Three Essays on the Economics of Education

Mariesa Ann Herrmann

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

Three Essays on the Economics of Education

Mariesa Ann Herrmann

This dissertation consists of essays on three inputs into the educational production function: curriculum, peers, and teachers. The chapters are linked by their focus on understanding the importance of these inputs for student achievement and by their exploitation of the exact timing of events (i.e., student mobility, receipt of special education services, teacher absences) to identify causal effects.

The first chapter examines how decentralized decision-making about schools' curricula affects student achievement. Decentralized decision-making in this context involves a tradeoff, since individual schools might be better able to target the needs of their student populations, but their choices might not be aligned with the goals of the larger community. Differences in curricula may also harm the achievement of mobile students, who have to adjust to new curricula when they change schools.

My analysis is based on quasi-experimental variation from New York City, which standardized math and reading curricula starting in Fall 2003. Schools received exemptions from the standardized curricula based on test score cutoffs, allowing me to use a regression discontinuity design to evaluate the causal effect of the policy change. I find that curriculum standardization has no significant effect on student achievement, either overall or for mobile students.

I also assess the extent to which curriculum standardization could benefit mobile students using a novel test that exploits the timing of student mobility relative to the achievement exam: I compare the achievement losses of students who could have been affected by having to change curricula — those who changed schools before the exam — to those of students

who may have similar personal circumstances but whose achievement could not have been affected by changing curricula — those who changed schools after the exam. The achievement losses of these groups of students are similar, suggesting that omitted variables associated with student mobility may explain the entirety of the mobility achievement gap and that curriculum standardization may have little scope to improve the achievement of mobile students.

The second chapter examines the effects of disabled peers on student achievement and behavioral outcomes. Although disabled students might be expected to have negative effects on their peers, this expectation has not been borne out in studies that focus on the peer effects of special education students. This chapter argues that this is likely because special education ameliorates negative peer effects.

To illustrate this argument, I exploit the longitudinal nature of the data to classify students into categories based on when they received special education services. I classify students who are ever observed in special education as "disabled" and those who currently receive special education services as "special education" students. Students who enter special education in the future are classified as having "undiagnosed disabilities," while those who exit special education have "declassified disabilities." I then use school-grade variation in the proportions of disabled peers to estimate their effects on non-disabled students.

Consistent with special education mitigating peer effects, I find that students with undiagnosed disabilities have significant negative effects on their non-disabled peers, but students who are receiving special education services do not. Students who have been declassified also have no effects on their peers, probably because they are positively selected among disabled students. I also find that segregation is the likely mechanism by which special education mitigates negative peer effects.

The third chapter, which is joint work with Jonah Rockoff and is forthcoming in the

Journal of Labor Economics, examines the effect of teacher absence on student achievement. To address potential sources of bias from the endogeneity of teacher absence, we use detailed data on the timing, duration, and cause of absences. Our main specification uses within teacher variation, and in support of a causal interpretation, we show that teacher absences before the exam have a significant negative effect on achievement while absences after the exam do not.

Our estimates suggest that the daily loss associated with one absence is 0.001 standard deviations in math and 0.0006 standard deviations in reading, the same as replacing an average teacher with one at the 10-20th percentile of teacher value-added. We also find evidence that the daily losses associated with an absence decline with the length of the absence spell, consistent with long-term substitutes being of higher quality or learning on the job.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation has benefited from helpful comments and suggestions from Doug Almond, Cristian Pop-Eleches, Leigh Linden, Brian Jacob, Sue Dynarski, Jeff Smith, Eric Verhoogen, Jacob Vigdor, Dick Murnane, and Damon Clark. I am also grateful for the comments of the seminar participants at Columbia University, the American Education Finance Association meeting, the University of Michigan, Mathematica Policy Research, MDRC, the American Institutes for Research, the SOLE/EALE conference, the Education Finance Resource Consortium, Harvard University, and the NBER Summer Institute.

Financial support was provided by a National Science Foundation Graduate Research Fellowship and the Educational Finance Research Consortium. Of course, any opinions or findings expressed are my own.

