5-2013

Essays on Academic Achievement and Student Behavior in Public Schools

Wael Soheil Moussa
Syracuse University

Abstract

This dissertation examines the student academic achievement through various mechanisms, put in place by the public school district, classroom student behavior, and negative external shocks to the students' living environment. I examine the impacts of various treatments on student short and long run academic outcomes such as math and English test scores, grade retention, special education diagnoses, as well as high school graduation. Each essay will be presented as a chapter of the dissertation.

The first essay uses student-level administrative data from New York City Public Schools to examine the impacts of school entry age on student academic outcomes (including test scores and high school graduation) and behavioral outcomes (such as suspensions and chronic absenteeism) for students in grades three through ten. My research design uses exogenous variation in students' month of birth, comparing the outcomes of students born just before to students born just after the school starting age cutoff date. I present evidence that entering school early increases the probability of high school graduation, among white, black, and Hispanic students. Starting school early has no effect on suspensions and chronic absenteeism. I find that starting school early has a negative effect on grade specific measures such as, test scores, GPA, retention, and special education.

The second essay uses exogenous variation on course scheduling in Chicago Public Schools to examine empirical implications of Lazear's (2001) educational production model. This essay investigates an underlying mechanism by which class size affects student performance, the behavioral composition of a classroom. Behavioral composition is defined as the number of non-disruptive students who are in attendance on a given school day, where attendance is not randomly assigned. To properly identify the effect of classroom behavioral

composition, we use random course scheduling to instrument for non-random attendance throughout the school day. Consistent with the Lazear framework, we find that an additional non-disruptive student in attendance increases the probability of passing English I and Algebra I, with larger effects for students in remedial versus regular classes. For regular English I students, we estimate a positive relationship between the number of non-disruptive students in attendance and own reading test score.

The final essay examines the impacts of the events following the terrorist attacks of September 11, 2001 where animosity geared towards the Arab community increased significantly. Specifically, we analyze the performance of Arab students in New York City public schools pre and post the terrorist attacks and compare the effects to non-Arab Muslim as well as Indian students in the same schools. We analyze the impacts of September 11 on the student achievement of Arab students enrolled in grades 3 through 8. We use a difference-in-differences approach using non-Arab Muslim students as a control group. We find that Arab students experience a decline in their scholastic performance post-9/11 by as much as 0.077 and 0.101 student level standard deviations in ELA and math test scores, respectively. We also find that retention and special education rates among Arab students post-9/11, increase by 62.8 and 16.8 percent, respectively.

ESSAYS ON ACADEMIC ACHIEVEMENT AND STUDENT BEHAVIOR IN PUBLIC SCHOOLS

by

Wael Soheil Moussa
B.A., American University of Beirut, 2006
M.A., American University of Beirut, 2008

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Graduate School of Syracuse University

May 2013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. The Impacts of School Entry Age on Student Achievement: Evidence from New York City Public Schools

Introduction

The age at which a child begins his or her schooling has attracted a lot of attention from policy makers and parents alike. Age at school entry is important especially in the United States, where almost all states have imposed an entry age requirement for kindergarten enrollment eligibility, which in turn affects a student's age at school start. A large body of literature studies the effect of a child's age at school entry on academic achievement during various stages of schooling, especially during primary school and middle school. Datar (2006) and Elder and Lubotsky (2009) find that starting school a year early leads to lower test scores in kindergarten and first grade. Black, Devereux, and Salvanes (2011) study the effect of early school entry on long-run outcomes, such as educational attainment and earnings, finding that starting school a year early does not affect educational attainment in Norway. This paper contributes to the current literature by estimating the impacts of starting school early on test scores, high school graduation, and other academic outcomes for elementary, middle, and high school students.

Many recent studies have found that starting school a year late can improve a student's test score in elementary school (Elder and Lubotsky, 2009; Robertson, 2010). Almost all of the school starting age studies account for the endogeneity of school starting age by using exogenous variation in children's date of birth, coupled with entry age requirements, as instruments to identify the causal effect on academic achievement. However, the analysis has been limited to a grade-by-grade basis, mostly due to data limitations. Consequently, the interpretation of the effects can be confounded, as mentioned in Black, Devereux, and Salvanes (2011). It is difficult to empirically distinguish between the effect of being among the youngest student in class and the effect of simply being young, which I will be referring to as the age at school entry effect and the age at test effect, respectively.

I use student-level administrative data from New York City Public Schools (NYCPS) to study the effect of school entry age on students' long run academic performance and behavioral outcomes spanning the school years 1999-2000 to 2006-2007. Specifically, I investigate the impact of school starting age on overall high school graduation rates as well as 4 and 5year graduation rates. This paper also estimates the impact of school starting age on behavioral outcomes such as chronic absenteeism and suspensions. Given that the data includes students from grades three to ten, I am able to control for student cohorts and net out the age at test effect, effectively comparing students who are of the same age but are born on different sides of the NYCPS cutoff. Although this solves the issue of getting rid of the age at test effect, another issue arises as a result. The issue that arises in this case would be that students in the treatment group and the control group are no longer in the same grade, making their performance on standardized test scores incomparable. To successfully isolate the effect of school starting age, the outcome variables would have to be unrelated or determined by a specific grade. For this reason I use high school graduation and student behavior as outcomes to identify the effect of school starting age.

The data include student transcript and demographic information, in addition to providing year and month of birth. The school age cutoff policy in NYCPS states that a child must turn five years old by December 31 to be eligible for kindergarten enrollment in the fall of that year. This provides a discontinuity regarding students' age at school entry. Due to a lack of availability of data on actual school starting age, I estimate the reduced form relationship between being born prior to the school age cutoff and long-run academic performance.

I estimate the impacts of an early school start on 4-year, 5-year, and overall high school graduation. I find that early school entry increases a student's probability of graduating from high school by 1.7 percentage points, and 4.1 percentage points among white students, using the ninth and tenth grade sample. In addition, I estimate the impact of school starting age on suspensions in school and chronic absenteeism, defined as being absent for at least ten percent of total school days, according to the NYCPS Department of Education[1] My estimates of the effect of an early school start age on these specific outcomes are the only such estimates in the literature. A few other studies analyze the impact of school starting age on non-academic outcomes, such as teenage fertility and child obesity, including Black, Devereux, and Salvanes (2008) and Anderson et al. (2011). I find that there are no significant effects of being born prior to the school age cutoff on student behavior.

Lastly, I estimate the reduced form regression of being born prior to the age cutoff on the traditional student performance measures: standardized test scores (Math, ELA), GPA, grade retention, and special education designations. This last exercise shows that when comparing the scholastic performance of students in different grades, I find results that are similar to those found in the literature when doing the analysis by grade. These results as with the previous literature are still not able to isolate the effects of school starting age on achievement.

Relevant Literature

The identification of the causal effect of school starting age on academic achievement is difficult in observational studies due to parents' abilities to manipulate school starting ages for their children. The school- or state-mandated age cutoff is enforceable for children born after the cutoff date, but school enrollment is not mandatory in the first year of eligibility. Parents of children with

[1] For more information visit http://www.nyc.gov/html/truancy/html/about/about.shtml

birthdays one to two months prior to the school mandated cutoff have the option to withhold their children from entering school as soon as they are eligible. “Redshirting” is a practice that has become increasingly popular among parents in the United States over the past 40 years. Deming and Dynarski (2008) show that the percentage of six year olds enrolled in first grade or above has dropped from 96 percent in 1968 to 84 percent in 2005.

Datar (2006), Elder and Lubotsky (2009), McEwan and Shapiro (2008), Cascio and Schanzenbach (2007), and Robertson (2011) account for the endogeneity of school starting age using exogenous variation in a student’s date of birth, as well as variation in state age cutoffs. These studies estimate the effect of starting school early on academic outcomes for students enrolled in the same grade, leading to another issue regarding the interpretation of their results. The authors address the endogeneity of school starting age by comparing outcomes of students born on different sides of the state policy cutoff. The students in the comparison groups, however, are not one to two months apart in age but are instead ten to eleven months apart in age, raising questions about the suitability of the control group. Consequently, the estimated effect of starting school early is a combination of the absolute age and age at school entry effects. The finding that students born after the cutoff have higher test performance can be attributed to these students are simpler older when taking the test, or older relative to their classmates. Additionally, the effect of school starting age is separately estimated for children in kindergarten, and grades three, five, and eight. Estimating separate effects for each grade imposes that students in the control group are no longer in the estimating sample. Individuals in the control group drop out of the analysis due to the fact that students born after the school cutoff started their schooling one year later. Therefore, students in the control group are no longer in the same grade as children born just before the cutoff date.

Cascio and Lewis (2006) address the issue of conditioning on a student's grade to identify the effect of starting school early/late by looking at outcomes of teenagers who took the Armed Forces Qualifying Test (AFQT). The study finds that minorities are most affected in terms of performance on the AFQT, finding that the effect of starting school a year late on blacks and Hispanics is positive and marginally significant in some specifications, while the effect on white individuals is not. Black, Devereux and Salvanes (2008) use a similar strategy to identify the effect of starting school early/late by estimating the school starting age on an IQ test administered in Norway at the time of military service enrollment. Using an examination that is outside the capacity of a school and adding in birth cohort fixed effects, the authors distinguish between the absolute age effect and school starting age. Black et al. (2008) find that the school starting age effects are modest when studying an out-of-school test compared to grade specific tests as those found in Elder and Lubotsky (2008), Datar (2006), Bedard and Dhuey (2006) and Robertson (2011). Black et al. (2008) find no effect of starting school early on other long-run outcomes such as teenage pregnancy for women, educational attainment and earnings.

The majority of studies have examined the effects of starting school late on academic performance measures such as test scores and grade progression. Even special education diagnoses can be attributed to students underperforming in their given grade, leading to a higher probability of being diagnosed with ADD or ADHD (Dhuey and Lipscomb, 2010; Elder and Lubotsky, 2009; Elder, 2010). Further, there is little evidence of the effect of starting school early on in-school behavioral outcomes or on the performance of students in high school.

Identification Strategy

I model the effects of the kindergarten entry age policy on test scores, grade retention, special education diagnoses, suspension and chronic absenteeism for students in grades 3-8. For students in the high-school sample, the same set of outcome variables will be used as well as yearly GPA and 4 year graduation. To estimate the causal effect of early entry to school on academic performance, I use plausibly exogenous variation in a student's date of birth to test the effect of age at school entry on academic outcomes. The ideal equation to be estimated would be the following:

$$(1) \qquad Y_{ist} = \alpha + \delta \mathrm{EntryAge}_i + \beta X_{ist} + \mu_s + \lambda_t + \epsilon_{ist}$$

Where Y is the outcome variable for student i in school s and year of birth t, EntryAge is the student's age at first entry into kindergarten, X is a vector of student specific control variables, $\mu_s$and λ_t denote school and year of birth fixed effects. The year of birth fixed effects are redefined as spanning from July of year t to June of year t+1. Consequently, every birth cohort is now centered on the New York City policy cutoff of December 31st. The effect of entry age in this case is however biased by collinearities with students' actual age, where students who are older tend to have acquired more human capital/skills at the time of the test. Hence, the inclusion of the cohort fixed effects ensures that the comparison of outcomes takes place between students who are at most two months apart in age at the time of test in a given NYC public school. This allows me to distinguish between the pure age effect and the age at school entry effect on student outcomes. The school entry age effect is the effect of starting school early and being among the youngest in class, while the absolute age effect is the effect of being a certain age. This is due to the mandated school age cutoff policy, where students who are born just before the cutoff are likely among the youngest in class, while those born just after are going to be among the oldest. Inclusion of the cohort fixed effects diminishes the likelihood that the treatment and control groups are systematically different in their unobservable characteristics, which might influence their academic achievement.

Variation in age at entry can however be endogenously determined by parents' decision to purposefully hold their child back and hence increase their child's age at entry by a year. Consequently, an instrument would be required to recover the causal effect of age at first entry. I use students' month of birth as a means to instrument for the endogenous variable, by assigning treatment to all students who were born prior to the policy cutoff and define students born after as the control group. I use a fuzzy regression discontinuity design, by restricting the sample to only those students born in December and January. Other specifications would include increasing the bandwidth around the cutoff to include November and February births respectively. The estimation is now a just identified two-stage least squares regression as follows:

(2) $$\mathrm{Y_{ist}} = \beta_1 + \beta_2 \mathrm{EntryAge_i} + \beta_3 \mathrm{X_{ist}} + \mu_s + \lambda_t + \epsilon_{ist}$$

(3) $$\mathrm{EntryAge_{ist}} = \gamma_1 + \gamma_2 \mathrm{BeforeCutoff_i} + \gamma_3 \mathrm{X_{ist}} + \mu_s + \lambda_t + \upsilon_{ist}$$

BeforeCutoff is an indicator variable that takes on a value of 1 if student s were born prior to the cutoff and 0 otherwise. The instrumental variable BeforeCutoff requires two conditions for the estimation strategy to yield causal effects. First, students' month of birth should be uncorrelated to the outcome variable Y except through the endogenous variable EntryAge. Second, the control variables in X have to be continuous at the point of discontinuity in the treatment.

Due to the lack of students' age at first entry into kindergarten in the data, I can only estimate the reduced form of the entry age effect. This should not affect the analysis in a radical manner, other than changing the way the coefficient of interest is interpreted. The estimating equation is now as follows:

(4) $$\mathrm{Y_{ist}} = \delta_1 + \delta_2 \mathrm{BeforeCutoff_i} + \delta_3 \mathrm{X_{ist}} + \mu_s + \lambda_t + \nu_{ist}$$

Although this specification is able to net out the age effect, the reduced form parameter of δ_2 is interpreted as a combination of the school entry age effect as well as a grade effect. In other words, the effect of being among the youngest in class in addition to being in a higher grade as a result of the policy cutoff is being identified, especially when estimating the effect on grades and test scores. The age effect is netted out when using this specification because it effectively compares the performance of students who are similar in their observable characteristics as well as in their unobservable ability. If we consider that children who are born one month apart are not inherently more or less skilled, then the absolute age effect is minimized. Therefore, the students on either side of the cutoff differ only in the assignment of the treatment. Note that, when estimating the effect of BeforeCutoff on outcomes that are not directly affected by the grade level such as chronic absenteeism, suspension and 4 year graduation, the effect retrieved from equation (4) is the school entry age effect since the grade effect would be zero in this case.

There is concern regarding the consistency of the estimates from equation (4) that would arise from noncompliance to the policy. In the case of the NYC public schools age policy, noncompliance can only take place on one side of the cutoff. In other words, only those children born in December or even November could potentially choose not to enroll in kindergarten when they become eligible and therefore delay entry to school. Buckles and Hungerman (2008) indicate that redshirting is most common among white parents of high educational attainment. The majority of students in NYCPS is of black and Hispanic origins and come from lower socio-economic backgrounds, which dampen the proportion of non-compliance before the cutoff. Consequently, estimation bias due to non-compliance in this case if any, would lead to an understatement of the effect of birth prior to the age cutoff. The reason is that one would expect students who enter school at an older age to have an advantage when taking exams as opposed to those who enrolled as soon as they became eligible.

Data and Sample Characteristics

Data

For this study I draw data from New York City Public Schools administrative student records. The data cover the universe of students who have ever enrolled in a NYCPS from 3rd through 10th grade, over the academic years 1999-2000 through 2006-2007. For students in grades 3 through 8, we observe their English Language Arts and Math test records. To compare test scores across years and grades, the test scores are transformed into z-scores, by subtracting the mean score of all students in the relevant school year and grade from each student's raw score and dividing by the standard deviation. The data allows me to identify students who have been retained in a given school year as well as observe their special education designation that school year. An interesting aspect of the data is the availability of student attendance and suspension records. This allows me to create indicators for whether a student was suspended from school for at least a day, as well as indentify students who display signs of chronic absenteeism. Additionally, the school attended is uniquely identified.

Data for students in high-school (grades 9-10) are drawn from the student transcript files that include, course titles/numbers[2] and standardized test scores in Math and Science. As in the elementary and middle school sample, I am able to identify those students who are diagnosed as requiring special education, whether a student has been retained in his/her grade, identify students who are chronically absent as well observing whether or not a student was suspended

[2] The data on courses includes the number of credits each course is worth as well as the final grade (numerical) received. I use this information to compute each student's overall GPA for each academic year the student appears in the data.

from school. Finally, I observe a common set of characteristics for all students in the sample that includes gender, ethnicity/race, Month and Year of Birth, English Language Learner (ELL) status, Special Education status, free meal status and the number of days absent, present and suspended. The standardized Math and Science test scores for students in high school are taken from the New York State Regents Examinations. Only the Math and Science exams are included in this study, due to the grade requirements for taking the other exams namely English, History and Geography. Lastly, I am able to identify students who eventually graduate from a NYCPS High School as well as the number of years it takes them to do so.

Descriptive Statistics

Summary statistics for the sample are presented in Table 1 and Table 2, for students in Grades 3 to 8 and students in grades 9 and 10 by month of birth, respectively. The sample consists of students who were enrolled and labeled active by the school. The sample is restricted to all students' first instance of each grade in the data. In other words, students who are repeating a grade are dropped from the analysis, to avoid issues of sample selection, where these students are arguably of a lower innate ability coupled with the fact that they are learning everything for the second time. Furthermore, I restrict the analytic sample to students who were born in the months of November, December, January and February to better exploit the discontinuity in the New York City kindergarten age requirement.

Panel A of Table 1 presents the outcome variables of interest for the sample of students enrolled in Elementary and Middle School are the scores on the standardized Math and ELA examinations, an indicator for being retained that given year and an indicator for having a Special Education designation. Further, I include an indicator for whether a student has been suspended that year and whether a student is considered “chronically” absent[3]. Table 1 shows that students who were born before the policy cutoff have lower test scores than those born after. The mean math z-score for students born in November and December are 0.076 and 0.057 student-level standard deviations respectively compared to 0.211 and 0.206 for students born in January and February respectively. This pattern is also observed in the ELA scores of November-December birth months who receive between 0.068 and 0.049 student-level standard deviations as opposed to 0.228-0.218 for January-February birth months. Retention rates for students born prior to the cutoff are 2.6 and 2.7 percent for students born in November and December, respectively. In contrast, students born in January and February have retention rates of 2.1-2.2 percent. 7.6-7.1 percent of the students born prior to the cutoff are given a special education designation, whereas 5.7-5.6 percent of those born after the cutoff are diagnosed as requiring special education services.

The discrepancy in elementary and middle school academic performance between those born before and those born after the cutoff can be seen in figure 1 as well. The figure shows that January-born students have the highest test scores and lowest retention rates even among students of all birth months. On the other hand students born in December have the lowest test scores and the highest retention rates among all students. This pattern is also visible when observing the rates of special education diagnoses.

When looking at outcome variables that are more behavioral in nature, the data shows that the students born on either side of the cutoff are not very different. Suspension rates for students born before and after the cutoff are on average between 5.4 percent and 5.6 percent.

[3] The New York City Department of Education defines students as chronically absent if they have missed more than 10 percent of the total number of school days in an academic year.

Likewise, the proportion of students who are absent for more than 10% of total school days is between 14.5 and 14.8 percent for students born before the cutoff, while that of students born after the cutoff is between 15.2 and 15.6 percent.

Panel A of Table 2 presents the outcome variables of interest for students enrolled in high-school. Note that the high-school sample undergoes the same restrictions as the elementary and middle school sample. In other words, only students enrolled in the 9^{th} and 10^{th} grade for the first time are kept in the analysis, in addition to restricting the sample to students born between November and February. I do not include students from grade 11 and 12 in the analysis primarily due to the high level of non-random attrition in the sample, seeing as students who progress to the 11^{th} and 12^{th} grades are likely to have a higher level of innate scholastic ability. Akin to the patterns observed in the younger sample, the high-school achievement of pre-cutoff students is much lower than that of post-cutoff students. The mean math score of the pre-cutoff sample is between 0.070 and 0.063 as opposed to 0.169 and 0.162 for the post-cutoff sample. Science scores are also much lower for the pre-cutoff sample, ranging between -0.025 and -0.0205 for students with a November and December birth month, respectively. In comparison, the mean science z-score for the post-cutoff sample is 0.106 and 0.109 for January and February born students, respectively. The average yearly GPA for students in the post-cutoff sample is between 2.00 and 2.01 whereas the mean GPA for pre-cutoff students is 1.867. Retention rates among the pre-cutoff group are between 28.1 and 28.3 percent, while post-cutoff retention rates are between 23.2 and 23.3 percent. Note that despite the differences in test score and GPA performance, the data shows that both sets of students graduate in 4 years at the same rate of around 21 percent. To clarify this number signifies that 21 percent of the students who are currently enrolled in 9^{th} and 10^{th} grade eventually graduate, and will have done so in 4 years. This observation lends support to the notion that the cutoff policy may not affect students' long run outcomes, especially when looking at a "finish-line" outcome such as the probability of graduating from high-school in 4 years.

Figure 2 displays the academic performance discontinuity, and a discrete jump can be observed in all of the student academic outcomes. Mathematics and Science Regents test scores are consistently lowest among December births, and highest among January births. Students born in December have the highest retention rates among all students in the high school sample, and January births have the lowest retention rates, as in the younger sample of students. The same goes for special education diagnoses among the students in the sample.

On the other hand, the behavioral outcomes of students in high-school who are born before and after the cutoff are similar. Between 24.1 and 22.1 percent of all students in the high-school sample show signs of chronic absenteeism and between 28.2 and 27.1 percent of all students in the sample have ever been suspended from school for at least a day in a given school year. While academic performance of students in the pre-cutoff sample is consistently lower than that of students in the post-cutoff sample, the student specific characteristics do not differ or vary in a systematic manner across observables.

To measure the effect of the policy as a result of only being born before or after the cutoff, it is important to observe that there is no selection into being born on either side. This is a necessary condition for the fuzzy RD design to yield non-trivial or spurious results (Imbens and Lemieux, 2008). Consequently, the vector of control variables has to satisfy the condition of continuity along the policy cutoff. Panel B of Table 1 and Table 2 show that around half the sample are male and is fairly stable across all months of births as well as different grades. The racial composition of the students is also stable across the different subsamples. Around 33 percent of the students in the sample are Black; Hispanic students account for 35-37 percent of the sample; Asian students account for around 13-15 percent of the sample; and around 0.3 percent are of American Indian decent and white students account for the remainder 17-18 percent of the sample. Additionally, the data shows that 55 percent of students in grades 3-8 receive free or reduced price meals and is consistent across the months of births that fall right before and after the policy cutoff. 58.6-58.9 percent of the high school sample receives free/reduced price meals. The figures in Table 1 and Table 2 show that there are no real discontinuities in the student characteristic profiles that arise from being born in December or January. This lends credibility to the exogeneity of the month of birth as a means to identify the effect of the policy on academic as well as non-academic outcomes.

Estimation Results and Discussion

In this section, I present the reduced form estimates of equation (4) on various academic and non-academic outcomes of students in grades 3-8 and grades 9-10 respectively. The specifications used in tables 3-5, include school fixed effects and year of birth (cohort) fixed effects. In addition, I run the same specification for each of the outcome variables using two different bandwidths around the age cutoff point. The first bandwidth includes only those students who were born in the months of December and January and is represented in the odd numbered columns of tables 3-6. The second is expanded to also include students born in November and February and is represented in the columns the same tables.

High School Graduation

I estimate the effect of being born prior to the school cutoff on the probability that students in 9th or 10th grade will have graduated in 4 years, 5years, or ever. For this scenario, the effect identified in this regression is only that of age at school entry. The reasoning behind this is that graduation is a finish-line outcome that is not determined by the current grade a student is in. Additionally, students in both the treatment and control group are of the same age. Therefore, the absolute age and grade effects are netted out of the equation, allowing for the isolation of the age at school entry effect on graduating in 4 years, 5years, or ever.

Using only the high school sample for this regression, column 1 of Table 3 shows that students born in December have on average a higher 4 year graduation rate than those born in January by 1.7 percentage points. Relative to a mean 4 year graduation rate of 21.4 percent, students born prior to the school cutoff are 7.9 percent more likely to graduate from high school in 4 years. Columns 3 and 4 of Table 3 report the coefficient estimates of the effect of being born prior to the cutoff on 5 year graduation. The estimates indicate that students who started school early are 2.6 percentage points more likely to graduate from high school within 5 years. This translates to an 8.5 percent higher probability of graduation for students born before the cutoff. The estimates are significant at the 1 percent level. Columns (5) and (6) of Table 3 report the coefficient estimates of the effect of being born prior to the cutoff on whether a student ever graduates. I find that students born before the cutoff are 3.7 percentage points more likely to ever graduate than students born after; the estimate is significant at the 1 percent level. This is equivalent to a 9.1 percent higher probability of high school graduation among students who started their schooling early.

Panels B, C, and D of Table 3 report the coefficient estimates of starting school early on various graduation measures, disaggregated by race. In columns (1) and (2), I find that the effect of starting school early on 4 year graduation is largest among white students. I estimate a 4.1 percentage point increase in graduation as a result of being born before the cutoff. Relative to a mean 4 year graduation rate of 30.2 percent for white students, this translates to a 13.5 percent higher probability of graduation within 4 years. The effect on black students is much smaller when considering 4 year graduation at 0.4 percentage points, and is statistically insignificant. Starting school early increases 4 year graduation rates among Hispanic students by 1.1 percentage points, and is statistically significant at the 5 percent level. This is equivalent to a 6.1 percent increase in the probability of graduating from high school in 4 years.

Columns (3) and (4) report the estimation results of the effect of being born before the school cutoff on 5 year graduation. The coefficient estimates suggest that starting school at an earlier age leads to a higher probability of graduating from high school within 5 years. I find that among white students, being born before the school cutoff leads to a 4.3 percentage point increase in the probability of graduating within 5years. This estimate is equivalent to a 10.8 percent increase in the probability of graduating within 5 years. The coefficient estimate on the black sample shows a 1.3 percentage point increase in 5 year graduation rates. Relative to a mean 5 year graduation rate of 27.3 percent, this translates to 4.8 percent increase in the probability of graduating within 5 years of starting high school. Similarly for the Hispanic sample, I find that being born before the cutoff leads to a 2.4 percentage point increase in 5 year graduation rates. This is equivalent to an increase of 8.9 percent in the probability of graduating within 5 years. All coefficient estimates are significant at the 1 percent level.

