PEABODY JOURNAL OF EDUCATION, *81*(4), 95–117

School Efficiency and Student Subgroups: Is a Good School Good for Everyone?

Dae Yeop Kim
Graduate School of Education
Portland State University

Jeffrey E. Zabel
Economics Department
Tufts University

Leanna Stiefel
Wagner Graduate School of Public Service
New York University

Amy Ellen Schwartz
Steinhardt School of Education
Wagner Graduate School of Public Service
New York University

State and federal accountability reforms are putting considerable pressure on schools to increase the achievement of historically low-performing groups of students and to close test score gaps. In this article, we exploit the

Support for this research was provided by the U.S. Department of Education grant number R30ST01011R. All results and interpretations are our own.

Correspondence should be sent to Dae Yeop Kim, Graduate School of Education, Portland State University, PO Box 751, Portland, OR 97207-0751. E-mail: yeop@pdx.edu

differences among the large number of elementary schools in New York City to examine how much schools vary in the efficiency of the education they provide to subgroups. In addition, we examine the extent to which observable school characteristics can account for the variation that exists.

We find that New York City elementary schools vary in how well they educate poor students compared to nonpoor students and Asian and White students compared to Black and Hispanic students. The disparities in school efficiency measures between boys and girls are lower than for the other subgroups. There is no conclusive evidence about which school resources and characteristics are associated with more or less efficient education across all subgroups.

An important and well-known provision of the No Child Left Behind Act is the requirement that schools report performance for specific subgroups of students as well as for all students in their school. The choice of required subgroups in the law reflects the historically lower performance of poor, Black, Hispanic, English language learner, disabled, and migrant students, compared to their peers, as well as the law's intent to ameliorate low performance of these groups by holding schools accountable for their progress. It is likely, however, that schools vary in how efficiently they educate specific subgroups, with some schools more efficient with poor children and others less, for example. What can we learn about how schools vary in the efficiency of their education of subgroups? How well will schools be able to fulfill No Child Left Behind's ambitious agenda to bring all subgroups of students to proficiency by 2014? Can we identify characteristics that describe schools that are more and less efficient at this task?

In this article, we exploit the differences among the large number of elementary schools in New York City to examine how much schools vary in the efficiency of the education they provide to subgroups. In addition, we examine the extent to which observable school characteristics can account for the variation that exists. With over 600 elementary schools, and a great diversity of students, New York City is an excellent place to study these issues. The city's schools cover a wide array of neighborhoods, with students who hail from over 160 countries, speak nearly 200 languages, represent all races and ethnicities, and have varying levels of income and educational needs.

To answer questions about variation in the efficiency of subgroup education across schools, we build on previous work in which we have developed a method for identifying the contribution of a school to student performance, controlling for the characteristics of the student body and school inputs. In this previous work, we have interpreted our school measures as efficiency indicators, and we continue to label the subgroup measures this way. Efficiency measures ideally control for the amount of resources and other inputs available to an organization and then measure how well the organization uses those resources to produce output. Our methodology operationalizes this concept of efficiency.

In brief, we find that New York City elementary schools vary in how well they educate poor students compared to nonpoor students and Asian and White students compared to Black and Hispanic students. Although there is some variation in the education of girls and boys, the disparities between these latter two groups are lower than for the other subgroups. We are also able to identify some characteristics that are associated with lower or higher gaps between subgroup efficiency measures, although most of our observable school variables have a statistically insignificant effect on variation in the subgroup efficiency gaps. We conclude with implications for policy and further research.

This article is organized as follows. First, we review the literature on measuring school efficiency. Then we summarize our method for estimating school subgroup efficiency measures. Next, we describe our data and present results on the variation within schools on the efficiency of subgroup education and on factors associated with the variation. Last, we draw conclusions for policymakers and researchers.

Review of the Literature on Measuring School Efficiency in Education

Research on how to measure efficiency in the production of educational outcomes has been minimal, although interest in such measures is now high as many states try to determine an adequate level of spending for each school district within their border, hoping to abstract from inefficiency in resource usage within districts.

Duncombe and Yinger (1999) used a nonparametric measure of efficiency, derived from data envelopment analysis (DEA), to control for efficiency differences among school districts in their estimations of school district cost functions. Their DEA measures in New York state show that large districts are not necessarily more inefficient than other districts, once other inputs that vary with costs are included in their equations. Rubenstein (2005) also explored the characteristics of DEA measures of school efficiency and found that they are sensitive to variables that are held constant and that a great many districts are deemed efficient because the method allows high performance on one output, but not others, leading to a high efficiency score. Bifulco and Bretschneider (2001) compared methods of measuring efficiency in the production of education, using DEA as one of their measures.

Schwartz and Zabel (2005) developed a method for estimating measures of school efficiency by using school-level production functions that capture the process by which the inputs to schooling are translated into outputs. Using school-level data on New York City elementary schools, they undertook an extensive evaluation of different specifications of education production functions (EPFs) and address various methodological challenges in estimating the EPFs. The authors found that school rankings generated from these models are somewhat similar but significantly different from the rankings given by average test scores without controlling for differences in school inputs and student characteristics.

Zabel, Stiefel, and Schwartz (2005) analyzed discrepancies in school efficiency measures based on student-level and school-level EPFs. They investigated how aggregation from the student-level to school-level data, sample selection, and measurement error associated with test scores affect estimates of school efficiency measures based on EPFs. They found that possible estimation biases caused by data aggregation, sample selection, and measurement error do not result in empirically meaningful changes in school efficiency measures, although they can significantly affect estimated regression coefficients.

We build on these studies by expanding the student-level EPF measures of efficiency to generate subgroup-specific school efficiency measures for students of different sexes, races or ethnicities, and poverty levels.

