.

a. Year-specific school level controls are total enrollment, number of teachers per 100 students, teachers' average number of days absent, the percentages of students who are black, Hispanic, Asian, free lunch eligible, LEP, recent immigrant, special education, resource room participant, and the percentages of teachers fully licensed and permanently assigned, with master's degrees, with more than five years of experience, working more than two years in the same school.

96, called the New York Network for School Renewal (NYNSR). The research strategy focuses attention on the academic gains of three cohorts of students originally attending either a fourth, fifth, or sixth Grade in a NYNSR school or one of a set of schools chosen randomly from the large set of New York City public schools. Students are credited to either the NYNSR (treatment) sample or the random (control) sample based on the school of their attendance in 1995–96, even if they moved schools over the study period. Thus, the analysis yields estimates akin to the intent to treat estimates provided in medical studies. While we estimate baseline models that include only the characteristics of the individual student as controls, fuller, undoubtedly more appropriate specifications, include variables describing the characteristics of the schools attended by students, to control for systematic differences between NYNSR and control schools. Both short-term and long-term models are estimated, to provide insight into the cumulative effect of these reforms over a longer time period.

Several limitations are worth noting. First, our analyses include only a subset of the full set of schools adopting NYNSR reforms. In particular, our evaluation only includes the set of founding NYNSR schools that educated fourth, fifth, or sixth grade students in 1995–96. Thus, a significant number of NYNSR schools are not included in our analysis, including both high schools and early childhood schools, as well as those that adopted reforms in later years. Second, our intent to treat estimates are, in some sense, conservative estimates, because we include in the treatment group many students who exited from NYNSR schools over the study period and therefore did not receive the full "NYNSR" treatment. Further, there were students in the control sample (a smaller group) who attended NYNSR schools at some point during the study period. Thus, our estimates are *not* estimates of the impact of treatment on the treated, but instead, are estimates of the impact of NYNSR participation on those attending a school in 1995–96 that adopted the NYNSR reform. The estimates may be viewed as reflecting the impact of a reform on all of the students targeted by the reform effort. Although estimating the impact of the reform treatment on the treated is attractive, it is quite difficult to disentangle the selection effects from the treatment effects, even given the unusually rich data available for this study.[25]

A wide range of alternative specifications has been investigated and only a representative sample reported and discussed at length here. In general, results are qualitatively insensitive to alternative specifications. Two are of particular interest. First, estimating fully specified models using only one prior test score, instead of two, yields qualitatively similar results. Second, using a different set of students as controls—students attending a set of comparison schools chosen purposively because they shared some of the characteristics of NYNSR schools—also yields qualitatively similar results.

Overall, the results provide little or no evidence that the NYNSR reforms are pernicious or serve to reduce performance on standardized tests of reading or math.

25. More specifically, student movement between schools over time—whether between NYNSR schools, between non-NYNSR schools, from NYNSR to non-NYNSR, or into NYNSR from non-NYNSR schools—is nonrandom, but is determined by the complex interplay of many largely unobserved variables. Thus, estimating the impact of treatment on the treated will be complicated by the potential for selection bias. Addressing the selection bias econometrically requires detailed data on the factors determining the selection process, many of which are likely to be unobserved, such as student motivation, parental involvement and tastes, or changes in residence.

Instead, there is some consistent evidence that there is a positive, long-term impact on both reading and math test performance of students attending a NYNSR school in the fourth Grade. There is, however, little or no evidence that there is any impact on test performance for sixth Grade students. The evidence for fifth Grade students is mixed. One interpretation of these results might be that NYNSR reforms are most effective at the elementary school level, and less effective at the middle school level. Another interpretation, however, is more agnostic. The mixed results could be seen as indicating uncertainty about the true impacts. There seems to be a positive impact, but it is not clear if it is specific to one cohort or not. This would lead to a lower impact, in the short term models of around 0.05 and in the long-term models of around 0.07—still positive but smaller. Further work, examining NYNSR-like reforms under different conditions and with additional data, would be warranted to understand fully how the NYNSR reforms changed student performance.

Finally, this study presents evidence on the feasibility of relatively low-cost evaluations of school reforms, ones that make use of existing administrative data rather than data obtained through expensive, large scale experiments. This lower cost evaluation, with sound methodology, is especially attractive in the area of school reforms, where randomized experimentation is difficult to implement and the costs of collecting new data for sufficiently large samples can be prohibitive.

References

Barnett, W. Steven. 1996. "Economics of School Reform: Three Promising Models." In *Holding Schools Accountable: Performance-Based Reform in Education,* ed. Helen F. Ladd, 299-326. Washington, D.C.: The Brookings Institution.

Berends, Mark, Sheila Nataraj Kirby, Scott Naftel, and Christopher McKelvey. 2001. *Implementation and Performance in New American Schools: Three Years into Scale-Up*. Santa Monica: RAND Corporation.

