
Amy Ellen Schwartz and Leanna Stiefel

PUBLIC EDUCATION IN THE DYNAMIC CITY: LESSONS FROM NEW YORK CITY

1. INTRODUCTION

The plight of urban schools and their failure to educate students adequately and efficiently have occupied the national discussion of public schools in America over the past quarter-century. While there is little doubt that failing schools exist in rural and suburban locations, the image of city school systems as underfinanced, inefficient, inequitable, and burdened by students with overwhelming needs is particularly well entrenched in the modern American psyche.

As the largest school district in the nation, New York City attracts particular attention to its problems. To some extent, this image reflects realities. New York City school children, like many urban students around the country, are more likely to be poor, nonwhite, and immigrant, with limited English skills and greater instability in their schooling. Moreover, the new waves of immigrants from around the world bring students with a formidable array of backgrounds, language skills, and special needs. The resulting changes in the student body pose special challenges for schools. At the same time, despite a decade of school finance litigation and reform, New York continues to have trouble affording the class sizes, highly qualified teachers, and other resources that its suburban neighbors enjoy. Finally, there is evidence of continuing segregation and disparities in performance between students of different races and ethnicities.

Nonetheless, not all the news is bad. As we describe in detail, our work on New York City's public schools—which includes extensive research on immigrant children—and our separate

work on school reform offer several reasons for optimism. First, immigrant students, who might be viewed as among those most seriously at risk of failure, are doing quite well. Our research suggests that although immigrants are somewhat segregated from the native born, this factor has little impact on the resources available in the schools they attend. Even more, immigrants in elementary and middle schools earn higher scores on average than do the native-born students who are otherwise similar to them, and the “immigrant advantage” increases over time, perhaps following the students’ acclimation or acquisition of English language skills.

Second, the school system is changing and not at all static. Each school year sees new schools open and old ones close, reorganization and reform of existing schools, and changes in curriculum, governance, and budgeting procedures, among other experiments. Whether these changes lead to improvements in test scores, more efficient use of resources, or greater equity is not always clear, but any notion that the system is intransigent and static seems inapt.

Third, advances in methods and the availability of data combined with increased public pressure for accountability have led to improvements in the quantity and quality of evaluations of the various reforms and a new emphasis on evidence to guide decision making. In some ways, New York City has been at the forefront of this movement by tracking expenditures at the school level, which allows for analysis of cost-effectiveness, and providing student-level data to researchers working to evaluate reforms in its schools.

Amy Ellen Schwartz is a professor of public policy, education, and economics and Leanna Stiefel a professor of economics at New York University.
<amy.schwartz@nyu.edu>
<leanna.stiefel@nyu.edu>

The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

Finally, some reforms and experiments are yielding positive, if modest, results. For example, evidence suggests that the first wave of small high schools created in the mid-1990s has been more successful at getting students to graduation without a significantly higher cost per graduate. In addition, budgeting reform introduced by Rudy Crew, the former chancellor of New York City’s public schools, and other whole-school reforms also seem to be yielding small positive effects on student test scores.

Our paper discusses all of these issues in greater detail, drawing lessons for urban schools in the conclusion.

2. NEW YORK CITY PUBLIC SCHOOL CHILDREN

As the largest school district in the nation, New York City educates more than 1.1 million students in roughly 1,300 public schools, with a student population that is diverse and challenging. To illustrate, we present the third-grade class of 2000-01 (Table 1, panel A). This cohort included roughly 72,000 native-born students, more than 33 percent of whom are black, nearly 40 percent Hispanic, roughly 9 percent Asian, and 14 percent white. Poverty is alarmingly common. More than 75 percent of the students are poor (as measured by eligibility for free lunch) and another 8 percent are near poor (as measured by eligibility for reduced-price lunch). Further, more than 33 percent of the students come from homes in which English is not the primary language and 5 percent have sufficiently limited English skills to be eligible for English as a Second Language or bilingual-education services.

At the same time, this cohort includes more than 10,000 students born outside the United States (hereafter referred to as immigrant or foreign-born students). That is, roughly one out of every eight third graders was foreign born in 2000-01. (Note, however, that because many of the native-born students are themselves children of immigrants, these figures in some sense understate the impact of immigrants on the public schools.) Immigrants differ noticeably from the native born in racial composition: more than 25 percent are Asian, less than 20 percent are black, 36 percent are Hispanic, and 18 percent are white. An even greater share of the foreign born are poor or near poor—in fact, only about 10 percent of foreign-born students are *not* poor. As one might expect, immigrant students are far more likely to come from homes in which English is not the primary language (more than 75 percent) and to be limited-English-proficient, or LEP (more than 25 percent). New York City’s immigrant population is extraordinarily

diverse, hailing from more than 200 countries and speaking more than 160 languages and dialects. While some arrive with strong academic backgrounds, rich and stable home lives, and poised for success in American schools, others arrive less well prepared, needing remediation, supplemental support, and special attention.

TABLE 1
Selected Characteristics of Third-Grade Students by Nativity Status

	Native Born	Foreign Born
Panel A: 2000-01		
Percentage of students who are		
Asian	0.089	0.277
Black	0.369	0.184
Hispanic	0.399	0.359
White	0.142	0.180
Female	0.499	0.495
Non-English-speaking at home	0.363	0.778
Eligible for free lunch	0.775	0.825
Eligible for reduced-price lunch	0.080	0.074
Limited-English-proficient	0.056	0.269
Test data		
Reading		
Mean score	-0.014	0.143
Percentage taking test	0.917	0.664
Math		
Mean score	-0.012	0.118
Percentage taking test	0.938	0.765
Number of students	71,931	10,428
Panel B: 1995-96		
Percentage of students who are		
Asian	0.072	0.231
Black	0.382	0.188
Hispanic	0.374	0.389
White	0.171	0.192
Female	0.504	0.496
Non-English-speaking at home	0.371	0.739
Eligible for free lunch	0.776	0.812
Eligible for reduced-price lunch	0.069	0.073
Limited-English-proficient	0.099	0.354
Test data		
Reading		
Mean score	0.006	-0.040
Percentage taking test	0.981	0.786
Math		
Mean score	0.006	-0.035
Percentage taking test	0.985	0.874
Number of students	62,513	10,845

Source: Authors’ calculations.

