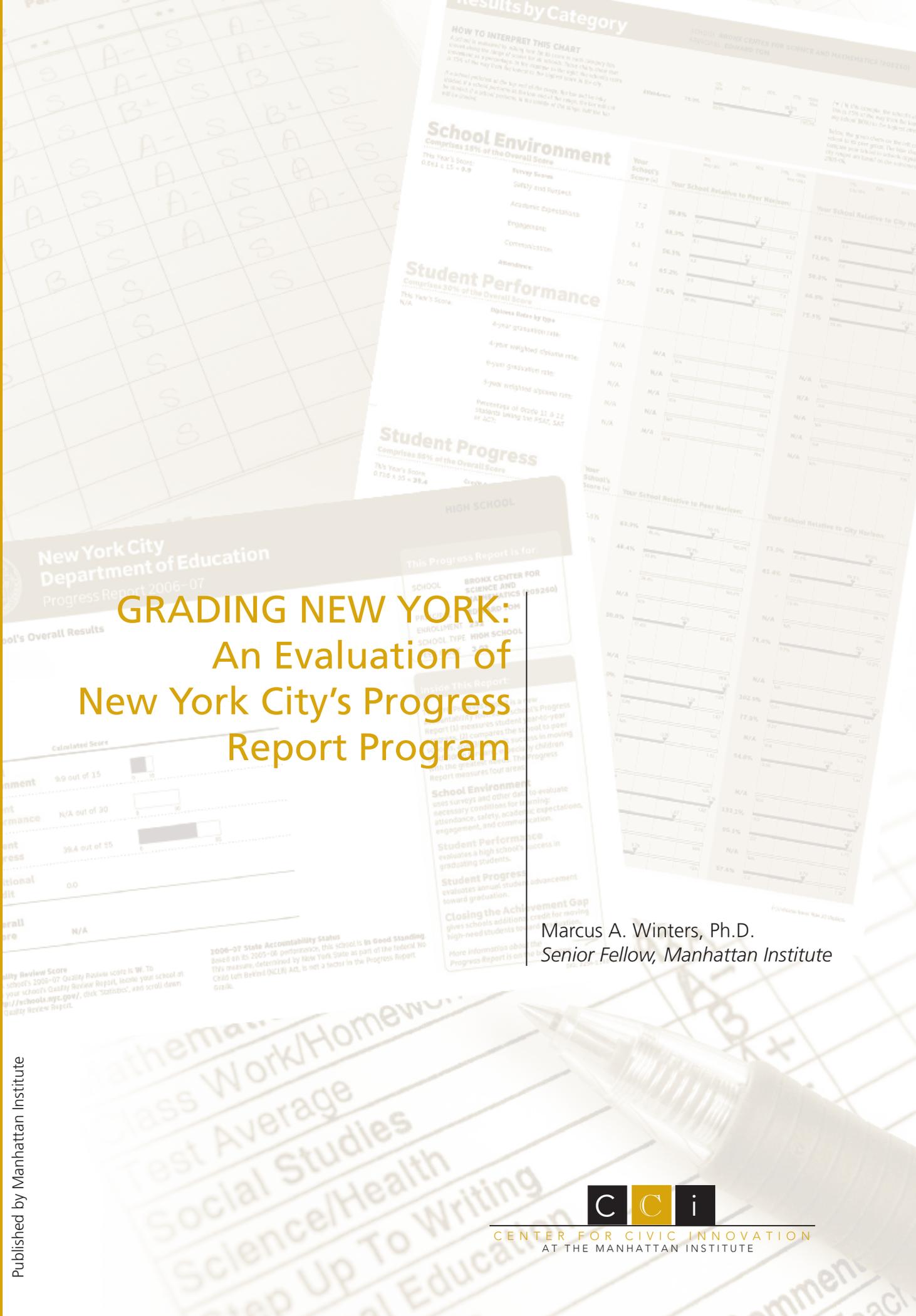


GRADING NEW YORK: An Evaluation of New York City's Progress Report Program

Marcus A. Winters, Ph.D.
Senior Fellow, Manhattan Institute



EXECUTIVE SUMMARY

In 2006-07, New York City, the largest school district in the United States, decided it would follow several other school systems in adopting a progress report program. Under its program, the city grades schools from A to F according to an accumulating point system based on the weighted average of measurements of school environment, students' performance, and students' academic progress.

The implementation of these progress reports has not been without controversy. While many argue that they inform parents about public school quality and encourage schools to improve, others contend that grades lower morale at low-performing schools. To date there has been too little empirical information about the program's effectiveness to settle these questions.