Theoretical Framework and Methods

The centerpiece of our model of school efficiency measurement is an EPF that links an individual student's educational outcome to various inputs, including individual student socioeconomic and educational characteristics, peer group characteristics, and school characteristics. We build on Schwartz and Zabel (2005), who specify an EPF for individual *i*, in school *s* and grade *g*, at time *t*:

$$\begin{aligned} TS_{isgt} = \beta_{0g} + \beta_{1g} TS_{ip,g-1,t-1} + \beta_{2g} ST_{isgt} + \beta_{3g} G_{isgt} \\ + \beta_{4g} SC_{isgt} + T_{gt} + S_{gs} + \varepsilon_{isgt} \end{aligned} \tag{1}$$

where TS_{isgt} is a vector of test scores for student *i* in grade *g* in school *s* in year *t*. Following the general specification of value-added EPFs, this model includes a lagged test score for student *i*, which is measured in the lagged year $t - 1$ and lagged grade $g - 1$. Note that a student may not have been in the same school in the previous year. Thus, the test from the previous grade and year is specified to be $TS_{ipg-1,t-1}$ where *p* is the school attended by student *i* in year $t - 1$, which may or may not be the same as her current school. ST_{isgct} is a vector of student characteristics including dummies for students who are female, Black, Hispanic, Asian, or others; are free/reduced-price lunch eligible; are a recent immigrant; and scored at less than or equal to the 40th percentile on the Language Assessment Battery (LAB).[1] G*isgt* represents the grade-level averages of the student characteristics, which serve as proxies of peer group characteristics. SC_{isgt} represents school-level inputs and characteristics, including a teacher–pupil ratio, nonteacher expenditures, and teacher characteristics. School-level controls other than school resources are school size, grade span, the percentage of full-time special education registration, and a dummy variable reflecting whether a school has students in grade *g* who took the LAB. T_{gt} and S_{gs} are time and school effects, which capture time and school contributions to education production that are not accounted for by the observable variables (ST_{isgt}, G_{isgt}, and SC_{isgt}). The model is estimated with a panel data set of fourth-grade students in 602 New York City schools over 5 years.

The school effect, S_{gs}, which measures a school's impact on education production, is the main focus of this article. We have interpreted this school effect as a measure of overall school efficiency. The term captures the effect of unmeasured characteristics of schools, such as unobserved teacher quality, principal leadership, or effectiveness of curriculum, as well as other unmeasured and unobserved student and school characteristics, after controlling for differences in observed student inputs and school resources.

On the basis of the EPF specified in Equation 1, we can also compute school efficiency measures for student subgroups (i.e., subgroup-specific school fixed effects). To do this, the regression equation (1) is modified to include fixed effects for each subgroup rather than a single school fixed effect (S_{gs}), and the variable in ST_{igst} that corresponds to the subgroup identifier is excluded. We estimate school effects for each school overall as well as for the following three categories of student subgroups, using panel data on fourth graders who have valid reading or math test scores: boys and girls, Whites/Asians and Blacks/Hispanics, and poor and nonpoor. The White and Asian students are combined because Asian

[1] According to the New York City Board of Education, the LAB helps identify students who are entitled to bilingual/ESL programs or to measure progress in developing English language proficiency.

students' academic performance in the New York City public schools is comparable to or slightly higher than that of White students. We assign students to the poor or nonpoor groups based on their eligibility for free or reduced-price lunch.

After estimating the subgroup efficiency measures, we then analyze the likelihood that schools that are efficient at educating one subgroup are also efficient at educating the other. To do so, we examine the correlations between the subgroup school effects. A correlation between the subgroup measures close to 1 is an indication that schools are comparably efficient with both subgroups, whereas a lower correlation is an indication that there is greater variation in the efficiency of education that a given school provides to the two subgroups. We also compare subgroup efficiency measures based on reading and math tests to see how closely tied these measures are for each subgroup. To better assess how large the disparities are, we rank schools on the basis of the two subgroup efficiency measures and then divide the rankings into deciles and compare the decile rankings of the two subgroup efficiency measures.

Finally, we examine whether there are school characteristics that describe differences in the school subgroup efficiency measures. To do this, we regress the difference in the subgroup efficiency measures on averages (over time) of grade-level student characteristics and school and teacher characteristics to see if particular factors are related to the differences. These analyses are meant to identify whether there are measurable school characteristics that differentially describe high efficiency with respect to certain subgroups versus others—for example, girls versus boys, or poor versus nonpoor, or Whites and Asians versus Blacks and Hispanics.

Data

Investigations of subgroup school efficiency have been significantly limited by data scarcity and quality in the past. In the near future, as states begin to assemble longitudinal student and school data to respond to reporting and accountability requirements of No Child Left Behind, this situation is likely to change. In the meantime, however, New York City has already made available a rich set of longitudinal data on students and schools in grades K–8. We make use of these data to study the variation and explanatory factors in student subgroup efficiency in elementary schools. Specifically, for this study, we use data for students attending 602 New York City elementary schools that served third, fourth, and fifth grades during a 5-year period, 1995–96 through 1999–2000. (The 1st year provides the lagged test score for the 1996–97 fourth graders.)

The data represent a rich source of information on student and school characteristics, measured at both student and school levels. Longitudinal student data are provided by special request from the New York City Department of Education (NYCDOE). Data on school resources and characteristics are obtained from two public databases maintained by NYCDOE: the School-Based Expenditure Reports and the Annual School Reports. The School- Based Expenditure Reports contain information on school expenditures, and the Annual School Reports provide schoolwide information on teacher resources, student enrollment, and average student demographics. In addition, we use the pupil-level data to create grade-specific student background characteristics, including percentages of each grade by gender; race/ethnicity; recent immigrant status; free/reduced-price lunch eligibility; and scores on the LAB, a test designed to assess English proficiency. Table 1 provides the characteristics of the 602 sample schools and their fourth-grade students for the academic year 1999–2000.