Notice, however, that many foreign-born students do quite well in school. Panel A of Table 1 reports the mean performance on standardized tests in reading and math. (For comparison purposes, these scores have been normalized for all students in a grade to produce a mean of zero and a standard deviation of 1.) Foreign-born students with sufficient English skills to take the standardized tests perform on average at a higher level than the native born. (Of course, many students do not take the tests, making it difficult to disentangle causality here. We return to this issue shortly.)

Finally, note that the student body changes over time, driven by differences in immigrants as well as in the native born. Consider the differences between this cohort and a similar cohort five years earlier. Panel B of Table 1 shows the characteristics of the third-grade cohort of 1995-96. Notice that there are considerably fewer native-born students in this cohort—nearly 9,000—but the number of immigrants is roughly constant. Thus, immigrants are even more important in this group. Further, the racial composition is different—fewer Asians, more Hispanics, more whites. While poverty rates are roughly similar, limited English proficiency is significantly more prevalent in both the native- and foreign-born populations in 1995-96. Finally, the proportion of students taking standardized tests is considerably higher in the earlier period, and the disparities in performance between the native- and foreign-born populations are almost zero.

This comparison of cohorts, however, ignores the change that occurs within a cohort over time, and our analysis suggests that this intra-cohort change is important. To illustrate, we consider the change in the third-grade cohort of 1995-96 by its eighth-grade year, 2000-01. As Table 2 shows, more than 20 percent of the students had left the New York City public school system (attritted) either to attend other public schools or private schools, and the attrittees are significantly less likely to be black and more likely to be white. Further, the attrittees are somewhat less likely to be poor, but more likely to be near poor, and they perform better on both reading and math tests than do continuing students.

Even more important than those who left are those who entered. Consider the eighth-grade cohort of 2000-01. Table 3 distinguishes between two groups of students in the cohort—those who entered in third grade or before and therefore were part of the third-grade cohort of 1995-96, and those who entered after third grade.¹ All told, nearly 33 percent of the eighth graders were not attending third grade in any public school in New York City five years earlier. The fraction entering after kindergarten is undoubtedly higher. Interestingly, while the differences between the attrittees and continuing students are relatively modest, the differences between the early and late entrants are stark. Nearly 45 percent of students entering after third grade are foreign born, compared with 15 percent of the

TABLE 2
Mean Characteristics of Third-Grade Students
by Attrition Status, 1995-96

	Continuing Students	Attritting Students
Percentage of students who are		
Asian	0.097	0.091
Black	0.360	0.326
Hispanic	0.376	0.376
White	0.166	0.204
Female	0.505	0.498
Non-English-speaking at home	0.424	0.429
Eligible for free lunch	0.789	0.740
Eligible for reduced-price lunch	0.068	0.076
Limited-English-proficient	0.135	0.139
Native born	0.861	0.828
Test data		
Mean score in reading	-0.015	0.076
Mean score in math	-0.010	0.053
Number of students	56,463	16,142
Percentage of all third graders in 1995-96	77.8	22.2

Source: Authors' calculations.

Notes: Continuing students are those registered in third grade in both 1995-96 and 2000-01. Students need not be continuously enrolled.

TABLE 3
Selected Means for Eighth-Grade Students
by Entrance Status, 2000-01

	Entered by Third Grade	Entered after Third Grade
Percentage of students who are		
Asian	0.113	0.142
Black	0.342	0.354
Hispanic	0.354	0.374
White	0.189	0.126
Female	0.523	0.458
Non-English-speaking at home	0.426	0.535
Eligible for free lunch	0.725	0.821
Eligible for reduced-price lunch	0.097	0.077
Limited-English-proficient	0.027	0.245
Native born	0.854	0.553
Test data		
Mean score in reading	0.111	-0.361
Mean score in math	0.119	-0.340
Number of students	46,566	21,711
Percentage of all eighth graders in 2000-01	68.202	31.798

Source: Authors' calculations.

Notes: The table presents 2000-01 means for New York City public school students enrolled in the eighth grade in 2000-01 and enrolled in the third grade in 1995-96. Students need not be continuously enrolled.

early entrants. Nearly 25 percent of the late entrants are limited-English-proficient in the eighth grade, compared with only 3 percent of the early entrants. Late entrants are significantly more likely to be poor and significantly less likely to be white. Finally, the late entrants score substantially lower on the standardized tests than do the early entrants.

The implications for policy are real. The success or failure of the public schools in delivering an eighth-grade class ready for high school hinges, in no small way, on the performance of students educated by schools outside the New York City public school system and, among those, a substantial number of schools in other countries. Put differently, this implies that there may be limits to the extent to which improving early childhood education, for example, can improve the high-school readiness of students at the end of middle school—an important goal for educators and parents. More generally, this turnover suggests that the implementation of school accountability for student performance may have to be done in a way that recognizes the particular difficulties of educating a student body that has high levels of turnover.

3. NEW YORK CITY PUBLIC SCHOOLS

Just as New York City’s public school students are diverse, so too are the city’s schools. To illustrate, we present descriptive statistics for 865 elementary and middle schools in 2000-01 (Table 4). The average elementary or middle school enrolled roughly 830 students and spent nearly \$11,000 per pupil, about \$6,200 of which was for expenses other than teachers (including administrators, support staff, books, and materials). The teacher-pupil ratio averaged .079, or roughly one teacher for every thirteen students. On average, about 80 percent of these teachers were licensed and permanently assigned, more than 70 percent had master’s degrees, nearly 60 percent had more than two years in their current school, and more than 50 percent had more than five years of experience. At the same time, the standard deviations on nearly all of these variables are substantial. While some schools enroll more than 1,000 students, others have only a couple of hundred. In some schools, virtually all teachers are licensed, while others have relatively few with licenses. School spending varies widely, driven by differences in teachers and the needs of students, as we discuss in greater detail.

Equally important is the variation in the characteristics of students. While the average school is roughly 16 percent white, the standard deviation is 23. Similar variability is seen in the other race groups. New York City public schools run the spectrum of racial diversity—there are schools that are virtually

homogeneously black, for example, and others that have very few blacks. The same can be said for many groups.

Just as the student population is dynamic and changing, so too are the public schools. To some extent, these changes reflect policy or economic changes affecting a wide range of schools. Labor market returns to education are ever-increasing, heightening the pressure to prepare students for the labor market and college. There has been an increasing focus on test scores and accountability across the nation, exemplified by the terms of the federal No Child Left Behind Act, which requires the tracking of test scores and gains in various ways. New York State has imposed its own set of accountability reforms, including high-stakes tests in fourth grade and eighth grade and rising standards for Regents high-school diplomas. The possibility of significant changes in school finance looms, as the state negotiates the implications of the Campaign for Fiscal Equity lawsuit, and policymakers and educators consider where the money will come from and how to spend it.