This paper incorporates student-level data in a regression-discontinuity design to study the impact of a school's receipt of a particular grade – A, B, C, D, or F — on student proficiency in math and English one year later.

The main findings of the paper are as follows:

- Students in schools earning an F grade made overall improvements in math the following year, though these improvements occurred primarily among fifth-graders.
- Students in F-graded schools did no better or worse in English than students in schools that were not graded F.

ABOUT THE AUTHOR

MARCUS A. WINTERS is a senior fellow at the Manhattan Institute. He has conducted studies of a variety of education policy issues including high-stakes testing, performance pay for teachers, and the effects of vouchers on the public school system. His research has been published in the journals *Education Finance and Policy*, *Economics of Education Review*, *Teachers College Record*, and *Education Next*. His op-ed articles have appeared in numerous newspapers, including the *Wall Street Journal*, *Washington Post*, and *USA Today*. He received a B.A. in political science from Ohio University in 2002 and a Ph.D. in economics from the University of Arkansas in 2008.

CONTENTS

1	Introduction
2	Previous Research
3	New York City’s Progress Report Policy Table 1. Total Point Range to Earn Particular School Grades by School Type
4	Overall Performance in New York City Table 2. Comparison of Mean Scores in New York City over Time
5	Data and Method Table 3. Descriptive Statistics Overall and by School Grade in 2006–07
7	Results Table 4. Replication of Rockoff and Turner (2008) Results from Aggregate Data Table 5. Impact of School Grades on Student Math Proficiency Table 6. Impact of School Grades on Student English Proficiency
9	Conclusions
10	Endnotes
11	References

GRADING NEW YORK: AN EVALUATION OF NEW YORK CITY'S PROGRESS REPORT PROGRAM

Marcus A. Winters

I) INTRODUCTION

Several public school systems in the U.S. have recently adopted policies intended to hold schools accountable for student outcomes, as measured typically by standardized math and English exams. One type of accountability policy centers on what are often referred to as “progress reports.” Programs using them provide schools with what are essentially public report cards, which grade them from A to F and often bring them material rewards or sanctions.

The New York City public school system, the largest school district in the United States, adopted a progress report program, which first graded schools on the basis of their performance at the end of the 2006–07 school year. Under the program, schools accumulate points based on their students’ performance on standardized exams and a variety of other factors. Besides risking public disgrace, schools that repeatedly receive F or D grades are subject to review and ultimately face takeover by the city. Schools that earn high grades are eligible to receive rewards.

The goal of the city’s policy is twofold: to inform parents about school quality; and to encourage schools to improve in response to incentives. Many argue, however, that the policy harms public schools

by depressing the morale of teachers and others. For example, a November 11, 2007, editorial in the *New York Times* argued that “the practice of giving, say, an F, to an otherwise high-performing school that lags in student improvement for a single year stigmatizes the entire school and angers parents.”

Unfortunately, the often vigorous debate between those who argue that the grading policy is essential to improving New York City’s public schools and those who believe that it is detrimental to them has thus far occurred in a data vacuum. This paper seeks to inform this debate with empirical evidence on the program’s effectiveness.

In particular, we follow the strategy of previous work on Florida’s schools in order to evaluate the impact of grading New York City schools on their students’ achievement one year later. The design of New York’s program allows for the use of a “regression discontinuity” approach, which, under certain reasonable assumptions, allows for a causal interpretation of the impact of earning a particular grade—A, B, C, D, or F—on school productivity as measured by student academic performance.

Our findings are somewhat mixed. Using data on students in grades four through eight, we found that students in schools that received an F grade in 2007 made academic improvements in English that were on a par with the improvements of students in schools that received better grades. However, we find that students in F- and D-graded schools made meaningful improvements in math relative to other schools, though this result appears to have been caused primarily by student progress in the fifth grade. In summary, our results suggest that schools may have responded to the F sanction by improving their performance and that there is no reason to believe that the sanction of a low grade harmed student achievement.

The paper continues in six parts. Section 2 provides a brief overview of previous research evaluating a similar progress report program in Florida. In Section 3, we discuss the design of New York City’s policy. In order to give our results on the relative improvement of D- and F-graded schools greater context, we pres-

ent some information about overall school progress in New York City in Section 4. We then devote Section 5 to discussing our methodology and data. We report results from estimation in Section 6, including a replication of results obtained by Rockoff and Turner (2008), who employed a similar design to study New York’s policy but used aggregate data. Section 7 states our conclusions.