Lastly, I study the effect of being born before the cutoff and starting school early on whether a student ever graduates. The coefficient estimates are reported in columns (5) and (6) of Table 3. Similar to the results from 4 and 5 year graduation, I find that the effects on white and Hispanic students are the largest. I estimate that starting school early leads to a 5.1percentage point increase in the graduation rates of white students. This estimate is equivalent to a 9.8

percent increase in the probability of graduating from high school. The coefficient estimate of being born prior to the cutoff in the black sample is +2.8 percentage points. This estimate means that, among black students, those born before the cutoff are 7.3 percent more likely to ever graduate. Starting school early, using the Hispanic sample, leads to a 3.6 percentage point increase in overall graduation rates. In other words, among Hispanic students, those who started school a year early are 9.7 percent more likely to graduate from high school.

These results can be attributed to exposure to a school environment from an earlier age, for instance students born in December of any given year potentially begin their school at the age of 4 years and 8 months. Garces and Currie (2002), find that white students from disadvantaged backgrounds up until age 5 who were assigned to the Head Start program are 28 percent more likely to graduate from high school relative to a sibling who was not assigned treatment. The results in my study are not as large, primarily because the Head Start program was designed with the purpose of increasing student achievement. The graduation effects I find can also be attributed to the fact that New York City compulsory attendance age is 17 and students are allowed to drop out only if their birthday falls before the beginning of the school year. This means students born in December would be closer to the finish line by the time they are 17 years of age than students born in January. For instance, a student who enrolled in kindergarten when they were 4 years and 8 months old would be 16 and 8 months old by the time they start grade 12, assuming no retention. On the other hand, students born in January will have been 16 years and 9 months when they start grade 11. It is possible, in this case, that being closer to the finish line increases a student’s willingness to complete their high school education.

Behavioral Outcomes – Middle School and High School

For the analysis on the school behavior outcomes, I drop the elementary school sample from the analysis, since the dependent variables have little to no variation for students in grades 3-5. Therefore, the analysis is carried out on the middle school and high school samples separately. Table 6 presents the estimation results on non-academic outcomes including probability of being suspended from school for at least a day as well as the probability of showing signs of chronic absenteeism, for students in grades 6-8 and 9-10 respectively. Interpretation of the effect of being born prior to the cutoff in this case is different from that of the findings on academic outcomes. Because the control and treatment group are only 1-2 months apart in terms of real age difference coupled with the fact that the outcomes measured in this specification are not grade related, I can interpret the effect of being born prior to the cutoff as the true effect of being relatively among the youngest in the class.

The coefficient estimates of the effect of early school entry on the probability of suspension and chronic absenteeism for the middle school sample are presented in the first panel of Table 3. I estimate that being relatively among the youngest in the class is associated with an almost negligible decrease in the probability of being suspended for at least a day of around 0.9 percentage points, which is also statistically insignificant. Increasing the bandwidth on the suspension regression does not change the magnitude of the effect and only makes the standard errors smaller, yet still insignificant. Regarding the effect on chronic absenteeism, I find that a positive effect of 1.1 percentage points, meaning that middle school students born prior to the cutoff are more likely to be chronically absent from school than those who are born after the cutoff. The coefficient estimate is slightly larger when using the larger bandwidth; however the estimates are statistically insignificant under both cases.

The empirical results from the high school sample are similar to those from the middle school sample. I estimate that being relatively younger in class does not affect a high school student's propensity to be suspended from school in a significant manner. The estimated coefficient on being born prior to the cutoff leads to a lower probability of suspension in high school by 0.8 percentage points and is statistically insignificant. The effect of being relatively younger in class has almost no effect on chronic absenteeism among high school students. The estimated coefficient is 0.001 and is statistically insignificant. When increasing the bandwidth for these regressions, the results are almost identical in terms of coefficient estimates and standard errors.

The findings support the idea that school starting age has little to no effect on the non-academic performance of students in both middle school as well as high school. This may be an indication of how a student's age at school entry affects her school performance overall. Seeing that I find little evidence as to its effect on school behavioral outcomes, this result provides some insight regarding the significance of age at school entry on academic performance.

Grade Specific Outcomes – Elementary and Middle School

Table 5 presents OLS estimates of the coefficient on cutoff for the elementary and middle school sample, on the following dependent variables: z-scores on the standardized math test score and ELA test score, as well as an indicator variable for whether a student had been diagnosed as special education and an indicator variable for whether a student has been retained in his/her grade.

As mentioned earlier, the specification includes school fixed effects as well as cohort fixed effects, ensuring that the students being compared before and after the cutoff are in fact of the same age. I find that students who were born prior to the cutoff, score on average 9.4 percent of a student-level standard deviation lower than those born after the cutoff, the coefficient however is marginally insignificant. Students born in December perform much worse on the ELA exam than those born in January by 12.9 percent of a student-level standard deviation and the coefficient is statistically significant at the 10 percent level. Increasing the bandwidth by one month provides for more accurate and smaller effect sizes of -8.1 and -11.3 percent of a student-level standard deviation on the math and ELA test scores respectively. These results roughly translate to test scores that are lower by around 3-5 percentile points for students in the pre-cutoff group.

Columns 5 through 8 present the coefficient estimates for students born prior to the school cutoff on the probability of being diagnosed as a special education student as well as the probability of being retained in a given grade. I estimate that being born in December leads to a higher probability of being given a special education designation by 2.6 percentage points. Considering that the proportion of December born students with a special education designation is 8.9 percent, this translates to a 29.2 percentage increase in the likelihood of special education designation relative to students born in January. Increasing the bandwidth leads to a slightly smaller effect size of 2.3 percentage points, or a 26.4 percent higher probability in special education diagnosis. The coefficient estimates in both columns are significant at the 1 percent level. The final two columns of Table 5 represent the effect of being born prior to the cutoff on grade retention. The coefficient estimate of being born prior to the school cutoff is +0.7 percentage points (relative to the mean of 2.7 percent), and +0.6 percentage points when including November and February birth months. This means that students born prior to the school cutoff are 25.9 percent more like to be retained as compared to those born after the cutoff.

To test for heterogeneous effects of starting school early, I disaggregate the effect by race as well as by gender4. For the most part, I find that there are no differential effects of starting school early or late on males and females. When analyzing the effect across race, I find that the effect on math test scores is lowest among white students at -0.083, and -0.070 when expanding the bandwidth. Both coefficients are statistically insignificant in this case. Black and Hispanic students display effect sizes that are larger in magnitude, of -0.101 and -0.098. The effect size is larger and statistically significant at the 10% level when including November and February births in the regression. The coefficient estimates of the effect on math z-scores are -0.084 and -0.081 for black and Hispanic students, respectively. The coefficient estimates on ELA test scores across the different races are for the most part similar in magnitude, and therefore I do not observe a heterogeneous effect there.

For the effects of birth prior to the cutoff on special education diagnoses and grade retention, I find that there exists a differential impact when looking at the magnitude of the estimated coefficients. Retention rates are on average highest among black and Hispanic students, with effect sizes of +0.8 and +0.9 percentage points, while the effect on white students is +0.3 percentage points. Relative to their respective mean retention rates, the effect size across race is effectively between 25 and 29 percent, indicating the difference across races is not large. However, when studying the differential impact on special education diagnoses, I find that the effect on white students is lowest among all students. White students are only 1 percentage point more likely to be diagnosed as a special education student, whereas black and Hispanic students are on average 2.7-3.0 percentage points more likely to be diagnosed. Even relative to the treatment group mean, white students are 18.8 percent more likely to be diagnosed and black and Hispanic students are 27.6-32.7 percent more likely.

These findings suggest that students who are born in November/December are placed at a disadvantage relative to those born in January/February due to the school cutoff policy. It is important to note that this result is driven by two effects, the effect of being relatively among the youngest in class as well as the effect of being in a higher grade, while the opposite is true for those born in January and February. Because the outcome variables measured for students in elementary and middle school are different in each grade, it is not possible to net out the grade effect from these regressions. Similarly, the grade effect cannot be netted out from retention regressions, by construction the retention variable is an explicit function of a student's performance on the standardized exams in each grade. The grade effect cannot be eliminated from the regressions on special education diagnoses, as students who fall behind in terms of their academic performance have a higher likelihood of being diagnosed with ADD/ADHD (Dhuey and Lipscomb, 2010).

Grade Specific Outcomes – High School (grades 9-10)

Results for the academic performance outcomes for the high school sample are presented in table 6. The dependent variables presented here are as follows: Yearly GPA and z-scores of the Math and Science Regents examinations. Again, I estimate each regression using two bandwidths and find that all the effects are approximately the same size when expanding the sample size to include those born in November and February. The Mathematics and Science Regents Examinations are required for all students who are in grades 9 and 10, after completing the subject specific sequence during the school year.

The first two columns show that students born prior to the cutoff are on average scoring 8.0-8.1 percent of a student-level standard deviation lower on the Math Regents Examination, which are slightly lower than the effect sizes from the elementary and middle school sample.

This translates to scores that are around 3.2 percentile points lower for those who fall in the pre-cutoff group. Columns (3) and (4) represent the coefficient estimates for the science exam are more or less in the same vein as the math exams, estimating that students in the pre-cutoff group are receiving 8.9 percent of a student-level standard deviation lower on their test scores. This is approximately a 3.6 percentile point drop when compared to the performance of students born in January and February. Columns (5) and (6) present the coefficient estimates of being born before the cutoff on students' yearly GPA. I find that students born in December/November are performing worse than their counterparts by 9.8 percent of a letter grade; this is relative to a mean GPA of 1.867. This is true for both sets of bandwidths.

Panels A, B and C display the effects across race and ethnicity. When testing for any differential effects of starting school a year early on race and gender, I find that gender does not play a role in affecting the performance of students who are born prior to the school cutoff. In contrast, I find that the effect on GPA and test scores is heterogeneous across race and ethnicity. White students are least affected by the school cutoff in terms of math and science test score performance, where the coefficient on math test scores is -0.032 while the effect on black and Hispanic students is almost four times as large and is statistically significant. The difference in science test score performance is not as stark as the math results, however the effect on white students is -0.080 student-level standard deviations, while the effect on black students is -0.092 and -0.113 for Hispanic students. All the coefficient estimates are significant at the 1% level in this case. In columns (5) and (6) of panels A, B and C, I find no heterogeneous effects on the racial composition of the students.

Although the results on GPA and test scores in the high school sample cannot be directly compared to the performance of students in the younger sample. It is important to note that this should be interpreted as the combined school entry age and grade effect of the school cutoff, and means that those falling in the post-cutoff group are only outperforming their counterparts in terms of grades and exam proficiency due to essentially taking different exams.

Robustness checks

To test the sensitivity of my empirical results, I run the regressions using several specifications. I run all of my regressions with and without student observable characteristics and find that my findings are insensitive to inclusion of student controls. Similarly, I find that my results are robust to the exclusion of school level fixed effects. Another specification check is the use of a larger bandwidth for the discontinuity sample. Inclusion of students born in November and February does not change the magnitude or the statistical significant of the estimation results.

To ensure that the effect identified is not just the effect of being 1-2 months older, or even spurious, I assign a false school cutoff date and run the same specification on the placebo cutoff date. I define the false cutoff date to be august 31 and limit the sample to include only students born in August and September. In this case, students born in august are assigned a treatment value of 1for the false cutoff variable and students born in September are assigned a value of zero. The placebo regressions are run on both the elementary-middle school as well as high school samples on all the dependent variables used in the earlier analysis.

Table A in the appendix display the results of the placebo test on the elementary-middle school sample. I find that the placebo cutoff date has no effect on the academic performance of students in both math and ELA exams. The coefficient estimate of the cutoff is +1.5 percent of a student-level standard deviation and is statistically insignificant. This means that students born in august obtain on average test scores that are 0.6 percentile points higher than those born in September, which is negligible. The coefficients on special education designation, retention, suspension and chronic absenteeism also show an insignificant result of the placebo cutoff, both statistically and in magnitude.

The placebo tests on the high school sample, represented in table B of the appendix, also provide a similar result. The effect on GPA is -0.017 and is statistically insignificant. The effect on Math and Science Regents exams is between -0.01 and -0.005, approximately between 0.4 and 0.2 percentile points lower for August-born versus September-born students. I find the same nil effect on retention, suspension, chronic absenteeism, 4 year, 5 year, and overall graduation. The results of the placebo tests reject the null hypothesis that the effects of being born prior to the cutoff are spurious and suggest that the estimates are indeed causal, for both the elementary-middle school and the high school samples.

Conclusion

Using student level administrative data from the New York City Public Schools, I study the effects of school entry age cutoff on students' academic and non-academic outcomes. I test the implications of starting school early on test scores of students in grades 3-8 and 9-10. I find that students, who start school early as a result of the age cutoff, are more likely to graduate from high school than students of the same age whose birthdays fell on the other side of the cutoff on standardized exams. I find that there is no effect of a student's school entry age on behavioral outcomes. This result suggests that the age at school entry effect may not be a significant determinant of student behavioral outcomes.

More importantly, I find evidence of sizeable benefits of starting school early on students' chances of graduating from high school. This lends support to the notion that the negative effects of starting school at an earlier age in the long-run are minimal or even non-existent. In fact, I estimate that students who were born prior to the cutoff are more likely to graduate from high-school by 1.7 percentage points. The effect is even larger in magnitude when allowing students more time to graduate. I estimate that students born before the cutoff are between 7.2 and 10.8 percent more likely to complete their high school education. This finding can be attributed to two possible mechanisms. First, students who are born prior to the cutoff start school as early as 4 years and 8 months in NYCPS, which leads to early exposure to a scholastic environment that may lead to beneficial effects later on in students' schooling. Currie and Garces (2002) find that students who are exposed to schooling earlier are more likely to graduate from high school. Second, students who are born prior to the school cutoff, are on average a year closer to completion which means that when students are faced with the choice of dropping out of school, the opportunity cost of staying and graduating is higher than those who are born after the cutoff.

These findings support the findings from Black, Devereux and Salvanes (2011) that show that starting school early increases earnings of 24 year olds, however these effects dissipate completely by the time an individual is 30 years of age. The findings of this study show that the long term effects of students starting their schooling early are significant in terms of high school graduation rates. Parents who want to send their children to school as early as possible can benefit from not having to take care of their children at home. Gelbach (2002) provides evidence that enrolling a child in public kindergarten leads to an increase in the labor supply of the mother by around 5 percent. This is also relevant in the context of public school districts in highly urbanized cities in the US, where baseline graduation rates are already very low.

References for Chapter 1

Anderson, P.M., Butcher, K.F., Cascio, E.U., Schanzenbach, D.W. 2011. Is being in school better? the impact of school on children's BMI when starting age is endogenous Journal of Health Economics 30, 977-986.

Angrist, J.D., Keueger, A.B. 1991. Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics 106, 979-1014.

Bedard, K., Dhuey, E. 2006. The persistence of early childhood maturity: International evidence of long-run age effects The Quarterly Journal of Economics 121, 1437-1472.

Black, S.E., Devereux, P.J., Salvanes, K.G. 2011. Too young to leave the nest? the effects of school starting age The Review of Economics and Statistics vol. 93, 455-467.

Black, S.E., Devereux, P.J., Salvanes, K.G. 2008. Staying in the classroom and out of the maternity ward? the effect of compulsory schooling laws on teenage births* The Economic Journal 118, 1025-1054.

Buckles, K., Hungerman, D.M. 2008. Season of birth and later outcomes: Old questions, new answers NBER Working Paper 14573.

Cascio, E.U., Lewis, E.G. 2006. Schooling and the armed forces qualifying test Journal of Human Resources XLI, 294-318.

Cascio, E.U., Schanzenbach, D.W. 2007. First in the class? age and the education production function NBER Working Paper 13663.

Cho, D. 2007. The role of high school performance in explaining women's rising college enrollment Economics of Education Review 26, 450-462.

Garces, E., Currie, J., & Thomas, D. (2002). Longer-Term Effects of Head Start. American Economic Review, 92(4), 999-1012.

Datar, A. 2006. Does delaying kindergarten entrance give children a head start? Economics of Education Review 25, 43-62.

Dhuey, E., Lipscomb, S. 2010. Disabled or young? relative age and special education diagnoses in schools Economics of Education Review 29, 857-872.

Dobkin, C., Ferreira, F. 2010. Do school entry laws affect educational attainment and labor market outcomes? Economics of Education Review 29, 40-54.

Elder, T.E. 2010. The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates Journal of Health Economics 29, 641-656.

Gelbach, J. B. (2002). Public schooling for young children and maternal labor supply. American Economic Review, 307-322.

Greene, J. P., & Forster, G. 2003. Public high school graduation and college readiness rates in the United States. Vol. 3.

Imbens, G.W., Lemieux, T. 2008. Regression discontinuity designs: A guide to practice Journal of Econometrics 142, 615-635.

Luobtsky, D.H., Elder, T.E. 2009. Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers Journal of Human Resources vol. 44, 641-683.

McEwan, P.J., Shapiro, J.S. 2008. The benefits of delayed primary school enrollment Journal of Human Resources 43, 1-29.

Robertson, E. 2011. The effects of quarter of birth on academic outcomes at the elementary school level Economics of Education Review 30, 300-311.

Table 1.1: Summary Statistics for NYCPS Elementary and Middle School (Grades 3-8) Students by Month of Birth

	November	December	January	February
Panel A: Outcome Variables				
Math test score (z-score)	0.0761	0.0572	0.2117	0.2064
	[0.983]	[0.991]	[0.971]	[0.974]
English Language Arts test score (z-score)	0.0686	0.0488	0.2280	0.2176
	[0.966]	[0.969]	[0.971]	[0.976]
Retention	0.026	0.027	0.021	0.022
	[0.145]	[0.162]	[0.145]	[0.145]
Special Education	0.084	0.089	0.065	0.068
	[0.258]	[0.266]	[0.230]	[0.233]
Chronic Absenteeism	0.145	0.148	0.156	0.152
	[0.349]	[0.351]	[0.360]	[0.356]
Suspension	0.054	0.054	0.055	0.056
	[0.222]	[0.222]	[0.224]	[0.225]
Panel B: Student Characteristics				
Male	0.502	0.502	0.505	0.505
	[0.500]	[0.500]	[0.500]	[0.500]
Female	0.498	0.498	0.495	0.495
	[0.500]	[0.500]	[0.500]	[0.500]
Black	0.323	0.329	0.339	0.334
	[0.467]	[0.470]	[0.473]	[0.471]
Hispanic	0.375	0.373	0.367	0.373
	[0.484]	[0.484]	[0.482]	[0.483]
Asian	0.140	0.136	0.129	0.129
	[0.347]	[0.342]	[0.335]	[0.335]
American Indian	0.00354	0.00339	0.00353	0.00325
	[0.0594]	[0.0581]	[0.0593]	[0.0569]
Received Free/Reduced Meals	0.552	0.549	0.543	0.545
	[0.497]	[0.498]	[0.498]	[0.498]
School Size	785.9	784.7	783.9	780.9
	[477.8]	[477.5]	[478.0]	[476.4]
Elementary School (Grades 3-5)	0.413	0.413	0.418	0.417
	[0.492]	[0.492]	[0.493]	[0.493]
Middle School (Grades 6-8)	0.587	0.587	0.582	0.583
	[0.492]	[0.492]	[0.493]	[0.493]
Number of Observations	219,842	228,558	226,987	205,082

Source: New York City Public School Transcript Data, 1999-2000 through 2006-2007
Notes: Columns represent summary statistics of enrolled students by their month of birth. Chronic Absenteeism is defined as having been absent for at least 10% of total school days. Suspension is defined as having been suspended from school for at least one day. White is the omitted racial category which accounts for the rest of the students' race/ethnicity. Figures in brackets indicate standard deviations.

Table 1.2: Summary Statistics for NYCPS High School (Grades 9 and 10) Students by Month of Birth

	November	December	January	February
Panel A: Outcome Variables				
Math test score (z-score)	0.0704	0.0627	0.1685	0.1620
	[0.931]	[0.946]	[0.891]	[0.891]
Science test score (z-score)	-0.0253	-0.0205	0.1063	0.1091
	[0.953]	[0.948]	[0.925]	[0.934]
Yearly GPA	1.867	1.867	2.004	2.014
	[0.880]	[0.875]	[0.877]	[0.867]
Suspension	0.282	0.279	0.271	0.273
	[0.446]	[0.448]	[0.445]	[0.447]
Chronic absenteeism rate	0.241	0.236	0.221	0.221
	[0.286]	[0.285]	[0.291]	[0.285]
Graduate in 4 years	0.209	0.208	0.210	0.213
	[0.486]	[0.486]	[0.481]	[0.482]
Graduate in 5 years	0.304	0.307	0.318	0.320
	[0.460]	[0.461]	[0.466]	[0.467]
Graduate ever	0.414	0.415	0.402	0.403
	[0.493]	[0.493]	[0.490]	[0.491]
Panel B: Student Characteristics				
Male	0.513	0.511	0.497	0.496
	[0.500]	[0.500]	[0.500]	[0.500]
Female	0.487	0.489	0.503	0.504
	[0.500]	[0.500]	[0.500]	[0.500]
Black	0.339	0.344	0.339	0.333
	[0.473]	[0.475]	[0.473]	[0.471]
Hispanic	0.364	0.361	0.345	0.352
	[0.481]	[0.480]	[0.475]	[0.478]
Asian	0.149	0.142	0.147	0.146
	[0.356]	[0.349]	[0.355]	[0.353]
American Indian	0.00351	0.00299	0.00309	0.00301
	[0.0592]	[0.0546]	[0.0555]	[0.0548]
Received Free/Reduced Meals	0.589	0.586	0.580	0.578
	[0.492]	[0.493]	[0.494]	[0.494]
School Size	21.08	21.00	21.02	20.95
	[13.47]	[13.45]	[13.59]	[13.72]
Number of Observations	50,081	52,874	67,823	61,430

Source: New York City Public School Transcript Data, 1999-2000 through 2006-2007
Notes: Columns represent summary statistics of enrolled students by their month of birth. Chronic Absenteeism is defined as having been absent for at least 10% of total school days. Suspension is defined as having been suspended from school for at least one day. White is the omitted racial category which accounts for the rest of the students' race/ethnicity. Figures in brackets indicate standard deviations.