TABLE 4
Mean Characteristics of Elementary
and Middle Schools, 2000-01

Variable	Mean	Standard Deviation
Total per-pupil expenditures	\$10,907	\$3,169
Nonteacher per-pupil expenditures	\$6,183	\$2,102
Teacher-pupil ratio	0.079	0.020
Average school enrollment	829.7	402.3
Percentage of teachers		
Licensed and permanently assigned	80.9	17.8
With master’s degree	72.6	15.7
With more than two years of experience	59.1	19.2
With more than five years of experience	51.5	15.1
Percentage of students in schools		
Female	49.2	3.2
White	16.5	23.1
Black	35.8	30.7
Hispanic	36.6	25.9
Asian and other	11.1	15.3
Eligible for free lunch	72.3	23.9
Eligible for reduced-price lunch	7.5	5.0
Native	86.0	10.0
Non-English-speaking at home	40.4	24.5

Source: Authors’ calculations.

Notes: The sample is 865 schools with students in either fifth or eighth grade (573 have only fifth graders; 194 have only eighth graders; 98 have both fifth and eighth graders). Schools serving only special-education students are excluded. Eligibility for free lunch is calculated only for students with nonmissing data. Native students are those born on U.S. soil.

The various pressures from within and from outside have yielded many changes in the New York City schools. For example, consider recent governance changes. Just a couple of years ago, New York City Mayor Michael Bloomberg gained control over the school district, earning the power to appoint the chancellor and assuming the responsibility for the district's performance. Chancellor Joel Klein quickly implemented a reorganization of the governance of the schools. The thirty-two community school districts, which had primary responsibility for elementary and middle schools in the city, were reorganized into ten considerably larger instructional regions. Curriculum reform soon followed along with changes in third- and fifth-grade promotion policies. The effort to build new small schools continued, following the belief that small schools are more successful, funded in part by the Gates Foundation. These are just some examples of the many changes affecting public education in New York City. Others include charter schools, vouchers, reforms to the high-school articulation process, teacher certification, and principal training.

Change and reform, however, are not new. As shown in Table 5, the period between 1996-97 and 2002-03 witnessed quite a bit of turnover in the schools. Every year in that period saw a set of schools close and an even larger set of schools open. By the end of the period, there were roughly 10 percent *more* schools than there were six years earlier and, of the 1,160 schools operating in 2002-03, roughly 15 percent had opened in the past five years. (These statistics exclude adult-education schools and special-education schools, among others.) Whole-school reforms and governance reforms were implemented during the terms of many previous chancellors, including Crew and Harold Levy.

TABLE 5
Schools Opening and Closing by Year

	Closed	Opened	Operating Schools
1996-97	—	—	1,052
1997-98	9	30	1,073
1998-99	8	11	1,076
1999-2000	18	59	1,117
2000-01	9	23	1,131
2001-02	25	24	1,130
2002-03	13	43	1,160

Source: Authors' calculations.

Notes: Closed is defined as no longer operating during that year. Citywide special-education schools, schools in prisons, adult-education schools, nonpublic schools, and community-based-organization schools are excluded from the sample. Only schools with nonzero registration are included.

In general, the motivation for the various reforms and changes can be characterized as aiming to improve the efficiency of resource use and/or the performance either of students overall or particular groups of students. Of course, not all changes and reforms are effective, and it is crucial to consider whether these programs are efficient. Doing so, however, is far from straightforward.

4. IS CHANGE GOOD?

Not all change is good, and distinguishing between which innovations are successful and worthy of replication and which are not is crucial to improving schools. Unfortunately, distinguishing between “what works” and “what doesn’t work” in education is particularly complicated compared with doing so in other settings and, while there has been woefully little attention paid to this in the past, there is quite a bit of attention being paid right now. For example, the U.S. Department of Education created and funds the What Works Clearinghouse (WWC) to provide answers and disseminate findings by reviewing and vetting evaluations based upon their scientific validity and reliability.²

To the economist, the fundamental criteria for evaluating reforms center on their effect on equity and efficiency, which must be carefully defined to be useful. Even then, applying these criteria requires confronting and resolving a host of conceptual and practical difficulties. Efficiency requires that resources be deployed in such a fashion that the greatest amount of output is produced with the inputs used. Figuring out what works requires assessing whether a reform or innovation had an effect on outputs and figuring out what works best requires an understanding of the impact on cost. Thus, we need to define and measure carefully changes in outputs, changes in inputs, and ultimately the relationship between these—the production function for education. (See Stiefel et al. [2005] for more on measuring school efficiency.)

In an ideal world, there is broad consensus on the appropriate measures of efficiency as well as abundant data tracking these measures across students, schools, and school districts over time. In addition, new programs and reforms are best implemented using randomized experiments that allow us to disentangle easily the causal relationship in the data. Unfortunately, these conditions are rarely met. Data on school resources are rarely tracked at the school level. There is only grudging consent to the use of test scores to measure output and little consensus on which subjects and what types of scores to use. (While No Child Left Behind has put the federal emphasis on a set of tests and statistics, it is not at all clear that

these will be broadly accepted by state education departments and school districts around the country.) However, administrative data on individual students that can be used to track their performance over time are increasingly available, and there are some jurisdictions in which expenditures and other variables are measured at the school level. New York City is one of these.

While few reforms are intentionally adopted in a randomized fashion, the complexity of the New York City system has often meant that reforms are not universally implemented at one moment and there is often some randomness in timing and/or implementation of the reforms, creating opportunities to disentangle causality in the impact estimate.

Another important criterion for assessing reforms revolves around equity, and again there are both practical and conceptual issues. If we agree that our concern is the equitable treatment of students (compared with, say, teachers), then we need to resolve several issues. First, equitable treatment for which students—low-performing, poor, black, Hispanic, girls or boys, disabled, English language learners?—to name just a few. While the ideal reform affects all equally, it is rarely, if ever, the case.

Second, how do we measure improvement in equity? What sort of measure is appropriate? If greater equity is achieved when a reform reduces disparities in performance between two groups—say, between blacks and whites—then it is almost certainly the case that the reform delivers greater improvements in performance for one group than the other. Put differently, are we looking for equity in levels or in gains?