2) PREVIOUS RESEARCH

For a thorough review of research evaluating accountability policies overall, see the recent survey by Figlio and Ladd (2008). For our current purposes, we focus on previous evidence evaluating progress reports as a specific form of accountability. In particular, New York City’s progress report policy is quite similar in its design to Florida’s A+ Program, which has graded schools in that state since 1998. Though the programs provide different incentives for schools that earn certain grades—in particular, until recently, students in Florida public schools that received more than one failing grade in a four-year period became eligible for private school vouchers—both utilize a point-based system that determines whether schools receive certain letter grades that carry important consequences.

Florida’s A+ Program has been the subject of several studies. Greene (2001) and Greene and Winters (2004) used aggregate school-level data to directly compare the educational gains made by differently graded schools. They found that schools that received an F grade made substantial academic improvements relative to other schools. These results were confirmed by Chakrabarti (2005), who went on to find evidence that the results were not driven by regression to the mean.

Some recent studies of the A+ Program have used student-level data and have taken advantage of its known grade thresholds to pursue a regression-discontinuity approach. West and Peterson (2006) limited their sample to students in schools that earned point totals barely qualifying them for an F grade or barely missing the benchmark and thus earning them a D

grade. Like other studies using aggregate data, West and Peterson's study found evidence that the incentives of schools earning an F grade had a positive impact on student academic proficiency.

Using a more flexible regression-discontinuity design, Rouse and others (2007) also drew upon individual-level data, but they included students in all schools throughout Florida. Their model incorporated a control for a cubic function of the point total earned by a school, which, under certain reasonable assumptions, allows for causal interpretations of the impact on a school of the receipt of a particular grade. They found additional evidence that the incentives of the F-grade sanction led to increased school performance. These findings were replicated by Winters, Greene, and Trivitt (2008), who also utilized this procedure and found that the school-grading policy improved student proficiency in science, which is not part of the grading process.

A recent study by Rockoff and Turner (2008) follows the procedure of Rouse and others (2007) in evaluating the impact of progress reports in New York using aggregated data. The earlier paper found that schools that received an F or a D grade in 2006–07 had statistically and substantially higher scores in math and reading in 2007–08. One value of the present paper is its replication of these previous results, which provides confidence in both papers' estimates.

Though there are other differences, the primary difference between the Rockoff and Turner (2008) work and the present paper is that here we utilize student-level data. Use of student-level data allows for more precise estimation and lends itself to a value-added approach to account for unobserved student heterogeneity that is not available in the school-aggregated data.

3) NEW YORK CITY'S PROGRESS REPORT POLICY

New York City's progress report policy rates schools on a variety of factors, according to an accumulating point system based on the weighted average of metrics intended to measure

school environment, student performance, and student academic progress. It then assigns grades from A to F to schools, according to certain benchmarks. Progress reports were first issued at the beginning of the 2006–07 school year. The city claims that the reports are “designed to help principals and teachers accelerate academic achievement” and that the policy “enables students, parents, and the public to hold the DOE and its schools accountable for student outcomes and improvement” (New York City Department of Education, 2007).

The first factor in determining the school's grade is school environment. This metric uses information from school and parent surveys about school safety and parental engagement. The environment index accounts for 15 percent of the total points that can be awarded.

The remainder of the points earned by a school are linked to performance on the state's standardized math and English exams. The value assigned to the percentage of students with test scores that meet the proficient or advanced benchmark on these tests is 30 percent of the total score that can be awarded. This measure rewards those schools in which students meet a particular academic level, although it may put schools in which students have lower beginning proficiency at a disadvantage. To compensate, 55 percent of the school's potential points are linked to the progress that students make on the standardized math and English tests during the year. This value-added measure takes into account the percentage of students making at least a year's worth of academic progress as well as the average change in proficiency scores of students who began the year with proficiency in the bottom third of the school. Schools can earn additional bonus points if students deemed “high need” make exemplary gains on the state exams.

The resulting scores on each of these factors are then further adjusted to account for the school's performance relative to the rest of the schools in the district and a grouping of schools with similar characteristics. The scores on each of these elements are weighted, as indicated above, in order to produce the school's total points under the system.