Table 1.3: OLS Results for 4 year, 5 year, and Overall Graduation for the High School Sample by Race

	[1]	[2]	[3]	[4]	[5]	[6]
	+/- 1 month	+/- 2 months	+/- 1 month	+/- 2 months	+/- 1 month	+/- 2 months
	Graduate in 4 years	Graduate in 4 years	Graduate in 5 years	Graduate in 5 years	Graduate Ever	Graduate Ever
Panel A: Full sample						
Treatment group sample mean	*0.214*	*0.214*	*0.307*	*0.306*	*0.415*	*0.414*
Born before school cutoff	0.017***	0.017***	0.026***	0.026***	0.038***	0.037***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	127,274	244,535	127,274	244,535	127,274	244,535
Panel B: White						
Treatment group sample mean	*0.302*	*0.302*	*0.396*	*0.400*	*0.521*	*0.523*
Born before school cutoff	0.041***	0.039***	0.043***	0.041***	0.051***	0.048***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	18,731	35,721	18,731	35,721	18,731	35,721
Panel C: Black						
Treatment group sample mean	*0.181*	*0.181*	*0.273*	*0.272*	*0.385*	*0.385*
Born before school cutoff	0.004	0.004	0.013**	0.014**	0.028***	0.029***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	43,375	82,697	43,375	82,697	43,375	82,697
Panel D: Hispanic						
Treatment group sample mean	*0.179*	*0.179*	*0.270*	*0.267*	*0.369*	*0.366*
Born before school cutoff	0.011**	0.010**	0.024***	0.023***	0.036***	0.035***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	44,726	86,570	44,726	86,570	44,726	86,570
Student Controls	yes	yes	yes	yes	yes	yes
Year of Birth (cohort) Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects	yes	yes	yes	yes	yes	yes

Notes: +/- 1 month and indicates the bandwidth of students who are born in December and January. +/- 2 months indicates that the bandwidth has been expanded to include students born in November and February. Math and ELA test scores are standardized by grade and school year. Treatment group sample means of outcomes are presented at the top of each column in each panel. All regressions include student controls (male, race/ethnicity, English Language Learner status, free/reduced price meals and school size). Year of birth (cohort) fixed effects indicate fixed effects for cohorts of students who are born within 2-4 months of each other, rather than born in the same calendar year. Numbers in parentheses represent standard errors clustered at the school year and month of birth level.***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 1.4: OLS Results for Non-Academic Outcomes for the Middle School and High School Samples by Race

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	+/- 1 month	+/- 2 months	+/- 1 month	+/- 2 months	+/- 1 month	+/- 2 months	+/- 1 month	+/- 2 months
	Suspended	Suspended	Chronically Absent	Chronically Absent	Suspended	Suspended	Chronically Absent	Chronically Absent
Panel A: Full sample								
Sample mean of treatment group	*0.055*	*0.055*	*0.148*	*0.152*	*0.279*	*0.281*	*0.236*	*0.238*
Born before school cutoff	-0.007	-0.007	0.011	0.018	-0.007	-0.007	0.001	0.001
	(0.019)	(0.013)	(0.024)	(0.015)	(0.045)	(0.044)	(0.026)	(0.026)
Observations	161,389	311,561	93,106	179,198	108,914	209,440	127,274	244,535
Panel B: White								
Sample mean of treatment group	*0.064*	*0.064*	*0.114*	*0.115*	*0.224*	*0.220*	*0.157*	*0.154*
Born before school cutoff	-0.001	-0.003	0.022	0.023	-0.019	-0.019	-0.003	-0.003
	(0.021)	(0.014)	(0.028)	(0.018)	(0.036)	(0.036)	(0.018)	(0.018)
Observations	24,876	47,897	14,275	27,493	15,657	29,917	18,731	35,721
Panel C: Black								
Sample mean of treatment group	*0.052*	*0.052*	*0.175*	*0.178*	*0.319*	*0.323*	*0.269*	*0.273*
Born before school cutoff	-0.005	-0.008	0.001	0.015	-0.005	-0.005	0.005	0.004
	(0.016)	(0.011)	(0.024)	(0.016)	(0.048)	(0.046)	(0.031)	(0.031)
Observations	53,188	101,922	30,152	57,474	36,977	70,518	43,375	82,697
Panel D: Hispanic								
Sample mean of treatment group	*0.049*	*0.182*	*0.177*	*0.049*	*0.283*	*0.287*	*0.287*	*0.289*
Born before school cutoff	-0.007	-0.006	0.022	0.025	-0.008	-0.008	-0.004	-0.004
	(0.016)	(0.011)	(0.027)	(0.018)	(0.047)	(0.045)	(0.032)	(0.031)
Observations	59,691	115,854	34,437	66,663	38,466	74,470	44,726	86,570
Student Controls	yes	yes	yes	yes	yes	yes	yes	yes
Year of Birth (cohort) Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
School Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes

Notes: +/- 1 month and indicates the bandwidth of students who are born in December and January. +/- 2 months indicates that the bandwidth has been expanded to include students born in November and February. Sample means of the treatment group are presented at the top of each column in each panel. All regressions include student controls (male, race/ethnicity, English Language Learner status, free/reduced price meals and school size). Year of birth (cohort) fixed effects indicate fixed effects for cohorts of students who are born within 2-4 months of each other, rather than born in the same calendar year. Numbers in parentheses represent standard errors clustered at the school year and month of birth level.***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 1.5: OLS Results for Academic Outcomes for the Elementary-Middle School Sample by Race

	[1] +/- 1 month	[2] +/- 2 months	[3] +/- 1 month	[4] +/- 2 months	[5] +/- 1 month	[6] +/- 2 months	[7] +/- 1 month	[8] +/- 2 months
	Math test score	Math test score	ELA test score	ELA test score	Special Education	Special Education	Retained	Retained
Panel A: Full sample								
Treatment group sample mean	0.0572	0.0665	0.049	0.059	0.089	0.087	0.027	0.027
Born before school cutoff	-0.094	-0.081*	-0.129*	-0.113**	0.026***	0.023***	0.007***	0.006***
	(0.066)	(0.046)	(0.063)	(0.044)	(0.004)	(0.003)	(0.002)	(0.001)
Observations	433,346	837,569	406,262	784,856	455,545	880,469	455,545	880,469
Panel B: White								
Treatment group sample mean	0.4559	0.4668	0.484	0.059	0.056	0.053	0.009	0.009
Born before school cutoff	-0.083	-0.070	-0.132*	-0.125***	0.011***	0.010***	0.003***	0.003***
	(0.061)	(0.043)	(0.064)	(0.044)	(0.004)	(0.003)	(0.001)	(0.001)
Observations	69,495	134,310	66,890	129,182	72,969	140,864	72,969	140,864
Panel C: Black								
Treatment group sample mean	-0.1452	-0.1393	-0.104	-0.097	0.112	0.111	0.036	0.035
Born before school cutoff	-0.101	-0.084*	-0.126**	-0.105**	0.031***	0.027***	0.009**	0.008***
	(0.065)	(0.046)	(0.058)	(0.041)	(0.005)	(0.004)	(0.003)	(0.002)
Observations	146,452	280,851	142,288	272,512	151,998	291,316	151,998	291,316
Panel D: Hispanic								
Treatment group sample mean	-0.1203	-0.1111	-0.133	-0.119	0.107	0.103	0.033	0.033
Born before school cutoff	-0.098	-0.086*	-0.127*	-0.109**	0.035***	0.030***	0.008***	0.008***
	(0.066)	(0.047)	(0.063)	(0.044)	(0.005)	(0.004)	(0.002)	(0.002)
Observations	160,567	311,756	144,961	281,644	168,603	327,521	168,603	327,521
Student Controls	yes	yes	yes	yes	yes	yes	yes	yes
Year of Birth (cohort) Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
School Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes

Notes: +/- 1 month and indicates the bandwidth of students who are born in December and January. +/- 2 months indicates that the bandwidth has been expanded to include students born in November and February. Math and ELA test scores are standardized by grade and school year. Treatment group sample means of outcomes are presented at the top of each column in each panel. All regressions include student controls (male, race/ethnicity, English Language Learner status, free/reduced price meals and school size). Year of birth (cohort) fixed effects indicate fixed effects for cohorts of students who are born within 2-4 months of each other, rather than born in the same calendar year. Numbers in parentheses represent standard errors clustered at the school year and month of birth level.***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 1.6: OLS Results for Test Scores and GPA for the High School Sample by Race

	[1] +/- 1 month Math test score	[2] +/- 2 months Math test score	[3] +/- 1 month Science test score	[4] +/- 2 months Science test score	[5] +/- 1 month Yearly GPA	[6] +/- 2 months Yearly GPA
Panel A: Full sample						
Treatment group sample mean	*0.0627*	*0.0665*	*-0.0205*	*-0.0229*	*1.867*	*1.867*
Born before school cutoff	-0.080**	-0.081***	-0.089***	-0.089***	-0.098***	-0.098***
	(0.027)	(0.028)	(0.021)	(0.023)	(0.022)	(0.023)
Observations	48,523	94,024	68,735	132,215	111,396	214,059
Panel B: White						
Treatment group sample mean	*0.4305*	*0.4409*	*0.3113*	*0.3127*	*2.305*	*2.322*
Born before school cutoff	-0.032	-0.029	-0.080***	-0.082***	-0.097***	-0.102***
	(0.022)	(0.021)	(0.016)	(0.021)	(0.022)	(0.023)
Observations	8,792	17,033	12,704	24,371	16,434	31,399
Panel C: Black						
Treatment group sample mean	*-0.2553*	*-0.2541*	*-0.2641*	*-0.2749*	*1.613*	*1.605*
Born before school cutoff	-0.120***	-0.119***	-0.092***	-0.088***	-0.087***	-0.087***
	(0.032)	(0.037)	(0.029)	(0.029)	(0.025)	(0.026)
Observations	14,484	27,778	21,100	40,179	38,299	72,896
Panel D: Hispanic						
Treatment group sample mean	*-0.2202*	*-0.2212*	*-0.2400*	*-0.2464*	*1.668*	*1.666*
Born before school cutoff	-0.122**	-0.126***	-0.113***	-0.115***	-0.107***	-0.108***
	(0.041)	(0.041)	(0.031)	(0.031)	(0.027)	(0.026)
Observations	14,707	28,715	20,519	39,716	38,483	74,561
Student Controls	yes	yes	yes	yes	yes	yes
Year of Birth (cohort) Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects	yes	yes	yes	yes	yes	yes

Notes: +/- 1 month and indicates the bandwidth of students who are born in December and January. +/- 2 months indicates that the bandwidth has been expanded to include students born in November and February. Math and ELA test scores are standardized by grade and school year. Treatment group sample means of outcomes are presented at the top of each column in each panel. All regressions include student controls (male, race/ethnicity, English Language Learner status, free/reduced price meals and school size). Year of birth (cohort) fixed effects indicate fixed effects for cohorts of students who are born within 2-4 months of each other, rather than born in the same calendar year. Numbers in parentheses represent standard errors clustered at the school year and month of birth level.***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Appendix I

Table 1.A: Placebo Test Results on Outcomes for the Elementary-Middle School Sample

	[1] Math test score	[2] ELA test score	[3] Special Education	[4] Retained	[5] Suspended	[6] Chronically Absent
Placebo cutoff	0.015	0.015	-0.003	-0.000	0.000	0.001
	(0.065)	(0.061)	(0.005)	(0.001)	(0.014)	(0.131)
Male	-0.028***	-0.182***	0.041***	0.009***	0.000	0.005**
	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)
Black	-0.343***	-0.280***	0.045***	0.008***	0.000	-0.002
	(0.007)	(0.008)	(0.002)	(0.001)	(0.002)	(0.015)
Hispanic	-0.284***	-0.276***	0.039***	0.006***	-0.000	0.010
	(0.007)	(0.007)	(0.004)	(0.001)	(0.002)	(0.019)
Asian	0.258***	0.091***	-0.011***	-0.002***	0.003	-0.043**
	(0.008)	(0.005)	(0.002)	(0.001)	(0.002)	(0.017)
American Indian	-0.417***	-0.411***	0.082***	0.016***	0.011**	0.021
	(0.021)	(0.018)	(0.006)	(0.004)	(0.004)	(0.013)
Free/Reduced meal	-0.228***	-0.233***	0.006	0.006***	0.010	0.029
	(0.049)	(0.038)	(0.005)	(0.002)	(0.009)	(0.110)
School size/100	0.031***	0.031***	-0.001	0.000	0.001	-0.037***
	(0.011)	(0.009)	(0.001)	(0.000)	(0.001)	(0.009)
Constant	0.264*	0.571***	0.091***	0.061***	-0.263***	3.088***
	(0.127)	(0.103)	(0.018)	(0.009)	(0.081)	(0.381)
Observations	444,890	416,162	467,248	467,248	250,526	467,248
Student Controls	yes	yes	yes	yes	yes	yes
Year of Birth (cohort) Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects	yes	yes	yes	yes	yes	yes

Notes: Placebo cutoff is defined as August 31, the treatment group are students born in August. Only students born in August and September are included. Year of birth (cohort) fixed effects indicate fixed effects for cohorts of students who are born within 2-4 months of each other, rather than born in the same calendar year. Numbers in parentheses represent standard errors clustered at the school year and month of birth level.***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

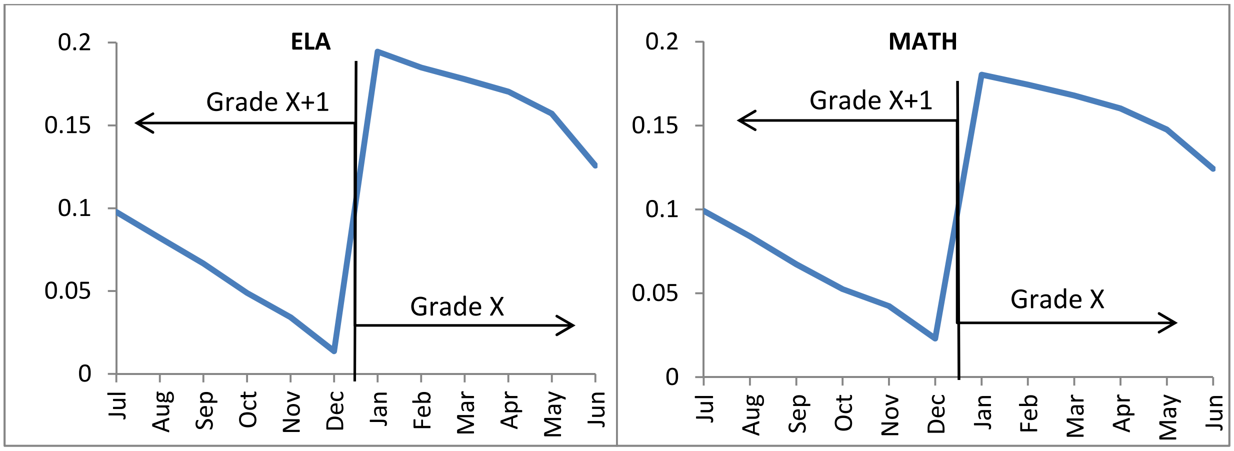
Table 1.B: Placebo Test Results on Outcomes for the High School Sample

	[1] Suspended	[2] Chronically Absent	[3] Graduated in 4 years	[4] Graduated in 5 years	[5] Graduated Ever
Placebo cutoff	-0.001	0.006***	-0.004	-0.004	-0.006*
	(0.006)	(0.002)	(0.004)	(0.004)	(0.004)
Male	-0.000	0.002	-0.043***	-0.047***	-0.047***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Black	0.005*	0.009***	-0.049***	-0.044***	-0.036***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)
Hispanic	0.004	0.034***	-0.057***	-0.061***	-0.052***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)
Asian	0.006*	-0.055***	0.028***	0.033***	0.037***
	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)
American Indian	0.034**	0.028	-0.051***	-0.041**	-0.003
	(0.014)	(0.019)	(0.019)	(0.020)	(0.020)
Free/Reduced meal	-0.004	0.002	-0.004	0.000	0.004
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
School size/100	0.010***	0.003***	0.002**	-0.003**	-0.005***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-1.491***	-2.384***	1.053***	1.451***	1.670***
	(0.168)	(0.048)	(0.057)	(0.060)	(0.061)
Observations	136,462	160,346	160,346	160,346	160,346
Student Controls	yes	yes	yes	yes	yes
Year of Birth (cohort) Fixed Effects	yes	yes	yes	yes	yes
School Fixed Effects	yes	yes	yes	yes	yes

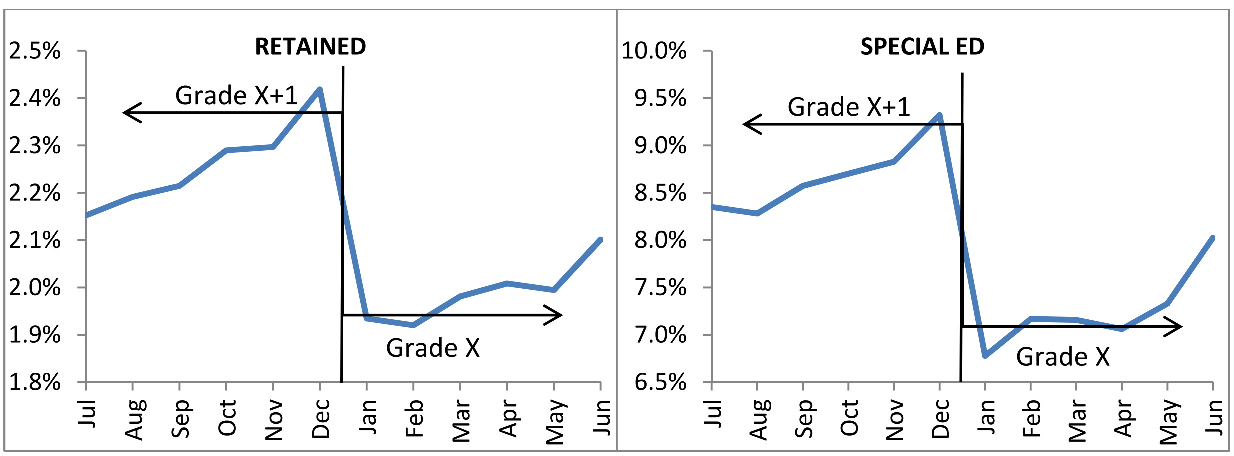
Notes: Placebo cutoff is defined as August 31, the treatment group are students born in August. Only students born in August and September are included. Year of birth (cohort) fixed effects indicate fixed effects for cohorts of students who are born within 2-4 months of each other, rather than born in the same calendar year. Numbers in parentheses represent standard errors clustered at the school year and month of birth level.***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Figure 1.1: Academic Performance of NYCPS Elementary and Middle School Students, Grades 3 through 8

Panel A: Average Math and ELA z-score by month of birth, with December 31st NYCPS cutoff

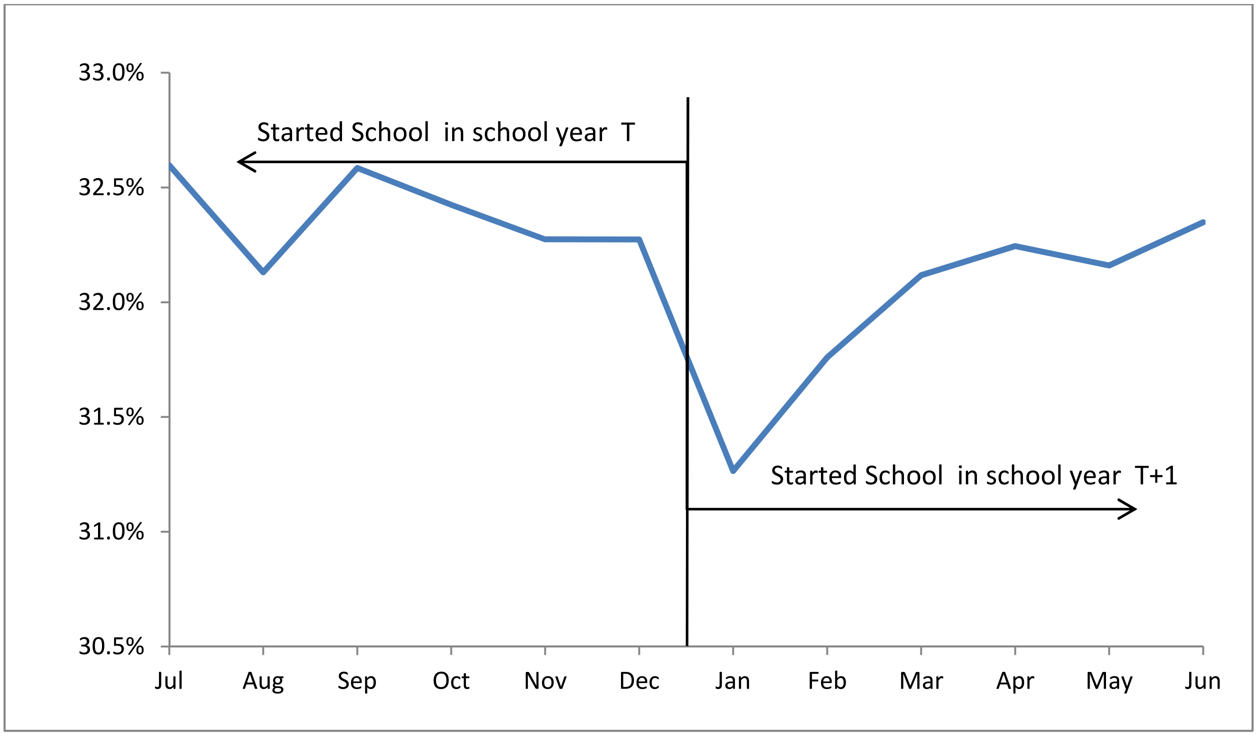


Panel B: Average Retention Rates and Special Education Diagnoses



Source: New York City Public School Transcript Data, 1999-2000 through 2006-2007

Figure 1.2: High School Graduation Rates of NYCPS High School Students, Grades 9-10



Source: New York City Public School Transcript Data, 1999-2000 through 2006-2007

CHAPTER 2. Making the Grade: The Impact of Classroom Behavior on Academic Achievement

Introduction

Lazear (2001) posits an elegant theoretical model of class size, in which students enrolled in smaller classes learn more because they experience fewer student disruptions during their class instruction. The Lazear framework hypothesizes that the mechanism behind the effect of class size on achievement is classroom behavior, whereby adding more students to a classroom increases the number of disruptions and consequently decreases the amount of time during which learning can take place. In other words, classroom education inherently has properties of a *public good*, in that if one student disrupts his or her class, the learning of all other students within the class is also harmed. Recent evidence suggests that there is considerable variation across students in the propensity to disrupt class and that this propensity is correlated with measurable student-level characteristics, such as socioeconomic status (Segal, 2008).

Our paper examines empirical implications of the Lazear educational production model. Using exogenous variation on course scheduling for ninth graders in Chicago Public Schools (CPS), we study heterogeneity in the impact of class size on student achievement in reading and mathematics. Our identification strategy allows us to analyze an underlying mechanism by which class size affects student performance, the behavioral composition of a classroom. Our classroom composition measure is constructed as the average number of non-disruptive students in attendance on a given school day; we characterize a classroom as being comprised of disruptive and non-disruptive students, where a student is considered disruptive if he or she dropped out of high school in any grade.

As one of the largest urban public school districts in the United States, currently serving over 400,000 students, CPS provides a unique opportunity for analyzing heterogeneity in class size effects for a population of predominantly racial/ethnic minority students that are largely from lower-income families. Approximately 75 percent of CPS students receive federal lunch subsidies, and the racial/ethnic makeup of the student body is approximately 55 percent black; 30 percent Hispanic; and the remaining 15 percent white, American Indian, or Asian.

Our data are taken from CPS administrative student transcript files, which include the ordering of classes over the day, student absences, course titles, grades, scores from standardized tests in reading and mathematics, and demographic characteristics for the universe of CPS high school students from the 1993-94 to the 2005-06 school years. To study the effects of the behavioral composition of a classroom on academic achievement, we use an instrumental variables approach, exploiting exogenous variation in the period of the day a course is offered in CPS. Our analysis focuses on students' course passing and test scores in four ninth grade course subjects: regular English I, remedial English I, regular Algebra I, and remedial Algebra I.

The Lazear theoretical model of educational production suggests that the behavioral composition of a classroom is an important determinant of educational achievement, implying that it should be included as an explanatory variable when estimating an educational production function. Our two-stage least squares (2SLS) results are in line with the Lazear framework, as we find that the behavioral composition of a classroom significantly affects student achievement.

For the full sample of students enrolled in regular English I, our 2SLS estimates indicate that an additional non-disruptive student in attendance on a given school day increases the probability of passing English I by 7.26 percentage points, relative to the mean passing rate of 76.8 percent. Disaggregating this effect by race and ethnicity, we find a larger increase in the probability of passing English I for black versus Hispanic students. We also observe that an additional non-disruptive student in attendance increases a student's own reading test score by

0.0222 student-level standard deviations for the full sample and by 0.0633 student-level standard deviations for Hispanic students; we estimate an insignificant effect on reading scores for black students.

Our full sample and Hispanic subsample effect sizes for reading test scores are similar in magnitude to the effect sizes reported in Finn and Achilles (1990) for first grade students that participated in Tennessee's Project STAR class-size randomized experiment. For the full sample, Finn and Achilles (1990) estimate an effect size for reading test scores of approximately 0.0275 student-level standard deviations for a one-student reduction in class size (based on Table 5, page 566). For minority students, Finn and Achilles (1990) estimate an effect size of approximately 0.0400 student-level standard deviations (based on Table 6, page 567). The full sample effect sizes are close in magnitude across the two studies, while the minority subsample effect size in Finn and Achilles (1990) is approximately two thirds of the magnitude of our Hispanic subsample effect size.

Our estimated impacts of the behavioral composition of a classroom on course passing for students enrolled in regular Algebra I are smaller in magnitude than the corresponding effects for regular English I. For the regular Algebra I sample, we find that an additional non-disruptive student in attendance increases the probability of passing Algebra I by 3.92 percentage points, relative to the mean passing rate of 72.4 percent. Estimates for the black and Hispanic subsamples are of approximately the same magnitude as the full sample estimate.

The Lazear model also suggests that the effects of the behavioral composition of a classroom on student outcomes should be larger for students enrolled in remedial versus regular courses because students in remedial courses have, on average, lower baseline academic performance than students in regular classes. Consistent with this, we find larger overall effects of classroom composition on English I and Algebra I course passing for students enrolled in remedial versus regular classes.

Overview of Empirical Literature on Class Size Effects and Mechanisms

Identifying the causal impact of class size on student attainment is difficult in observational studies due to nonrandom sorting of students across schools and classrooms by students, parents, teachers, and administrators, as well as heterogeneity in financial and educational resources. As a result, studies that estimate class size effects have generally used experimental or quasi-experimental research designs. For example, many papers have used data from Tennessee's Project STAR class-size randomized experiment to examine the effect of smaller class sizes on student achievement, whereby students and teachers in participating elementary schools were randomly assigned to one of three class types: small (13-17 student) classes, regular (22-25 student) classes, and regular classes with a teacher aide. Finn and Achilles (1990), Word et al. (1990), Krueger (1999), Nye, Hedges, and Konstantopoulos (1999, 2000), Finn, Gerber, Achilles, and Boyd-Zaharias (2001), Krueger and Whitmore (2001), and McKee, Rivkin, and Sims (2010) find statistically significant effects of attending a smaller class on student achievement and educational attainment.

Other work has used quasi-experimental research, isolating plausibly exogenous variation in class sizes in earlier grades (elementary and/or middle) from non-linear relationships between enrollment and class sizes (class-size rules) in a regression discontinuity design framework and/or idiosyncratic population compositions due to random variation in the timing of births. Such studies have been conducted using data from Israel (Angrist and Lavy (1999)), Connecticut (Hoxby (2000)), Texas ((Rivkin, Hanushek, and Kain (2005)), and California (Babcock and

Betts (2009) and Jepsen and Rivkin (2009)). Other than Hoxby (2000), this research finds statistically significant impacts of smaller class sizes on student outcomes primarily for elementary school students.[4]

A recent study, McKee, Rivkin, and Sims (2010), extends Lazear's (2001) theoretical framework with the goal of empirically investigating heterogeneity in class size effects by income and prior achievement. As with Lazear (2001), McKee, Rivkin, and Sims (2010) assume that the amount of time available for teaching depends on the level of classroom disruption, implying that class size effects are largest in classrooms with students that have higher propensities to disrupt. This would lead to larger benefits of reduced class sizes in poorer schools if the likelihood of disruption were larger at the lower end of the income distribution. The amount of time available for learning also depends on the quality of learning, which is itself a function of baseline academic achievement and class size. The authors then discuss how smaller class sizes may be more or less beneficial to higher-achieving versus lower-achieving students, concluding that the magnitude of class size effects across the achievement distribution is ambiguous. Using data from Tennessee's Project STAR class-size randomized experiment, the authors empirically test the predictions of their model, finding that greater benefits from reduced class sizes accrue to students with higher baseline achievement, as well as to students in lower-income schools.