Test Scores

Students' fourth-grade reading and math test scores as well as their lagged values are reported in Table 1. The NYCDOE administered to fourth graders the CTB/McGraw-Hill Test of Basic Skills (CTB) in reading between 1995–96 and 1997–98 and New York State English Language Arts reading test in 1998–99 and 1999–2000. To make these tests comparable over the years, we normalize test scores (i.e., create z scores) to have a mean of zero and standard deviation of 1, based on citywide averages and standard deviations for the fourth grade.[2] In 1999–2000, about 10% of fourth graders in the sample did not have their third-grade reading or math test scores from the previous year. In the estimations of EPFs, an indicator for a missing third-grade score is included as a right-hand side variable to avoid biasing results by excluding these students from the estimation.

Student Characteristics

Table 1 shows that 80% of fourth-grade students are eligible for free or reduced-price lunch and about 50% of fourth graders are female, 35% are Black, 37% are Hispanic, and 13% are Asian or other. On average, 11% of the fourth-grade students take the LAB, and 9% score less than or equal to the 40th percentile on that test, making them eligible for special English

[2]We used different types of test scores in different years. For the CTB reading tests administered between 1995–96 and 1997–98, we used normal curve equivalent scores to compute z scores. For the state English Language Arts reading test for 1998–99 and 1999–00, we converted scale scores to z scores.

Table 1

Descriptive Statistics of Education Production Function Variables, New York City Public Schools, Grade 4, Academic Year 1999–2000

Variable	*N*	*M*	*Min*	*Max*
Test scores				
Reading (*z* score)	65,261	.02	–5.29	3.91
Lagged third-grade reading	58,502	.13	–2.48	2.85
Math (*z* score)	66,045	.02	–7.21	4.95
Lagged third-grade math	58,428	.15	–2.32	2.63
Student characteristics				
Eligible for free lunch	69,397	.78	.00	1.00
Eligible for reduced-price lunch	69,397	.08	.00	1.00
Female	72,145	.50	.00	1.00
Black	72,145	.34	.00	1.00
Hispanic	72,145	.37	.00	1.00
Asian or other	72,145	.13	.00	1.00
Took LAB	72,145	.11	.00	1.00
LAB ≤ 40th percentile	72,145	.09	.00	1.00
LAB percentile	8,080	21.15	1.00	99.00
Resource room participant	72,145	.08	.00	1.00
Recent immigrant	72,145	.06	.00	1.00
School characteristics and resources				
Total enrollment as of 10/31 in hundreds	602	802.86	100.00	2,200.00
% full-time special education students	602	.05	.00	.18
School serves third–fifth grades	602	.57	.00	1.00
School serves third–sixth grades	602	.34	.00	1.00
School serves third–seventh grades	602	.00	.00	1.00
School serves third–eighth grades	602	.09	.00	1.00
Nonteacher expenditure per pupil in $1,000	602	5.29	2.64	16.19
Teacher–pupil ratio	602	.07	.05	.13
% teachers fully licensed/permanently assigned	595	.83	.40	1.00
% teachers with > 5 years of experience	595	.58	.13	.94
% teachers more than 2 years in this school	595	.78	.33	1.00
% teachers with master's degree or higher	593	.65	.13	.90

Note. $N = 72{,}145$. All test scores are measured in *z* scores. Test scores in all years are from the CTB/McGraw-Hill Test of Basic Skills (reading) or California Achievement Test (math) normal curve equivalents, except for the 1998–99 and 1999–2000 fourth-grade reading and math scores. Fourth-grade students were given state reading and math tests in these 2 years, and New York City Department of Education reported scaled scores for these tests. According to the New York City Department of Education, the Language Assessment Battery (LAB) is given to identify students who are entitled to bilingual/English as a second language programs or to measure progress in developing English language proficiency. All expenditure figures are measured in 1997 dollars, deflated using the regional consumer price index. *Resource room* includes students receiving resource room, consultant teacher, and related services. *Recent immigrants* are students who entered a U.S. school system within the last 3 years.

language instruction. The average LAB percentile score is 21. Finally, 8% of the students in the fourth grade participate in part-time special education (resource room, consultant teachers, or related services), and about 6% are recent immigrants, having been enrolled in a U.S. school for 3 or fewer years.

School Characteristics and Resources

The average enrollment of the schools in our sample is 803 students. The largest school is almost three times the average, with 2,200 students, whereas the smallest enrolls only 100. On average, schools have approximately 5% of their students in various full-time special education programs. The sample schools contain different grade spans—although a vast majority of schools (91%) serves either third through fifth or third through sixth grades, 9% of schools serve third through eighth grades.[3]

Elementary schools in our sample spent nearly $5,300 per pupil on expenditure items other than classroom teacher expenditures in 1999–2000.[4] In terms of teacher resources, these schools have an average teacher:pupil ratio of 0.07, or seven teachers per 100 students. In a typical school, 83% of the teachers are fully licensed and permanently assigned, 58% have more than 5 years' experience, 65% have worked more than 2 years in their current schools, and about 78% have master's or higher degrees.

Results

Estimates of Schoolwide Efficiency Measures

We begin with results of the estimation of overall school efficiency measures (school fixed effects) for schools with fourth grades, separately for reading and math. Although the emphasis in this article is on the size, sign, and statistical significance of the school efficiency measures for subgroups, we show results for the entire school to provide information on how student, grade, and school characteristics enter into the EPF that we use to estimate the fixed effects. The results in Table 2 present estimates for coefficients of Equation 1. In this estimation, we instrument for possible

[3]Some of these elementary schools also serve second grade or lower grades.

[4]Classroom teacher expenditures include full-time teacher salaries, preparation period payments, per diem per session payments for substitute teachers, and fringe benefits. An average New York City school spent about $4,000 per pupil on these expenditure items.