Third, we need to decide whether to focus attention on the equity in the distribution of resources (inputs measured in dollars, teacher counts, teacher qualifications, say) or in the distribution of outputs (such as test scores or graduation outcomes), as we have implicitly assumed in our earlier discussion. Finally, there are the usual difficulties inherent in distributional analyses—alternative measures are available and they are not always consistent. (See Berne and Stiefel [1984] for more on equity measurement in education.)

Despite these difficulties, recent experience indicates that progress is being made in evaluating school reforms, in assessing changes in both efficiency and equity. New York City in many ways is an excellent “laboratory” for studying schools—the student body is large and diverse; the many schools vary widely in size, composition, organization, and the like; and schools change over time. Further, the Department of Education collects (and has been willing to provide to researchers) detailed data on students, including test scores, socio-demographics, language skills, and nativity, along with

comprehensive school-level expenditure data. Thus, we have been able to explore in some detail the treatment and experience of immigrant students in the New York City public schools and to assess the effects of recent reform efforts. We now turn to a brief discussion of some examples from our research on New York City’s public school students.

5. THE EDUCATION OF IMMIGRANT STUDENTS

How well immigrant students fare in New York City public schools reflects, in large part, how well the school system responds to change. New countries, new languages, and new challenges are the norm, rather than the exception. One particular concern regarding immigrants derives from their propensity to settle in communities with others from their own country. This strong link between residential location and elementary school attendance may well mean that immigrant children will go to segregated schools with few native-born students, which carries with it concerns about access to social networks, peers, English language acquisition, and, to the extent that immigrants are less active politically, about the prospect of creating school communities that are insufficiently funded.

Measures of exposure and isolation show that this concern may be misplaced (Table 6). In fact, immigrants are not very segregated at all. As of 1998–99, the typical elementary or middle-school student went to a school in which 76.3 percent of his or her schoolmates are native born. The isolation index of .237 is not very high either. To be sure, some specific groups of foreign born, such as those from the Dominican Republic, the former Soviet Union, or China, are more highly isolated—their own-group isolation indexes are 10.5 percent, 17.5 percent, and 13.4 percent, respectively. However, certainly compared with the racial segregation of nonwhite (at 90.4 percent) or free- and reduced-lunch-eligible (also 90.4 percent) students, this level of segregation is mild.

Of course, the native-born peers with whom immigrant students attend school may be children of immigrants themselves, leaving open the possibility that their schools will be less well supported than other schools. Immigrant advocates, for example, often do not distinguish between the foreign born and children of immigrant communities, and claim that immigrants do not receive their fair share of spending. At the same time, parents of native-born students wonder if immigrants are taking resources from their children. How do resources vary with the representation of immigrants?

TABLE 6

Exposure of New York City Public Elementary and Middle-School Students, Immigrant and Native Born, 1998-99

	Exposure to Native Born	Isolation Index	Percentage of Total Students
Native born	0.854	0.854	0.839
Foreign born	0.763	0.237	0.161
Recent immigrant	0.767	0.117	0.073
Limited English skills	0.750	0.106	0.050
Born in Dominican Republic	0.803	0.105	0.031
Born in Mexico, Central America, or Spanish South America	0.758	0.071	0.026
Born in other Caribbean	0.811	0.093	0.024
Born in former Soviet Union	0.669	0.175	0.017
Born in South Asia	0.723	0.066	0.016
Born in China, Taiwan, or Hong Kong	0.696	0.134	0.012
Nonwhite	0.841	0.904	0.844
Eligible for free or reduced-price lunch	0.836	0.904	0.866

Source: Ellen et al. (2002, Table 4).

Our examination of the distribution of spending suggests that immigrant students receive the same level of most school resources that native-born students receive. To be specific, we estimated school-level expenditure regressions for New York City elementary and middle schools in the late 1990s. In these models, we controlled for features of the school population that traditionally garner more resources for schools—the percentage of poor, special-education, LEP students, for example—and found that, *ceteris paribus*, the percentage of immigrants in the schools rarely affects the per-pupil amount devoted to students as a whole. The representation of immigrants was significant only for nonclassroom expenditures and the percentage of teachers who have permanent teaching certification, and these work in opposite directions. Put differently, immigrants seem to draw resources in just the same way that native-born students do—because of their poverty status, English proficiency status, and special-education needs. Thus, we conclude that there is no “smoking gun” suggesting that immigrants are treated inequitably.

Of course, this equity concern about resources is closely tied to the question of how immigrants perform in the New York City schools. That immigrants receive resources equal to those of similar, native-born students may or may not be an efficient use of resources, depending on how immigrants do in school.

Performance significantly below that of the native born might suggest that resources would be more efficiently used by redistributing toward immigrants. Thus, we consider the academic performance of immigrant students; in brief, our findings suggest that this concern is unnecessary.

We estimate the nativity gap in performance—the difference in average test scores of foreign- and native-born students—for different grades and years in the late 1990s, using various specifications of a regression model to control for other differences between immigrant and native-born students. Table 7 presents representative results for fifth- and eighth-grade reading and math test scores. (As before, test scores have been normalized to a mean of zero and a standard deviation of 1.) Column 1 shows the unadjusted mean differences in performance; column 2 shows the size of the nativity gap once we control for the previous year’s performance (a value-added specification); column 3 shows the estimated nativity gap once we include a full set of control variables. On the whole, the evidence suggests that foreign-born students outperform native-born students, *ceteris paribus*.

Of course, while foreign-born students might do better on the whole, there may well be significant differences among the immigrants masked in the overall category. As Table 8 shows, there are marked differences in the characteristics of students from different regions of the world. For example, while nearly all of the Dominican students are poor, poverty is less common among Europeans. Again, while 25 percent of the Dominican students are LEP, only 1 percent of Caribbean students require English remediation. Further, special-education rates differ significantly across regions. Finally, there are differences in the length of time students have attended the New York City public schools. While native-born students have been enrolled for nearly five years, which is consistent with kindergarten entry, foreign-born students average more than one year less in the schools. While students from some regions differ marginally from the native born, students from other regions are significantly more recent additions. Do these differences translate into differences in performance across regions? As Table 9 illustrates, we find that once we control for differences in the underlying characteristics of students, there are relatively few differences across regions, although Russian and Chinese students perform particularly well. (We present results for reading tests; similar results are obtained for math.)