[4] Many of the studies above have also examined whether the impact of attending a larger class is heterogeneous across student demographics. They generally find evidence of heterogeneity, with larger class size effects for black and lower-income students. Using quantile regression analysis, other research has looked at whether there is heterogeneity across the distribution of prior student achievement. Three non-experimental studies, Eide and Showalter (1998) with class size data from the United States, and Levin (2001) and Ma and Koenker (2006) with data from The Netherlands, find little to no heterogeneity across the prior achievement distribution in the benefits of attending a smaller class. However, Konstantopoulos (2008), McKee, Rivkin, and Sims (2010), and Ding and Lehrer (2011), which use data from Tennessee's Project STAR class-size randomized experiment, do find that smaller class sizes yield larger benefits to students with higher past achievement.

Babcock and Betts (2009) also examine mechanisms, investigating whether class size effects for elementary school students in San Diego vary depending on two separately identified student classifications: baseline student effort, as measured by teacher assessments of students' conduct in the classroom, and baseline achievement, as calculated by letter grades in academic subjects. Exogenous variation in class size follows from a state policy that legislatively lowered class sizes in kindergarten through third grade only, allowing the authors to study the impact of class size on test scores using the transition from third to fourth grade. The results indicate that class size effects are larger for students with lower baseline effort, consistent with an implication of the Lazear model that students with more behavioral problems benefit more from smaller classes, while there is no evidence of heterogeneity across high- and low-achieving students.

Empirical Strategy

We model the effects of the behavioral composition of a classroom on course passing and test scores for students in four ninth grade CPS course subjects: regular English I, remedial English I, regular Algebra I, and remedial Algebra I. We define a classroom to be composed of two types of students, disruptive and non-disruptive, where a student is defined as disruptive if he or she ever left high school due to one of the following reasons reported in CPS administrative records: legally committed to a correctional facility, lost (truant officer cannot locate), excessive absences, and uniform discipline code violation (infringement of the CPS code of conduct). Consider the following linear specification:

$$(1)\ Y_{ict} = \beta_0 + \beta_1 Composition_{ct} + \delta' X_{it} + \eta_{jkt} + \varepsilon_{ict},$$

where Y_{ict} denotes one of two outcome variables for student i enrolled in class c in the fall or spring semester of academic year t: the receipt of a grade of D or better in a particular course or the test score in a subject-relevant (reading or mathematics) standardized examination. The explanatory variable of interest is $Composition_{ct}$, the average number of non-disruptive students in attendance on a given school day in the classroom that student *i* is enrolled in:

(2) $Composition_{ct} = \frac{1}{90}\sum_{i \in ct}(90 - Absences_{ict})(1 - Disruptive_i)$,

where $Disruptive_i$ is an indicator for whether student *i* is disruptive, and $Absences_{ict}$ is the number of days student *i* was absent in a particular semester in academic year *t* for class *c*. With 90 school days in each semester, $(90 - Absences_{ict})$ is the number of days that student *i* was in attendance in a particular semester in academic year *t* for class *c*. This variable is then summed over every non-disruptive student in the class. This sum is divided by 90 to obtain the daily average number of non-disruptive students in attendance in class *c* in a given semester in academic year *t*, $Composition_{ct}$.

X_{it} is a vector of observable student-specific characteristics, which includes the subject-specific eighth grade Iowa Test of Basic Skills (ITBS) test score measured in student-level standard deviations, demographic variables, and neighborhood variables measured at the level of the Census block group. η_{jkt} represents our fixed effects, where subscript *j* is for a high school or a high school teacher, and *k* is for a middle school. Three sets of fixed effects are used to capture time-variant, unobserved high school (school attended in ninth grade), middle school (school attended in eighth grade), and/or high school teacher (ninth grade teacher) quality. Specifically, we include, in separate specifications, high school-by-semester fixed effects (a separate fixed effect for each high school in each semester of each academic year), middle school-by-high school-by-semester fixed effects (a separate fixed effect for each combination of middle school and high school in each semester of each academic year), and teacher-by-semester fixed effects (a separate fixed effect for each high school teacher in each semester of each academic year).[5]

The inclusion of these fixed effects ensures that we estimate the effect of the behavioral composition of a classroom on student achievement using only variation in classroom compositions within a given high school in a given semester (first set of fixed effects); within a given middle and high school combination in a given semester (second set of fixed effects), or for a given high school teacher in a given semester (third set of fixed effects). ε_{ict} represents the idiosyncratic error term.

Ordinary least squares (OLS) estimation of equation (1) leads to biased estimates of the effects of the behavioral composition of a classroom on student achievement because our explanatory variable of interest, $Composition_{ct}$, is a function of the number of student absences, which are not randomly assigned across students. For example, number of absences is negatively correlated with prior student achievement when students with lower past achievement have higher probabilities of being absent on a given school day.[6] This implies that the coefficient on classroom composition from an OLS regression has a downward bias.

To eliminate this bias, and hence to estimate the causal effect of classroom composition on student outcomes, we use an instrumental variables regression approach, exploiting exogenous variation in the ordering of classes over the school day. Our set of excluded instrumental variables for $Composition_{ct}$ is $Period_{ct}$ and $Period_{ct}^2$, a quadratic function in the

[5] We do not estimate models with middle school-by-teacher-by-semester fixed effects because of the very large number of singleton groups (i.e., fixed effects with exactly one observation), and thus much smaller effective sample sizes, when using these fixed effects.

[6] Our data lend support to this hypothesis. For the regular English I sample, the raw correlation between eighth grade ITBS reading score and the number of ninth grade absences is -0.1569. For the regular Algebra I sample, the raw correlation between eighth grade ITBS mathematics score and the number of ninth grade absences is -0.1918. The corresponding raw correlations for the remedial English I and the remedial Algebra I samples are -0.1325 and -0.1850, respectively.

period of the school day in which a class is scheduled. This identification strategy is similar to the one used by Cortes, Bricker, and Rohlfs (2012), in which they use effectively random variation in course scheduling to measure how the returns to classroom learning vary by course subject and how attendance in one class spills over into learning in other subjects. In Section IV, we show graphically that student absences in a particular class vary depending on the period of the day in which the class is offered. Moreover, our measure of the behavioral composition of a classroom generally exhibits an inverted U-shaped pattern when plotted against the period of the school day in which the course is offered, implying that the excluded instrumental variables are highly correlated with classroom composition.

Equation (1) is now the second-stage equation, and the first-stage equation is:

(3) $Composition_{ct} = \alpha_0 + \alpha_1 Period_{ct} + \alpha_2 Period_{ct}^2 + \theta' X_{it} + \eta_{jkt} + \upsilon_{ict}$,

where the endogenous variable, $Composition_{ct}$, is a function of the excluded instruments, as well as the control variables and fixed effects that appear in equation (1). In Section V, we show that period of the day and period of the day squared are statistically significant and strong predictors of the behavioral composition of a classroom in almost all of our empirical specifications.

Data and Sample Characteristics

Data Source

The data for this study come from CPS administrative student records. Our data cover the universe of ninth grade students in CPS from the 1993-94 to the 2005-06 school years. We link each student's record to his or her individual transcript file. The transcript data include course titles and numbers, period of the day, absences by class period, and unique teacher identifiers for each class taken by students. The CPS data also include multiple standardized test scores, a detailed set of descriptive variables about each student, and 1990 and 2000 neighborhood characteristics for the Census block group in which each student resides.

The standardized tests that were administered, and the scores of the students who took them, vary from year to year in our sample. Consequently, the samples for the test score regressions are smaller than the samples for the course passing regressions. For the majority of students, eighth grade reading and mathematics test scores are available from the Iowa Test of Basic Skills (ITBS) for each year of our data. The ninth grade test score data for reading and mathematics are taken from the TAP (Test of Achievement and Proficiency) for the 1993-94 to the 2001-02 school years and from the EXPLORE test for the 2002-03 to the 2005-06 school years. To compare observations from different years in our sample, each test score is converted into a z-score, whereby each student's raw test score is standardized using the mean and standard deviation across all students in CPS that took the relevant examination in a given year.

Descriptive Statistics

The summary statistics for the analytic samples are given in Table 1. We focus on students enrolled in general (i.e., non-vocational, non-magnet, and non-alternative) high schools.[7] The student-level outcome variables of interest are an indicator for passing regular or remedial English I, an indicator for passing regular or remedial Algebra I, the score on the standardized reading examination, and the score on the standardized mathematics examination. In accordance with CPS policy, we defined a student as having passed a course if he or she received a grade of D or better in that course.[8] Panel A of Table 1 shows that students enrolled in remedial classes have lower course passing rates in both English I and Algebra I, as compared to

[7] We also restrict the analysis to first-time ninth grade students; for consistency, if a student repeated ninth grade one or more times, we only use his or her first instance of ninth grade in the data.
[8] Chicago Public Schools Policy Manual Board Report 04-0128-PO1 (January 28, 2004)

students in regular classes. The passing rates of students enrolled in regular English I and Algebra I are 77 and 72 percent, respectively, while the passing rates of students enrolled in remedial English I and Algebra I are 75 and 69 percent, respectively.

The lower academic performance of students in remedial classes can also be observed in their ninth grade reading and mathematics test scores. The average ninth grade reading score for the regular English I sample is 0.080 student-level standard deviations versus -0.327 student-level standard deviations for the remedial English I sample. Likewise, the average ninth grade mathematics score for the regular Algebra I sample is 0.076 student-level standard deviations, as compared to -0.424 student-level standard deviations for the remedial Algebra I sample.

This difference in academic performance is also seen in the students' baseline (eighth grade) performance on reading and mathematics examinations, as shown in the last two rows of panel C. The average baseline reading score for the regular English I sample is 0.080 student-level standard deviations, in contrast to -0.352 student-level standard deviations for the remedial English I sample. Similarly, the average baseline mathematics score for the regular Algebra I sample is 0.042 student-level standard deviations, as opposed to -0.545 student-level standard deviations for the remedial Algebra I sample.

While student achievement is lower for students enrolled in remedial versus regular classes, it is important to note that student-specific characteristics do not differ in a systematic manner across observables for the different course subjects. The mean age of ninth graders in all course subjects is 14.3 years, and classes are comprised of 50 percent male. The racial composition is stable across course subjects. Black students account for between 52 and 57 percent of the students enrolled in any given class; Hispanic students account for between 34 and 37 percent; and white, American Indian, and Asian students together account for the remaining eight to 12 percent. Eighty-four percent of ninth graders receive free or reduced lunch, and their proportion across course subjects is fairly stable, ranging between 81 and 87 percent. We find that the proportion of students in special education programs is higher in remedial versus regular classes (21 percent compared to 15 percent). Lastly, the neighborhood characteristics of a student's residence (shown in panel D) are similar for both students enrolled in regular classes and for students enrolled in remedial classes.

Instrumental Variables

To measure the causal effects of the behavioral composition of a classroom on student academic performance, we now make the case for our excluded instrumental variables: period of the day and period of the day squared. After students select the courses that they will take in a semester, the ordering of classes over the day is a computerized and essentially random process that is determined based on scheduling constraints.[9] Moreover, our analysis focuses on English I and Algebra I, required courses that are offered multiple times during every period of the day. A testable implication of the course scheduling process is that student and classroom characteristics should be similar between classes that meet in a particular period of the day and those that meet at other times. In other words, we assert that students enrolled in a given course subject in a given period are otherwise similar to students who take a course in that subject at another time during the day. Tables 2A and 2B present strong evidence of this premise, lending credibility to the use of differences in course scheduling in CPS as an exogenous source of variation in classroom composition to identify the effect of classroom composition on student achievement.

Table 2A shows, separately by period of the day, the fraction of courses offered in each subject. Though we focus on English and mathematics courses for this study, it is still instructive

[9] In private discussions, school administrators have indicated that the process is computerized and essentially random.

to look at all course subjects to validate period of the day as a viable instrument. Table 2A is calculated from unweighted student-level data, and the fractions in each column sum to one. As Table 2A shows, the breakdown of classes by subject is generally stable over the course of the day; for a particular subject, the percentage of course offerings in that subject differs by, at most, two percentage points across periods. This implies that schools do not appear to systematically schedule academic subjects in certain periods, such as those with lower absence rates (we return to this point later in this section).

Even stronger evidence of the validity of using period of the day as an instrument is provided by regressing an indicator for the period of the day that the student took (regular or remedial) English I or Algebra I on the student- and neighborhood-level control variables used in our outcome regressions, as well as teacher-by-semester fixed effects. Table 2B reports coefficients from such linear probability models for the sample of students enrolled in regular English I; each column is for a different period of the day.[10] Almost all coefficients are statistically insignificant, and the coefficients that are significant show no apparent pattern across columns.

We next examine the raw, reduced-form (i.e., first-stage) effects of having English I and Algebra I in a particular period on the average number of non-disruptive students in attendance in that period. Panels A and B of Figure 1 show the relationship between period of the day and the endogenous variable of our model, the behavioral composition of a classroom. Panel A gives the average numbers of non-disruptive students in attendance for regular and remedial English I as functions of the period of the day during which these classes meet. The corresponding Algebra I graphs are presented in panel B. In both panels, the solid lines show the means for the full

[10] See Appendix Tables A1-A3 for the regression results for the other samples: Appendix Table A1 for regular Algebra I, Appendix Table A2 for remedial English I, and Appendix Table A3 for remedial Algebra I.

(pooled black, Hispanic, white, Asian, and American Indian) sample, the longer dashed lines show the means for black students, and the shorter dashed lines show the means for Hispanic students.

These figures indicate that the average numbers of non-disruptive students in attendance for English I and Algebra I are generally at their lowest levels in first period, gradually rise until approximately fourth period, and then gradually decline over the remainder of the school day. This inverted U-shaped pattern is most pronounced for the regular English I, remedial English I, and regular Algebra I full samples, as well as their corresponding black and Hispanic subsamples, providing compelling evidence of a strong relationship between our excluded instrumental variables and our measure of classroom composition. For the full sample, the average number of non-disruptive students in attendance varies across periods from 16.8 to 18.7 for regular English I, from 16.3 to 18.2 for remedial English I, from 17.2 to 18.9 for regular Algebra I, and from 16.9 to 17.7 for remedial Algebra I.

Empirical Results and Discussion

In this section, we present our empirical analysis of Lazear's (2001) model, which suggests that the behavioral composition of a classroom is a key contributing factor to educational attainment. Tables 3-6 report the OLS and the two-stage least squares (2SLS) regression results for equation (1), which gives the effects of classroom composition on course passing and test scores. These tables show the estimated coefficient on our measure of classroom composition, as well as the F-statistic for the test of the predictive power of the excluded instruments in first-stage equation (3).

Table 3 presents the results for the regular English I sample; the dependent variable in panel A is an indicator for whether a student received a passing grade in his or her English I course, and the dependent variable in panel B is his or her z-score on the standardized reading examination. Table 4 reports the findings for the regular Algebra I sample; the dependent variable in panel A is an indicator for whether a student received a passing grade in his or her Algebra I class, and the dependent variable in panel B is his or her z-score on the standardized mathematics examination. Tables 5 and 6 display the estimates for the remedial English I and remedial Algebra I samples, respectively. Within each panel, the topmost set of results is for the full (pooled black, Hispanic, white, Asian, and American Indian) sample, the next set is for the black subsample, and the bottommost set is for the Hispanic subsample. We focus on these subgroups because they together comprise 90 percent of our analytic sample (Table 1).

Each column reports the results for a different regression specification. All specifications include student- and neighborhood-level characteristics as baseline controls (listed in the footnotes to Tables 3-6, as well as summarized in Table 1); however, different sets of fixed effects are included in Columns (1)-(3). Specifically, the column layouts are as follows: Column (1) contains high school-by-semester fixed effects, Column (2) replaces the previous set of fixed effects with middle school-by-high school-by-semester fixed effects, and Column (3) includes only teacher-by-semester fixed effects.[11] Since teachers have discretion in determining course grades, for the course passing results we focus on the empirical specification in Column (3) because it controls for teacher fixed effects. For the test score results we focus on the empirical specification in Column (2) because it contains high school fixed effects and also controls for

[11] The numbers of observations are different across columns for a given sample because singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing numbers of observations to those shown.

unobserved characteristics of the student's middle school, therefore making the specification in Column (2) more conservative than that in Column (1).

First-Stage Results

Before turning to our estimates of equation (1), we provide further evidence for the validity of our excluded instruments. Table 2C presents estimates of first-stage equation (3) using the full student samples and teacher-by-semester fixed effects. The first two columns of Table 2C display the coefficients on period of the day and period of the day squared using the full regular English I sample from the course passing regression (first column) and from the reading test score regression (second column). These first-stage coefficients correspond to the 2SLS estimates in the top row of Column (3) in each panel of Table 3. The third and fourth columns of Table 2C are for the full regular Algebra I sample (corresponding to Table 4), the fifth and sixth columns are for the full remedial English I sample (corresponding to Table 5), and the last two columns are for the full remedial Algebra I sample (corresponding to Table 6).

In each case, the coefficient on the squared term is negative, implying an estimated inverted-U relationship between period of the day and the behavioral composition of a classroom. The F-statistics for the tests of the joint significance of the excluded instruments are statistically significant at the one percent level and are almost always larger than 10. Consistent with Figure 1, each turning point in the relationship between period of the day and classroom composition (i.e., the particular period of the day at which classroom composition is at its maximum, based on the coefficient estimates for period and period squared) is approximately fourth period.[12] Overall, Table 2C, along with our discussion in the previous section, provides

[12] Each turning point is computed as the negative of the coefficient on the linear term divided by twice the coefficient on the squared term.

convincing support for our use of a quadratic function in the period of day as an instrument for classroom composition.

English I Course Passing and Reading Test Score Results for Regular Classes

We begin with the OLS results in panel A of Table 3, observing in all three columns a very small but mostly statistically significant association between the behavioral composition of a classroom and the probability that a student passes regular English I. In each case, we find that an additional non-disruptive student in attendance on a given school day is associated with a change in the probability of passing English I of less than one half of one percentage point (-0.16 to +0.33 percentage points).

As discussed in Section III, the endogeneity of the behavioral composition of a classroom implies that OLS estimation of equation (1) leads to downwardly biased estimates of the effects of classroom composition on student achievement. As a result, we now turn to our 2SLS estimation results. For each 2SLS regression, we report the F-statistic for the test of the predictive power of the excluded instruments in the first-stage equation. For the full sample and black subsample, we observe large and statistically significant first-stage F-statistics; all are greater than 10. Due to the smaller sample sizes, the first-stage F-statistics for the Hispanic subsample are not as large, but they are still statistically significant at the five percent level or better.

After instrumenting with period of the day and period of the day squared, all of the 2SLS coefficients on classroom composition are larger in magnitude relative to the corresponding OLS coefficients and are statistically significant at the five percent level or better. These findings are consistent with the Lazear framework in that the behavioral composition of a classroom is an important determinant of the likelihood that a student passes regular English I.

For the full sample, we estimate that an additional non-disruptive student in attendance increases the probability of passing English I by 6.36 to 7.26 percentage points. As stated earlier, our preferred model specification is Column (3), which includes teacher-by-semester fixed effects. Relative to the mean passing rate of 76.8 percent, the estimated coefficient of 0.0726 in Column (3) translates into an increase of 9.45 percent (0.0726/0.768=0.0945) in the probability of passing English I. When we break down this effect by race and ethnicity, we find a larger increase in the probability of passing English I for black versus Hispanic students: based on the results in Column (3), an extra non-disruptive student in attendance increases the probability of passing English I by 8.53 percentage points for black students (relative to the mean of 75.3 percent) and by 5.58 percentage points for Hispanic students (relative to the mean of 77.6 percent).

Turning next to the OLS results in panel B of Table 3, we observe in all three specifications a positive and at least marginally significant association between the behavioral composition of a classroom and reading test scores for regular English I students. For the full sample, we find that an additional non-disruptive student in attendance is associated with an increase of 0.0060 to 0.0081 student-level standard deviations in the student's reading score, which is less than one percentile point. The OLS coefficients on classroom composition are larger in magnitude for black students and smaller in magnitude for Hispanic students.

Moving to the 2SLS estimation results, for the full sample and black subsample, we again observe large and statistically significant first-stage F-statistics; all are above 10. The F-statistics for the Hispanic subsample are again smaller, but they are still significant at the five percent level or better. All of the 2SLS coefficients on classroom composition for the full sample and Hispanic subsample are larger in magnitude, as compared to the corresponding OLS coefficients, and are statistically significant at the five percent level (except in Column (3) for the full sample).

Focusing on our preferred model specification in Column (2), which controls for middle school-by-high school-by-semester fixed effects, we estimate that an additional non-disruptive student in attendance leads to a 0.0222 student-level standard deviation increase in reading test scores for the full sample (approximately one half to one percentile point) and a 0.0633 student-level standard deviation increase in reading test scores for Hispanic students (approximately two percentile points). The estimated impacts on reading test scores for black students are smaller in magnitude than the estimated effects for Hispanic students and are statistically insignificant in all model specifications.

Algebra I Course Passing and Mathematics Test Score Results for Regular Classes

We focus on the 2SLS results in this subsection because the OLS estimates for the regular Algebra I sample have approximately the same magnitude as the OLS estimates for the regular English I sample. The course passing results are reported in panel A of Table 4. All of the 2SLS coefficients on classroom composition are positive and larger in magnitude than the analogous OLS coefficients and are statistically significant at the five percent level or better, again in accordance with the Lazear model. For the full sample, we see that an additional non-disruptive student in attendance increases the probability of passing Algebra I by 3.92 to 5.19 percentage points. The estimated coefficient of 0.0392 in Column (3), our preferred specification, implies an increase of 5.41 percent in the probability of passing Algebra I relative to the mean passing rate of 72.4 percent. This percent increase in the probability of passing regular Algebra I is approximately half the magnitude of the corresponding percent increase in the probability of passing regular English I (9.45 percent). Estimates for the black and Hispanic subsamples in Column (3) are roughly the same as the estimates for the full sample: an additional non-disruptive student in attendance increases the probability of passing Algebra I by 4.21 percentage points for black students (relative to the mean of 70.2 percent) and by 4.32 percentage points for Hispanic students (relative to the mean of 74.4 percent).

The 2SLS results for mathematics test scores are shown in panel B of Table 4. While all first-stage F-statistics are statistically significant at the one percent level and are above 20 for the full sample and black subsample, we estimate small and statistically insignificant effects of the behavioral composition of a classroom on mathematics test scores in all samples and specifications. The one exception is the positive and marginally significant effect of classroom composition on mathematics test scores for black students in Column (3) when including teacher-by-semester fixed effects. This estimated effect size of 0.0114 student-level standard deviations is smaller than the statistically significant effect sizes for reading test scores in panel B of Table 3.

Course Passing and Test Score Results for Remedial Classes

Another empirical implication of the Lazear theoretical framework is that the effects of the behavioral composition of a classroom on student achievement should be larger for students enrolled in remedial versus regular courses because students in remedial courses have lower average baseline academic performance. Table 5 presents the results for students enrolled in remedial English I. We again concentrate on the 2SLS estimates.

Starting with the course passing results in panel A of Table 5, we find that all first-stage F-statistics are statistically significant at the one percent level and are greater than 10 for the full sample and black subsample, other than in Column (3) for the black subsample. Each 2SLS coefficient on classroom composition is larger in magnitude than the corresponding OLS coefficient and is always significant at the one percent level. For the full sample in Column (3), we estimate that an additional non-disruptive student in attendance increases the probability of passing English I by 10.50 percentage points. Relative to the mean passing rate of 75.0 percent, the estimated coefficient of 0.1050 translates into an increase of 14.00 percent in the probability of passing English I. This percent increase in the probability of passing remedial English I is approximately 50 percent larger than the magnitude of the analogous percent increase in the probability of passing regular English I (9.45 percent), in line with the Lazear model.

As with the regular English I course passing analysis, we find a larger increase in the probability of passing remedial English I for the black versus Hispanic subsamples: the results in Column (3) imply that an extra non-disruptive student in attendance increases the probability of passing English I by 14.81 percentage points for black students (relative to the mean of 72.4 percent) and by 5.41 percentage points for Hispanic students (relative to the mean of 77.9 percent). For black students, the effects on course passing are larger for remedial versus regular English I, while for Hispanic students the impacts are approximately the same.

The 2SLS results for reading test scores are in panel B of Table 5. While all first-stage F-statistics are statistically significant at the one percent level and are above 10 for the full sample, we estimate small and statistically insignificant effects of classroom composition on reading test scores.

We now discuss the results for students enrolled in remedial Algebra I, which are presented in Table 6. In both panels, the first-stage F-statistics are generally smaller than the corresponding first-stage F-statistics in Tables 3-5, reflecting a weaker relationship between classroom composition and a quadratic function in the period of the day for the remedial Algebra I sample. This is evident from the right graph in panel B of Figure 1. Most first-stage F-statistics are statistically significant at the five percent level or better for the full sample and Hispanic subsample, whereas only the first-stage F-statistics in Column (3) are at least marginally significant for the black subsample.

Focusing on the 2SLS estimates, we begin with the course passing results in panel A of Table 6. For the full sample in Column (3), we estimate that an additional non-disruptive student in attendance increases the probability of passing Algebra I by 6.84 percentage points. Relative to the mean passing rate of 68.5 percent, the estimated coefficient of 0.0684 translates into an increase of 9.99 percent in the probability of passing Algebra I. This percent increase in the probability of passing remedial Algebra I is approximately twice the magnitude of the corresponding percent increase in the probability of passing regular Algebra I (5.41 percent), in agreement with the Lazear model.

We find a larger increase in the probability of passing remedial Algebra I for the black versus Hispanic subsamples: the estimated coefficients in Column (3) imply that an extra non-disruptive student in attendance increases the probability of passing Algebra I by 11.60 percentage points for black students (relative to the mean of 67.3 percent) and by 2.08 percentage points for Hispanic students (relative to the mean of 70.2 percent). The former effect is significant at the five percent level, while the latter effect is significant at the 10 percent level. For black students, the effects on course passing are larger for remedial versus regular Algebra I, while the opposite is true for Hispanic students.