Table 1. Total Point Range to Earn Particular School Grades by School Type

Grade	Elementary		K-8		Middle	
	Min	Max	Min	Max	Min	Max
A	64.0	100+	64.0	100+	65.2	100+
B	49.9	64.0	50.3	64.0	50.5	65.2
C	38.8	49.9	38.1	50.3	38.8	50.5
D	30.9	38.8	29.4	38.0	30.9	38.8
F	0.0	30.9	0.0	29.4	0.0	30.9

As reported in New York City Department of Education (2007). Point system is a weighted function of factors related to school environment, student performance, and student progress relative to the city as a whole and matched peer schools.

Schools are then assigned grades from A to F on the basis of the number of total points earned. The range of overall points that yield particular grades is reported in Table 1. The table shows that there are slightly different point requirements for elementary, middle, and K-8 schools, for which we account in the analysis.

Many commentators in New York have argued that the grading system does not accurately measure school “quality.” For example, it has often been pointed out in the popular press that many of the same schools earning poor grades under the city’s system receive high marks under the different accountability system that the No Child Left Behind Act calls for, and vice versa.

This study is not particularly interested in the extent to which the progress report policy accurately measures school quality. Under no circumstances should this paper be interpreted as suggesting that the program is accurately or inaccurately identifying successful or “failing” schools. The identification of those factors that underlie a successful school involves value judgments that only communities, the school district, and elected representatives can make.

Nor is this paper concerned with whether progress reports have improved school effectiveness generally in New York. It may be that every school responds positively or negatively to the grading policy, regardless of whether it receives a high grade or a low grade at the outset. Our procedure does not lend itself to measuring general improvements throughout the school system.

Rather, the goal of this paper is to measure how schools that are officially and publicly deemed to be failing and thus face sanction respond to that designation. In particular, we evaluate whether student proficiency in such schools suffers or increases as a consequence of the schools’ success in earning higher grades. As we will see in the next section, the point system used to grade schools allows us to control for unobserved differences in school quality during estimation. We emphasize, however, that for our purposes, it is the grade and points themselves that were earned under the system that are important, not the particular factors that are responsible for a higher score or grade.

Schools that receive a poor grade under the program face unspecified sanctions and even restructuring or closure if they fail to improve.¹ However, the act of stigmatizing schools as “failing” could have a motivating effect on them. Several researchers have speculated that accountability policies could “shame” schools into better performance (Figlio and Rouse 2005, Ladd 2001, Carnoy 2001, Harris 2001). In this paper, we are not particularly concerned with the causes of improvements in student performance, though inquiry into the causes is a clear avenue for future research.

4) OVERALL PERFORMANCE IN NEW YORK CITY

As mentioned above, our research did not directly measure whether the progress report program led to general improvements or declines in the performance of public schools in

New York City. However, reviewing some summary statistics about overall progress in the school system can help place any results about the relative performance of schools earning particular grades into a broader context.

Table 2 summarizes the performance of New York City schools on fourth- and eighth-grade math and English exams in 2006, 2007, and 2008, using data aggregated to the school level in our data set.² Between 2006 and 2008, schools made statistically significant progress in both grades and in both math and English. The gains were largest in eighth-grade math and smallest in fourth-grade English. In fact, scores in English were actually statistically lower in 2007 than in 2006, but schools made progress in the next year.

	2006	2007	2008
Grade 4 ELA	652.8	651.7 #	654.4 ***, **+
Grade 8 ELA	632.7	639.2 **+	642.9 ***, **+
Grade 4 Math	667.3	672.2 **+	676.5 ***, **+
Grade 8 Math	633.4	641.6 **+	652.5 ***, **+

*** Greater than 2006 at $p < 0.01$
 **+ Greater than 2007 at $p < 0.01$
 # Note: For ease of interpretation, only a one-tailed test for whether later year has greater value than prior years is reported. In this case, a one-tailed test indicates that the 2007 score is statistically lower than the 2006 score, $p = 0.0284$.

Thus, the overall story on the state tests is one of relative improvement from 2006 to 2008. Though these gains are statistically significant (that is, we can have high confidence that the true gain, once we take into account measurement error, is greater than zero), we are not able to say that these overall gains are directly related to the progress report program, nor is it the place of this paper to conclude that such improvements are substantial enough to warrant overall optimism about the city's schools.

5) DATA AND METHOD

We utilize a student-level data set provided by the New York City Department of Education. The data set includes demographics

and test scores on the state's standardized math and English exams for the universe of New York City public school students enrolled in grades three through eight from the 2006–07 through the 2007–08 school years. We are also able to link students to the schools they attend and thus school grades and points earned under the policy at the end of the 2006–07 school year.

