Grading New York: An Evaluation of New York City's Progress Report Program

Marcus A. Winters, Ph.D.
Senior Fellow
Manhattan Institute for Policy Research
52 Vanderbilt Ave, 3rd Floor
New York, NY 10017
Phone: 212-599-70003
Fax: 212-599-3494
mwinters@manhattan-institute.org

JEL Codes: H11; I28;

Keywords: Accountability, High-stakes Testing;

Abstract:

This paper utilizes student level data to evaluate the short-run impact of schools earning an F grade under New York City's progress report policy. Use of a regression discontinuity design allows for causal interpretation of the impact of earning particular grades on student academic proficiency. The results suggest that students in schools earning an F grade at the end of the 2006-07 school year made overall improvements in math the next year, though this is primarily driven by students in the fifth grade. Students in F graded schools did not make improvements in English relative to students in schools that earned higher grades.

1) Introduction

Several school systems have recently adopted policies intended to hold schools accountable for their students meeting particular goals on outcome measures such as standardized math and reading exams. One important type of such accountability policies centers on what are often referred to as "progress reports." Such programs provide schools with what are essentially public report cards grading them from A to F, and oftentimes tie these grades to further rewards or sanctions.

The New York City public school district, the largest school district in the United States, adopted such a progress report system and first graded schools based on 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. Along with the potential public shame, schools that receive multiple F or D grades are subject to review and potential takeover by the city. Schools that earn high grades are eligible to receive additional rewards.

Since its adoption, the progress report program has been the focus of heated debate in New York. Some argue that the policy is essential in order to give parents understandable information about the quality of the schools in their area and to provide low performing schools with an incentive to improve. However, those opposed to the program argue that identifying schools as "failing" discourages rather than encourages schools and thus will have a negative impact on academic proficiency.

Following prior work in Florida, this paper evaluates the impact of receiving particular grades under New York's policy on student achievement one year later. The design of New York's program allows for 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 measured by student academic performance.

Our findings are somewhat mixed. We find that students in schools that received an F grade in 2007 made academic improvements in reading that were on par with students in schools that received other grades. However, we find that students in F and D graded schools made significant and meaningful improvements in math relative to other schools, though this result appears to be primarily driven by student progress in the fifth grade.

The paper continues in four parts. Section 2 provides a brief overview of prior research evaluating a similar progress report program in Florida. In Section 3, we discuss the design of New York's policy. We then discuss the method and data in Section 4. We report results from estimation in Section 5, including a replication of previous results reported by Rockoff and Turner (2008) using aggregate data. Section 6 concludes.

2) Previous Research

For a thorough review of research evaluating accountability polices overall see the recent survey by Figlio and Ladd (2008). For our current purposes, we focus on prior evidence evaluating progress reports as a specific form of accountability. In particular, New York's progress report policy is quite similar in its design to Florida's A+ Program, which has graded schools in that state since 1999. 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 (2008), 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. (Chakrabarti 2007; Chakrabarti 2008; West and Peterson 2006; Rouse and others 2007).

In the design most directly related to the current paper, Rouse and others (2007) also used 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 about the impact of a school receiving 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 model developed by Rouse and others (2007) to evaluate the impact of progress reports in New York using aggregated data. That previous paper found that schools that received an F or D grade in 2006-07 had statistically and substantially higher scores in math and reading in 2007-08. One value of the current paper is its replication of these previous results.

Though there are some 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 individual 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'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 to schools from A to F 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 it "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 taken into account in the school's grade is school environment. This metric uses information from school and parent surveys about issues regarding safety and parental engagement within the school. The environment index accounts for 15 percent of the school's total points.

The remainder of the points earned by a school are linked to performance on the state's standardized math and reading exams. The percentage of students with test scores that meet the proficient or advanced benchmark on these tests account for 30 percent of the school's overall score. This measure rewards schools for students meeting a particular academic level, but may put schools where students have lower beginning proficiency at a disadvantage. To account for this, 55 percent of the school's points are linked to the progress that students make on the standardized math and reading tests during the year. This value-added measure takes into account the percent of students making at least a year's worth of academic progress and the average change in proficiency scores for students who began the year with proficiency in the bottom third in 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 city 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.