Special thanks go to Joanna Lack, Raelene Stroom, Marsha Modeste, and Jennifer Bell-Ellwanger for their help in obtaining the data and for helpful discussions. Skanda Amarnath provided excellent research assistance, and Veronica Cabezas provided invaluable help in the preparation of the data.

I owe special gratitude to my advisors, Jonah Rockoff, Janet Currie, and Miguel Urquiola, who have provided me with exceptional advice and mentorship throughout my years in graduate school. Jonah provided me with access to the data and has spent countless hours training me to be a better economist. Janet broadened my knowledge base, helped sharpen my skill set, and always reminded me of the bigger picture. Miguel was always on-hand to answer my questions and provide useful feedback.

I would also like to thank my friends and colleagues at Columbia for their helpful discussions and for making my time in graduate school memorable. At the risk of forgetting some names, I thank David DeRemer, Takakazu Honryo, Cecilia Machado, Marcos Nakaguma, Christine Pal, Maria Jose Prados, Johannes Schmieder, Anukriti Sharma, Hitoshi Shigeoka, Lesley Turner, Reed Walker, and Zhanna Zhanabekova.

Finally, I thank my husband Eric Milliken for his technological assistance and editing services, and most importantly, his love and support.

Dedication

For Eric.

Chapter 1

One Size Fits All? The Effect of Curriculum Standardization on Student Achievement

1.1 Introduction

Economists have long debated the merits of decentralized decision-making in organizations. The arguments for decentralization center on better information and adaptability to local conditions, while the arguments against emphasize the need for coordination and greater economies of scale. These tradeoffs have been the focus of a large theoretical literature in organizational economics (e.g., Dessein 2002, Acemoglu et al. 2007), and a growing number of empirical studies have examined the relationship between decentralization and firms' productivity (Janod and Saint-Martin 2004), labor skill mix (Caroli and Van Reenen 2001), and product variety (Thomas 2011).

These tradeoffs have also recently been highlighted in education in a growing debate over the decentralization of curricular decisions in American public schools. The U.S. does not currently have a national curriculum. This stands in contrast to many countries, many of which (e.g., Japan, South Korea, Finland), outperform the U.S. on international measures of student achievement (TIMSS 2008). Partly motivated by this evidence, the National Governors Association and the Council of Chief State School Officers sponsored the creation of new, national guidelines for the knowledge and skills expected of students from kindergarten through grade 12 in mathematics and English. Released in June 2010, these Common Core State Standards have been adopted in nearly all of the 50 states, and the U.S. Department of Education is spending \$360 million to develop assessments and curricular supports for the "Common Core," as it has come to be known. However, this shift towards national standards has prompted concerns that the U.S. is moving towards adopting a "one-size-fits-all, centrally controlled curriculum" ("Closing the Door on Innovation" 2011).

Whether the adoption of a nationally standardized curriculum would improve student achievement in the U.S. is theoretically ambiguous. Curriculum standardization may prevent schools from choosing curricula that are better aligned with the needs of their particular student populations but may align instruction with the priorities of the larger community. In addition, standardization may benefit mobile students, who may have difficulty adjusting to different curricula when they change schools. These benefits could be considerable since students who change schools already tend to suffer numerous educational disadvantages.[1]

Understanding how centralizing curricular decision-making affects student achievement is important not only for informing debates over a national curriculum but also because curriculum standardization has been at the heart of a number of recent educational reforms. Many large, urban school districts (e.g., Chicago, New York City, Los Angeles, and Houston) have already implemented district-wide standardized curricula, but there has not yet been any empirical study of how this affects student achievement.

In this paper, I examine how the use of a standardized curriculum affects overall achievement, and, in particular, the achievement of mobile students. I base my analysis on quasi-experimental variation caused by curricular reforms in New York City, which began implementing standardized math and reading curricula in all but a set of exempt schools in fall 2004. Exemptions from the standardized curricula were primarily based on whether schools' test scores met or exceeded certain thresholds, allowing me to estimate the causal effect of curriculum standardization on achievement using a regression discontinuity design. Schools that barely missed the exemption cutoff had to implement the standardized curricula, and schools that barely made the threshold did not, thus providing a comparison of students in schools that used different curricula but were otherwise similar.