Panel B of Table 6 displays the results for mathematics test scores. In all cases, we estimate statistically insignificant effects of classroom composition on mathematics test scores. The estimated effect sizes are generally small for models with more precisely estimated first-stage relationships.

Comparisons of Effect Sizes

It is instructive to compare our estimated effects of classroom composition on test scores with the estimated effects of class size on test scores from Tennessee's Project STAR class-size randomized experiment, as reported in Finn and Achilles (1990). An important difference between our study and that of Finn and Achilles (1990) is the measurement of the classroom-level variable of interest: our classroom composition measure is a specific type of "effective" class size that is based on the observed attendance records of non-disruptive students in a given class, whereas the class sizes studied in Finn and Achilles (1990) are "roster" class sizes based on the number of students officially enrolled in/assigned to a particular class. We generally estimate positive effects of classroom composition on test scores, while Finn and Achilles (1990) find negative effects of class size on test scores, implying that the analysis below will be a comparison of the magnitudes of the effect sizes across the two studies.

We focus on the effect sizes for our regular English I full sample and Hispanic subsample because, for all other samples, we estimate statistically insignificant effects of classroom composition on test scores when using our preferred 2SLS specification. For the regular English I full sample, the estimated 2SLS coefficient on classroom composition for the reading test score regression is 0.0222 (Column (2) in panel B of Table 3), which implies an effect size of 0.0222 student-level standard deviations for a one-student *increase* in the number of non-disruptive students in attendance. For the regular English I Hispanic subsample, the estimated 2SLS coefficient on classroom composition for the reading test score regression is 0.0633 (Column (2) in panel B of Table 3), which indicates an effect size of 0.0633 student-level standard deviations for a one-student *increase* in the number of non-disruptive students in attendance.

Finn and Achilles (1990) report effect sizes for first grade students, disaggregated by examination subject. Because their reported effect sizes are based on an approximately eight-student reduction in class size (moving from a regular class or a regular class with an aide to a small class), we divide these effect sizes by eight before comparing them with ours. Finn and Achilles (1990) present results for three standardized examinations in reading: the Basic Skills First (BSF) reading, the Stanford Achievement Test (SAT) word study skills, and the SAT reading examinations. For these examinations, the authors find full sample effect sizes of 0.21, 0.22, and 0.23 student-level standard deviations, respectively, for an approximately eight-student reduction in class size (Table 5, page 566). Dividing these effect sizes by eight gives 0.0263, 0.0275, and 0.0288 student-level standard deviations for an approximately one-student *reduction* in class size. These numbers are very similar in magnitude to our full sample effect size of 0.0222 student-level standard deviations for a one-student increase in the number of non-disruptive students in attendance.

Finn and Achilles (1990) also report effect sizes by minority status (Table 6, page 567), finding larger effect sizes for minority students than for white students.[13] For the BSF reading, the SAT word study skills, and the SAT reading examinations, the effect sizes for minority students are 0.35, 0.32, and 0.35 student-level standard deviations, respectively, for an eight-student reduction in class size, translating into effect sizes of 0.0438, 0.0400, and 0.0438 student-level standard deviations for a one-student reduction in class size. These effect sizes are approximately two thirds of the magnitude of our Hispanic subsample effect size of 0.0633 student-level standard deviations for a one-student increase in the number of non-disruptive students in attendance. The effect sizes for white students in Finn and Achilles (1990) are 0.10,

[13] Finn and Achilles (1990) do not provide results for more disaggregated racial or ethnic breakdowns.

0.16, and 0.15 student-level standard deviations, respectively, for an eight-student reduction in class size, translating into 0.0125, 0.0200, and 0.0188 student-level standard deviations for a one-student reduction in class size.

We also compare our estimated full sample effect size for regular English I to Israeli class size effects reported in Angrist and Lavy (1999). Angrist and Lavy (1999) estimate that a one-student reduction in class size leads to a 0.275 point increase in reading test scores for fifth graders (Table IV, Column (2), page 554), translating into an effect size of about 0.0225 student-level standard deviations for a one-student reduction in class size.[14] The fourth-grade effect size is approximately half this magnitude. We see that the fifth-grade effect size is very similar in magnitude to our full sample effect size of 0.0222 student-level standard deviations for a one-student increase in the number of non-disruptive students in attendance.

Conclusion

Using administrative student transcript files from CPS, we analyze empirical implications of the Lazear educational production model. The Lazear framework suggests that the behavioral composition of a classroom is a central determinant of educational attainment, signifying that it should be included as an explanatory variable when estimating an educational production function. To that end, we exploit exogenous variation on course scheduling in CPS to study heterogeneity in the effect of class size on student achievement in reading and mathematics. Most importantly, our research design permits us to explore an underlying mechanism by which class size affects student achievement, the behavioral composition of a classroom.

---

[14] To obtain the 0.0225 student-level standard deviation effect size for a one-student class size reduction, we divided by eight the 0.18 student-level standard deviation effect size for an eight-student class size reduction reported on page 567.

In accordance with the theoretical predictions of the Lazear model, we find that, for students enrolled in regular English I, an additional non-disruptive student in attendance increases the probability of passing English I by 7.26 percentage points and raises a student's own reading test score by 0.0222 student-level standard deviations. The estimated impacts of the behavioral composition of a classroom on course passing for students enrolled in regular Algebra I are smaller than the corresponding effects for students enrolled in regular English I. Also consistent with the Lazear framework, we observe larger overall effects of classroom composition on English I and Algebra I course passing for students enrolled in remedial versus regular classes.

References for Chapter 2

Angrist, Joshua D., and Victor Lavy. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics, 114*(2), 533-575.

Babcock, Philip, and Julian R. Betts. (2009). Reduced-Class Distinctions: Effort, Ability, and the Education Production Function. *Journal of Urban Economics, 65*(3), 314-322.

Chicago Public Schools. (2004, January 28). *High School Promotion Policy* (Chicago Public Schools Policy Manual, Section: 605.1, Board Report: 04-0128-PO1). Retrieved from http://policy.cps.k12.il.us/documents/605.1.pdf.

Cortes, Kalena E., Jesse Bricker, and Chris Rohlfs. (2012). The Role of Specific Subjects in Education Production Functions: Evidence from Morning Classes in Chicago Public High Schools. *The B.E. Journal of Economic Analysis & Policy, 12*(1) (Contributions), Article 27.

Ding, Weili, and Steven F. Lehrer. (2011). Experimental Estimates of the Impacts of Class Size on Test Scores: Robustness and Heterogeneity. *Canadian Labour Market and Skills Researcher Network Working Paper No. 77*.

Eide, Eric, and Mark H. Showalter. (1998). The Effect of School Quality on Student Performance: A Quantile Regression Approach. *Economics Letters, 58*(3), 345-350.

Finn, Jeremy D., and Charles M. Achilles. (1990). Answers and Questions About Class Size: A Statewide Experiment. *American Educational Research Journal, 27*(3), 557-577.

Finn, Jeremy D., Susan B. Gerber, Charles M. Achilles, and Jayne Boyd-Zaharias. (2001). The Enduring Effects of Small Classes. *Teachers College Record, 103*(2), 145-183.

Hoxby, Caroline M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics, 115*(4), 1239-1285.

Jepsen, Christopher, and Steven Rivkin. (2009). Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size. *Journal of Human Resources, 44*(1), 223-250.

Konstantopoulos, Spyros. (2008). Do Small Classes Reduce the Achievement Gap between Low and High Achievers? Evidence from Project STAR. *Elementary School Journal, 108*(4), 275-291.

Krueger, Alan B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics, 114*(2), 497-532.

Krueger, Alan B., and Diane M. Whitmore. (2001). The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR. *Economic Journal*, *111*(468), 1-28.

Lazear, Edward P. (2001). Educational Production. *Quarterly Journal of Economics*, *116*(3), 777-803.

Levin, Jesse. (2001). For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement. *Empirical Economics, 26*(1), 221-246.

Ma, Lingjie, and Roger Koenker. (2006). Quantile Regression Methods for Recursive Structural Equation Models. *Journal of Econometrics, 134*(2), 471-506.

McKee, Graham J., Steven G. Rivkin, and Katharine R. E. Sims. (2010). Disruption, Achievement and the Heterogeneous Benefits of Smaller Classes. *NBER Working Paper No. 15812*.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. (1999). The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment. *Educational Evaluation and Policy Analysis, 21*(2), 127-142.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. (2000). The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment. *American Educational Research Journal*, *37*(1), 123-151.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. (2005). Teachers, Schools, and Academic Achievement. *Econometrica, 73*(2), 417-458.

Segal, Carmit. (2008). Classroom Behavior. *Journal of Human Resources, 43*(4), 783-814.

Word, Elizabeth, John Johnston, Helen Pate Bain, B. DeWayne Fulton, Jayne Boyd-Zaharias, Charles M. Achilles, Martha Nannette Lintz, John Folger, and Carolyn Breda. (1990). *The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Final Summary Report 1985-1990*. Nashville, TN: Tennessee State Department of Education. Retrieved from http://www.heros-inc.org/summary.pdf.

Table 2.1: Summary Statistics for 9th Grade Chicago Public School Students by Course Subject

	Regular English I		Remedial English I		Regular Algebra I		Remedial Algebra I	
Panel A: Outcome Variables								
Course pass rate	0.768	(0.422)	0.750	(0.433)	0.724	(0.447)	0.685	(0.465)
Reading test score (z-score)	0.080	(0.921)	-0.327	(0.718)	-0.008	(0.862)	-0.393	(0.679)
Math test score (z-score)	0.084	(0.912)	-0.274	(0.836)	0.076	(0.880)	-0.424	(0.755)
Panel B: Classroom Characteristic								
Classroom composition	18.111	(6.122)	17.766	(5.873)	18.275	(5.863)	17.565	(5.843)
Panel C: Student Characteristics								
Age	14.225	(0.560)	14.324	(0.554)	14.224	(0.528)	14.319	(0.574)
Male	0.490	(0.500)	0.514	(0.500)	0.496	(0.500)	0.507	(0.500)
White	0.090	(0.286)	0.061	(0.240)	0.083	(0.276)	0.062	(0.242)
Black	0.542	(0.498)	0.550	(0.497)	0.523	(0.499)	0.574	(0.494)
Hispanic	0.342	(0.474)	0.370	(0.483)	0.366	(0.482)	0.350	(0.477)
Asian	0.024	(0.153)	0.017	(0.129)	0.025	(0.157)	0.012	(0.111)
American Indian	0.002	(0.040)	0.001	(0.033)	0.002	(0.039)	0.001	(0.032)
Free or reduced lunch	0.814	(0.390)	0.868	(0.338)	0.841	(0.366)	0.846	(0.361)
Classified as disruptive	0.158	(0.364)	0.104	(0.305)	0.136	(0.343)	0.112	(0.315)
Bilingual education	0.412	(0.585)	0.428	(0.586)	0.450	(0.606)	0.408	(0.600)
Lives with biological parent	0.843	(0.363)	0.802	(0.399)	0.824	(0.381)	0.830	(0.375)
Special education	0.144	(0.351)	0.214	(0.410)	0.148	(0.355)	0.203	(0.402)
8th grade ITBS reading test score (z-score)	0.080	(0.924)	-0.352	(0.783)	0.002	(0.883)	-0.400	(0.765)
8th grade ITBS math test score (z-score)	0.030	(0.928)	-0.314	(0.813)	0.042	(0.888)	-0.545	(0.640)
Panel D: Neighborhood (Census Block Group) Characteristics								
Median family income	31,180	(15,128)	34,119	(16,229)	32,446	(15,781)	32,425	(14,595)
Percent school age (5-18)	0.236	(0.074)	0.242	(0.073)	0.237	(0.074)	0.241	(0.071)
Percent Hispanic	0.259	(0.315)	0.287	(0.333)	0.278	(0.324)	0.270	(0.328)
Percent black	0.487	(0.450)	0.498	(0.448)	0.474	(0.449)	0.514	(0.449)
Mean education	11.916	(1.129)	11.640	(1.357)	11.817	(1.226)	11.734	(1.259)
Percent in poverty	0.249	(0.199)	0.253	(0.189)	0.248	(0.195)	0.251	(0.186)
Observations	237,912		292,136		279,050		134,336	

Source: Chicago Public Schools High School Transcript Data, 1993-94 through 2005-06

Notes: Regular English I (regular Algebra I) represents the sample of students enrolled in only one regular Algebra I (regular English I) course per semester. Remedial English I (remedial Algebra I) represents the sample of students enrolled in at least one remedial English I (remedial Algebra I) course per semester; some students take regular English I (regular Algebra I) in addition to the remedial class. Numbers in parentheses indicate standard deviations.

Table 2.2A: Distribution of 9th Grade Course Offerings by Period of the Day

	Period of the day is …						
Course Subject	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**
English	0.22	0.21	0.21	0.20	0.20	0.20	0.20
Mathematics	0.16	0.16	0.17	0.16	0.16	0.17	0.16
Social Studies	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Science	0.13	0.13	0.13	0.14	0.13	0.14	0.14
Foreign Language	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Shop	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Business	0.05	0.06	0.05	0.06	0.05	0.05	0.05
Vocational	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Art, Music, and Physical Education	0.24	0.25	0.25	0.25	0.26	0.25	0.24
Other	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Source: Chicago Public Schools High School Transcript Data, 1993-94 through 2005-06
Note: The fractions in each column sum to one.

Table 2.2B: Determinants of Period of the Day (Regular English I Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent Variable: Period of the day is …						
	1st	2nd	3rd	4th	5th	6th	7th
Age	-0.0000	-0.0040***	-0.0023*	0.0010	-0.0004	0.0022*	0.0035***
	(0.0011)	(0.0013)	(0.0012)	(0.0011)	(0.0011)	(0.0012)	(0.0013)
Male	0.0004	0.0024	-0.0000	0.0022*	-0.0031**	-0.0033**	0.0021
	(0.0013)	(0.0016)	(0.0014)	(0.0013)	(0.0012)	(0.0014)	(0.0015)
White	-0.0260*	0.0114	0.0187	0.0215	0.0110	-0.0211*	0.0039
	(0.0139)	(0.0146)	(0.0125)	(0.0140)	(0.0121)	(0.0125)	(0.0144)
Black	-0.0221	0.0078	0.0149	0.0229*	0.0139	-0.0182	0.0041
	(0.0138)	(0.0145)	(0.0121)	(0.0138)	(0.0123)	(0.0127)	(0.0141)
Hispanic	-0.0230*	0.0143	0.0168	0.0172	0.0139	-0.0207*	0.0054
	(0.0137)	(0.0147)	(0.0124)	(0.0140)	(0.0120)	(0.0126)	(0.0144)
Asian	-0.0262*	0.0156	0.0176	0.0154	0.0167	-0.0148	-0.0011
	(0.0144)	(0.0151)	(0.0131)	(0.0144)	(0.0128)	(0.0131)	(0.0153)
Free or reduced lunch	-0.0023*	0.0056***	-0.0035**	-0.0023	-0.0020	-0.0019	0.0042***
	(0.0014)	(0.0018)	(0.0016)	(0.0015)	(0.0014)	(0.0015)	(0.0016)
Classified as disruptive	0.0038**	0.0012	-0.0004	0.0034**	-0.0017	-0.0017	-0.0021
	(0.0016)	(0.0017)	(0.0017)	(0.0017)	(0.0016)	(0.0015)	(0.0016)
Bilingual education	-0.0001	-0.0028*	0.0003	0.0026	-0.0012	0.0010	-0.0035**
	(0.0015)	(0.0016)	(0.0018)	(0.0018)	(0.0015)	(0.0014)	(0.0017)
Lives with biological parent	-0.0016	-0.0025	0.0021	-0.0002	-0.0023	0.0025*	0.0021
	(0.0014)	(0.0016)	(0.0018)	(0.0016)	(0.0015)	(0.0014)	(0.0015)
Special education	0.0043	-0.0065	0.0030	0.0026	0.0006	-0.0134*	0.0044
	(0.0058)	(0.0079)	(0.0080)	(0.0080)	(0.0069)	(0.0069)	(0.0063)
Special education x year trend	-0.0009	0.0003	0.0013	-0.0010	0.0012	0.0015	-0.0014
	(0.0009)	(0.0012)	(0.0012)	(0.0011)	(0.0010)	(0.0010)	(0.0010)
8th grade ITBS reading z-score	0.0052***	0.0004	0.0014	0.0024	-0.0039**	0.0013	-0.0059***
	(0.0015)	(0.0018)	(0.0018)	(0.0016)	(0.0016)	(0.0017)	(0.0015)
Neighborhood med. Fam. Inc.	-0.0010	0.0009	-0.0003	-0.0000	0.0004	-0.0002	-0.0002
	(0.0007)	(0.0009)	(0.0008)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Neighborhood pct. sch age 5-18	-0.0254***	0.0044	0.0189	-0.0101	-0.0054	-0.0083	0.0250**
	(0.0096)	(0.0124)	(0.0118)	(0.0100)	(0.0099)	(0.0110)	(0.0121)
Neighborhood percent Hispanic	0.0074*	0.0003	0.0059	-0.0019	-0.0126***	-0.0036	-0.0014
	(0.0041)	(0.0047)	(0.0049)	(0.0047)	(0.0043)	(0.0045)	(0.0051)
Neighborhood percent black	0.0038	0.0008	0.0049	-0.0082**	-0.0075**	0.0016	-0.0020
	(0.0036)	(0.0039)	(0.0040)	(0.0036)	(0.0033)	(0.0038)	(0.0042)
Neighborhood mean education	0.0006	-0.0013	0.0017*	-0.0005	0.0001	-0.0009	-0.0003
	(0.0009)	(0.0010)	(0.0010)	(0.0009)	(0.0009)	(0.0009)	(0.0010)
Neighborhood pct. in poverty	-0.0000	0.0070	-0.0011	0.0036	0.0009	-0.0046	-0.0050
	(0.0051)	(0.0061)	(0.0057)	(0.0050)	(0.0048)	(0.0052)	(0.0052)
Observations	235,853	235,853	235,853	235,853	235,853	235,853	235,853

Notes: Each specification includes teacher-by-semester fixed effects. Neighborhood median family income is measured in \$10,000s. Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample size to that shown. Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 2.2C: First-Stage Results for Full Sample Column (3) Regression Specifications of Tables 3-6

	Table 3: Regular English I		Table 4: Regular Algebra I		Table 5: Remedial English I		Table 6: Remedial Algebra I	
	Panel A: Course Passing	Panel B: Reading Test Score	Panel A: Course Passing	Panel B: Math Test Score	Panel A: Course Passing	Panel B: Reading Test Score	Panel A: Course Passing	Panel B: Math Test Score
Period	0.5315***	0.5089***	0.7497***	0.9204***	0.4250***	0.4389***	0.5410***	0.4874***
	(0.0977)	(0.1004)	(0.0985)	(0.1109)	(0.0907)	(0.0945)	(0.1339)	(0.1393)
Period^2	-0.0678***	-0.0646***	-0.0911***	-0.1121***	-0.0497***	-0.0514***	-0.0661***	-0.0591***
	(0.0116)	(0.0119)	(0.0118)	(0.0132)	(0.0106)	(0.0110)	(0.0157)	(0.0164)
F-Statistic	18.18***	15.45***	29.80***	35.96***	11.08***	10.92***	8.86***	6.53***
Turning Point	3.92	3.94	4.11	4.11	4.28	4.27	4.10	4.12
Observations	235,853	202,158	278,624	195,980	291,037	250,663	133,957	112,120

Notes: Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample sizes to those shown. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The first-stage F-statistic is the F-statistic on the excluded instruments for classroom composition. The turning point is the period of the day at which classroom composition is at its maximum, based on the coefficient estimates for Period and Period^2.

Table 2.3: OLS and 2SLS Results for Course Passing and Reading Test Scores for Regular English I Sample

	Sample Mean	(1) OLS	(1) 2SLS	(2) OLS	(2) 2SLS	(3) OLS	(3) 2SLS
		Panel A: English I Course Passing					
Full Student Sample							
Classroom Composition	0.768	0.0001	0.0657***	-0.0005	0.0636***	0.0028***	0.0726***
		(0.0003)	(0.0113)	(0.0004)	(0.0117)	(0.0004)	(0.0129)
First-Stage F-Statistic		21.55***		19.89***		18.18***	
Observations		237,912		204,976		235,853	
Black Student Subsample							
Classroom Composition	0.753	0.0009**	0.0760***	0.0003	0.0747***	0.0033***	0.0853***
		(0.0004)	(0.0145)	(0.0005)	(0.0147)	(0.0005)	(0.0181)
First-Stage F-Statistic		18.25***		18.17***		12.50***	
Observations		129,042		102,867		127,096	
Hispanic Student Subsample							
Classroom Composition	0.776	-0.0013***	0.0554***	-0.0016***	0.0320**	0.0020***	0.0558***
		(0.0005)	(0.0202)	(0.0005)	(0.0163)	(0.0006)	(0.0171)
First-Stage F-Statistic		5.00***		4.19**		6.92***	
Observations		81,260		71,404		79,900	
		Panel B: Reading Test Score					
Full Student Sample							
Classroom Composition		0.0081***	0.0246**	0.0075***	0.0222**	0.0060***	0.0142
		(0.0007)	(0.0106)	(0.0007)	(0.0112)	(0.0010)	(0.0111)
First-Stage F-Statistic		16.59***		15.44***		15.45***	
Observations		203,926		174,200		202,158	
Black Student Subsample							
Classroom Composition		0.0099***	0.0097	0.0097***	0.0089	0.0083***	0.0159
		(0.0008)	(0.0128)	(0.0009)	(0.0130)	(0.0012)	(0.0141)
First-Stage F-Statistic		13.75***		13.43***		10.15***	
Observations		108,152		84,758		106,482	
Hispanic Student Subsample							
Classroom Composition		0.0052***	0.0670**	0.0052***	0.0633**	0.0025*	0.0413**
		(0.0010)	(0.0281)	(0.0012)	(0.0290)	(0.0015)	(0.0205)
First-Stage F-Statistic		4.81***		4.45**		6.66***	
Observations		71,556		62,423		70,415	
Fixed Effects:							
High School-by-Semester		Yes	Yes	-	-	-	-
Middle School-by-High School-by-Semester		-	-	Yes	Yes	-	-
Teacher-by-Semester		-	-	-	-	Yes	Yes

Notes: The samples include students enrolled in only one regular English I course per semester. The sample mean is the mean course pass rate for the sample in Column (1). Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample sizes to those shown. All regressions include student controls (age, male, race/ethnicity, free or reduced lunch, classified as disruptive, bilingual education, guardian status, special education, special education x year trend, 8th grade ITBS score, and enrolled in 8th+ period) and neighborhood controls (median family income, percent school age (5-18), percent Hispanic, percent black, mean education, and percent in poverty). Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The first-stage F-statistic is the F-statistic on the excluded instruments for classroom composition.

Table 2.4: OLS and 2SLS Results for Course Passing and Math Test Scores for Regular Algebra I Sample

	Sample Mean	(1) OLS	(1) 2SLS	(2) OLS	(2) 2SLS	(3) OLS	(3) 2SLS
Panel A: Algebra I Course Passing							
Full Student Sample							
Classroom Composition	0.724	-0.0020***	0.0518***	-0.0023***	0.0519***	0.0009**	0.0392***
		(0.0003)	(0.0097)	(0.0004)	(0.0103)	(0.0004)	(0.0064)
First-Stage F-Statistic		20.49***		18.99***		29.80***	
Observations		279,050		242,419		278,624	
Black Student Subsample							
Classroom Composition	0.702	-0.0020***	0.0502***	-0.0027***	0.0502***	0.0009*	0.0421***
		(0.0005)	(0.0110)	(0.0006)	(0.0119)	(0.0005)	(0.0075)
First-Stage F-Statistic		16.81***		15.95***		25.62***	
Observations		146,064		116,921		145,302	
Hispanic Student Subsample							
Classroom Composition	0.744	-0.0022***	0.0672***	-0.0023***	0.0609***	0.0002	0.0432**
		(0.0005)	(0.0225)	(0.0005)	(0.0227)	(0.0006)	(0.0174)
First-Stage F-Statistic		5.88***		4.98***		5.61***	
Observations		102,093		91,120		101,495	
Panel B: Math Test Score							
Full Student Sample							
Classroom Composition		0.0054***	0.0037	0.0044***	0.0053	0.0040***	0.0068
		(0.0005)	(0.0059)	(0.0005)	(0.0071)	(0.0007)	(0.0054)
First-Stage F-Statistic		25.33***		21.37***		35.96***	
Observations		196,369		168,071		195,980	
Black Student Subsample							
Classroom Composition		0.0057***	0.0073	0.0048***	0.0076	0.0042***	0.0114*
		(0.0007)	(0.0063)	(0.0008)	(0.0077)	(0.0009)	(0.0060)
First-Stage F-Statistic		26.69***		22.74***		35.44***	
Observations		100,741		78,481		100,078	
Hispanic Student Subsample							
Classroom Composition		0.0047***	0.0010	0.0036***	0.0110	0.0031***	-0.0050
		(0.0006)	(0.0127)	(0.0007)	(0.0151)	(0.0010)	(0.0146)
First-Stage F-Statistic		6.60***		5.66***		6.15***	
Observations		73,238		64,640		72,749	
Fixed Effects:							
High School-by-Semester		Yes	Yes	-	-	-	-
Middle School-by-High School-by-Semester		-	-	Yes	Yes	-	-
Teacher-by-Semester		-	-	-	-	Yes	Yes

Notes: The samples include students enrolled in only one regular Algebra I course per semester. The sample mean is the mean course pass rate for the sample in Column (1). Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample sizes to those shown. All regressions include student controls (age, male, race/ethnicity, free or reduced lunch, classified as disruptive, bilingual education, guardian status, special education, special education x year trend, 8th grade ITBS score, and enrolled in 8th+ period) and neighborhood controls (median family income, percent school age (5-18), percent Hispanic, percent black, mean education, and percent in poverty). Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The first-stage F-statistic is the F-statistic on the excluded instruments for classroom composition.