Notice, however, that these cross-sectional snapshots may be misleading. Suarez-Orozco (2001) argues that “among immigrants today, length of residence in the United States seems associated with declining health, school achievement, and aspirations.” This argument is shared by other researchers. While the hypothesis that the superior academic performance

TABLE 7

Selected Regression Results for Reading and Math Tests, Foreign-Born Students by Grade and Year

	Reading			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Fifth grade, 1997-98						
Foreign born	0.122*** (0.019)	0.126*** (0.010)	0.070*** (0.010)	0.061*** (0.022)	0.105*** (0.010)	0.050*** (0.009)
Number of observations	64,971	64,971	64,971	66,629	66,629	66,629
R ²	0.00	0.54	0.57	0.00	0.58	0.60
Fifth grade, 2000-01						
Foreign born	0.083*** (0.018)	0.089*** (0.010)	0.020 (0.018)	0.115*** (0.021)	0.108*** (0.012)	0.043*** (0.014)
Number of observations	71,141	71,141	71,141	72,509	72,509	72,509
R ²	0.00	0.47	0.45	0.00	0.55	0.55
Eighth grade, 1997-98						
Foreign born	-0.004 (0.024)	0.037*** (0.010)	0.023* (0.014)	-0.029 (0.028)	0.062*** (0.012)	0.026* (0.014)
Number of observations	57,465	57,465	57,465	59,749	59,749	59,749
R ²	0.00	0.58	0.60	0.00	0.56	0.58
Eighth grade, 2000-01						
Foreign born	0.014 (0.027)	0.058*** (0.013)	0.035*** (0.013)	0.099*** (0.027)	0.148*** (0.013)	0.065*** (0.013)
Number of observations	57,152	57,152	57,152	59,024	59,024	59,024
R ²	0.00	0.54	0.59	0.00	0.59	0.62
Prior-year test score	No	Yes	Yes	No	Yes	Yes
Additional variables	No	No	Yes	No	No	Yes

Sources: Schwartz and Stiefel (forthcoming, Table 5); authors' calculations.

Notes: The sample is New York City public school students who took a reading or math test. Robust standard errors are in parentheses. Demographic characteristics include age and a set of dummies indicating eligibility for free lunch, eligibility for reduced-price lunch, sex, race, and the existence of missing data. Educational characteristics are language other than English frequently spoken at home, took the language assessment battery (LAB), percentile on the LAB, scored at or below the 40th percentile on the LAB, part-time special-education participation, prior-year test score, and whether the student took the test in the prior year. School resources are nonteacher expenditures (in thousands of dollars), teacher-pupil ratio, percentage of teachers with more than five years of experience, percentage of teachers with more than two years in the school, percentage of teachers licensed and permanently assigned, percentage of teachers with a master's degree, enrollment (in hundreds), and dummy variables indicating that teacher characteristic and expenditure data are nonmissing. Cohort variables are dummies for the number of years in the New York City public schools. The teacher-pupil ratio is instrumented with the prior-year enrollment and enrollment squared.

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

TABLE 8

Characteristics of Fifth- and Eighth-Grade Students by Region, 1997-98

Region	Number of Students	Percentage Eligible for Free Lunch	Percentage Eligible for Reduced-Price Lunch	Percentage Limited-English-Proficient	Percentage Special Education	Percentage Female	Years in New York City Public Schools
Panel A: Fifth grade							
Africa	206	77.7	7.3	6.3	4.9	51.0	2.7
Caribbean	1,911	83.5	6.4	1.0	5.5	54.0	3.1
China	459	69.1	10.2	6.5	3.3	50.1	4.1
Dominican Republic	1,409	94.7	2.3	25.1	5.3	49.7	4.4
East Asia	471	45.2	21.0	2.8	4.0	50.5	3.9
Eastern Europe	329	59.0	14.9	6.1	3.0	51.7	3.7
Guyana	729	84.1	8.2	0.0	6.6	52.8	3.1
Latin America	1,296	86.1	5.6	18.5	6.9	46.6	4.4
Russia	1,127	56.8	10.6	2.0	4.7	49.1	3.9
South Asia	638	71.3	10.2	5.8	5.8	45.9	4.0
West Asia	219	68.0	7.3	3.2	8.2	51.1	4.3
Western Europe	252	56.3	12.7	3.6	6.7	46.0	3.8
All foreign born	9,046	76.6	8.1	8.2	5.5	50.2	3.8
All native born	55,925	73.5	7.0	4.3	9.7	51.1	4.9
Panel B: Eighth grade							
Africa	224	70.1	8.9	6.3	2.2	49.6	3.4
Caribbean	2,890	74.7	7.4	2.0	4.6	52.7	4.2
China	678	66.4	12.4	10.8	3.7	49.0	6.0
Dominican Republic	1,667	92.2	2.1	35.5	4.4	49.2	6.0
East Asia	665	46.6	17.4	7.4	2.4	52.3	5.3
Eastern Europe	382	61.3	12.6	6.0	3.4	52.9	5.2
Guyana	956	77.2	7.6	0.5	4.0	53.2	4.3
Latin America	1,784	84.1	5.5	21.2	5.3	46.9	6.3
Russia	1,230	49.8	13.4	2.0	2.5	48.9	4.7
South Asia	693	70.9	11.4	8.1	2.0	47.0	5.5
West Asia	257	68.5	5.4	5.1	4.7	37.0	5.9
Western Europe	266	59.4	10.2	4.1	6.4	55.6	5.7
All foreign born	11,692	72.9	8.3	10.9	4.0	50.1	5.2
All native born	45,773	66.8	7.9	3.1	8.6	50.9	7.7

Source: Schwartz and Stiefel (forthcoming, Table 8).

TABLE 9

Regional Regression Results for Reading Education Production Functions, Foreign-Born Students

	Fifth Graders		Eighth Graders	
	1997-98	2000-01	1997-98	2000-01
Russia	0.135*** (0.037)	-0.116 (0.119)	0.157*** (0.045)	0.315*** (0.073)
Eastern Europe	0.082* (0.043)	0.017 (0.055)	0.116*** (0.038)	0.151*** (0.057)
Western Europe	0.123** (0.048)	0.044 (0.048)	0.058 (0.041)	0.087* (0.051)
China	0.161*** (0.042)	0.143*** (0.044)	0.097*** (0.034)	0.080 (0.051)
East Asia	0.083*** (0.030)	0.068 (0.043)	0.090*** (0.028)	-0.042 (0.036)
South Asia	0.045* (0.027)	-0.039 (0.033)	-0.023 (0.038)	0.034 (0.039)
West Asia	0.100** (0.046)	0.079 (0.053)	-0.026 (0.038)	-0.084* (0.045)
Africa	0.082 (0.051)	0.190*** (0.053)	0.043 (0.050)	0.079 (0.054)
Dominican Republic	0.121*** (0.021)	0.065*** (0.024)	0.053** (0.021)	0.071*** (0.020)
Caribbean	0.033* (0.018)	-0.016 (0.030)	-0.006 (0.020)	-0.057*** (0.019)
Guyana	-0.155*** (0.029)	-0.037 (0.035)	-0.135*** (0.043)	-0.102*** (0.038)
Latin America	0.106*** (0.019)	0.067*** (0.024)	0.015 (0.018)	-0.004 (0.021)
Constant	0.109 (0.081)	-0.379 (0.494)	1.124** (0.445)	1.876*** (0.582)
Number of observations	64,971	71,141	57,465	57,152
R ²	0.57	0.45	0.60	0.57

Source: Authors' calculations.