Table 2
Parameter Estimates for Fourth-Grade Education Production Function, New York City, 1995–96 to 1999–2000, Instrument Variable Regression With School Fixed Effects

Variable	*(1) Reading*	*(2) Math*
Lagged 3rd-grade test score	0.9196*** (0.0027)	0.8845*** (0.0025)
Have lagged 3rd-grade test score	0.3827*** (0.0046)	0.3279*** (0.0042)
Student characteristics		
FL eligible	–0.0654*** (0.0047)	–0.0613*** (0.0044)
RPL eligible	–0.0164*** (0.0063)	–0.0171*** (0.0058)
Female	0.0564*** (0.0027)	0.0120*** (0.0025)
Black	–0.0952*** (0.0057)	–0.1160*** (0.0053)
Hispanic	–0.0428*** (0.0052)	–0.0445*** (0.0049)
Asian/other	0.0758*** (0.0058)	0.0967*** (0.0054)
Took LAB	–1.3155*** (0.0298)	–0.8386*** (0.0242)
LAB ≤ 40th percentile	0.4030*** (0.0243)	0.3031*** (0.0203)
LAB percentile	0.0225*** (0.0005)	0.0155*** (0.0004)
Resource room participant	–0.0788*** (0.0051)	–0.1057*** (0.0046)
Recent immigrant	0.1603*** (0.0089)	0.1552*** (0.0069)
Average student characteristics		
Prop. students FL eligible	–0.0021 (0.0390)	0.0559 (0.0362)
Prop. students RPL eligible	–0.0102 (0.0572)	–0.0569 (0.0531)
Prop. female students	–0.0725** (0.0320)	–0.1014*** (0.0297)
Prop. Black students	–0.1184** (0.0563)	–0.3010*** (0.0525)
Prop. Hispanic students	–0.1516*** (0.0543)	–0.0818 (0.0504)
Prop. Asian/other students	0.1463** (0.0610)	–0.1675*** (0.0563)
Prop. students who took LAB	–0.2416 (0.3273)	–0.4584* (0.2650)
Prop. students with LAB ≤ 40th percentile	0.4002 (0.2717)	0.1113 (0.2275)
Average LAB percentile	0.0096* (0.0051)	0.0174*** (0.0041)
Prop. students participating resource room	0.0922* (0.0515)	–0.0082 (0.0480)
Prop. students who are recent immigrants	0.1751* (0.1007)	–0.0840 (0.0781)
School characteristics and resources		
Total enrollment as of 10/31 in hundreds	0.0094*** (0.0025)	0.0001 (0.0022)
School serves 3rd–6th grades	0.0252** (0.0100)	0.0102 (0.0093)
School serves 3rd–7th grades	0.0060 (0.0205)	0.0827*** (0.0189)
School serves 3rd–8th grades	0.0227 (0.0208)	0.0795*** (0.0191)
Prop. full-time special educ. students	–0.1209* (0.0633)	0.0659 (0.0583)
Nonteacher expenditure per pupil in $1,000	0.0115 (0.0139)	–0.0095 (0.0127)
Nonteacher expenditure per pupil in $1,000 squared	–0.0010 (0.0017)	–0.0125*** (0.0015)
Teacher–pupil ratio	3.0305*** (1.0228)	–1.6523* (0.9424)
Teacher–pupil ratio squared	–1.4500 (0.9197)	–5.9690*** (0.8487)

(continued)

Table 2 *(Continued)*

Variable	*(1) Reading*	*(2) Math*
Nonteacher Expenditure × Teacher–Pupil Ratio	0.0874 (0.3072)	2.1851*** (0.2833)
Prop. teachers fully licensed/permanently assigned	–0.0598* (0.0337)	0.1551*** (0.0307)
Prop. teachers > 5 years' experience	–0.0276 (0.0313)	–0.1798*** (0.0288)
Prop. teachers > 2 years in this school	–0.0513** (0.0232)	–0.1033*** (0.0211)
Prop. teachers MA or higher	0.0442 (0.0391)	–0.0832** (0.0358)
Year dummies		
Academic year 1997–98	0.0508*** (0.0063)	–0.0242*** (0.0057)
Academic year 1998–99	0.0208*** (0.0074)	–0.0601*** (0.0068)
Academic year 1999–2000	–0.0170* (0.0097)	–0.1047*** (0.0090)
Constant	–0.7145*** (0.1030)	–0.0969 (0.0951)
Observations	254,119	260,076
No. schools	602	602
R^2	.56	.61

Note. All dependent variables are measured in z scores. Standard errors are in parentheses. Test scores in all years are from the CTB/McGraw-Hill Test of Basic Skills (reading) or California Achievement Test (math) normal curve equivalents, except for the 1998–99 and 1999–2000 fourth-grade reading and math scores. Fourth-grade students were given state reading and math tests in these 2 years, and the New York City Department of Education (NYCDOE) reported scaled scores for these tests. According to the NYCDOE, the Language Assessment Battery (LAB) is given to identify and evaluate English language proficiency for students whose home language is other than English. Students earning a score less than or equal to the 40th percentile are eligible for English as a second language and bilingual services. Regression equations include school fixed effects and a group of missing value indicators for student-level free lunch and school-level teacher resource variables. FL = free lunch; RPL = reduced-price lunch; Prop. = proportion; educ. = education; MA = master's degree.

$^*p < .10$. $^{**}p < .05$. $^{***}p < .01$.

measurement error in the lagged test scores,[5] and we include indicators for students who are missing data for variables such as free or reduced-lunch eligibility or prior test scores.

Column 1 in Table 2 shows regression results for reading tests, and column 2 shows the results for math tests. The overall $R2$ values for the two equations reported in columns 1 and 2 are .56 and .61, respectively.