Table 2.5: OLS and 2SLS Results for Course Passing and Reading Test Scores for Remedial English I Sample

	Sample Mean	(1) OLS	(1) 2SLS	(2) OLS	(2) 2SLS	(3) OLS	(3) 2SLS
		Panel A: English I Course Passing					
Full Student Sample							
Classroom Composition	0.750	-0.0004	0.0819***	-0.0011***	0.0828***	0.0031***	0.1050***
		(0.0004)	(0.0151)	(0.0004)	(0.0153)	(0.0004)	(0.0226)
First-Stage F-Statistic		16.65***		16.78***		11.08***	
Observations		292,136		283,765		291,037	
Black Student Subsample							
Classroom Composition	0.724	0.0009*	0.1086***	0.0002	0.1069***	0.0038***	0.1481***
		(0.0004)	(0.0250)	(0.0005)	(0.0246)	(0.0005)	(0.0434)
First-Stage F-Statistic		10.25***		10.57***		5.84***	
Observations		160,726		154,352		159,494	
Hispanic Student Subsample							
Classroom Composition	0.779	-0.0025***	0.0375***	-0.0029***	0.0427***	0.0014**	0.0541***
		(0.0005)	(0.0110)	(0.0005)	(0.0123)	(0.0007)	(0.0167)
First-Stage F-Statistic		9.76***		8.49***		6.84***	
Observations		108,173		105,560		107,200	
		Panel B: Reading Test Score					
Full Student Sample							
Classroom Composition		0.0038***	-0.0061	0.0025***	-0.0057	0.0024***	0.0013
		(0.0005)	(0.0073)	(0.0004)	(0.0068)	(0.0006)	(0.0088)
First-Stage F-Statistic		14.82***		14.29***		10.92***	
Observations		251,668		244,362		250,663	
Black Student Subsample							
Classroom Composition		0.0037***	-0.0054	0.0017***	-0.0065	0.0023***	0.0107
		(0.0006)	(0.0089)	(0.0005)	(0.0080)	(0.0008)	(0.0125)
First-Stage F-Statistic		8.38***		8.54***		5.21***	
Observations		134,674		129,195		133,541	
Hispanic Student Subsample							
Classroom Composition		0.0030***	0.0020	0.0026***	0.0052	0.0012	-0.0037
		(0.0007)	(0.0096)	(0.0007)	(0.0100)	(0.0010)	(0.0134)
First-Stage F-Statistic		9.45***		7.40***		7.14***	
Observations		96,814		94,526		95,908	
Fixed Effects:							
High School-by-Semester		Yes	Yes	-	-	-	-
Middle School-by-High School-by-Semester		-	-	Yes	Yes	-	-
Teacher-by-Semester		-	-	-	-	Yes	Yes

Notes: The samples include students enrolled in at least one remedial English I course per semester; some students take regular English I in addition to the remedial class. The sample mean is the mean course pass rate for the sample in Column (1). Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample sizes to those shown. All regressions include student controls (age, male, race/ethnicity, free or reduced lunch, classified as disruptive, bilingual education, guardian status, special education, special education x year trend, 8th grade ITBS score, and enrolled in 8th+ period) and neighborhood controls (median family income, percent school age (5-18), percent Hispanic, percent black, mean education, and percent in poverty). Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The first-stage F-statistic is the F-statistic on the excluded instruments for classroom composition.

Table 2.6: OLS and 2SLS Results for Course Passing and Math Test Scores for Remedial Algebra I Sample

	Sample Mean	(1) OLS	(1) 2SLS	(2) OLS	(2) 2SLS	(3) OLS	(3) 2SLS
		Panel A: Algebra I Course Passing					
Full Student Sample							
Classroom Composition	0.685	-0.0003	0.1158***	-0.0007	0.1056***	0.0014**	0.0684***
		(0.0007)	(0.0423)	(0.0007)	(0.0390)	(0.0007)	(0.0175)
First-Stage F-Statistic		4.29**		4.42**		8.86***	
Observations		134,336		124,070		133,957	
Black Student Subsample							
Classroom Composition	0.673	0.0005	0.2553	0.0006	0.2669	0.0018**	0.1160**
		(0.0009)	(0.1983)	(0.0009)	(0.2491)	(0.0008)	(0.0457)
First-Stage F-Statistic		0.84		0.58		3.19**	
Observations		77,123		68,885		76,658	
Hispanic Student Subsample							
Classroom Composition	0.702	-0.0012	0.0424**	-0.0012	0.0421***	0.0010	0.0208*
		(0.0009)	(0.0172)	(0.0009)	(0.0163)	(0.0011)	(0.0113)
First-Stage F-Statistic		7.38***		8.65***		9.64***	
Observations		46,948		43,880		46,646	
		Panel B: Math Test Score					
Full Student Sample							
Classroom Composition		0.0033***	0.0320	0.0027***	0.0029	0.0010	-0.0123
		(0.0007)	(0.0254)	(0.0007)	(0.0201)	(0.0008)	(0.0138)
First-Stage F-Statistic		2.53*		2.97*		6.53***	
Observations		112,447		103,586		112,120	
Black Student Subsample							
Classroom Composition		0.0041***	0.0491	0.0028***	-0.0326	0.0013	-0.0149
		(0.0008)	(0.0721)	(0.0009)	(0.0596)	(0.0010)	(0.0232)
First-Stage F-Statistic		0.44		0.58		2.51*	
Observations		64,619		57,541		64,207	
Hispanic Student Subsample							
Classroom Composition		0.0022**	0.0183	0.0018	0.0185	-0.0006	-0.0050
		(0.0011)	(0.0197)	(0.0011)	(0.0168)	(0.0014)	(0.0149)
First-Stage F-Statistic		5.25***		6.45***		7.19***	
Observations		39,420		36,726		39,117	
Fixed Effects:							
High School-by-Semester		Yes	Yes	-	-	-	-
Middle School-by-High School-by-Semester		-	-	Yes	Yes	-	-
Teacher-by-Semester		-	-	-	-	Yes	Yes

Notes: The samples include students enrolled in at least one remedial Algebra I course per semester; some students take regular Algebra I in addition to the remedial class. The sample mean is the mean course pass rate for the sample in Column (1). Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample sizes to those shown. All regressions include student controls (age, male, race/ethnicity, free or reduced lunch, classified as disruptive, bilingual education, guardian status, special education, special education x year trend, 8th grade ITBS score, and enrolled in 8th+ period) and neighborhood controls (median family income, percent school age (5-18), percent Hispanic, percent black, mean education, and percent in poverty). Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The first-stage F-statistic is the F-statistic on the excluded instruments for classroom composition.

Appendix II: Additional Regressions for Testing Instrument Validity

Table 2.A1: Determinants of Period of the Day (Regular Algebra I Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent Variable: Period of the day is …						
	1st	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**
Age	-0.0022**	0.0030**	0.0016	-0.0020**	-0.0004	-0.0003	0.0005
	(0.0009)	(0.0012)	(0.0011)	(0.0010)	(0.0009)	(0.0010)	(0.0011)
Male	0.0011	0.0031**	-0.0010	0.0015	-0.0008	-0.0020	-0.0020
	(0.0012)	(0.0014)	(0.0012)	(0.0013)	(0.0011)	(0.0012)	(0.0013)
White	0.0040	-0.0246*	0.0118	-0.0042	-0.0029	-0.0074	0.0215
	(0.0095)	(0.0141)	(0.0124)	(0.0125)	(0.0117)	(0.0123)	(0.0136)
Black	0.0042	-0.0271*	0.0120	-0.0109	-0.0004	-0.0078	0.0247*
	(0.0096)	(0.0141)	(0.0122)	(0.0123)	(0.0116)	(0.0119)	(0.0134)
Hispanic	0.0021	-0.0253*	0.0098	-0.0061	0.0004	-0.0094	0.0230*
	(0.0096)	(0.0140)	(0.0123)	(0.0124)	(0.0118)	(0.0123)	(0.0134)
Asian	0.0008	-0.0165	0.0042	-0.0104	-0.0014	-0.0054	0.0237*
	(0.0098)	(0.0150)	(0.0125)	(0.0126)	(0.0120)	(0.0128)	(0.0144)
Free or reduced lunch	0.0063***	-0.0018	0.0035**	-0.0001	-0.0011	-0.0035**	-0.0025*
	(0.0014)	(0.0017)	(0.0016)	(0.0015)	(0.0014)	(0.0015)	(0.0014)
Classified as disruptive	0.0005	-0.0029	-0.0006	-0.0043***	0.0037**	-0.0013	0.0024
	(0.0016)	(0.0018)	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0017)
Bilingual education	-0.0005	0.0004	-0.0003	-0.0007	0.0012	0.0010	0.0022
	(0.0015)	(0.0016)	(0.0015)	(0.0017)	(0.0015)	(0.0015)	(0.0016)
Lives with biological parent	0.0021*	0.0003	0.0001	0.0016	0.0000	-0.0007	-0.0024*
	(0.0011)	(0.0015)	(0.0014)	(0.0014)	(0.0012)	(0.0013)	(0.0014)
Special education	-0.0044	-0.0023	0.0137*	-0.0097	0.0028	0.0171**	-0.0077
	(0.0062)	(0.0078)	(0.0080)	(0.0072)	(0.0066)	(0.0077)	(0.0075)
Special education x year trend	0.0003	0.0009	-0.0017*	0.0017*	0.0005	-0.0017	-0.0001
	(0.0009)	(0.0010)	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0010)
8th grade ITBS math z-score	-0.0046***	-0.0013	0.0017	-0.0026*	0.0033**	0.0031**	-0.0017
	(0.0014)	(0.0017)	(0.0018)	(0.0016)	(0.0015)	(0.0015)	(0.0017)
Neighborhood median family income	-0.0002	0.0005	0.0008	-0.0015**	-0.0009	0.0001	0.0009
	(0.0006)	(0.0006)	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Neighborhood percent school age (5-18)	-0.0087	-0.0184*	-0.0163	0.0156	-0.0126	0.0341***	-0.0059
	(0.0095)	(0.0111)	(0.0105)	(0.0108)	(0.0095)	(0.0096)	(0.0103)
Neighborhood percent Hispanic	-0.0016	0.0043	0.0097**	0.0001	0.0075*	-0.0043	-0.0100**
	(0.0035)	(0.0047)	(0.0047)	(0.0040)	(0.0041)	(0.0042)	(0.0044)
Neighborhood percent black	-0.0003	0.0057	0.0087**	0.0042	0.0026	-0.0099***	-0.0036
	(0.0032)	(0.0038)	(0.0038)	(0.0033)	(0.0037)	(0.0033)	(0.0036)
Neighborhood mean education	0.0003	0.0003	-0.0001	0.0004	0.0008	0.0011	-0.0028***
	(0.0007)	(0.0010)	(0.0009)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Neighborhood percent in poverty	0.0028	0.0052	0.0056	-0.0121**	-0.0059	-0.0050	0.0023
	(0.0048)	(0.0053)	(0.0054)	(0.0051)	(0.0043)	(0.0044)	(0.0050)
Observations	278,624	278,624	278,624	278,624	278,624	278,624	278,624

Notes: Each specification includes teacher-by-semester fixed effects. Neighborhood median family income is measured in $10,000s. Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample size to that shown. Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 2.A2: Determinants of Period of the Day (Remedial English I Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent Variable: Period of the day is …						
	1st	2nd	3rd	4th	5th	6th	7th
Age	0.0005	0.0015	0.0004	-0.0014	-0.0017*	0.0011	0.0030***
	(0.0008)	(0.0011)	(0.0009)	(0.0009)	(0.0009)	(0.0010)	(0.0009)
Male	-0.0007	-0.0010	0.0019	-0.0019	0.0010	0.0004	0.0018
	(0.0013)	(0.0013)	(0.0013)	(0.0012)	(0.0011)	(0.0012)	(0.0013)
White	0.0126	-0.0178	0.0399***	-0.0074	-0.0130	-0.0053	0.0254*
	(0.0121)	(0.0130)	(0.0137)	(0.0158)	(0.0129)	(0.0158)	(0.0138)
Black	0.0117	-0.0190	0.0304**	-0.0041	-0.0127	-0.0029	0.0271*
	(0.0124)	(0.0130)	(0.0141)	(0.0156)	(0.0129)	(0.0156)	(0.0140)
Hispanic	0.0095	-0.0162	0.0405***	-0.0057	-0.0104	-0.0068	0.0229
	(0.0120)	(0.0129)	(0.0139)	(0.0156)	(0.0128)	(0.0156)	(0.0140)
Asian	0.0127	-0.0119	0.0473***	-0.0007	-0.0156	-0.0135	0.0192
	(0.0123)	(0.0137)	(0.0155)	(0.0160)	(0.0135)	(0.0159)	(0.0141)
Free or reduced lunch	0.0011	0.0009	0.0006	-0.0002	-0.0005	0.0016	-0.0024
	(0.0014)	(0.0017)	(0.0015)	(0.0014)	(0.0014)	(0.0014)	(0.0016)
Classified as disruptive	-0.0007	0.0022	-0.0005	-0.0001	-0.0001	0.0006	-0.0004
	(0.0017)	(0.0018)	(0.0019)	(0.0016)	(0.0017)	(0.0016)	(0.0018)
Bilingual education	0.0017	0.0016	-0.0043***	-0.0001	-0.0022	-0.0009	0.0029*
	(0.0012)	(0.0014)	(0.0015)	(0.0014)	(0.0016)	(0.0015)	(0.0016)
Lives with biological parent	0.0007	0.0006	-0.0024*	0.0011	-0.0050***	0.0007	0.0023*
	(0.0011)	(0.0013)	(0.0013)	(0.0012)	(0.0013)	(0.0012)	(0.0012)
Special education	0.0001	-0.0004	-0.0045	-0.0017	-0.0081	0.0090	-0.0009
	(0.0051)	(0.0070)	(0.0062)	(0.0061)	(0.0052)	(0.0066)	(0.0054)
Special education x year trend	0.0003	-0.0001	0.0003	0.0005	0.0003	-0.0006	-0.0002
	(0.0005)	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0006)
8th grade ITBS reading z-score	0.0004	-0.0006	0.0009	-0.0004	0.0015	-0.0001	-0.0022*
	(0.0010)	(0.0012)	(0.0011)	(0.0012)	(0.0010)	(0.0011)	(0.0012)
Neighborhood median family income	0.0004	0.0010	-0.0005	0.0016***	-0.0012**	-0.0005	-0.0002
	(0.0005)	(0.0007)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Neighborhood percent school age (5-18)	-0.0144	-0.0059	0.0234**	0.0064	0.0050	0.0095	0.0265***
	(0.0088)	(0.0103)	(0.0101)	(0.0094)	(0.0090)	(0.0102)	(0.0096)
Neighborhood percent Hispanic	0.0037	0.0082*	-0.0041	0.0056	0.0007	-0.0126***	-0.0036
	(0.0032)	(0.0045)	(0.0046)	(0.0040)	(0.0043)	(0.0043)	(0.0043)
Neighborhood percent black	0.0023	0.0118***	0.0053	0.0005	-0.0009	-0.0130***	-0.0022
	(0.0029)	(0.0037)	(0.0041)	(0.0035)	(0.0036)	(0.0037)	(0.0035)
Neighborhood mean education	-0.0010	-0.0005	0.0008	0.0005	0.0007	-0.0006	-0.0001
	(0.0006)	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Neighborhood percent in poverty	0.0082*	0.0040	-0.0009	-0.0012	0.0073*	-0.0040	-0.0131***
	(0.0045)	(0.0057)	(0.0052)	(0.0046)	(0.0044)	(0.0046)	(0.0044)
Observations	291,037	291,037	291,037	291,037	291,037	291,037	291,037

Notes: Each specification includes teacher-by-semester fixed effects. Neighborhood median family income is measured in $10,000s. Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample size to that shown. Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

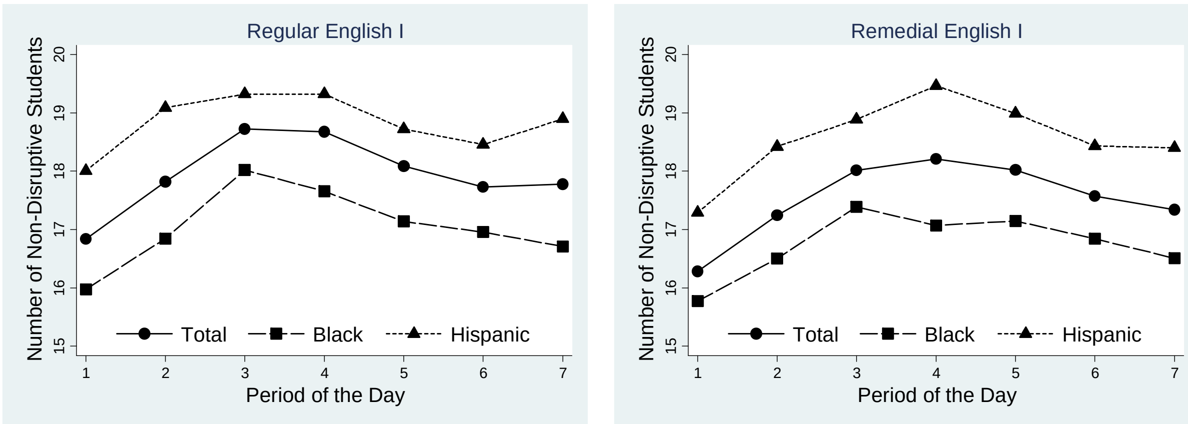
Table 2.A3: Determinants of Period of the Day (Remedial Algebra I Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent Variable: Period of the day is …						
	1st	2nd	3rd	4th	5th	6th	7th
Age	0.0017	-0.0026*	-0.0022*	0.0005	-0.0001	-0.0022	0.0029**
	(0.0012)	(0.0014)	(0.0013)	(0.0012)	(0.0014)	(0.0013)	(0.0012)
Male	-0.0010	0.0013	-0.0003	-0.0035**	-0.0012	0.0018	0.0001
	(0.0017)	(0.0015)	(0.0016)	(0.0017)	(0.0018)	(0.0016)	(0.0019)
White	0.0236	-0.0028	-0.0301	0.0102	-0.0173	0.0197	-0.0256
	(0.0184)	(0.0217)	(0.0245)	(0.0146)	(0.0242)	(0.0153)	(0.0195)
Black	0.0228	-0.0067	-0.0227	0.0046	-0.0149	0.0130	-0.0226
	(0.0185)	(0.0222)	(0.0240)	(0.0139)	(0.0238)	(0.0156)	(0.0192)
Hispanic	0.0221	-0.0032	-0.0223	0.0058	-0.0160	0.0203	-0.0293
	(0.0186)	(0.0215)	(0.0243)	(0.0141)	(0.0235)	(0.0154)	(0.0192)
Asian	0.0235	-0.0034	-0.0318	0.0172	-0.0155	0.0099	-0.0245
	(0.0188)	(0.0220)	(0.0240)	(0.0150)	(0.0242)	(0.0152)	(0.0207)
Free or reduced lunch	-0.0033	0.0007	0.0019	-0.0058***	0.0022	-0.0022	0.0032
	(0.0020)	(0.0023)	(0.0021)	(0.0019)	(0.0021)	(0.0020)	(0.0022)
Classified as disruptive	-0.0037	0.0050**	0.0021	-0.0017	-0.0003	0.0041*	-0.0046*
	(0.0028)	(0.0026)	(0.0028)	(0.0024)	(0.0029)	(0.0024)	(0.0027)
Bilingual education	-0.0024	-0.0001	0.0015	0.0012	0.0011	-0.0015	-0.0022
	(0.0018)	(0.0020)	(0.0021)	(0.0017)	(0.0023)	(0.0021)	(0.0019)
Lives with biological parent	-0.0019	0.0014	0.0044**	0.0003	0.0011	0.0052***	-0.0054***
	(0.0016)	(0.0018)	(0.0017)	(0.0017)	(0.0018)	(0.0016)	(0.0019)
Special education	-0.0027	0.0119*	-0.0159**	0.0070	0.0001	0.0023	-0.0121*
	(0.0057)	(0.0063)	(0.0068)	(0.0061)	(0.0050)	(0.0044)	(0.0066)
Special education x year trend	0.0004	-0.0007	0.0021***	-0.0003	0.0004	-0.0007	0.0008
	(0.0006)	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0006)	(0.0007)
8th grade ITBS math z-score	-0.0023	-0.0014	0.0012	-0.0011	-0.0007	0.0007	0.0022
	(0.0017)	(0.0019)	(0.0018)	(0.0015)	(0.0018)	(0.0015)	(0.0019)
Neighborhood median family income	0.0006	-0.0006	0.0008	-0.0008	-0.0004	0.0004	-0.0004
	(0.0008)	(0.0010)	(0.0009)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Neighborhood percent school age (5-18)	0.0214	-0.0169	0.0045	0.0254*	-0.0325**	-0.0234*	0.0098
	(0.0133)	(0.0156)	(0.0140)	(0.0134)	(0.0133)	(0.0123)	(0.0141)
Neighborhood percent Hispanic	0.0011	-0.0058	-0.0074	-0.0098**	0.0131**	-0.0002	0.0105*
	(0.0050)	(0.0064)	(0.0062)	(0.0050)	(0.0065)	(0.0055)	(0.0061)
Neighborhood percent black	0.0001	0.0079	-0.0076	-0.0053	0.0033	-0.0041	0.0043
	(0.0046)	(0.0049)	(0.0051)	(0.0043)	(0.0051)	(0.0053)	(0.0050)
Neighborhood mean education	0.0013	-0.0011	-0.0008	-0.0012	0.0008	-0.0006	0.0010
	(0.0011)	(0.0012)	(0.0011)	(0.0009)	(0.0011)	(0.0010)	(0.0010)
Neighborhood percent in poverty	-0.0039	-0.0054	0.0045	-0.0077	0.0020	0.0179***	-0.0020
	(0.0078)	(0.0079)	(0.0071)	(0.0073)	(0.0070)	(0.0064)	(0.0071)
Observations	133,957	133,957	133,957	133,957	133,957	133,957	133,957

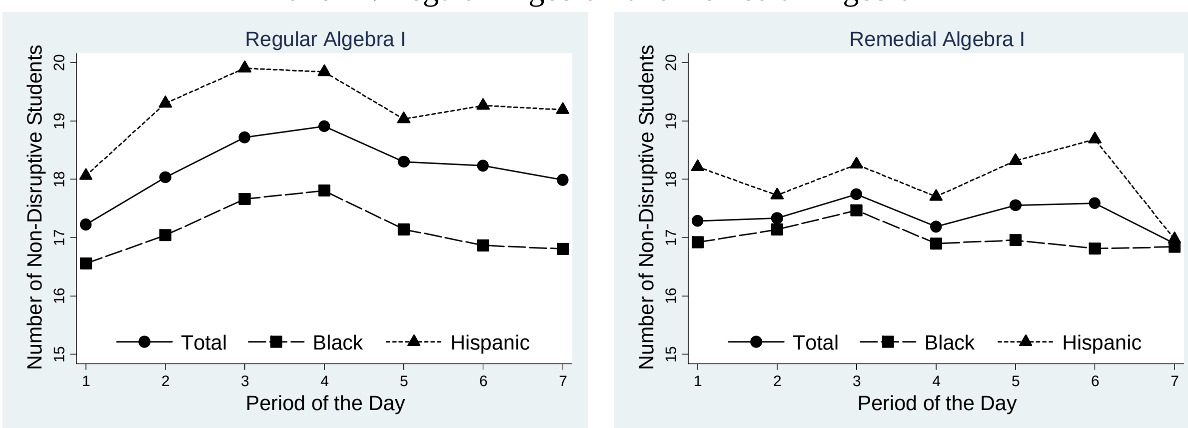
Notes: Each specification includes teacher-by-semester fixed effects. Neighborhood median family income is measured in $10,000s. Singleton groups (i.e., fixed effects with exactly one observation) were dropped in the estimation, reducing the sample size to that shown. Numbers in parentheses represent standard errors clustered at the high school-by-semester level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Figure 2.1: Average Number of Non-Disruptive Students in Attendance by Period

Panel A: Regular English I and Remedial English I



Panel B: Regular Algebra I and Remedial Algebra I



Notes: Figures represent the average number of non-disruptive students in attendance in a given classroom during a given period of the day for regular English I, regular Algebra I, and their respective remedial courses.

Chapter 3. The Effects of September 11, 2001 on the Academic Achievement of Arab Students: Evidence from New York City Public Schools

Introduction

On September 11, 2001 the City of New York suffered one of the worst terrorist attacks in the nation's history. These unanticipated attacks claimed almost 3,000 lives and 100s of billions of dollars of economic damage. Following the attacks, anti-Islamic sentiment significantly increased within the United States. In 2003, an ABC poll reported over 1/3 of respondents believed Islam promoted violence against the West. A Pew Research Center report from the same year released similar findings, 49 percent of Americans believed a significant portion of Muslims held anti-American views.

This increase in anti-Islamic sentiment had a broad reaching effect on the Arab American community, because of the close association with the ethnicity of the 9/11 terrorists and the community's connection to Islam. Kaushal et al. (2007) estimate that post 9/11, first and second generation immigrants from Arab and Muslim countries[15] experienced a 10 percent decline in real wages and biweekly earnings relative to non-Muslim comparison groups. Using a similar methodology and the same treatment group, Mora and Davila (2005) find a systematic decline in the wages of Arab immigrants shortly between 2000 and 2002. Property values within 1,000 feet of a Mosque decreased in values by over 17 percent (Bogin, 2010). In New York City, Arab and Muslim school children were increasingly bullied and harassed. 17 percent of the Arab and Muslim public school students surveyed reported being the object of bigotry, 28 percent reported being stopped by the police as a result of racial profiling, and 11 percent reported being physically assaulted (Cristillo, 2008). The total number of hate crimes against

[15] Among the 23 countries that Kaushal et al. (2007) include in their treatment group, 19 are Arab in addition to Afghanistan, Bangladesh, Indonesia and Pakistan that are predominantly Muslim.