Notes: The model includes controls for free-lunch eligibility, reduced-price-lunch eligibility, gender, age, ethnicity/race, English proficiency, language assessment battery scores, special-education status, prior-year reading and math scores, teacher-pupil ratio, teacher experience, teacher tenure, teacher licensing, teacher education, and school enrollment. Cohort dummies control for the number of years in the New York City public schools. Students who have zero to one year in the New York City public schools entered the system in the 1997-98 school year. Specifically, they entered on or after November 1, 1996. Students who have at least one but less than two years entered between November 1, 1995, and October 31, 1996. Fifth graders with five or more years in the New York City public schools entered on or before October 31, 1992. The teacher-pupil ratio is instrumented with the prior-year enrollment and enrollment squared.

*Statistically significant at the 10 percent level.

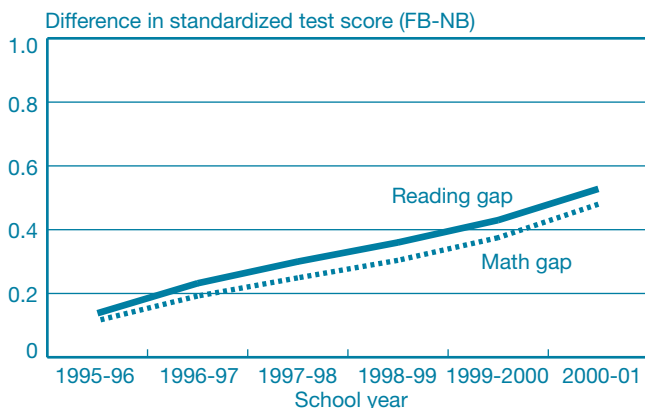
**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

of immigrant students “disappears” with time in the United States (that is, performance converges to the lower performance of native-born students) has intuitive appeal and surface validity, there is relatively little statistical evidence to support it. To address this concern, we investigate the evolution of performance of a cohort of students attending New York City schools from third through eighth grades, using a regression model to control for a range of time-varying characteristics and student-fixed effects to capture unobserved time-invariant characteristics. We find that the performance of immigrants *diverges* from that of native-born students (Chart 1). Separate analysis by race group suggests that the time path differs across groups (Chart 2). White immigrants diverge the most from their white native-born counterparts, while Hispanic immigrants show some early divergence but then begin to converge back in later grades. Overall, we find little evidence for convergence.

We have examined several dimensions of the treatment of immigrant students in the New York City public schools—a group that presents special challenges because of the students’ late entry into the schools, limited English proficiency, and the like, and that may well be at particular risk because of the group’s potentially low level of political clout. Our results are

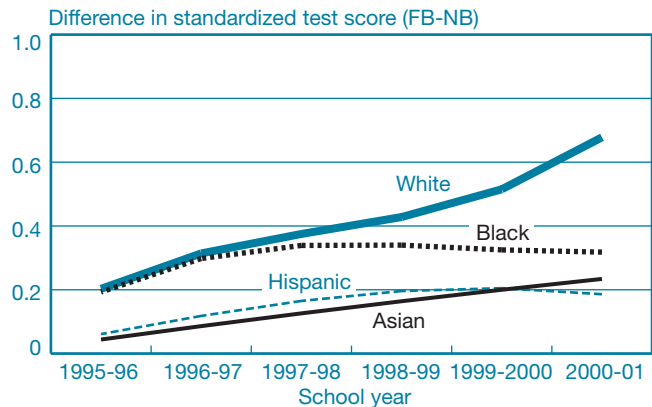
CHART 1
Regression-Adjusted Nativity Gap by Year
Standard Academic Progress Cohort; Calendar Time



Source: Schwartz and Stiefel (2005, Figure 3).

Notes: The standard academic progress cohort includes students originally enrolled in the third grade in 1995-96 who remained enrolled every year through the 2000-01 school year and it progresses one grade each year. The nativity gap is defined as the difference between the average z-score of foreign-born (FB) and native-born (NB) students. It is interpreted as the number of standard deviations by which foreign-born students outperform native-born students. The performance of students tested outside the indicated year and grade is not included. Models include student-fixed effects.

CHART 2
Regression-Adjusted Nativity Gap by Year and Race
Standard Academic Progress Cohort; Calendar Time;
Reading Scores



Source: Schwartz and Stiefel (2005, Figure 7).

Notes: The standard academic progress cohort includes students originally enrolled in the third grade in 1995-96 who remained enrolled every year through the 2000-01 school year and it progresses one grade each year. The nativity gap is defined as the difference between the average z-score of foreign-born (FB) and native-born (NB) students. It is interpreted as the number of standard deviations by which foreign-born students outperform native-born students. The performance of students tested outside the indicated year and grade is not included. Models include student-fixed effects.

encouraging. Segregation is relatively mild, resource allocation seems equitable, and, perhaps most importantly, immigrant student performance is good and trending upward. In the end, it seems that immigrants may well be *good* for the New York City public schools.

6. EVALUATING SCHOOL REFORMS IN NEW YORK CITY: SOME EXAMPLES

The dynamic nature of New York City’s public schools provides a natural laboratory for new educational policies and reforms. How well do these work? We examine three recent reforms, using data provided by the New York City Department of Education. The first, the Performance-Driven Budgeting (PDB) initiative, changed the way that resources are allocated within schools. The second, the New York Networks for School Renewal (NYNSR) project, is an example of whole-school reform, not unlike others implemented elsewhere, such as Success for All. The third is the small-schools initiative,

which continues as new small schools are opening each year in New York City and elsewhere. The methodology is relatively straightforward and replicable and, because it relies upon administrative data, it is relatively inexpensive. The implication is that evaluation is both possible and affordable and needs to be integral to policymaking. As we observed, our findings are generally positive. Reforms yield positive, if small, effects on student outcomes.