[5]As discussed by Zabel et al. (2005), because the lagged test score may not reflect the actual level of cumulative education up to year $t - 1$ but measure it with error, the inclusion of the lagged test score as a regressor will lead to attenuation bias. Following the authors' suggestion about how to fix the measurement error, we estimate Equation 1 by instrumenting the lagged test score with lagged test scores on different subjects. That is, the lagged reading test score is used as an instrument for the lagged math test score, and vice versa.

The estimated coefficients on the independent variables reported for reading and math test scores are similar in magnitude and level of significance. The lagged test score for both reading and math tests are positive and significant, indicating a strong relationship between prior and current test scores. The fourth graders with third-grade lagged test score data performed significantly better on both reading and math tests than did students without those scores.

Coefficients on the student characteristic variables between the reading and math regressions also show consistent signs. In general, students who are eligible for free or reduced-price lunch, non-White, and resource room program participants have significantly lower test scores than their counterparts. Those who took the LAB also have significantly lower reading and math test scores, although their scores tend to increase as they have higher LAB percentile scores. In addition, being eligible for English language services (LAB ≤ 40) results in higher test scores. Recent immigrant and female students score higher on both reading and math tests.

Coefficients on the average fourth-grade student characteristics generally show signs similar to those of individual student characteristics variables but have much lower levels of statistical significance. Furthermore, the signs and magnitude of estimated coefficients are generally consistent with those in the literature on peer effects, although no effort is made here to distinguish endogenous and exogenous peer effects.

Some school characteristics and resources have statistically significant coefficients, but the signs and particular ones are not consistent for math and reading. For example, larger schools are associated with higher fourth-grade reading scores, but not math scores, whereas schools that serve third through seventh graders or third through eighth graders (as compared to third through fifth graders) are associated with higher math scores but not reading scores. In general, the year dummies show a pattern, but again a somewhat different one for reading than for math. Finally, the school fixed effects in this equation, representing school efficiency measures, are statistically significant as a group ($F = 9.78$, $p < .001$).

We next estimate the alternative specifications of Equation 1 where we include separate school fixed effects for each subgroup and exclude the relevant subgroup indicator from the set of student-level variables. As described earlier, three subgroup analyses are examined: (a) boys and girls, (b) Whites/Asians and Blacks/Hispanics, and (3) poor and non-poor. In the interest of brevity, we do not continue to show the details of the regression equations that generate these fixed effects. The coefficients' signs and significance are similar to those shown in Table 2.

How Much Do School Efficiency Measures Differ Between Subgroups?

Table 3 reports descriptive statistics for the school subgroup efficiency measures and shows the differences in the size of the measures between groups. The relevant information on these scores is their difference from one another and the statistical significance of this difference, also shown. The higher mean efficiency scores for female, White/Asian, and nonpoor students indicate that the sample schools are more efficient at educating these groups of students than they do with male, Black/Hispanic, and poor students, who are otherwise similar. Furthermore, the table shows that a relatively larger number of schools have significantly different efficiency effects between White/Asian and Black/Hispanic or between poor and nonpoor subgroups, compared to the male and female subgroups.

Table 3
Summary Statistics for School Efficiency Measures by Student Subgroups

				No./% Schools With Significantly Different Subgroup Efficiency Measures at		
Subgroup	*N*	*M*	*SD*	*1%*	*5%*	*10%*
A. Reading						
Female	602	–0.68	0.13			
Male	602	–0.71	0.13			
Diff: female – male	602	0.03	0.07	12/2.0	39/6.5	70/11.6
White/Asian	596	–0.71	0.26			
Black/Hispanic	596	–0.79	0.13			
Diff: White/Asian – Black/Hispanic	596	0.08	0.24	52/8.7	109/18.3	154/25.8
Nonpoor	594	–0.71	0.26			
Poor	594	–0.75	0.12			
Diff: nonpoor – poor	594	0.05	0.22	29/4.9	68/11.5	107/18.0
B. Math						
Female	602	–0.04	0.16			
Male	602	–0.11	0.14			
Diff: female – male	602	0.07	0.08	6/1.0	26/4.3	44/7.4
White/Asian	597	–0.11	0.25			
Black/Hispanic	597	–0.21	0.14			
Diff: White/Asian – Black/Hispanic	597	0.11	0.23	85/13.7	153/25.6	189/31.7
Nonpoor	594	–0.12	0.26			
Poor	594	–0.17	0.14			
Diff: nonpoor – poor	594	0.05	0.23	39/6.6	99/16.7	131/22.0

Note. Diff. = difference.

Thus, New York City fourth-grade schools seem to be approximately equally efficient in their education of boys and girls but exhibit fairly irregular performance with respect to students of different races, ethnicities, and income levels. At the 10% significance level, for example, between 22% and 26% of the schools educate observationally equivalent White and Asian students more effectively than Black and Hispanic students. The percentages are almost as high for poor versus nonpoor students (between 18% and 22%).

Another way to obtain a sense of the differences in school efficiency across subgroups is through an analysis of Pearson and Spearman rank correlations between the subgroup-specific school efficiency measures for reading and math for the three subgroup categories. Table 4 presents such correlations. Note that although the correlations between female and male performance are quite high (reading and math correlations are .89 and .87, respectively), the correlations across subjects is quite a bit lower (around .5). This pattern of correlations would seem to imply that schools specialize by subject. A school that is relatively efficient with its students in math (for both sexes) will not do as well in reading (for either sex).