Schools are then assigned grades from A to F based on 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.

[TABLE 1 ABOUT HERE]

There are two potentially important treatments for schools that receive poor grade under the program. First, the city has warned that schools that receive F or D grades may be subject to unspecified sanctions and even restructuring or closure if they fail to improve.[1] Secondly, the act of distinguishing schools as "failing" could have a motivational effect on schools. In particular, several researchers have suggested that such policies are effective because they "shame" schools into better performance (Figlio & Rouse 2005; Ladd 2001, Carnoy 2001, Harris 2001). In this paper we are not particularly concerned with which of these or any other aspects drive increases in student performance, though this is a clear avenue for future research.

This paper is concerned with the short-run effects of the policy as measured by the impact of the first set of school grades on student proficiency on year later. One limitation of this need to focus only on the very short run is that schools may not have had time to respond to the program in time to see an impact on student achievement. Though the policy and grading methodology was made public in April of 207, schools first received their progress report scores and grades in late September, and these were made public in early November. Students were then administered the state English exam only two months later -- in mid January. Schools had more time to have an impact on student math proficiency, which was evaluated on the state test in early March.

4) Overall Performance in New York

The analysis in this paper only allows us to measure the improvements of schools earning particular grades under the progress report program relative to other schools. It

[1] New York City Department of Education website, accessed 9-22-08: http://schools.nyc.gov/Accountability/SchoolReports/ProgressReports/Consequences/default.htm

does not directly measure whether the progress report program has led to general improvements or declines in the performance of public schools in New York. However, it is useful to review some summary statistics about overall progress in the school system in order to place our results into a broader context.

[TABLE 2 ABOUT HERE]

Table 2 summarizes the performance of New York City schools on the grades 4 and 8 math and English exams in each year from 2006-2008 from data aggregated to the school level in our dataset.[2] Between 2006 and 2008, schools made statistically significant progress in both grades and in both math and English. The results are largest in 8th grade math and are 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.

Thus, the overall story on the state tests is one of relative improvement from 2006-2008. Though these gains are statistically significant, they cannot be traced directly back to the progress report program, nor is it the place of this paper to discuss whether such improvements are substantial enough to warrant praise. However, this general improvement in performance indicates that any difference in the performance of schools earning certain grades found in the below analysis should be kept in the context of an environment of overall average improvement on the state exams.

5) Data and Method

[2] We are restricted to only grades 4 and 8 because the school system only began testing other grades in 2005. We are restricted to begin with the 2006 school year because scale scores prior to 2006 are not comparable to those in latter years.

We utilize a student-level dataset provided by the New York City Department of Education. The dataset 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 3 through 8 from the 2005-06 through the 2007-08 school years. We are also able to link students to schools, 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 dataset overall and broken out by the letter grade earned by the school at the end of 2006-07. These descriptive statistics are not identical to, but closely match those reported by Rockoff and Turner (2008), who used data aggregated by the Department of Education.

[TABLE 3 ABOUT HERE]

One difficulty with the dataset is that the state does not claim that the result of its math and reading exams are "vertically aligned" across grades. That is, the same score in fifth grade may represent different information about a student's proficiency than it would in the fourth or sixth grades. This means that the relationship between a student's previous year's score and current score could also vary by grade level. Specification checks (not reported here) suggested that there are slight but significant differences in the relationship between prior and current student proficiency across grades. To account for this, along with estimating models that include all grade levels we also 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 better fit the particular design of the New York program.