Arabs increase by 1,160 cases in 2001 and by 453 cases against Muslims according to Rubenstein (2004).

In this study we examine the impact of the increase in the anti-Islamic sentiment on the schooling outcomes of Arab students, including performance on standardized tests, grade retention and special education designations. There are a number of different mechanisms through which the events of September 11 may have influenced schooling outcomes. Interpersonal relationships among students may have changed. Non-Arab and non-Muslim students may have become less cooperative with their Arab and Muslim peers (indirect bullying), and the amount of teasing, taunting, threatening, hitting, or name-calling (direct bullying) may have increased. Being a victim of bullying is associated with low self-esteem, self-harm, suicidal intention, depression, lowliness, and physical ill-health (Barker et al. 2008; Beaty and Alexeyev 2008; Fekkes et al. 2006; Ferguson, Beautrais, and Horwood 2003). Brown and Taylor (2008) found that these factors reduce educational attainment for students at the ages of 7, 11, and 16. If Arab and Muslim students were more likely to have been bullied after the September 11 terrorist attacks, then they might suffer more physical and mental health problems, resulting in lower student performance and educational attainment.

Second, students' out-of-school environments may have changed as a result of the September 11 terrorist attacks. Arab students may have been more discriminated against in their neighborhoods and surroundings. Further, friends with similar ethnic and religious background may have experienced prejudice and discrimination. A large body of empirical research has tried to identify the effect of socio-economic environments such as neighborhoods on children and teenagers. There is consensus that youth outcomes such as crime, problem behavior, and health vary dramatically across neighborhoods (Sampson, Morenoff, and Gannon-Rowley 2002). The

September 11 terrorist attacks may have led to environmental changes for Arab students, possibly lowering their academic performance. Given these mechanisms, we expect that the events of 9/11 may lead to a decline in the academic performance of Arab students, including lower test scores and higher retention rates.

Using a natural experiment design and restricted administrative data from New York City Public Schools, we estimate the impact of 9/11 on the educational outcomes of Arab students. Relative to first and second generation immigrants from non-Arab Muslim countries, Arab student test scores declined in the school years following the events of 9/11. Depending on the specification, we find that math scores decline by 0.05-0.10 student level standard deviations and by 0.02-0.08 student level standard deviations for English scores. When we include student fixed effects we identify the effect of 9/11 from Arab students who were present in the data pre- and post- 9/11, we find that grade retention by a percentage point and special education designations increased by 0.7 percentage points relative to non-Arab Muslim students.

Our paper is organized as follows: in the next section we describe the data and student trends; we then move on to a discussion of our empirical strategy; this is followed by a presentation of the results and a conclusion.

Empirical Strategy and Identification

In this study, we wish to estimate the impact of the events of September 11, 2001 on the academic performance of Arab students in a New York City Public School from grades 3 through 8. For this study we define the treatment group as students who were born in an Arab[16] country and students who speak Arabic at home. These countries include Algeria, Bahrain, Comoros, Djibouti, Egypt, Iraq, Jordan, Kuwait, Lebanon, Lybia, Morocco, Mauritania, Oman,

[16] The list of Arab countries is based on the list of member states of the Arab League.

Qatar, Somalia, Saudi Arabia, Sudan, Syria, United Arab Emirates, Tunisia and Yemen. We focus on immigrants from Arab countries as our treatment group, because of their perceived alignment with Islam and shared ethnicity with the terrorists. A simple comparison of Arab student performance pre- and post- 9/11 would not suffice, as the estimates can be confounded with other factors contemporaneous to the 9/11 attacks. To circumvent this difficulty, we use a difference-in-differences (DD) approach to obtain estimates that are plausibly causal. For the DD strategy to succeed, the comparison group must be similar in terms of observable characteristics to Arab students, but were likely not affected by September 11 in the same manner.

We define the control group as students who were born in a Muslim majority country (Pew Research Center, 2011), as well as students who speak the language of these countries, respectively. These students share many similarities with the treatment group, but, at least in the eyes of many Americans[17], are not as closely associated with Islam or the terrorist attacks. We define a country as having a Muslim majority if the Muslim population exceeds 50% of the total population, according to estimates from the Pew Research Center[18] (2011). An initial estimating equation is provided below:

$$(1) \qquad Y_{it} = \alpha_0 + \alpha_1 Arab_i + \alpha_2 Arab_i \times Post9/11_t + X_{it}\beta + \lambda_t + \varepsilon_{it}$$

Where Y_{it} is the outcome variable for student *i* in school year *t*, *Arab* is an indicator variable that takes on a value of 1 if the student is born in an Arab country or speaks Arabic at home, *Post-9/11* indicates school years occurring after the attacks of September 11, 2001[19]. X_{it} is a vector of

[17] Rubenstein (2004) shows that the number of hate crimes directed at Arabs increased by 1,160 cases in 2001, while the number directed at Muslims went up by 453 cases.
[18] The list of Muslim countries are as follows: Albania, Azerbaijan, Bangladesh, Brunei, Chad, Gambia, Guinea, Indonesia, Iran, Kazakhstan, Kyrgyzstan, Mali, Maldives, Malaysia, Niger, Nigeria, Pakistan, Senegal, Sierra Leone, Tajikistan, Turkmenistan, Turkey and Uzbekistan.
[19] Because the terrorist attack occur in the beginning of the fall of 2001/02 and exams are administered in the spring semester of every year, our post-period starts with the school year 2001/02.

student specific characteristics, λ_t denotes school year fixed effects and ε_{it} is the idiosyncratic disturbance.

School environment plays a large role in determining student outcomes. A school's environment including the racial composition and diversity of the student body may impact the Arab American student experience, and as such moderate or exacerbate the impact of 9/11. To control for the possible differential impact, stemming from heterogeneity of schools, of 9/11 on Arab students, we include school fixed effects in our second specification:

(2) $$Y_{it} = \alpha_0 + \alpha_1 Arab_i + \alpha_2 Arab_i \times Post9/11_t + X_{it}\beta + \lambda_t + \mu_j + \varepsilon_{it}$$

where μ_j represents fixed effects for the school each student is attending.

Although we control for observable student traits - gender, ell status, whether a student receives test modifications, receipt of free or reduced price meals – student-level unobservables may impact the effects of 9/11. In specifications (1) and (2) our sample includes all elementary and middle school students. Although we control for many student specific characteristics, it is possible that Arab students in our sample have fundamentally different unobservables pre- and post- 9/11. We control for this in specification (3) by employing individual student fixed effects. By using this specification the effect of 9/11 is identified using only students observed in both the pre- and post- periods. This is because after a within transformation of the data, we only observe variation in $Arab_i \times Post9/11_t$ for this particular cohort of students. The new estimating equation is as follows:

(3) $$Y_{it} = \alpha_0 + \alpha_1 Arab_i + \alpha_2 Arab_i \times Post9/11 + X_{it}\beta + \lambda_t + \phi_i + \eta_k + \varepsilon_{it}$$

Across all our specifications, endogenous sample selection may lead us to underestimate the effect of 9/11 on student academic performance. It is possible that a certain percentage of the

Arab students that exit our sample post-9/11, were induced to leave the New York City Public Schools as a result of excessive bullying or harassment. We do not observe changes in these students' schooling outcomes, which may have been significant. To maintain that these estimates are not influenced by the location of the schools attended, we include fixed effects for the ZIP Codes of each school at the 5-digit level.

Data and Sample Characteristics

Data

The data are drawn from the student level New York City Public Schools administrative records. The data span the academic years 1998/99 through 2004/05 for students in 3rd through 8th grade. Since the September 11, 2001 took place in the beginning of the fall semester of the academic year 2001/02, we consider the academic years 2001/02-2004/05 as the post-period in our analysis. A key characteristic of the data is the ability to identify students' place of birth as well as language spoken at home.

Another important aspect of the data is the ability to track students over time and observe their progress at different stages of their schooling. The NYCPS data also include 5-digit ZIP codes in addition to unique identifiers for the school each student is attending, which allows us to control for school location and compute the enrollment size of each school in every year the data is available, respectively.

The student records data includes a set of student specific characteristics that includes, gender, ethnicity, English Language Learner (ELL) status, free or reduced price meals, English Language Arts (ELA) test scores, standardized math test scores, and whether they receive any test modifications. All students in elementary and middle school are subject to taking the ELA as well as the state math exam in the spring semester of each year, results of which help determine grade promotion or summer school placement. To make the test scores comparable across different years in the sample, each test score is standardized using the mean and standard deviation of test scores across all students who took the same exam in a given grade and year. We are also able to observe whether a student has been retained at any point in the data, in addition to special education diagnoses. All parents or guardians of students in a NYCPS are required to fill out the Home Language Identification Survey (HLIS) upon first school entry.

Student Trends

The number of Arab and non-Arab Muslim students in elementary and middle school was just over 15,000 in the school year 1998/98, which translates to roughly around 3.5 percent of the total student population in these grades. Over the last fifteen years, the percentage of Arab and non-Arab Muslim students in grades 3 through 8 has steadily increased despite the events of September the 11 to around 4.5 percent in 2004/05.

The trends in Arab and non-Arab Muslim students' academic performance in elementary school over the period 1998/99 – 2004/05 are presented in figure 1. Panel A shows the trends of Arab and non-Arab Muslim students in the English Language Arts (ELA) and state math test scores. Although the baseline ability of both sets of students is fairly different in terms of test score performance, the mean test score of Arab and non-Arab Muslim students has followed a rather similar trajectory up until the academic year 2001/02. Following September 11, 2001, the gap in performance between Arab and non-Arab Muslim students appears to have increased slightly in both test subjects. Although the trends in ELA test scores are not generally declining for Arab students following the events of 9/11, we believe that the trajectory of the test scores has been altered. The change in the progress of Arab students' test scores can be attributed to the increased negative attention that the students may have faced in and out of school.

Grade retention and special education rates over time for both Arab and non-Arab students are presented in Panel B of figure 1. Grade retention among non-Arab Muslim students has steadily decreased from 2.03 percent in 1998/99 to 0.76 percent in 2004/05, while grade retention among Arab students has been more volatile over the years but generally declining from 2.29 percent in 2000/01 to 1.57 percent in 2004/05. Arab and non-Arab Muslim students show a converging trend in special education designations prior to 9/11 and a diverging one post 9/11; where Arab students were receiving special education services at a higher rate than their counterparts.

The trends of Arab and non-Arab Muslim students enrolled in middle school (grades 6-8) from 1998/99 to 2004/05 are presented in figure 2. Student trends in ELA and state math test scores are displayed in Panel A of figure 2, again showing that there is a discrepancy in the baseline test score of Arab and non-Arab Muslim students. The trends for both groups of students appear to follow a similar trajectory prior to 9/11 in both the ELA and state math test scores. After 9/11 the difference in test scores between the two groups of students has increased unfavorably for Arab students.

Panel B presents the trends in grade retention and special education rates for the two groups of students in middle school over the same time period. Even though the trends in grade retention of both groups share a similar progression prior to 9/11, the trends post-9/11 are not dissimilar either. However, in this case it is unclear if Arab students were more or less affected by the events of 9/11 relative to their non-Arab Muslim counterparts. We would require a more detailed inspection of the data to ascertain whether 9/11 had an effect on grade retention.

Looking at the progression of the percentage of students who receive special education services, we see that the trends for Arab and non-Arab Muslim students were somewhat parallel prior to the events of 9/11, whereas the post-9/11 trends diverge for both groups of students with Arabs receiving special education designations at a higher rate than non-Arabs.

Summary Statistics and Baseline Difference-in-Differences

Descriptive statistics for Arab and non-Arab Muslim students enrolled in grades 3-8 in the pre- and post- 9/11 periods are presented in columns [1]-[4] of Table 1. The sample means of the treatment and control group in the pre- and post-periods allow us to compute a baseline difference-in-differences (DD) estimate of the effect of 9/11 on student outcomes, which are presented in column [5]. The student outcomes of interest are the State Math and ELA test scores, both of which are standardized by grade and school year, an indicator for whether a student was retained in a given year and an indicator variable for whether a student received special education services in a given year.

Table 1 shows that Arab students have lower math and English test scores in both the pre- and post-periods. The average math test score for Arab students is between 0.20 and 0.22 student level standard deviations, while the mean math test score for non-Arab Muslim students is between 0.34 and 0.41 student level standard deviations. Average ELA test scores of Arab students increased from 0.06 to 0.09 student level standard deviations. Alternatively, average ELA test scores for non-Arab Muslim students increased from 0.23 to 0.30 student level standard deviations. Retention rates dropped for both groups from the pre-period to the post-period by 0.7 percentage points. Lastly, we observe a rise in the percentage of students receiving special education services for Arab students by 0.5 percentage points compared to only a rise 0.06 percentage points for non-Arab Muslim students.

The DD calculations from sample means, as presented in column (5) of Table 1, indicate that math test scores of Arab students have declined by 0.066 standard deviations relative to non-Arab Muslim students. ELA test scores of Arab students have declined by 0.054 standard deviations relative to the control group. While we calculate a zero effect on Arab students' retention rates, we find that special education designations have increased by 0.5 percentage points in comparison to the control group. With the exception of grade retention, all DD calculations are significant at the 1 percent level.

Table 1 also presents summary statistics of student specific characteristics including: gender, ELL status[20], whether a student receives any test modifications, whether a student receives free or reduced price meals and the location of the school they are attending. For the most part, we observe that the characteristics of the treatment group do not differ in any systematic manner to the control group. Whereby between 53 and 56 percent of the sample are males, 95-98 percent are English language learners, 3.4-5.9 percent receive test modifications and 86.4-88.2 percent receive free or reduced price meals. In addition to the student characteristics being similar across the treatment and control groups, our DD calculations on the control variables show that the composition of the students was not altered as a result of 9/11. The fact that the treatment and control groups are similar in their observable characteristics as well as their composition before and after 9/11 adds to the validity of our chosen comparison group.

The only difference we do observe in the data is in the location of the school attended, pre- and post- 9/11. The calculated DD on the boroughs shows that the only big change in school

[20] English Language Learner is an indicator variable if a student was eligible and has taken an English proficiency exam or is a continuing English language learner.

location is an increase in the percent of students who attend a school in Staten Island. As a sensitivity check, we run our analysis excluding all schools in Staten Island and it has no impact on our results.

Results

To estimate the impacts of 9/11 on student outcomes we employ three different specifications, equations (1)-(3), which work to control for potentially confounding factors. Table 3 and Table 4 present results for the Difference-in-Differences estimates on Arab students' ELA and State Math test scores as well as grade retention and Special Education designation, relative to non-Arab Muslim students. Table 5 presents results using an alternate control group, to test the sensitivity of our estimates. Table 6 presents the results of our falsification test where we constrain the analysis to pre-9/11 data and assign an erroneous attack date occurring prior to the school year 1999/2000.

DD Results on ELA and Math Test Scores

Columns (1) and (4) of Table 2 present results from equation (1) which includes student controls and year fixed effects. Panel A presents our findings on both elementary and middle school students. We estimate a 0.052 student level standard deviation decline in the Math test scores of Arab students relative to non-Arab Muslim students, statistically significant at the 1 percent level. This translates to an effect size of approximately 2 percentile points. We estimate an ELA test score effect of -0.028 student level standard deviations, approximately half the size of the math score effect, statistically significant at the 5% level. This represents a change of approximately 1 percentile point. Columns (2) and (5) present results from equation (2), which adds school level fixed effects. We estimate a -0.045 student level standard deviation decrease in Arab students' math test scores, which is statistically significant at the 5 percent level. The coefficient estimate on the ELA test score is half the magnitude of the math test score, -0.021 student level standard deviations, but is statistically insignificant. Estimates from equations (1) and (2) identify an average treatment effect, alternatively, results from equation (3), presented in columns (3) and (6) allow us to identify the effect of 9/11 using within student, across year variation. This eliminates any potential confounding factors related to changes in student characteristics, pre- and post- 9/11.

Column (3) presents the coefficient estimates of the effect of 9/11 on math test scores, using equation (3). We estimate a -0.101 student level standard deviation drop in math test scores of Arab students who are observed pre- and post- 9/11, approximately a 4 percentile point decrease. This is statistically significant at the 1 percent level. Column (6) presents the estimates of the impact of 9/11 on ELA test scores, using the same specification. The coefficient estimate on ELA test scores is -0.077 student level standard deviations, which is statistically significant at the 1 percent level. This translates to a 3 percentile point decline in the performance of Arab students on the ELA exam.

Panel B presents our results using the elementary school sample. The coefficient estimate, presented in the column (1), shows a 0.047 student level standard deviation decrease in math test scores of Arab students when using equation (1), and is statistically significant at the 5 percent level. This translates to a 2 percentile point drop in math test performance of Arab students. Column (4) presents the impact on ELA test scores from equation (1). We estimate a 0.023 student level standard deviation decline in the ELA test scores of Arab students, an effect size of just under 1 percentile point, although the estimate is statistically insignificant. When we add school-level fixed effects, the results do not change much, which are presented in columns (2) and (5). We find that the effect on Math test scores is -0.041 standard deviations and is significant at the 5 percent level, while the effect on ELA test scores is virtually the same as in column (4) and insignificant.

Including individual student fixed effects to our regression specification yields results that are larger in magnitude and are more accurately estimated. Column (3) of Panel B presents the coefficient estimates on Math test scores. Where we find a 0.067 standard deviation decline in Arab student performance relative to non-Arab Muslim students, which is statistically significant at the 1 percent level. Whereby, this estimate translates to a 3 percentile point drop in the test scores of Arab students in elementary school. The coefficient estimate on ELA test scores is -0.051 standard deviations and is statistically significant at the 1 percent level. This is an effect size of approximately 2 percentile points.

Test score results for students in our middle school sample are presented in Panel C of Table 2. We find a statistically significant decrease in math test scores of Arab students of 0.047 student level standard deviations, from the estimation of equation (1). When we include school fixed effects, the effect size increases in magnitude to -0.057 student standard deviations, but is marginally insignificant at the 10% level. The coefficient estimate is smaller in size and more precise when we estimate using student-level and ZIP Code fixed effects. We find that the effect of 9/11 on math test scores of Arab students in middle school is -0.031 student standard deviations, and is statistically significant at the 10 percent level. The effect of 9/11 on ELA test scores of Arab students in middle school is -0.023 when using the simplest specification, -0.039 when we include school fixed effects and -0.011 when we use student and ZIP code fixed effects in our estimation equation. All three estimates are statistically insignificant.

The effect of September 11 is most pronounced when we employ specification (3). These larger effects are either the result of limiting identifying variation to students enrolled in NYCPS before and after 9/11, or eliminating student level time-invariant heterogeneity. Additionally, we find that among those students, the effects were biggest among students in grades 3 through 5.

DD Results on Grade Retention and Special Education

Another set of student outcomes that we study in this paper are grade retention and special education designation. It should follow naturally from the effect of 9/11 on test scores, that grade retention among Arab students may be affected in a similar fashion. In this study, we view special education diagnoses as an outcome, since students' mental and emotional health may have been impacted by the events or even the aftermath of 9/11.

Table 3 presents the results of the effect of 9/11 on Arab students' retention and special education rates, relative to non-Arab Muslim students. The first column of Panel A presents the DD estimates, controlling only for student observable characteristics. We estimate that, on average, 9/11 had no effect on grade retention of students in the full sample. Similarly, we find no effect on grade retention when including school fixed effects to the estimating equation. The coefficient estimate of the effect of 9/11 on Arab students' grade retention when using individual and ZIP Code fixed effects is rather sizeable and statistically significant at the 1 percent level. Using equation (3), we estimate a 1 percentage point increase in retention rate among Arab students as a result of 9/11, relative to a mean retention rate of 1.59 percent. This means that Arab students were 62.8 percent more likely to be retained post-9/11 than non-Arab Muslim students.

Next, we disaggregate the effect of 9/11 by students in elementary and middle school. When estimating equations (1) and (2), we find small positive and insignificant results of 9/11's effect on grade retention in either the elementary or middle school sample. Including student fixed effects and ZIP Code fixed effects, we find that 9/11 had a larger effect on Arab students who were attending a NYCPS pre- and post- 9/11. We find a 0.5 percentage point increase in the retention rate of Arab students on grades 3-5, although the point estimate is statistically insignificant, as shown in column (3) of Panel B. Interestingly, we find that 9/11 had a larger effect on students in middle school where we estimate a 0.8 percentage point increase in retention rates, relative to a mean retention rate of 1.40 percent among Arab middle school students.

When we study the impact of 9/11 on special education designation, we find that the effects are small and insignificant when using specifications (1) and (2). Similar to our findings on grade retention, our estimates are largest and most precise when using a student fixed effects specification. We estimate a 0.7 percentage point increase in special education designations of Arab students as a result of the events of 9/11, this estimate is relative to a mean special education designation rate of 4.2 percent. We then disaggregate the effect of 9/11 by students in elementary and middle school and find that the effect on special education is insignificant when using equations (1) and (2) in our estimation. When using student fixed effects in our estimation, we find that the effect of 9/11 on special education diagnoses of Arab students in elementary school is +0.5 percentage points relative to a mean of 3.5 percent, the estimate is significant at the 10% level. The coefficient estimate on the middle school sample, when using the student fixed effects specification, is +0.8 percentage points. This estimate translates to a 16.9 percent higher probability of being designated as a special education student, relative to non-Arab Muslim students.

Similar to the findings from the test score regressions, we find that the effects of 9/11 are most prominent among students who were enrolled in a NYCPS before and after the events of the terrorist attacks. However, in the case of grade retention and special education, the effects are biggest among Arab students in middle school.

Alternative Control Group

First, to ensure that our analysis is not sensitive to the control group selected, we run the same analysis discussed earlier using an alternative control group. For this exercise we chose student who were born in India or speak any of the official languages of India at home. Students from India may share certain similarities with Arab as well as non-Arab Muslim students, with the exception of religion, and should not have been affected by the events of 9/11 in the same manner as students from Arab countries. Tables 4 and 5 present the DD results of the effects of 9/11 on Arab students with Indian students as the comparison group.

Results of the DD estimates on state math and ELA examinations are presented in Table 4. Panel A presents the regression results using the full sample. The coefficient estimate of the impact of 9/11 on state math test scores of Arab students appear to be robust to the choice of control group. We estimate a decline in the math test scores of Arab students 0.054 standard deviations when using equation (1), a 0.060 standard deviation decline when we include school fixed effects and a 0.108 decrease when estimating with student and ZIP Code fixed effects. The coefficient estimates of the effect of 9/11 on ELA test scores are presented in columns (4) through (6). We find a -0.022 student level standard deviation effect when estimating equation (1), the coefficient estimate when we include school fixed effects is -0.027 student level standard deviations, both of which are statistically insignificant. Lastly, we find that ELA test scores of Arab students decline by 0.078 student level standard deviations relative to Indian students, when estimating equation (3), and is statistically significant at the 1 percent level. These estimates are almost equal in size to the estimates found using non-Arab Muslim students as the control group.

Panels B and C display results of the effect of 9/11 on Arab students relative to students from India using the elementary school and middle school samples separately. We find that the effect of 9/11 is negative across the board again in this case for student in elementary and middle school. However, we find that the effect of 9/11 on math test scores was biggest among students in grades 3-5, as shown in the columns (1) through (3) of Table 4. The point estimate is -0.044 student level standard deviations when using student fixed effects. This estimate is smaller than the -0.067 coefficient when using the first control group in the analysis. The effect on Arab students in middle school is -0.028 standard deviations and is statistically insignificant. When we analyze the impact of 9/11 on ELA test scores, we find that Arab students in middle school were most affected. Column (6) in Panel C of table 4 shows that ELA test scores of Arab students in middle school decreased by 0.076 standard deviations relative to Indian students. This estimate translates to a 3 percentile point decline in ELA test score performance.

Table 5 presents DD estimates on grade retention and special education designation as dependent variables using Indian students as the control group. As in tables 3 through 4, Panel A presents results from the full sample, Panel B presents results using the elementary school sample and Panel C presents the results using the middle school sample. The findings in this case are also very similar to the results using non-Arab Muslim students as the comparison group. We find no effect of 9/11 on either outcome when using equations (1) and (2) to obtain the DD estimates. We also find no effect when restricting the analysis to either the elementary or middle school sample. Similar to the findings presented in table 3, we find that the effect of 9/11 on grade retention and special education is most pronounced when including student and ZIP Code fixed effects as presented in columns (3) and (6) of Table 5. In fact, we find effect sizes that are almost equal to those found when using non-Arab Muslim students as the control group. We estimate a 1 percentage point increase in retention rates of Arab students, relative to the mean retention rate of 1.59 percent. We estimate that special education designations increase by 0.7 percentage points as a result of the events of 9/11, relative to a mean of 4.16 percent.

Next, we disaggregate the effect of 9/11 on retention and special education by elementary and middle school. Once more, we find that the effect of 9/11 was largest among students in middle school. As such, we find no effect on retention of students in elementary school in any of the specifications used, results of which are presented in Panel B of Table 5. As is the case when using non-Arab Muslims as a control group, we find a +0.7 percentage point effect of 9/11 on special education designation when including student fixed effects. We estimate that retention rates increased by 1.1 percentage points, relative to a mean of 1.4 percent retention rate among Arab students in middle school. Lastly, we estimate a 0.9 percentage point increase in special education designation among Arab students in middle school, relative to a mean of 4.9 percent.

These results add to the robustness of our estimates, as the choice of control group has not altered our findings in a drastic manner. We find that the pattern of the effect of 9/11 on student outcomes are, for the most part, consistent with the pattern found when using non-Arab Muslim students as a control group.

Pre-Trend Verification

To ensure that the DD estimates are valid, we need to address concerns that school performance trends of Arab and non-Arab Muslim students prior to the events of 9/11 were the same. This is a necessary condition for the DD estimates in order to yield effects that are indeed causal. To test whether the treatment and control group did not have differing academic performance trends, we restrict our analysis to pre-9/11 data and implement a false treatment date of September 2000. For the control group we have chosen to be valid, we expect that the false treatment would not affect the performance of Arab students in any of the specifications used. Consequently, we obtain DD estimates of the effects of the false treatment date on Math and ELA test scores, grade retention and special education designation using the specifications (1), (2) and (3)[21].