6.1 Performance-Driven Budgeting

In 1996, New York City Schools Chancellor Rudy Crew initiated an effort to move budgeting decisions toward school-level decision makers and to tie the new budgeting practice to school performance. Termed Performance-Driven Budgeting, the underlying logic was that decision makers closer to the student are better able to align resources with academic needs.

The centerpiece of our analysis of this reform is a school-level production function linking student performance on fourth- and fifth-grade tests to school inputs (teacher resources and expenditures). The effect of the PDB reform was identified as the difference in school performance before and after the PDB intervention, relative to the schools that did not implement PDB—in essence, a difference-in-difference design. As shown in Table 10, the coefficients on the “implemented PDB” variable indicate a positive, albeit small, effect of around .06 standard deviations in reading and math (in fourth-grade) test scores. To put this effect size in context, we note that educators as a rule-of-thumb aspire to sizes of .25 when initiating specific curriculum reforms; racial test-score gaps between white and black or Hispanic students are around .7. Thus, .06 is indeed small, but it is also positive.

6.2 The New York Networks for School Renewal

The New York Networks for School Renewal project had a somewhat different genesis, beginning in 1995-96 with eighty founding schools. Representing a model of whole-school reform, which involves voluntary networks and small school sizes, the project was initiated with a \$25 million, five-year grant from the Annenberg Foundation.³ Our analysis of NYNSR uses student-level data to estimate the effect of the reform on students attending fourth, fifth, or sixth grade in 1995-96, as well as an “intent-to-treat design” to disentangle the effect of the reform from all other changes. Table 11

TABLE 10
The Effect of PDB Participation on Standardized Tests

Dependent variables	Fourth Grade		Fifth Grade	
	Reading	Math	Reading	Math
Participation variable				
Implemented PDB in 1997-98	0.0557** (0.0254)	0.0599** (0.0269)	0.0568** (0.0247)	0.0187 (0.0263)
Number of observations	2,436	2,436	2,436	2,436
R ²	0.9234	0.9290	0.9252	0.9304

Source: Stiefel et al. (2003, Table 5).

Notes: PBD is the Performance-Driven Budgeting initiative. All regressions are weighted by enrollment share. All dependent variables are measured in z-scores. Test scores in all years are from the CTB (reading) or CAT (math) normal curve equivalents, except for 1998-99 fourth-grade reading and math scores. Fourth-grade students were given new state reading and math tests in 1998-99, and the Board of Education reports their scaled test scores. Regression equations include a set of teacher characteristics (percentage licensed, with more than five years of experience, with more than two years of experience, with a master’s degree; average number of days absent per year) and a set of school characteristics (percentage students female, Asian and other, black, Hispanic; average daily attendance; percentage eligible for free lunch, limited-English-proficient, resource room participant, special-education, recent immigrant) as well as school- and year-fixed effects and a group of missing value indicators, the log of expenditures, and enrollment and a constant term. Standard errors are in parentheses.

*p<.10.

**p<.05.

***p<.01.

illustrates our results, showing two- or three-year (long-term) changes in reading and math test scores in two differently specified models. On the whole, the impact estimates are positive, with many statistically different from zero, and no evidence exists of any negative effect. In addition, the size of the effects, when significant and positive, is between .16 and .25, considerably higher than those found for the PDB reforms.

6.3 Small-Schools Initiative

In the mid-1990s, reformers turned their attention to improving the performance of American high-school students. While various initiatives have been attempted—including offering child care on school sites and imposing graduation test requirements—one of the most enduring, visible, and well-funded initiatives is the “small-schools” movement. Headlines have trumpeted New York City’s (and Chicago’s) efforts to

TABLE 11

Long-Term-Impact Analysis of NYNSR Participation on Standardized Reading and Math Scores, by Cohort

	Fourth-Grade Cohort 1998-99 (Grade 7)	Fifth-Grade Cohort 1998-99 (Grade 8)	Sixth-Grade Cohort 1997-98 (Grade 8)
Baseline reading regressions			
NYNSR	0.161*** (0.036)	0.165*** (0.063)	0.064 (0.044)
R ²	0.627	0.634	0.646
Including school characteristics ^a			
NYNSR	0.155** (0.063)	0.029 (0.065)	0.062 (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*** (0.045)	0.039 (0.048)	0.047 (0.040)
R ²	0.666	0.678	0.645
Including school characteristics ^a			
NYNSR	0.229*** (0.056)	-0.113* (0.062)	0.001 (0.077)
R ²	0.680	0.699	0.667
Number of observations in all models	5,024	4,977	6,153

Source: Schwartz, Stiefel, and Kim (2004, Table 4).

Notes: NYNSR is the New York Networks for School Renewal project. Test scores are measured in z-scores transformed from normal curve equivalents for the CTB (reading) or CAT (math) exams, except for the DRP reading test scores in 1994-95 and state reading (ELA) and math test scores for the eighth grade in 1998-99. Huber's robust standard errors are reported in parentheses. All regressions include 1994-95 and 1995-96 test scores. Dummies are used for students who are female; exposed to a language other than English; Asian, Hispanic, black, and recent immigrant; and, for each year, attendance rates, language assessment battery percentiles, free- or reduced-price-lunch eligibility, resource room participation, grade retention, and advancement to a grade higher than typical; and a set of missing-value indicators. Regressions with school variables include the number of consecutive years a student has been in the same school. "Recent immigrant" and "advancement to a grade higher than typical" are dropped from the 1998-99 regressions. As of 1998-99, no recent immigrant student in 1995-96 retained that status. None of the fourth- and fifth-grade-cohort students who advanced to a higher grade than typical in 1998-99 had valid reading or math test scores for that year.

^aYear-specific school controls are total enrollment; number of teachers per 100 students; teachers' average number of days absent; the percentage of students who are black, Hispanic, Asian, free-lunch-eligible, limited-English-proficient, recent immigrants, special education, and resource room participants; and the percentage of teachers fully licensed and permanently assigned, with a master's degree, with more than five years of experience, and working more than two years in the same school.