We estimate an education production function taking the form:

$$(1)\quad \begin{aligned} Y_{ist} &= \beta_0 + \beta_1 f(Y_{ist-1}) + \beta_2 X_{ist} + \beta_3 School_{st} + \beta_4 Type_s + \beta_5 g(Points_{st-1}) \\ &+ \beta_6 (Type_s * g(Points_{st-1})) + \beta_7 Grade_{st-1} + \varepsilon_{ist} \end{aligned}$$

Where Y_{ist} is the test score of student i in school s at the end of year t; $f(Y_{ist-1})$ is a cubic function of the student's prior test score; X is a series of observed characteristics about the student; School is a series of aggregate demographics for the school; Type is an indicator for the school type (elementary, K-8, middle) as reported on the school's progress report; $g(Points_{st-1})$ is a cubic function for the final number of points earned by the student's school on each of the components of the overall points system at the end of the 2006-07 school year; Grade is the letter grade earned by the school at the end of the 2006-07 school year; ε is a stochastic term clustered by school; and the $\beta_0 - \beta_7$ are parameters to be estimated. The model is estimated using OLS.

The identifying assumption of (1) is that there is no difference in school performance that is not conveyed in (a cubic function of) the number of points that a public school earned under the formula. If this assumption holds, then we can interpret the estimated effect of earning a particular grade (β_7) as the causal influence of a student's school receiving that grade (A, B, C, D, or F), on his academic proficiency.

The basic idea behind this technique is to take advantage of the known cutoffs across 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 effectiveness that is not already represented in the point total. However, though schools with similar point totals are likely also similar in their effectiveness, where their score falls according to the grading policy matters greatly to their incentives under the policy. For schools with point totals near the cutoffs, which side of the cutoff they end up on (and thus, which grade they receive) is largely determined by randomness.

We make a few sampling restrictions that are worth mention. First, when estimating (1) we exclude students who were observed to be 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 systematically differ from students in other grades. We also exclude students whose school is listed as a high school on its progress report. The dataset contains some observations of students taking the seventh and eighth grade exams listed as attending a high school, though the progress report definition suggests that such identified schools would teach grades nine through twelve. From looking at the data it appears that these are specialty schools (for the arts, etc.), and so we choose to eliminate them from the dataset. However, our results are robust when these sampling restrictions are relaxed.

One difficulty with measuring (1) is that use of a lagged dependent variable as an independent regressor introduces a potential autocorrelation problem when calculating standard errors. Many papers avoid this issue by moving the prior test score measure to the left-hand-side of the equation and using the difference between achievement in year t and t-1 as the dependent variable (that is, such papers estimate $\Delta Y_{it} = Y_{it} - Y_{it-1}$). However, this procedure restricts the parameter of the student's prior proficiency to equal

1, which is not likely to be accurate because it does not allow for student knowledge to decay from one year to the next. This procedure would be even more problematic in our present case because the test scores are not vertically aligned.

Equation (1) follows a less restrictive approach. In fact, controlling for a cubic function of the student's prior proficiency even relaxes the assumption made by some other papers that the relationship between prior and current year proficiency is linear. We account for the potential for autocorrelation bias by calculating standard errors using the bootstrap method with 300 iterations. Prior research outside of education has used this method to account for autocorrelation in standard error estimates (see for example Cameron and Trivedi 2006; MacKinnon 2002). Following recent papers, we further account for school heterogeneity by allowing standard errors to be clustered at the school level.

Use of an interaction term in (1) accounts for the fact that the cutoffs from the point system vary somewhat by school type, as shown in Table 1 above. Though perhaps not entirely clear from the equation, our procedure is to create one variable interacting school type with the points earned on the progress report in one of the input factors (environment, etc.), another variable interacting school type with the square of points earned on that factor, and a third interaction with the number of points in that factor cubed. We generate such variables for each input factor used to produce school points.

One potential issue with estimating (1) worth mention here is that focusing on treatments that disproportionally affect students in low-performing schools, as does this study, may be impacted by regression-to-the-mean. Schools and students at the bottom of the achievement distribution may have such low scores in part because of random error.

If a negative impact of random error was more present in these schools then improvements made on those tests in later years could be an inflated measure of the child's academic progress. In their similar study in Florida, Rouse and Others (2007) present a series of tests indicating that regression-to-the-mean is not the driving force of their results. Unfortunately, the timing of when wide scale testing began in New York does not allow us to adopt similar tests here. Thus, it remains possible that our results are impacted by regression to the mean.