Table 3 presents descriptive information about the schools in our data set overall and disaggregated by the letter grade earned by the school at the end of 2006–07. These descriptive statistics are not identical to, but do closely match, those reported by Rockoff and Turner (2008), who used data aggregated by the Department of Education.

One difficulty with the data set is that the state does not claim that the results of its math and English exams are “vertically aligned” across grades. When results are vertically aligned, a particular score should indicate a certain level of proficiency regardless of the grade for which the exam was prepared. So a fifth-grade student with a score of 600 would have the same level of reading proficiency, as measured by a fifth-grade test, as a third-grade student with a score of 600, as measured by a third-grade test, and so on. Our lack of a vertically aligned score is important for measurement because it means that the relationship between a student's previous year's score and current score could be affected by grade level. This causes a difficulty in estimation because our method is to pool students across grades into a single regression equation. Specification checks (not reported here) suggested that there are slight but significant differences in the relationship between previous and current student proficiency across grades. To account for these, along with estimating models that include all grade levels, we report models restricted to each grade level tested individually.

We follow the regression-discontinuity method first presented in this context by Rouse and others (2007) to study Florida's similar school-grading program. This method takes advantage of the discrete cutoffs in the continuous point system utilized to assign schools particular letter grades. We slightly modify this procedure to fit better the particular design of the New York program.³

Table 3. Descriptive Statistics Overall and by School Grade in 2006–07

	All Schools	F	D	C	B	A
Number of Schools	977	41	84	253	373	226
School Type						
Elementary	585	26	51	156	225	127
K–8	120	5	9	31	49	26
Middle	272	10	24	66	99	73
Percent Black	33.5%	45.5%	44.5%	36.3%	32.2%	26.4%
Percent Hispanic	40.1%	39.2%	40.0%	38.2%	39.7%	42.9%
Percent White	14.0%	10.3%	10.3%	15.3%	14.5%	13.8%
Percent Asian	11.8%	4.3%	4.61%	9.6%	13.0%	16.4%
Average Score Math 2007	671.8	658.5	661.7	668.6	673.3	679.1
Average Score Math 2008	674.2	663.1	664.8	670.4	675.7	681.8
Average Score English 2007	653.1	642.3	644.5	650.5	654.3	659.4
Average Score English 2008	656.2	648.8	648.7	654.0	657.0	661.6
Overall Points	54.0	23.6	35.0	44.7	56.6	72.6
Environmental Points	7.6	4.9	5.6	6.8	7.9	9.3
Performance Points	13.6	8.4	9.9	12.1	13.9	17.1
Progress Points	30.2	9.0	18.3	24.1	32.2	42.0
Additional Points	2.3	0.4	0.7	1.2	2.4	4.3
School-level descriptive statistics aggregated from student-level data set.						

We use a cross-sectional regression model to measure how the relationship between student and school characteristics affects the student’s math or English score on the 2007–08 administration of the exam. In particular, we run regressions where the student’s 2007–08 test score is the dependent variable and independent variables include a cubic function of the student’s score on the exam in 2006–07,⁴ observable characteristics about the student (race, ethnicity, special-education status, etc.), observable characteristics about the school (percentage of students who are of a particular race or ethnicity, etc.), and whether the school is listed as an elementary, K–8, or middle school. The model also controls for a cubic function of the number of points earned by the student’s school in each of the categories of the overall point system at the end of the 2006–07 school year and the letter grade earned by the school at the end of that year. Finally, we include an interaction between points earned on each of the input factors and the school type (elementary, K–8, or middle school). These interactions account for the fact that the cutoffs from the point system vary somewhat by school type, as shown in Table 1.

The central assumption of our procedure is that there is no difference in school quality that is conveyed in the school’s grade that is not also accounted for in (a cubic function of) the number of points that a public school earned in each category under the formula. If this assumption holds, we can interpret the estimate as the impact of a school’s receipt of a particular grade on a student’s academic proficiency.

The basic idea behind this technique is to take advantage of the known cutoffs above or below which schools are assigned different letter grades according to the policy. The continuous point system provides a direct measure of the quality of each public school as determined by the school system. Though important for policy purposes, the cutoffs on the point scale at which a school earns an A, B, C, D, or F grade are set at somewhat arbitrary points and thus convey little to no additional information about the school’s performance that is not already represented in the point total. Schools with similar point totals are likely to be similar in their effectiveness, but whether their score falls on one side or the other

of a cutoff will determine whether they receive the F-grade sanction.