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

convert large comprehensive high schools with up to 5,000 students into small schools with 500 or fewer students. Whether the small-schools initiative succeeds depends on its effectiveness with its own students, the impact on district costs associated with smaller units and more of them, and the effects on the larger high schools that remain. Our analysis of the small schools created in the early phases of the initiative attempts to address the first issue, using data on school expenditures and

cohort graduation rates in New York City high schools. The use of cohort graduation rates is key. The New York City Department of Education tracks students for up to seven years, beginning in ninth grade—recording whether they graduate, transfer to another school or system, drop out, or continue past four years. Thus, we can construct, for each school, the budget per graduate and examine the way it varies with school size. The findings are compelling. The small academic high schools,

most like the ones being replicated now, have a better performance record, deliver a higher cohort graduation rate, and in the end have similar per-pupil expenditures as the large schools. Put differently, the small high schools have higher graduation rates to balance their higher expenditures per pupil.

7. LESSONS

New York City, like cities around the United States and the world, faces particular difficulties providing public education efficiently and equitably. The student body is heterogeneous and dynamic. Poverty is common, and limited proficiency in English challenges many. Further, turnover is high. Each year, thousands of new students enter the New York City public schools midway through their school career, many of them from schools outside the United States. New York City schools include substantial numbers of students from dozens of countries, speaking many languages. Together, these factors pose a formidable challenge to the school system. That said, we still find much cause for optimism. Our research shows that, other things equal, immigrant students fare reasonably well. Their performance on standardized tests is good, their schools receive resources in the same measure as schools with more native-born students, and their performance seems to improve over time as they adjust to their schools and new homes. Thus,

the programs and interventions that the New York City school system has in place to address the difficulties faced by immigrant students seem to be working.

Further, the school system itself seems quite dynamic. Each year brings a wide range of reforms—in curriculum, school organization, governance, testing, and accountability—and, while not all of them work, our research on earlier reforms suggests that it is possible to use evaluation to disentangle those programs that work from those that do not. Administrative data are increasingly available, allowing relatively low-cost evaluations. Even more important, advances in econometric methods are facilitating efforts to disentangle causality and distinguish good programs and good schools from bad ones.

At the same time, there is much room for improvement. While evaluation is possible, it is still far from universal. Too many reforms are implemented and declared successes or failures without any investigation, and the largest and most sweeping reforms are rarely subject to careful evaluation.⁴ Further, evaluation can be simplified. We make too little use of randomization and access to data, and the ease of using and interpreting the data is more limited than it should be. Finally, there are many inequities and inefficiencies that continue. For instance, disparities persist between blacks, Hispanics, whites, and Asians, as well as in the allocation of teachers and resources across schools, despite significant efforts to close these gaps. Much more work remains to be done.

ENDNOTES

1. Notice that the group of students who entered by third grade is a subset of the continuing students in Table 2 because only a fraction, roughly 82 percent, of the continuing students from the third-grade cohort of 1995-96 were in eighth grade (others were in seventh grade, in special education, or elsewhere).

2. The WWC was established in 2002 by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of "what works" in education. It aims to promote informed decision making on education through a set of easily accessible databases and user-friendly reports that provide consumers with ongoing, high-quality reviews of the effectiveness of replicable educational interventions (programs, products, practices, and policies) that aim to improve student outcomes. The WWC is administered by the Institute of Education Sciences through a contract to a joint venture of the American Institutes for Research and the

Campbell Collaboration. Both organizations are nationally recognized leaders in education research and rigorous reviews of scientific evidence. Subcontractors to the project are Aspen Systems Corporation, Caliber Associates, Duke University, and the University of Pennsylvania. (See <<http://www.whatworks.ed.gov/whoware/overview.html#key>>.)

3. Other examples of whole-school reform are Success for All, Accelerated Schools, Edison Schools, Comer Schools, and New American Schools. All of these reforms aim to change many parts of the school at once (some combination of components such as curriculum, teacher attitudes, time devoted to subjects, use of technology).

4. The New York City Department of Education has requested proposals from outside evaluators for reform of its promotion/retention policy.

REFERENCES

- Berne, R., and L. Stiefel. 1984. *THE MEASUREMENT OF EQUITY IN SCHOOL FINANCE: CONCEPTUAL AND EMPIRICAL DIMENSIONS*. Baltimore: Johns Hopkins University Press.
- Ellen, I. G., K. O'Regan, A. E. Schwartz, and L. Stiefel. 2002. "Immigrant Children and Urban Schools: Lessons from New York on Segregation, Resources, and School Attendance Patterns." In W. G. Gale and J. R. Pack, eds., *BROOKINGS-WHARTON PAPERS ON URBAN AFFAIRS*. Washington, D.C.
- Schwartz, A. E., and L. Stiefel. 2004. "Immigrants and the Distribution of Resources within an Urban School District." *EDUCATIONAL EVALUATION AND POLICY ANALYSIS* 26, no. 4 (winter): 303-27.
- . 2005. "Testing the Convergence Hypothesis in Immigrant Academic Achievement: A Longitudinal Analysis." Unpublished paper, New York University.
- . Forthcoming. "Is There a Nativity Gap? New Evidence on the Academic Performance of Immigrant Students." *EDUCATION FINANCE AND POLICY*.
- Schwartz, A. E., L. Stiefel, and D. Y. Kim. 2004. "The Impact of School Reform on Student Performance: Evidence from the New York Networks for School Renewal Project." *JOURNAL OF HUMAN RESOURCES* 39, no. 2 (spring): 500-22.
- Stiefel, L., R. Berne, P. Iatarola, and N. Fruchter. 2000. "High-School Size: Effects on Budgets and Performance in New York City." *EDUCATIONAL EVALUATION AND POLICY ANALYSIS* 22, no. 1 (spring): 27-40.
- Stiefel, L., R. Rubenstein, A. E. Schwartz, and J. Zabel. 2005. *MEASURING SCHOOL PERFORMANCE AND EFFICIENCY: IMPLICATIONS FOR PRACTICE AND RESEARCH*. 2005 Yearbook of the American Education Finance Association. New York.
- Stiefel, L., A. E. Schwartz, C. Portas, and D. Y. Kim. 2003. "School Budgeting and School Performance: The Impact of New York City's Performance-Driven Budgeting Initiative." *JOURNAL OF EDUCATION FINANCE* 28, no. 3 (winter): 403-24.
- Suarez-Orozco, M. M. 2001. "Globalization, Immigration, and Education: The Research Agenda." *HARVARD EDUCATIONAL REVIEW* 71, no. 3 (fall): 345-65.

The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.