Table 4
Correlations of Subgroup Efficiency Measures for Reading and Math

	Pearson			*Spearman*		
Variable	*Male Reading*	*Male Math*	*Female Reading*	*Male Reading*	*Male Math*	*Female Reading*
Male math	.56	—		.53	—	
Female reading	.89	.51	—	.89	.48	—
Female math	.53	.87	.50	.49	.85	.48
	W/A Reading	*W/A Math*	*B/H Reading*	*W/A Reading*	*W/A Math*	*B/H Reading*
W/A math	.36	—		.39	—	
B/H reading	.42	.24	—	.50	.27	—
B/H math	.26	.44	.55	.31	.50	.52
	Nonpoor Reading	*Nonpoor Math*	*Poor Reading*	*Nonpoor Reading*	*Nonpoor Math*	*Poor Reading*
Nonpoor math	.45	—		.46	—	
Poor reading	.50	.34	—	.61	.40	—
Poor math	.24	.49	.56	.32	.62	.52

Note. W/A = White/Asian; B/H = Black/Hispanic.

Comparing White/Asian to Black/Hispanic subgroups or poor to nonpoor subgroups, however, reveals correlations of school efficiency scores that are quite low, for either math or reading (ranging from .42 to .62). Across subjects, the correlations are even lower (ranging from .24 to .56). Schools differ for given subjects and even more so across subjects in their performance with these subgroups.

A final way to compare the subgroup efficiency measures is to look at the rankings by the size of the school subgroup effect. To do so, we group the schools into the top 10%, middle 80%, and bottom 10%, in order of the size of their school effects for each subgroup individually. The top and bottom 10% are one way of classifying "good" and "bad" schools on the basis of their efficiency. We then calculate the percentage of schools that are assigned to different groups by each pair of subgroup efficiency measures. For example, if a school is in the top 10% of schools on the basis of its efficiency measure for girls, but in the bottom 10% on the basis of its efficiency measure for boys, we would describe it as changing its ranking across three groups for the girl-to-boy comparison. Second, we group the schools into 10 deciles. We then calculate the percentage of schools that differ by more than 1 decile for each pair of subgroup efficiency measures.

The results, given in Table 5, are largely consistent with those from the previous correlation analysis. Comparing subgroup efficiency measures between boys and girls, only 12% to 13% of the sample schools change their rankings across the three groups—top 10%, middle 80%, and bottom 10%. However, the percentage of schools changing their performance rankings across the three groups rises to 23% to 25% between the poor

Table 5
Percentages of Schools That Differ Across Subgroup Efficiency Measures

Measure	*Differ Across Three Groups*[a]	*Differ by More Than 1 Decile*
Reading		
Male versus female	13.0	18.9
White/Asian versus Black/Hispanic	27.0	52.9
Poor versus nonpoor	25.3	47.3
Math		
Male versus female	12.3	36.9
White/Asian versus Black/Hispanic	25.3	51.7
Poor versus nonpoor	22.9	44.4

[a]The three groups are the top 10%, the middle 80%, and the bottom 10%, ranked by size of subgroup efficiency measure (subgroup fixed effect).

and nonpoor subgroups and 25% to 27% between the White/Asian and Black/Hispanic groups.

Similar results are given by the comparison of decile rankings between the three categories of student subgroups. For both reading and math, the decile rankings show the lowest level of discrepancies between boys and girls, 18.9% for reading and 36.9% for math. Between the poor and nonpoor subgroups, 44.4% to 47.3% of the sample schools changed more than 1 decile ranking. The largest discrepancies in the decile rankings of subgroup efficiency are found between the White/Asian and Black/Hispanic subgroups. The school efficiency rankings for these two racial/ethnic groups differ by at least 1 decile in more than 50% of the schools.

Apparently schools vary widely in their effectiveness in educating students by race and by income but not so much by gender. Next, we take advantage of this variation to explore whether there are school characteristics that are associated with better and worse outcomes for the gap in efficiency measures between subgroups.

Determinants of Differences in Subgroup Efficiency Measures

We estimate regressions of the differences in the subgroup efficiency measures on the averages across all years of peer group, school, and teacher characteristics. The results may shed light on the observable factors that differentiate the relative performance of one subgroup relative to the other. We weight all observations by the standard errors for the differences in the subgroup efficiency measures derived from estimating the student-level EPFs, which provides more weight to the more precisely estimated differences. We do this because the standard errors will be positively influenced by small numbers of students in any subgroup, which means that using them as weights will lower the influence of schools with small numbers of one subgroup. Unfortunately, segregation in New York City's schools means that schools with only a few students in a subgroup are not uncommon, and weighting means they will not have disproportionate influence on the results.

Regression results for the three subgroup categories are presented in Table 6. The dependent variables in these six regressions are the difference in the subgroup efficiency measures. Columns 1 and 2 show the difference between female and male (female minus male), columns 3 and 4 show Black/Hispanic minus White/Asian, and columns 5 and 6 show poor minus nonpoor. Note that although there are a total of 602 schools, only 596 and 594 are included in the regressions for the Black/Hispanic versus White/Asian and the poor versus nonpoor subgroups because some schools had no students in one of the two subgroups under consideration.

Table 6

Regression Results for Differences in Subgroup Efficiency Measures, New York City, Fourth Graders

	Female – Male		*Black/Hispanic – White/Asian*		*Poor – Nonpoor*	
	(1) Reading	*(2) Math*	*(3) Reading*	*(4) Math*	*(5) Reading*	*(6) Math*
Student characteristics:						
Grade average						
% FL	0.0134	–0.0411	0.0047	–0.1663	–0.1204	–0.1247
	–0.0221	–0.0367	–0.1394	–0.1364	–0.1637	–0.1529
% RPL	0.0510	–0.0594	–0.0523	0.5451	–0.3366	–0.3191
	–0.0639	–0.1013	–0.3807	–0.3747	–0.4465	–0.4217
% female	0.1883**	–0.4463***	0.5653	–0.2505	0.0452	–1.6677***
	–0.0795	–0.1040	–0.3922	–0.3849	–0.4581	–0.4328
% Black	–0.0162	0.0038	–0.0428	0.0452	–0.0674	–0.0113
	–0.0183	–0.0313	–0.1159	–0.1159	–0.1362	–0.1302
% Hispanic	–0.0249	–0.0141	0.0449	–0.0160	–0.1105	–0.3118**
	–0.0228	–0.0353	–0.1350	–0.1311	–0.1584	–0.1468
% Asian/other	–0.0377	–0.0634	–0.1188	–0.2163	–0.0656	–0.0511
	–0.0249	–0.0405	–0.1506	–0.1501	–0.1772	–0.1687
% took LAB	0.0941	0.0021	–0.5967**	0.2507	0.0736	0.5761*
	–0.0597	–0.0724	–0.2789	–0.2707	–0.3241	–0.3014
% use resource room	–0.1332*	0.1383	0.8034**	0.1094	–0.0582	–1.0242***
	–0.0734	–0.0943	–0.3720	–0.3494	–0.4170	–0.3926
% recent immig.	–0.0308	0.5041**	1.2403	0.1633	1.4299	–0.9510
	–0.1392	–0.2028	–0.7615	–0.7498	–0.8961	–0.8452