For instance, a public elementary school earning a point total of 31.0 under the grading system is likely educating its students just as well as another public elementary school that earns 30.8 points. However, these schools face very different incentives, since the former would receive a D grade and the latter an F grade. By controlling for each school’s point total, we are thus able to measure the independent impact of earning a particular grade on the school’s previous productivity level.

We make a couple of sampling restrictions that are worth mentioning. First, we exclude students who were tested in the third grade in 2007–08. Since we utilize a lagged dependent variable, and testing begins in the third grade, these students must have been retained in the third grade and thus may categorically differ from students in other grades. Second, we exclude students whose school is listed as a high school on its progress report. The data set contains observations of students taking the seventh- and eighth-grade exams who are listed as attending a high school, though the progress report definition suggests that such identified schools would teach grades nine through twelve. The data appear to indicate that these are specialty schools (for the arts, etc.), so we chose to eliminate them from the data set. However, our results remain robust when these sampling restrictions are relaxed.

It is possible that focusing on treatments that disproportionately affect students in low-performing schools, as this study does, may be affected by regression to the mean. Schools and students at the bottom of the achievement distribution may have such low scores partly because of random error. If a negative random error were more present in these schools, improvements made on tests in later years could be an inflated measure of a child’s academic progress.⁵ In their similar study in Florida, Rouse and others (2007) present a series of specification tests indicating that regression to the mean is not the driving force behind their results. Unfortunately, the timing of the beginning of wide-scale testing in New York City

does not allow us to adopt similar tests there. Thus, it remains possible that our results are affected by regression to the mean.

6) RESULTS

We first aggregate our data set in order to replicate the recent results reported by Rockoff and Turner (2008). The results of this test are reported in Table 4. Though not identical, the coefficient and standard-error estimates in Table 4 closely mirror those reported in Rockoff and Turner’s paper.

Table 4. Replication of Rockoff and Turner (2008) Results from Aggregate Data

VARIABLES	Math	English
School Grade 2006–07		
A	-1.555 [1.401]	-1.383 [1.067]
B	0.0409 [0.779]	-0.422 [0.593]
D	2.322** [0.930]	0.291 [0.708]
F	4.927*** [1.706]	2.250* [1.299]
Observations	977	977
R-squared	0.911	0.917
* significant at p < 0.10 ** significant at p < 0.05 *** significant at p < 0.01		
Robust standard errors in brackets. Models additionally control for school level, cubic functions of the school’s peer index, environmental performance, progress performance, additional score, and an interaction between school type and these functions.		

This replication suggests that the Rockoff and Turner paper’s finding that F- and D-graded schools made bigger improvements than higher-graded schools continues to hold. It also lends some confidence that the data utilized to estimate our models of primary interest using student-level data are accurate. This confirmation is particularly important in view of the fact that Rockoff and Turner (2008) rely on data that were reported at the school-aggregated level, while we utilize data aggregated from our individual-level data set.

Table 5 reports the results of estimation in math overall and for each particular grade level. As in the aggregate data, the overall model that includes all grade levels continues to find a statistically significant and substantial positive effect after a school has received an F or a D grade. Here we find that students in F-graded schools made test-score improvements that were 3.5 scale points higher than students in C-graded schools. In standard-deviation terms, attending an F school has a positive one-year impact of about a 0.18 standard deviation in math proficiency relative to students in schools that earned a C grade.

The additional columns of Table 5 report results in math in regressions restricted only to the particular grade level. We do find evidence that the result varies across grade levels. In particular, we find a strong positive impact in students attending an F or a D school in the fifth grade. However, the results in other grades are

statistically insignificant, and the coefficients of interest are negative in the fourth grade. The reasons for such different effects across grades are unclear. However, it is worth noting that the coefficient estimates on each of the cubic factors of the student's lagged math score are similar, though the small differences are statistically significant. This gives some confidence in our overall estimate (column 1) because it indicates that lack of vertical alignment in the test scores across grades is probably not having a large impact on estimation of the model.

Table 6 reports our results in reading, following a similar format. We find no significant difference between the reading performance of students in F- or D-graded schools and that of students in schools receiving better grades. A result lacking statistical significance is found both in the overall regression and in each of the grade-level regressions.