I find that curriculum standardization has no significant effect on student achievement. "Intent-to-treat" estimates indicate that falling below the cutoff and being mandated to

[1]A 1994 GAO report found that around 16% of 3rd graders had attended at least three different schools since 1st grade, and data from the 1998 National Assessment of Educational Progress found that 34% of 4th graders, 21% of 8th graders, and 10% of 12th graders had changed schools at least once in the previous year (National Research Council and Institute of Medicine 2010). Mobile students are disproportionately likely to be black or Hispanic and receive free lunch (Hanushek et al. 2004).

adopt the standardized curricula results in statistically insignificant test score changes of -0.004 standard deviations in math and -0.007 standard deviations in reading. The 2SLS estimates are somewhat imprecise but rule out, at a 95% confidence level, effects below -0.22 standard deviations in math and -0.26 standard deviations in reading and effects above 0.17 standard deviations in math and 0.22 standard deviations in reading, the type of large improvements found in some prior studies of curricula.

I also find no evidence that standardization improved the achievement of mobile students. In addition to the quasi-experimental variation caused by the reforms, I use a novel test to assess how much of the mobility achievement gap could be causally related to changing schools — comparing the achievement of students who move before the exam (or school year) to those who move after the exam (or school year). Since any declines in the achievement of students who move after the exam (or school year) cannot be causally related to changing schools, these declines likely provide a lower bound on the amount of the achievement gap that can be attributed to unobservable factors that tend to be correlated with student mobility (e.g. parents' job loss, divorce, etc.).[2] Results from this test suggest that negative selection could explain the entire achievement gap between mobile and non-mobile students, suggesting that decentralization of curricular decisions has little role to play in explaining low achievement among mobile students.

This paper continues as follows. Section 1.2 presents the conceptual framework, and Section 1.3 discusses the institutional background. Section 1.4 describes the data, Section 1.5 presents the estimation strategy, and Section 1.6 reports the results. Section 1.7 concludes.

[2] This assumes that students who move after the exam (or school year) are less negatively selected than students who move before the exam (or school year). This assumption is supported by results for selection on observable characteristics; it is more difficult to say whether this is true for unobservable characteristics.

1.2 Conceptual Framework

The potential effects of curriculum standardization on student achievement can be illustrated using a simple model in which achievement is produced from a combination of student, family, and school inputs. In the model, schools are assumed to have chosen curriculum optimally given their objective functions, which might maximize students' expected achievement or might focus on other outcomes altogether.[3] The notion that schools may have objectives other than maximizing student test scores is one rationale for school accountability reforms.

Since the effectiveness of a particular curriculum may vary across students, schools may choose different curricula either because they have different student populations or different objective functions.[4] Moreover, schools may be discouraged from coordinating their curricular decisions by high costs of communication. These differences in curricula may disrupt the continuity of instruction for students who change schools, harming their achievement.

The fundamental assumption of the model is that students' decisions to change schools are exogenous to schools' choices of curricula. This assumption seems reasonable to the extent that moves happen for reasons that are likely unrelated to children's schooling, and parents would have to spend considerable time and effort to collect the curricula-related information needed to inform these decisions.[5]

[3] Schools may assign different weights to the achievement of different groups of students; for example, they may assign higher weights to students at the margin of passing the standardized exams.

[4] The effectiveness of various curricula for different groups of students is the subject of a large observational and experimental literature, but many of these studies suffer from methodological problems. Many cross-sectional studies compare schools that endogenously chose different curricula, and randomized control trials often include only one teacher in the treatment or control group, possibly confounding curriculum with teaching quality. While a complete discussion of this literature is outside the scope of this paper, I refer the interested reader to the U.S. Department of Education's What Works Clearinghouse (WWC), which assesses research on various curricula (see http://ies.ed.gov/ncee/wwc/), and to the reports of the National Mathematics Advisory Panel (2008) and National Reading Panel (2000). The latter reports primarily focus on effective teaching practices (e.g., teacher-directed vs. student-centered math instruction, the importance of phonemic awareness and phonics for reading instruction) but also provide some discussion of curriculum materials related to these approaches.