(continued)

Table 6 *(Continued)*

	Female – Male		*Black/Hispanic – White/Asian*		*Poor – Nonpoor*	
	(1) Reading	*(2) Math*	*(3) Reading*	*(4) Math*	*(5) Reading*	*(6) Math*
School characteristics and resources						
School efficiency	0.0052	0.0673**	–0.1583	0.0545	–0.1774	–0.0937
	–0.0219	–0.0283	–0.1115	–0.1050	–0.1310	–0.1180
Grade span: 3–6	–0.0021	0.0072	0.0190	–0.0370	0.0481	0.0185
	–0.0058	–0.0077	–0.0301	–0.0285	–0.0350	–0.0321
Grade span: 3–7	–0.0041	–0.0812	0.2003	–0.0049	–0.1798	0.2720
	–0.0495	–0.0647	–0.2444	–0.2398	–0.2859	–0.2691
Grade span: 3–8	0.0028	0.0149	–0.0077	0.0924*	–0.0600	0.0169
	–0.0105	–0.0133	–0.0509	–0.0497	–0.0585	–0.0554
% special educ.	–0.0621	0.0093	–0.8876**	–0.0798	0.1527	0.5361
	–0.0925	–0.1098	–0.4364	–0.4062	–0.4884	–0.4570
Total enrollment	–0.0007	0.0016	–0.0066	–0.0054	0.0103*	0.0157***
	–0.0009	–0.0013	–0.0051	–0.0049	–0.0059	–0.0055
Nonteacher exp.	0.0130**	0.0066	0.0139	–0.0503**	–0.0413	–0.0639**
	–0.0051	–0.0064	–0.0261	–0.0238	–0.0285	–0.0268
Teacher–pupil ratio	–0.7022*	0.2594	–4.7652***	1.3069	3.4531*	5.1052***
	–0.4164	–0.4726	–1.8143	–1.7501	–2.0830	–1.9678

Teacher characteristics						
% licensed	0.0914**	–0.0737	–0.5334**	–0.5889***	–0.6998***	–0.7806***
	–0.0449	–0.0576	–0.2151	–0.2134	–0.2526	–0.2396
% >5 years exp.	–0.0126	–0.0282	0.3969**	0.2285	0.4271**	0.1906
	–0.0358	–0.0464	–0.1764	–0.1725	–0.2054	–0.1933
% > 2 years in school	–0.0444	0.0759	–0.1247	0.1707	0.3640	0.9003***
	–0.0505	–0.0677	–0.2550	–0.2507	–0.2980	–0.2818
% MA	0.0013	–0.0507	–0.2423	–0.0561	–0.3819	–0.3672
	–0.0418	–0.0553	–0.2108	–0.2078	–0.2418	–0.2305
Observations	602	599	593	596	599	598
R^2	.07	.12	.09	.08	.07	.16

Note. The dependent variable is the difference in subgroup efficiency measures. The regression is weighted using the standard errors for the difference in subgroup efficiency measures. Standard errors are reported in parentheses. FL = free lunch; RPL = reduced-price lunch; LAB = Language Assessment Battery; immig. = immigrant; educ. = education; exp. = expenditure; MA = master's degree.

$^*p < .10$. $^{**}p < .05$. $^{***}p < .01$.

For all six columns of Table 6, a positive sign for a coefficient estimate indicates that an increase in the corresponding variable is favorable toward the performance of the female, Black/Hispanic, and poor subgroups relative to the male, White/Asian and nonpoor subgroups, respectively. A negative coefficient implies the opposite result. Also note that the regressions control for overall school efficiency to determine whether schools that are relatively efficient with girls, Blacks and Hispanics, or poor students are also ones that are doing an inefficient job with their students overall.

Beginning with the female/male differential (Table 6, columns 1 and 2), we see that for the full sample of 602 schools, the female percentage of students in the grade has a positive impact on the efficiency of a school for female reading and a negative impact for female math; schools are more efficient with girls in reading but worse in math relative to boys when the concentration of girls is higher. This result controls for variation across time in the concentration of girls because it is based on differences in fixed effects that include this variation in their estimation. That is, on average, over time, this concentration affects efficiency. Another way to understand the result is to note that the estimation of the EPF in Table 2 shows that the percentage of girls in the grade has a negative impact on the performance of all students' individual test scores in both reading and math but that females do significantly better than boys, on average, in both subjects. Of the other student characteristics in Table 6, the percentage of recent immigrants in the grade has a significantly positive impact on the relative school efficiency for girls in math.

Of the school characteristics and resources, the overall school efficiency measure is positive and statistically significant for girls in math (the only time this variable has a significant coefficient), implying that efficient performance with girls is associated with efficient performance for all students. Not only is there no trade-off (which would be indicated by an insignificant coefficient), but also higher school efficiency overall seems to help girls relatively. Also, higher nonteacher expenditures have a positive impact on the relative reading performance of girls, whereas the teacher–pupil ratio has a negative impact. Of the teacher characteristics, only the percentage of licensed teachers has a positive impact on the relative reading performance of girls. None of the school or teacher characteristics appear to have a significant impact on the relative math performance of girls.