However, while we cannot completely eliminate the possibility that regression-to-the-mean exists here, use of a regression discontinuity framework makes it quite unlikely that this phenomenon is driving any findings. Regression-to-the-mean occurs by schools at the bottom of proficiency tending to make greater gains, and schools at the top of proficiency tending to make declines due to randomness. However, in essence, this model compares the performance of students in schools near the grade level cutoffs that fell on either side, largely due to randomness. It is implausible that the grade level cutoffs were set at points that are particularly important for regression to the mean. That is, the schools we are comparing to one another are actually very similar in their prior productivity, and there is no reason to believe that a school that fell just below the cutoff for earning a particular grade would be more or less impacted by regression-to-the-mean than a school whose points fell just above the cutoff.

5) Results

We first aggregate our dataset in order to replicate the recent results reported by Rockoff and Turner (2008). The result of this test are reported in Table 4. Though not identical, the coefficient and standard error estimates in Table 3 closely mirror those reported in that prior paper.

[TABLE 4 ABOUT HERE]

This replication suggests that the prior paper's finding that F and D graded schools made improvements relative to higher graded schools continue to hold. It also lends some confidence that the data utilized to estimate our models of primary interest using student level data is accurate. This confirmation is particularly important given that Rockoff and Turner (2008) rely on data that was reported at the school aggregated level while we utilize data aggregated from our individual level dataset.

Table 5 reports the results of estimation of various forms of (1) in math overall and for each particular grade level. As in the aggregate data, the overall model including all grade levels continues to find a statistically significant and substantial positive effect from a school earning an F or 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, this result translates to a positive one year impact from attending an F school of about 0.18 standard deviations in math proficiency relative to students in schools that earned a C grade.

[TABLE 5 ABOUT HERE]

The additional columns of Table 4 report results in math in regressions restricted only to the particular grade level. We do find that the result varies across grade levels. In particular, we find a strong positive impact from attending an F or D school for students 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 unlikely having a large impact on its estimation.

Table 6 reports our results in reading following a similar format. We find no significant difference in reading performance for students in F or D graded schools relative to students in other schools. This insignificant result is found both in the overall regression and in each of the grade level regressions.

[TABLE 6 ABOUT HERE]

Though we are unable to test such a hypothesis here, one plausible reason for the null effect in English while we find a positive effect in math is that the schools had less time to react to its grade before students were administered the English exam. As mentioned above, there was only about a three month period between when schools first received their grade and when students were administered the English exam. There was an even shorter period between when the grades were made public and the English exam, which is potentially important if we believe that part of the treatment of the F sanction comes from parent and community pressure due to the grade.

6) Conclusion

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 taking advantage of the city's continuous point system for assigning school grades utilized here 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 earning an F or 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 driven by fifth grade students. We find no evidence that earning any particular school grade has a significant impact on student reading proficiency.

References

Cameron, A. C. and P.K. Trivedi. (2005) Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Carnoy, M. (2001). Do school vouchers improve student performance? American Prospect. 12(1): 42-45.

Chakrabarti, R. (2007). Vouchers, public school response, and the role of incentives: evidence from Florida. Federal Reserve Bank of New York, Staff Report no. 37.

Chakrabarti, R. (2008). Impact of voucher design on public school performance: evidence from Florida and Milwaukee voucher programs. Federal Reserve Bank of New York, Staff Report no. 315.

Figlio, D. N. and H. F. Ladd (2008). School accountability and student achievement. In: Helen F. Ladd and Edward B. Fiske eds. Handbook of Research in Education Finance and Policy. Routlage Press.

Figlio, D. N. and C. Rouse (2005). Do accountability and voucher threats improve low-performing schools? Journal of Public Economics. 90: 239-255.

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, (3): 66-71.

Harris, D. (2001). What caused the effects of the Florida A+ Program: Ratings or vouchers? In: Margin Carnoy ed. School Vouchers: Examining the Evidence. Economic Policy Institute.