Table 5. Impact of School Grades on Student Math Proficiency

	All Grades	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Prior Math Score	-18.10*** [0.302]	-19.05*** [1.048]	-17.09*** [0.685]	-20.42*** [0.767]	-10.75*** [0.712]	-23.44*** [0.707]
Prior Math Score Squared	0.0287*** [0.000451]	0.0305*** [0.00155]	0.0272*** [0.000999]	0.0325*** [0.00115]	0.0171*** [0.00110]	0.0366*** [0.00107]
Prior Math Score Cubed	-1.4e-05*** [2.24e-07]	-1.5e-05*** [7.65e-07]	-1.3e-05*** [4.85e-07]	-1.6e-05*** [5.75e-07]	-8.5e-06*** [5.70e-07]	-1.8e-05*** [5.39e-07]
School Grade 06–07						
A	-1.288 [1.110]	1.299 [1.721]	-3.441** [1.727]	0.844 [2.333]	0.524 [2.235]	-4.209 [2.689]
B	-0.372 [0.615]	0.551 [0.915]	-1.919** [0.926]	1.207 [1.249]	-0.0103 [1.218]	-1.431 [1.454]
D	1.653** [0.677]	0.798 [1.050]	3.288*** [1.089]	2.748 [1.820]	-1.21 [1.677]	2.847 [1.829]
F	3.537** [1.587]	-2.176 [2.494]	8.179*** [2.353]	5.712 [4.196]	2.989 [3.402]	5.778 [3.621]
Observations	317531	65492	65434	60988	62721	62895
R-squared	0.671	0.621	0.663	0.674	0.677	0.696
* significant at p < 0.10 ** significant at p < 0.05 *** significant at p < 0.01						
Dependent variable is the student's score on the New York State math exam. Bootstrapped standard errors clustered by school in brackets. Models additionally control for borough, school percent Indian, school percent Asian, school percent Hispanic, school percent black, school percent multiple race, school percent English-language learner, student race, whether the student is an English-language learner, whether the student is disabled, school level, cubic functions of the school's peer index, environmental performance, progress performance, additional score, and an interaction between school type and these functions. The All Grades model additionally accounts for student's grade level.						

Table 6. Impact of School Grades on Student English Proficiency

	All Grades	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Prior English Score	-12.63*** [0.268]	-15.71*** [0.514]	-7.463*** [0.484]	-13.73*** [0.482]	-16.42*** [0.663]	-11.96*** [0.612]
Prior English Score Squared	0.0209*** [0.000421]	0.0261*** [0.000787]	0.0124*** [0.000777]	0.0222*** [0.000739]	0.0267*** [0.00102]	0.0198*** [0.000981]
Prior English Score Cubed	-1.0e-05*** [2.19e-07]	-1.3e-05*** [4.01e-07]	-6.3e-06*** [4.15e-07]	-1.1e-05*** [3.77e-07]	-1.3e-05*** [5.18e-07]	-1.0e-05*** [5.22e-07]
School Grade 06-07						
A	-0.104 [0.786]	0.204 [1.591]	0.0586 [1.329]	0.265 [1.269]	-0.0745 [1.334]	1.246 [1.902]
B	-0.149 [0.397]	-0.701 [0.887]	-0.459 [0.698]	0.845 [0.717]	-0.0862 [0.695]	0.396 [1.054]
D	0.0777 [0.479]	0.314 [0.989]	1.103 [0.756]	0.345 [0.880]	-0.462 [1.041]	-1.268 [1.298]
F	1.096 [1.086]	0.70 [2.080]	1.517 [1.664]	0.242 [2.009]	-0.77 [2.078]	-0.14 [2.309]
Observations	312349	64398	64308	60223	61622	61797
R-squared	0.595	0.611	0.553	0.585	0.618	0.619
* significant at p < 0.10 ** significant at p < 0.05 *** significant at p < 0.01						
Dependent variable is the student's score on the New York State English exam. Bootstrapped standard errors clustered by school in brackets. Models additionally control for borough, school percent Indian, school percent Asian, school percent Hispanic, school percent black, school percent multiple race, school percent English-language learner, student race, whether the student is an English-language learner, whether the student is disabled, school level, cubic functions of the school's peer index, environmental performance, progress performance, additional score, and an interaction between school type and these functions. The All Grades model additionally accounts for student's grade level.						

7) CONCLUSIONS

In this paper, we have evaluated the impact of schools earning particular grades under New York City's progress report policy on student academic proficiency. The regression-discontinuity methodology, by taking advantage of the city's continuous point system for assigning school grades, allows us to make causal interpretations of the impact of such school grades on student progress.

Our results can be construed as indicating a mixed-positive effect from receipt of an F or a D grade under the policy. We find that students in F-graded schools made significant and substantial improvements in math, though these results appear to be primarily the result of progress made by fifth-grade students. We find no evidence that a school's grade has a significant impact on student proficiency in English.