For the Black/Hispanic to White/Asian differential (columns 3 and 4 of Table 6), the percentage of students who took the LAB has a negative impact, and the percentage of students who use the resource room has a positive impact, on the relative school efficiency for the Black/Hispanic subgroup in reading. None of the student characteristics appear to significantly affect the relative school efficiency for the Black/Hispanic subgroup in math. The strongest impact among the school and teacher characteristics on the relative school efficiency for the Black/Hispanic subgroup in both reading and math comes from the percentage of licensed teachers in the school; the higher the percentage, the worse the Black/Hispanic subgroup efficiency relative to the White/Asian subgroup.

The strongest results are for the poor versus nonpoor subgroups (columns 5 and 6 in Table 6), particularly in math. The percentage of females, Hispanics, recent immigrants, and students who use the resource room have a negative impact on school efficiency for math of the poor subgroup relative to the nonpoor subgroup. None of the student characteristics appear to significantly affect the relative school efficiency for the poor subgroup in reading. Among the school characteristics and resources, total enrollment has a positive impact on the relative school efficiency for the poor subgroup in both reading and math. Additional resources appear to have an ambiguous impact on the nonpoor subgroup; increases in nonteacher expenditures, the percentage of licensed teachers, and the percentage of teachers who have been in the school for more than 2 years blunt the relative math school efficiency, whereas increases in the teacher–pupil ratio and the percentage of teachers with master's degrees enhance the relative efficiency for math for the poor subgroup. Further, an increase in the percentage of licensed teachers reduces the relative efficiency in reading, whereas an increase in the percentage of teachers with more than 5 years of experience adds to the relative efficiency in reading for the poor subgroup.

Conclusions

State and federal accountability reforms are putting considerable pressure on schools to increase the achievement of historically low-performing groups of students and to close test score gaps. In this article we have presented evidence on the variation in how efficiently schools educate particular groups of students in a large urban district just before the passage of the No Child Left Behind Act. We explore the possibility that differences in school resources and other factors are associated with differences in measured efficiency.

Using a production function model of student performance, we estimated school efficiency in the education of various student subgroups. We found significant variation in the contribution of New York City elementary schools to the efficient production of student test scores by race/ethnicity and poverty status. Schools that do an efficient job with their White and Asian students are not particularly likely to do the same with their Black and Hispanic students. Likewise, schools that efficiently educate nonpoor students often do less well with poor students. School efficiency for boys and girls is more highly correlated within schools. Also, schools are inconsistent in their contribution to math and to reading; doing well with students in one subject does not correlate with doing well in the other subject. Schools with fourth grades have a long way to go before they are consistently efficient in the education of all subgroups across multiple subjects. Although we looked only at elementary schools, and within those schools at the performance of fourth graders, we have no reason to think that other levels of school or grades would show widely different results.

This variation in the efficiency of educating subgroups of students affords an opportunity to learn whether there are school factors that are associated with better and worse school contributions to the efficiency of reducing gaps across groups of students. Although our results do not reveal many such factors, they do indicate a few that are worth pursuing further by policymakers as well as researchers. For example, across racial subgroups, the percentage of part-time special education students positively affects efficiency in producing reading results for Blacks and Hispanics, perhaps indicating returns to having high concentrations of such students to work with in a school. On the other hand, the percentage of certified teachers has a negative impact on this group's efficiency compared to Whites and Asians in both math and reading. This might indicate a need to increase the skills of certified teachers in teaching reading to Black and Hispanic students. For poor students, efficiency in producing math results is negatively affected by more peers who are female, Hispanic, recent immigrants, and who are in part-time special education programs. School size, on the other hand, positively affects the Black–Hispanic relative efficiency, whereas teacher characteristics show a mixed effect. In sum, there is not a clear message about which school resources are associated with more or less efficient education across all subgroups.

To the extent that these results continue to apply to schools in urban areas, the success of No Child Left Behind's strategy to rely on schools to "fix" the test score gaps will have to involve increased school efficiency in producing subgroup success. Differences in efficiency are not clearly related to any particular resource configuration, and this means that schools will have to look to the ways in which they use the resources at their disposal.

References

Bifulco, R., & Bretschneider, S. (2001). Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, *20*, 417–429.

Duncombe, W., & Yinger, J. (1999). Performance standards and educational cost indexes: You can't have one without the other. In H. Ladd, R. Chalk, & J. Hansen (Eds.), *Equity and adequacy in education finance: Issues and perspectives* (pp. 260–282). Washington, DC: National Academy Press.

No Child Left Behind Act of 2001, Pub L. No. 107–110, 115 Stat. 1425 (2002).

Rubenstein, R. (2005). The reliability of school efficiency measures using data envelopment analysis. In L. Stiefel, R. Rubenstein, A. E. Schwartz, & J. E. Zabel (Eds.), *Measuring performance and efficiency: Implications for practice and research* (pp. 93–117). New York: Eye on Education.

Schwartz, A. E., & Zabel, J. E. (2005). The good, the bad and ugly: Measuring school efficiency using school production functions. In L. Stiefel, R. Rubenstein, A. E. Schwartz, & J. E. Zabel (Eds.), *Measuring performance and efficiency: Implications for practice and research* (pp. 37–66). New York: Eye on Education.

Zabel, J. E., Stiefel, L., & Schwartz, A. E. (2005, May). *Student-level and school-level production functions: Effects of aggregation and sample composition on efficiency measures and coefficient estimates.* Paper presented at the annual conference of the Urban Institute Conference on Uses of Longitudinal Data in Education.