MacKinnon, J. G. (2002). Bootstrap inference in econometrics. The Canadian Journal of Economics. 35(4): 615-645.

New York City Department of Education (2007). Educator Guide: The New York City Progress Report.

Ladd, H. F. (2001). School-based educational accountability systems: The promise and the pitfalls. National Tax Journal. 54(2): 385-400.

Rockoff, J.E and L.J. Turner (2008)."Short run impacts of accountability on school quality. 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, (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.

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

	Elementary		K-8		Middle	
Grade	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.

Table 2
Comparison of Mean Scores in New York City Over Time

	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-tail test indicates that the 2007 score is statistically lower than the 2006 score, $p = 0.0284$.

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%	44.51%	36.3%	32.2%	26.4%
Percent Hispanic	40.1%	39.2%	39.98%	38.2%	39.7%	42.9%
Percent white	14.0%	10.3%	10.30%	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 dataset.

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

VARIABLES	Math	English
School Grade 06-07		
A	-1.555	-1.383
	[1.401]	[1.067]
B	0.0409	-0.422
	[0.779]	[0.593]
D	2.322**	0.291
	[0.930]	[0.708]
F	4.927***	2.250*
	[1.706]	[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 level and these functions.

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***	-19.05***	-17.09***	-20.42***	-10.75***	-23.44***
	[0.302]	[1.048]	[0.685]	[0.767]	[0.712]	[0.707]
Prior Math Score Squared	0.029***	0.031***	0.027***	0.033***	0.017***	0.037***
	[0.001]	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]
Prior Math Score Cubed	-1.4e-05***	-1.5e-05***	-1.3e-05***	-1.6e-05***	-8.5e-06***	-1.8e-05***
	[2.24e-07]	[7.65e-07]	[4.85e-07]	[5.75e-07]	[5.70e-07]	[5.39e-07]
School Grade 06-07						
A	-1.288	1.299	-3.441**	0.844	0.524	-4.209
	[1.110]	[1.721]	[1.727]	[2.333]	[2.235]	[2.689]
B	-0.372	0.551	-1.919**	1.207	-0.0103	-1.431
	[0.615]	[0.915]	[0.926]	[1.249]	[1.218]	[1.454]
D	1.653**	0.798	3.288***	2.748	-1.21	2.847
	[0.677]	[1.050]	[1.089]	[1.820]	[1.677]	[1.829]
F	3.537**	-2.176	8.179***	5.712	2.989	5.778
	[1.587]	[2.494]	[2.353]	[4.196]	[3.402]	[3.621]
Observations	317,531	65,492	65,434	60,988	62,721	62,895
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 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***	-15.71***	-7.463***	-13.73***	-16.42***	-11.96***
	[0.268]	[0.514]	[0.484]	[0.482]	[0.663]	[0.612]
Prior English Score Squared	0.021***	0.026***	0.012***	0.022***	0.027***	0.020***
	[0.000421]	[0.000787]	[0.000777]	[0.000739]	[0.00102]	[0.000981]
Prior English Score Cubed	-1.0e-05***	-1.3e-05***	-6.3e-06***	-1.1e-05***	-1.3e-05***	-1.0e-05***
	[2.19e-07]	[4.01e-07]	[4.15e-07]	[3.77e-07]	[5.18e-07]	[5.22e-07]
School Grade 06-07						
A	-0.104	0.204	0.059	0.265	-0.075	1.246
	[0.786]	[1.591]	[1.329]	[1.269]	[1.334]	[1.902]
B	-0.149	-0.701	-0.459	0.845	-0.086	0.396
	[0.397]	[0.887]	[0.698]	[0.717]	[0.695]	[1.054]
D	0.0777	0.314	1.103	0.345	-0.462	-1.268
	[0.479]	[0.989]	[0.756]	[0.880]	[1.041]	[1.298]
F	1.096	0.7	1.517	0.242	-0.77	-0.14
	[1.086]	[2.080]	[1.664]	[2.009]	[2.078]	[2.309]
Observations	312,349	64,398	64,308	60,223	61,622	61,797
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 grade level student.