Bifulco, Robert. 2001. "Do Whole-School Reform Models Boost Student Performance: The Case of New York City." Syracuse University. Unpublished.

Cook, Thomas D., Farah-naaz Habib, Meredith Phillips, Richard A. Settersten, Shobha C. Shagle, and Serdar M. Degirmencioglu. 1999. "Comer's School Development Program in Prince George's County, Maryland : A Theory-Based Evaluation." *American Educational Research Journal* 36(3):543–97.

Cook, Thomas D., H. David Hunt, and Robert F. Murphy. 2000. "Comer School Development Program in Chicago: A Theory-Based Evaluation." *American Educational Research Journal* 37(2):535–97.

Finn, Jeremy D. and Kristin E. Voelkl. 1993. "School Characteristics Related to Student Engagement." *Journal of Negro Education* 62(3):249–68.

Fowler, William J., Jr. and Herbert J. Walberg. 1991. "School Size, Characteristics and Outcomes." *Educational Evaluation and Policy Analysis* 13(2):189–202.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66(3):361–69.

Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24(3):1141–77.

Hanushek, Eric A. John F. Kain, and Steven G. Rivkin. 2001. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." National Bureau of Economics Research Working Paper No. 8479.

Herman, Rebecca, Daniel Aladjam, Patricia McMahon, Erik Masem, Ivor Mulligan, Amy Smith O'Malley, Sherri Quinones, Alison Reeve, and Darren Woodruff. 1999. *An Educator's Guide to Schoolwide Reform*. Washington, D.C.: American Institute for Research.

Iatarola, Patrice. 2002. "The Cost-Effectiveness of New York Networks for School Renewal." Florida State University. Unpublished.

Jones, Elizabeth M., Gary D. Gottfredson, and Denise C. Gottfredson. 1997. "Success for Some: An Evaluation of a Success for All Program." *Evaluation Review* 21(6):643–70.

King, Jennifer A. 1994. "Meeting the Educational Needs of At-Risk Students: A Cost Analysis of Three Models." *Educational Evaluation and Policy Analysis* 16(1):1–19.

Lee, Valerie E. and Julia B. Smith. 1993 "Effects of School Restructuring on the Achievement and Engagement of Middle-Grade Students." *Sociology of Education* 66(3):164–87.

———. 1995. " Effects of High School Restructuring and Size on Early Gains in Achievement and Engagement." *Sociology of Education* 68(4):241–70.

———. 1997. "High School Size: Which Works Best, and for Whom?" *Educational Evaluation and Policy Analysis* 19(3):205–27.

Lee, Valerie E. and Susanna Loeb. 2000. "School Size in Chicago Elementary Schools: Effects on Teachers' Attitudes and Students' Achievements." *American Educational Research Journal* 37(1):3–31.

Lindsay, Paul. 1982. "The Effect of High School Size on Student Participation, Satisfaction, and Attendance." *Education Evaluation and Policy Analysis* 4(1):57–65.

Lopus, Jane S. 1990. "Do Additional Expenditures Increase Achievement in the High School Economics Class?" *Journal of Economic Education* 21(3):277–86.

Monk, David H. 1984a. "The Conception of Size and the Internal Allocation of School District Resources." *Education Administration Quarterly* 20(1):39–67.

———. 1984b. "Interdependencies Among Educational Inputs and Resource Allocation in Classrooms." *Economics of Education Review* 3(1):65–73.

Morgan, David L. and Duane F. Alwin. 1980. "When Less is More: School Size and Student Social Participation." *Social Psychology Quarterly* 43(2):241–52.

Northwest Regional Educational Laboratory. 1999. *Catalog of School Reform Models*. Portland: Northwest Regional Educational Laboratory.

Odden, Allan, Sarah Archibald, and Anita Tychsen. 2000. "Can Wisconsin Schools Afford Comprehensive School Reform?" *Journal of Education Finance* 25(3):323–42.

Pittman, Robert B. and Perri Haughwout. 1987. "Influence of High School Size on Dropout Rate." *Education Evaluation and Policy Analysis* 9(4): 337–43.

Pradl, Gordon, Christine Donis-Keller, Monica Martinez, and Carmen Arroyo. 2001. *New York Networks for School Renewal (NYNSR): An Implementation Study*. The NYNSR Research Consortium's Publication, Institute for Education and Social Policy. New York: New York University.

Sander, William. 1993. "Expenditures and Student Achievement in Illinois: New Evidence." *Journal of Public Economics* 52: 403–16.

Schneider, Barbara L. 1985. "Further Evidence of School Effects" *Journal of Educational Research* 78(6):351–56.

Stiefel, Leanna, Robert Berne, Patrice Iatarola, and Norm Fruchter. 2000. "High School Size: Effects on Budgets and Performance in New York City." *Educational Evaluation and Policy Analysis* 22(1):27–39.

Todd, Petra E. and Kenneth I. Wolpin. Forthcoming. "Towards a Unified Approach for Modeling the Production Function for Cognitive Achievement" *Economic Journal*.