Table 6 presents the results of our falsification test, using non-Arab Muslim students as the chosen control group. We find that the false treatment date has no effect on any of the student outcomes. The effect of the false treatment on ELA test scores, when including student fixed effects is -0.007 student level standard deviations. The effect on math test scores is even smaller in size, -0.003 standard deviations. These estimates are insignificant both statistically and in magnitude. We also find small and insignificant results of the false treatment date on grade retention and special education. The effect of the false date on retention and special education range between -0.002 and 10^{-4}, all of which are statistically insignificant.

Results of the falsification test using Indian students as the control group are presented in Table 7. The results in this case also show that there is almost no effect of the fake treatment on student outcomes of Arab students relative to Indian students. We find a rather small change in the math scores of Arab students, post-fake treatment, of -0.008 standard deviations when including student fixed effects. Similarly, we only find a slight change in ELA test scores of

[21] For the sake of brevity only the results of equations (2) and (3) are presented, we find that the results of the falsification test are robust to the specification used.

Arab students relative to Indian students of -0.006 student level standard deviations. Both ELA and math estimates are statistically insignificant. Lastly, we estimate very small negative, but statistically insignificant, effect of the false treatment on grade retention and special education designation ranging between 0.1 and 0.3 percentage points.

In addition to our treatment and control groups being similar along observables, as well as in composition pre- and post- 9/11, we find that both control groups used in the analysis do not violate the DD assumption of parallel pre-trends in order to retrieve causal effects of the events of the September 11 terrorist attacks. As a result of the falsification test on both sets of control groups, we fail to reject the null hypothesis that the control groups are indeed valid.

Conclusion

Using administrative student record data from NYCPS, we analyze the impact of the events of September 11 on the academic performance of Arab students in elementary and middle school. We find that the increased negative attention directed at people who share the terrorist's ethnic identities, i.e. Arab, resulted in a decline in the academic performance of Arab students. Using specification (3) on the full sample, we estimate a 4 percentile point drop in the performance of Arab students in state math examinations post-9/11 and a 3 percentile decline in ELA test scores. Subsequently, we estimate an increase in grade retention by 57 percent, among Arab students who were enrolled in a NYCPS pre- and post- 9/11. Another effect of 9/11 we found was on special education designation. We find that the rate at which Arab students receive special education services have increased by 14 and 17 percent among students in elementary and middle school, respectively.

In this study we estimate the detrimental effects of the post September 11, 2001 increase in animosity towards the Arab community on the academic performance of Arab students enrolled in a New York City Public School, using several specifications. We analyze the effect of 9/11 on student outcomes using two alternative control groups and find that our estimates are robust in that respect. Additionally, we implement a falsification test on the pre-9/11 data to assess the validity of our DD assumptions, where we fail to reject the validity of either comparison group chosen. Our findings suggest that these effects were most apparent among Arab students who were present in New York City before and after the terrorist attacks took place.

References for Chapter 3

Barker, E.D., Arseneault, L., Brendgen, M., Fontaine, N., Maughan, B. 2008. Joint development of bullying and victimization in adolescence: Relations to delinquency and self-harm. Academy of Child and Adolescent Psychiatry Vol. 47, 1030-1038.

Beaty, L.A., Alexeyev, E.B. 2008. The problem of school bullies: What the research tells us Adolescence Vol. 443, 1-11.

Bogin, A. 2010. Anti-Islamic sentiment and its impact on residential property values. Syracuse University

Brown, S., Taylor, K. 2008. Bullying, education and earnings: Evidence from the national child development study Economics of Education Review 27, 387-401.

Cristillo, L. 2008. Religiosity, education and civic belonging: Muslim youth in New York City public schools. Research Report.

Davila, A., Mora, M.T. 2005. Changes in the earnings of Arab men between 2000 and 2002 Journal of Population Economics vol. 18, 587-601.

Fekkes, M., Pijpers, F.I.M., Fredriks, A.M., Vogels, T., Verloove-Vanhorick, S.P. May 2006. Do bullied children get ill, or do ill children get bullied? A prospective cohort study on the relationship between bullying and health-related symptoms Pediatrics. Vol. 117, 1568-1574.

FERGUSSON, D.M., BEAUTRAIS, A.L., HORWOOD, L.J. 2003. Vulnerability and resiliency to suicidal behaviours in young people Psychological Medicine 33, 61.

Kaushal, N., Kaestner, R., Reimers, C. 2007. Labor market effects of September 11th on Arab and Muslim residents of the united states The Journal of Human Resources vol. 42, pp. 275-308.

Rubenstein, W.B. 2004. The real story of U.S. hate crime statistics: An empirical analysis Tulane Law Review Vol. 78, 1213-1246.

Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T. 2002. Assessing ‘Neighborhood effects’: Social processes and new directions in research Annual Review of Sociology Vol. 28, 443-478.

Table 3.1: Pre- and post- September 11, 2001 summary statistics for Arab and non-Arab Muslim students in NYCPS in elementary and middle school

	Arab		Non-Arab Muslim		
	Pre-9/11	Post-9/11	Pre-9/11	Post-9/11	DD
Dependent Variables:	[1]	[2]	[3]	[4]	[5]
Math Scale Z Score	0.205	0.217	0.345	0.412	-0.066***
	[0.915]	[0.926]	[0.892]	[0.908]	(0.014)
ELA Scale Z Score	0.0586	0.0924	0.228	0.301	-0.054***
	[0.904]	[0.934]	[0.875]	[0.915]	(0.014)
Retained	0.019	0.011	0.014	0.007	0.000
	[0.124]	[0.100]	[0.101]	[0.0783]	(0.002)
Special Education	0.0336	0.0385	0.0200	0.0206	0.005***
	[0.180]	[0.192]	[0.140]	[0.142]	(0.002)
Control Variables:					
Male	0.559	0.552	0.528	0.528	-0.007
	[0.497]	[0.497]	[0.499]	[0.499]	(0.007)
Female	0.441	0.448	0.472	0.472	0.007
	[0.497]	[0.497]	[0.499]	[0.499]	(0.007)
English Language Learner	0.988	0.970	0.972	0.953	0.001
	[0.109]	[0.162]	[0.165]	[0.203]	(0.002)
Received Test Modification	0.0538	0.0591	0.0339	0.0358	0.003
	[0.226]	[0.236]	[0.181]	[0.186]	(0.003)
Received Free/Reduced Meals	0.864	0.870	0.877	0.882	0.001
	[0.353]	[0.456]	[0.339]	[0.449]	(0.005)
Brooklyn	0.539	0.523	0.286	0.293	-0.023***
	[0.499]	[0.499]	[0.452]	[0.455]	(0.007)
Manhattan	0.0665	0.0561	0.0421	0.0382	-0.007**
	[0.249]	[0.230]	[0.201]	[0.192]	(0.003)
Queens	0.267	0.255	0.559	0.529	0.018**
	[0.443]	[0.436]	[0.496]	[0.499]	(0.006)
Staten Island	0.0551	0.0941	0.0201	0.0308	0.028***
	[0.228]	[0.292]	[0.140]	[0.173]	(0.003)
Bronx	0.0723	0.0715	0.0930	0.109	-0.017***
	[0.259]	[0.258]	[0.290]	[0.312]	(0.004)
Observations	11,098	20,158	38,255	75,393	144,904

Source: New York City Public Schools Transcript Data, 1998-1999 through 2004-2005.

Notes: English Learner Status is awarded to students who do not speak English as the first language at home. Test modifications are accommodations made primarily for students who are in special education. Figures in brackets denote standard deviations. Figures in parentheses denote standard errors for the difference-in-differences measure. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Table 3.2: DD Results of the effects of 9/11 on ELA and state math test scores

	[1]	[2]	[3]	[4]	[5]	[6]
	Math test score	Math test score	Math test score	ELA test score	ELA test score	ELA test score
Panel A: Full sample						
Arab X Post 9/11	-0.052***	-0.045**	-0.101***	-0.028**	-0.021	-0.077***
	(0.014)	(0.018)	(0.009)	(0.014)	(0.018)	(0.011)
Observations						
	106,633	106,633	106,633	95,279	95,279	95,279
Panel B: Elementary School Sample						
Arab X Post 9/11	-0.046**	-0.041**	-0.067***	-0.023	-0.023	-0.051***
	(0.019)	(0.020)	(0.015)	(0.021)	(0.019)	(0.019)
Observations	57,630	56,348	56,348	43,460	43,460	43,460
Panel C: Middle School Sample						
Arab X Post 9/11	-0.047**	-0.057	-0.031*	-0.023	-0.039	-0.011
	(0.024)	(0.035)	(0.019)	(0.021)	(0.032)	(0.020)
Observations	50,285	50,285	50,285	52,965	52,965	52,965
Student Controls	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects		yes			yes	
Student Fixed Effects			yes			yes
ZIP Code Fixed Effects			yes			yes

Notes: Control group consists of students from non-Arab Muslim majority countries. Student controls (male, English Language Learner status, test modifications, free/reduced price meals, borough location and school size), Arab indicator variable and school-year fixed effects. Figures in columns [1] and [4] are based on equation (1), figures in columns [2] and [5] are based on equation (2) and figures in columns [3] and [6] are based on equation (3). Math and ELA test scores are standardized by school year and grade level, reported as z-scores. The full sample consists of students in grades 3-8, while the elementary and middle school samples consist of students in grades 3-5 and 6-8, respectively. Numbers in parentheses represent standard errors clustered at the school level. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 3.3: DD results of the effects of 9/11 on grade retention and special education designation

	[1]	[2]	[3]	[4]	[5]	[6]
	Retained	Retained	Retained	Special Education	Special Education	Special Education
Panel A: Full sample						
Sample Mean		*0.0159*			*0.0416*	
Arab X Post 9/11	0.001	0.001	0.010***	0.003	0.001	0.007***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
Observations	125,232	125,232	125,232	125,232	125,232	125,232
Panel B: Elementary School Sample						
Sample Mean		*0.0192*			*0.0350*	
Arab X Post 9/11	0.002	0.002	0.005	0.005	0.002	0.005*
	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)
Observations	65,394	65,394	65,394	65,394	65,394	65,394
Panel C: Middle School Sample						
Sample Mean		*0.0140*			*0.0490*	
Arab X Post 9/11	0.002	0.001	0.008*	0.005	0.001	0.008***
	(0.003)	(0.002)	(0.004)	(0.005)	(0.007)	(0.003)
Observations	59,838	59,838	59,838	59,838	59,838	59,838
Student Controls	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects		yes			yes	
Student Fixed Effects			yes			yes
ZIP Code Fixed Effects			yes			yes

Notes: Control group consists of students from non-Arab Muslim majority countries. Student controls (male, English Language Learner status, test modifications, free/reduced price meals, borough location and school size), Arab indicator variable and school-year fixed effects. Figures in columns [1] and [4] are based on equation (1), figures in columns [2] and [5] are based on equation (2) and figures in columns [3] and [6] are based on equation (3). Grade retention takes on a value of 1 if a student had been retained in a given school year. Special education takes on a value of 1 if a student was provided with special education services in a given school year. The full sample consists of students in grades 3-8, while the elementary and middle school samples consist of students in grades 3-5 and 6-8, respectively. Numbers in parentheses represent standard errors clustered at the school level. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 3.4: DD results of the effects of 9/11 on ELA and state math test scores using control group II

	[1] Math test score	[2] Math test score	[3] Math test score	[4] ELA test score	[5] ELA test score	[6] ELA test score
Panel A: Full sample						
Arab X Post 9/11	-0.054***	-0.060**	-0.108***	-0.022	-0.027	-0.078***
	(0.019)	(0.028)	(0.013)	(0.019)	(0.026)	(0.015)
Observations	38,903	38,903	38,903	35,526	35,526	35,526
Panel B: Elementary School Sample						
Arab X Post 9/11	-0.060**	-0.050	-0.044**	-0.042	-0.027	-0.024
	(0.026)	(0.032)	(0.022)	(0.029)	(0.034)	(0.027)
Observations	20,339	20,339	20,339	19,096	19,096	19,096
Panel C: Middle School Sample						
Arab X Post 9/11	-0.027	-0.059	-0.028	-0.021	-0.048	-0.076***
	(0.032)	(0.050)	(0.024)	(0.032)	(0.046)	(0.027)
Observations	18,564	18,564	18,564	16,430	16,430	16,430
Student Controls	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects		yes			yes	
Student Fixed Effects			yes			yes
ZIP Code Fixed Effects			yes			yes

Notes: Control group II consists of students from India. Student controls (male, English Language Learner status, test modifications, free/reduced price meals, borough location and school size), Arab indicator variable and school-year fixed effects. Figures in columns [1] and [4] are based on equation (1), figures in columns [2] and [5] are based on equation (2) and figures in columns [3] and [6] are based on equation (3). Math and ELA test scores are standardized by school year and grade level, reported as z-scores. The full sample consists of students in grades 3-8, while the elementary and middle school samples consist of students in grades 3-5 and 6-8, respectively. Numbers in parentheses represent standard errors clustered at the school level. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 3.5: DD results of the effects of 9/11 on grade retention and special education designation using control group II

	[1]	[2]	[3]	[4]	[5]	[6]
	Retained	Retained	Retained	Special Education	Special Education	Special Education
Panel A: Full sample						
Sample Mean		*0.0159*			*0.0416*	
Arab X Post 9/11	-0.002	-0.002	0.010***	0.001	0.000	0.007**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)
Observations	44,006	44,006	44,006	44,006	44,006	44,006
Panel B: Elementary School Sample						
Sample Mean		*0.0192*			*0.0350*	
Arab X Post 9/11	0.001	-0.000	0.003	0.003	0.004	0.007**
	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.003)
Observations	22,780	22,780	22,780	23,303	22,780	22,780
Panel C: Middle School Sample						
Sample Mean		*0.0140*			*0.0490*	
Arab X Post 9/11	-0.003	-0.003	0.011***	0.003	0.000	0.009***
	(0.002)	(0.002)	(0.004)	(0.006)	(0.007)	(0.003)
Observations	21,226	21,226	21,226	21,226	21,226	21,226
Student Controls	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes
School Fixed Effects		yes			yes	
Student Fixed Effects			yes			yes
ZIP Code Fixed Effects			yes			yes

Notes: Control group II consists of students from India. Student controls (male, English Language Learner status, test modifications, free/reduced price meals, borough location and school size), Arab indicator variable and school-year fixed effects. Figures in columns [1] and [4] are based on equation (1), figures in columns [2] and [5] are based on equation (2) and figures in columns [3] and [6] are based on equation (3). Grade retention takes on a value of 1 if a student had been retained in a given school year. Special education takes on a value of 1 if a student was provided with special education services in a given school year. The full sample consists of students in grades 3-8, while the elementary and middle school samples consist of students in grades 3-5 and 6-8, respectively. Numbers in parentheses represent standard errors clustered at the school level. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Table 3.6: Falsification Test on Math and ELA Test Scores, Grade Retention and Special Education

	[1]	[2]	[3]	[4]
Panel A: Math and ELA test scores	Math test score		ELA test score	
Arab X Post 9/11	-0.012	-0.003	0.004	0.007
	(0.014)	(0.010)	(0.014)	(0.011)
Observations	36,004		35,902	
Panel A: Math and ELA test scores	Retained		Special Education	
Arab X Post 9/11	-0.001	-0.002	-0.000	0.000
	(0.003)	(0.002)	(0.003)	(0.001)
Observations	49,353		49,353	
Student Controls	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes
School Fixed Effects	yes		yes	
Student Fixed Effects		yes		yes
ZIP Code Fixed Effects		yes		yes

Notes: Control group consists of students from non-Arab Muslim majority countries. False treatment date assigned is September 1st, 1999. Columns [1] and [3] present falsification results using equation (2). Columns [2] and [4] present falsification results using equation (3). Results from using equation (1) are not presented for brevity and redundancy of the estimates. Numbers in parentheses represent standard errors clustered at the school level. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

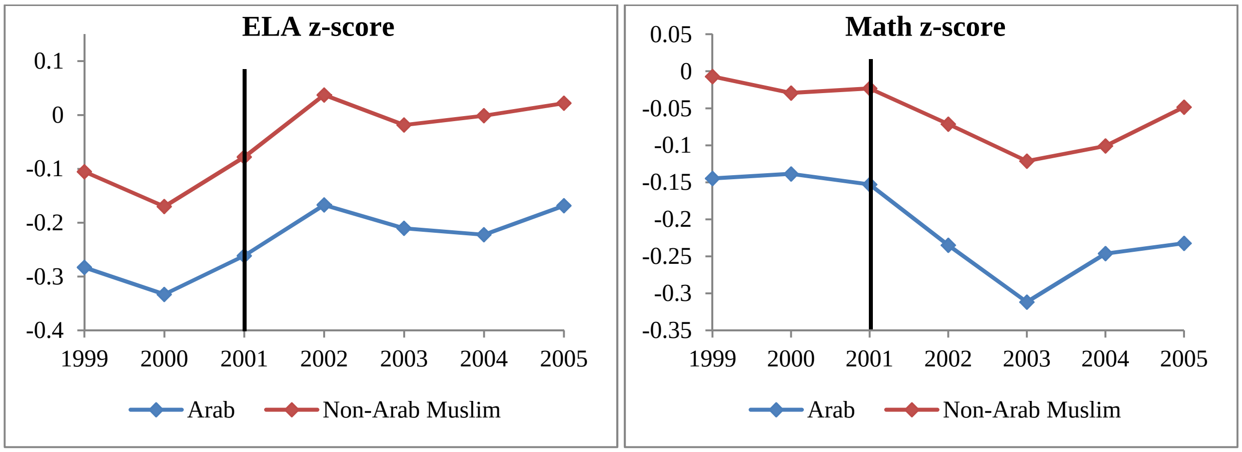
Table 3.7: Falsification Test on Math and ELA Test Scores, Grade Retention and Special Education using Control Group II

	[1]	[2]	[3]	[4]
Panel A: Math and ELA test scores	Math test score		ELA test score	
Arab X Post 9/11	-0.017	-0.008	-0.017	-0.006
	(0.020)	(0.014)	(0.019)	(0.015)
Observations	14,742		14,710	
Panel A: Math and ELA test scores	Retained		Special Education	
Arab X Post 9/11	-0.001	-0.003	-0.001	-0.001
	(0.003)	(0.002)	(0.003)	(0.001)
Observations	18,375		18,375	
Student Controls	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes
School Fixed Effects	yes		yes	
Student Fixed Effects		yes		yes
ZIP Code Fixed Effects		yes		yes

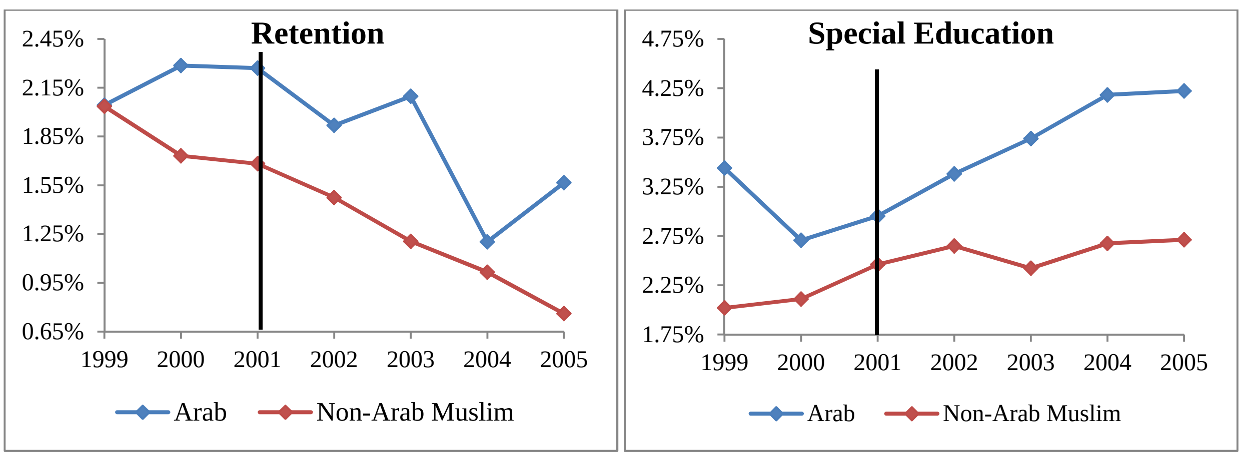
Notes: Control group consists of students from India. False treatment date assigned is September 1st, 1999. Columns [1] and [3] present falsification results using equation (2). Columns [2] and [4] present falsification results using equation (3). Results from using equation (1) are not presented for brevity and redundancy of the estimates. Numbers in parentheses represent standard errors clustered at the school level. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively.

Figure 3.1: Academic Performance of NYCPS Elementary School Students, Grades 3 through 5

Panel A: Average Math and ELA z-score by school year, solid black line denotes September 11, 2001



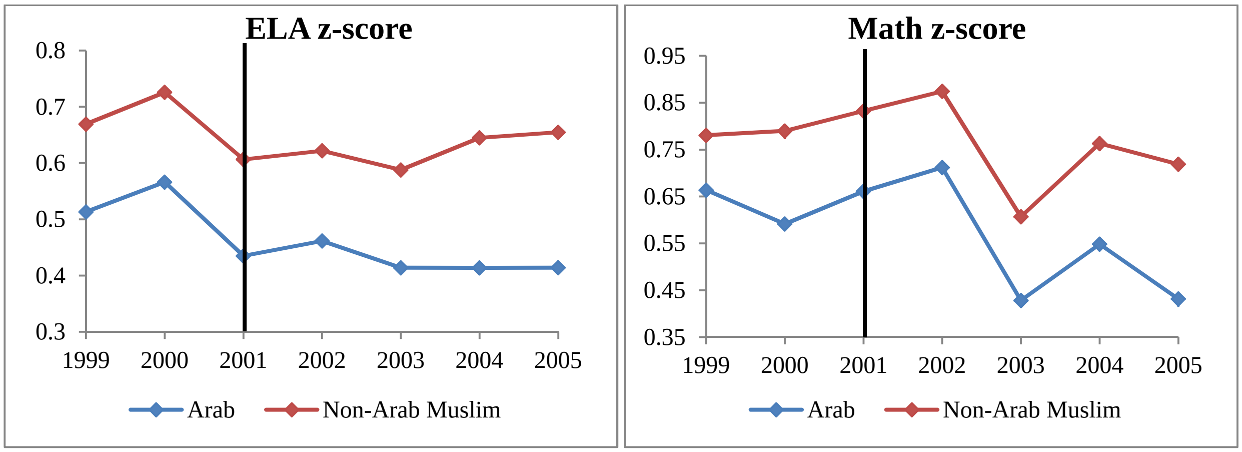
Panel B: Average retention rates and special education diagnoses by school year



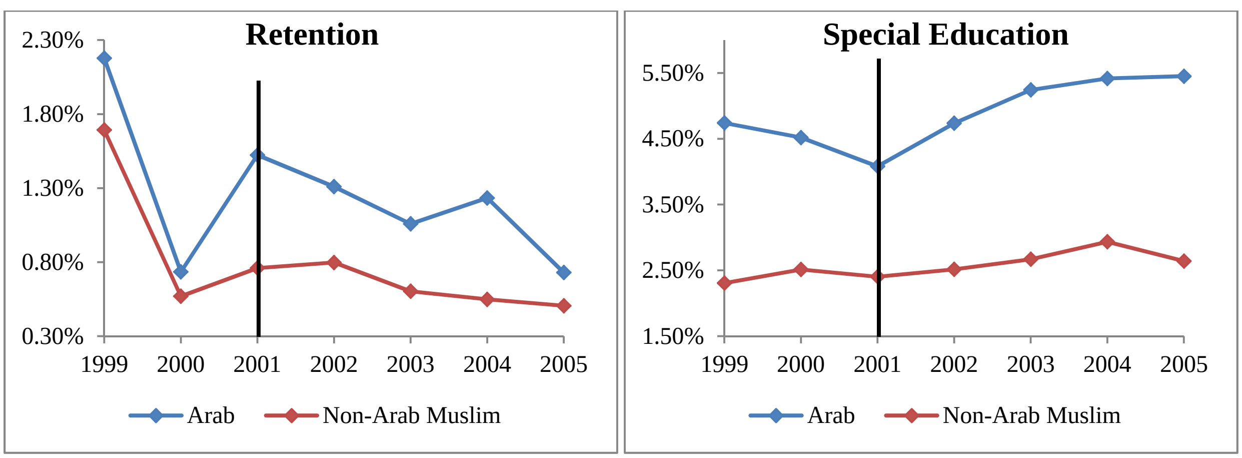
Source: New York City Public School Transcript Data, 1999-2000 through 2004-2005

Figure 3.2: Academic Performance of NYCPS Middle School Students, Grades 6 through 8

Panel A: Average Math and ELA z-score by school year, solid black line denotes September 11, 2001



Panel B: Average retention rates and special education diagnoses by school year



Source: New York City Public School Transcript Data, 1999-2000 through 2006-2007

VITA

NAME OF AUTHOR: Wael Soheil Moussa

PLACE OF BIRTH: Beirut, Lebanon

DATE OF BIRTH: 10 May, 1984

EDUCATION:

M.A. Economics, American University of Beirut, 2008

B.A. Economics, American University of Beirut, 2006

RESEARCH EXPERIENCE AND EMPLOYMENT:

2006-08	Research Assistant, American University of Beirut, Beirut, Lebanon
2012	Instructor, Department of Economics, Syracuse University, Syracuse, NY

AWARDS, SCHOLARSHIPS AND FELLOWSHIPS:

2008-2012	Maxwell Summer Research Fellowship, Department of Economics, Syracuse University
2008-2013	Syracuse University Graduate Assistantships for the PhD in Economics, Syracuse University