ENDNOTES

1. New York City Department of Education website, accessed September 22, 2008:
<http://schools.nyc.gov/Accountability/SchoolReports/ProgressReports/Consequences/default.htm>.
2. We had to begin with the 2006 school year because scale scores prior to 2006 are not comparable with those in the later years.
3. For a more technical treatment of our procedure, see http://www.manhattan-institute.org/pdf/cr_55_tech_version.pdf.
4. A cubic function simply means that we included a variable for the score, another variable for the score squared, and another for the score cubed. Use of the cubic function allows for a more flexible model because it relaxes the assumption of linearity in measuring the impact of school points on student proficiency. That is, only controlling for the student's prior score makes the strong assumption that every point has the same impact on the student's proficiency the next year. The cubic function allows us to account for any nonlinearities in this relationship. This same basic argument holds for our use of a cubic function for each of the components of the school's overall points under the progress report system.
5. Think of a child who took a test near a window and became distracted by a loudly barking dog. The child's test score on the exam would be lower than his true proficiency due to the accident of his location. When he took the exam the next year, the child was not distracted and posted a score that better reflected his true proficiency. However, in the data set, it will appear that he made a larger proficiency gain than he truly did. Since F schools have students with relatively low scores, it is possible that a disproportionate number of students in these schools had scores that, for some reason, were lower than their true level. If such a result were due to random error, we would be worried about regression to the mean.

- Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Carnoy, M. 2001. "Do School Vouchers Improve Student Performance?" *American Prospect* 12, no. 1:42–45.
- Chakrabarti, R. 2005. "Do Public Schools Facing Vouchers Behave Strategically? Evidence from Florida." Manuscript. Program on Education Policy and Governance.
- Figlio, D. N., and H. F. Ladd. 2008. "School Accountability and Student Achievement." In *Handbook of Research in Education Finance and Policy*, ed. Helen F. Ladd and Edward B. Fiske. New York: Routledge.
- Figlio, D. N., and C. Rouse. 2005. "Do Accountability and Voucher Threats Improve Low-Performing Schools?" *Journal of Public Economics* 90: 239–55.
- Greene, J. P. 2001. "An Evaluation of the Florida A-Plus Accountability and School Choice Program." Manuscript. Manhattan Institute.
- Greene, J. P., and M. A. Winters. 2004. "Competition Passes the Test." *Education Next* 4, no. 3: 66–71.
- Harris, D. 2001. "What Caused the Effects of the Florida A+ Program: Ratings or Vouchers?" In *School Vouchers: Examining the Evidence*, ed. Martin Carnoy. Economic Policy Institute.
- Ladd, H. F. 2001. "School-Based Educational Accountability Systems: The Promise and the Pitfalls." *National Tax Journal* 54, no. 2: 385–400.
- New York City Department of Education. 2007. "Educator Guide: The New York City Progress Report."
- Rockoff, J. E., and L. J. Turner. 2008. "Short-Run Impacts of Accountability on School Quality." Unpublished manuscript.
- Rouse, C. E., J. Hannaway, D. Goldhaber, and D. N. Figlio. 2007. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." National Center for Analysis of Longitudinal Data in Education Research, Working Paper 13.
- West, M. R., and P. E. Peterson. 2006. "The Efficacy of Choice Threats within School Accountability Systems: Results from Legislatively Induced Experiments." *Economic Journal* 116, no. 510: C46–62.
- Winters, M. A., J. P. Greene, and J. Trivitt. 2008. "Building on the Basics: The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects." Manuscript. Manhattan Institute.

CENTER FOR CIVIC INNOVATION

Stephen Goldsmith,
Advisory Board Chairman Emeritus
Howard Husock,
Vice President, Policy Research

FELLOWS

Edward Glaeser
Jay P. Greene
George L. Kelling
Edmund J. McMahon
Peter Salins
Fred Siegel
Marcus A. Winters

The Center for Civic Innovation's (CCI) mandate is to improve the quality of life in cities by shaping public policy and enriching public discourse on urban issues. The Center sponsors studies and conferences on issues including education reform, welfare reform, crime reduction, fiscal responsibility, immigration, counter-terrorism policy, housing and development, and prisoner reentry. CCI believes that good government alone cannot guarantee civic health, and that cities thrive only when power and responsibility devolve to the people closest to any problem, whether they are concerned parents, community leaders, or local police.

www.manhattan-institute.org/cci

The Manhattan Institute is a 501(C)(3) nonprofit organization. Contributions are tax-deductible to the fullest extent of the law. EIN #13-2912529