[5] According to the CPS, 26% of within county moves are for family-related related reasons, 6% are for

Consider a simple linear representation of the student achievement function:

$$A_{ist} = X_{it}[\alpha + \beta(c_{st})] + M_{it}\lambda_{it} + \epsilon_{it} \tag{1.1}$$

where A_{ist} is the achievement of student i who attends school s at the time of the exam in year t. Achievement depends on student and family characteristics X_{it} and the school's curriculum c_{st}, which affects achievement through its interaction with student characteristics, i.e., $\beta(c_{st})$. Achievement also depends on student mobility, which is represented by M_{it}, an indicator for whether the student moved to school s from a different school. ϵ_{it} represents all other factors that affect student achievement.

λ_{it} is a coefficient that measures the effect of changing schools on achievement, and while it could be positive, research generally supports the notion that it is negative on average. For instance, a meta-analysis of 16 studies by Reynolds et al. (2009) finds that student mobility is significantly associated with lower achievement and dropping out. However, it is unclear whether these estimates should be interpreted causally since they may be confounded by factors that tend to be correlated with changing schools (e.g., parents' job loss, divorce, eviction). Nonetheless, a number of studies that include controls for prior achievement and demographic characteristics still find significant negative effects of mobility on achievement, especially for students who move multiple times (e.g., Temple and Reynolds 1999, Strand 2002, Hanushek et al. 2004).[6] Potential channels for this effect can be illustrated by ex-

---

work-related reasons, and 65% are for housing-related reasons (Schacter 2001a,b). Family and work-related reasons may be even more important determinants of mobility for New York City students, the majority of whom receive free lunch; Hanushek et al. (2004) report "among economically disadvantaged students [in the NLSY], . . . the prevalence of a divorce or a job loss by the father was more than 50% greater in the year of a school move than in years of non-moves." In New York City, while the list of schools that received exemptions from the standardized curricula was publicly available, information whether schools actually implemented it was not.

[6] In the U.K., which has a national curriculum, Strand (2002) finds that conditional on prior test scores and student characteristics, the math test scores of transfer students within one London Local Education Authority (LEA) are 0.12 standard deviations lower than those of their non-mobile counterparts. Lyle (2006)

pressing λ_{it} as follows:

$$\lambda_{it} = X_{it}[\beta(c_{\hat{s}t}) - \beta(c_{st})]\rho + \gamma 1[c_{\widehat{s}t} \neq c_{st}] + \delta_{it} \tag{1.2}$$

where $\rho \in$[0,1] is a weighting factor that accounts for the amount of time the student spent before the exam at her previous school $\hat{s}$ and was affected by its curricula $c_{\hat{s}t}$, and γ is the part of the mobility achievement gap that can be attributed to students having to change curricula when they change schools and is assumed to be less than or equal to 0. δ_{it} captures all the other effects of changing schools on achievement (e.g., loss of social networks).

Standardization changes the curriculum for all students and improves the continuity of instruction for mobile students. For non-movers, the difference between a student's achievement under a standardized curriculum $\tilde{c}$ and schools' chosen curriculum c_{st} can be shown to equal

$$A_{ist}(X_{it}, c_{st} = \tilde{c}, M_{it} = 0, \epsilon_{it}) - A_{ist}(X_{it}, c_{st} = c_{st}, M_{it} = 0, \epsilon_{it}) = X_{it}[\beta(\tilde{c}) - \beta(c_{st})] \tag{1.3}$$

while for movers, the difference equals

---

exploits a more plausibly exogenous source of variation in mobility — moves caused by military relocations — and finds reduced achievement among military children who have experienced 3 or more moves. While Hanushek et al. (2004) and Gibbons and Telhaj (2011) find evidence of peer effects of student mobility, the model excludes peer effects for two reasons. First, using five years of NYC data, I did not find any evidence of such peer effects in standard regressions of student achievement on peer characteristics and school-grade fixed effects. Second, it is unlikely that curriculum standardization could benefit the peers of mobile students without benefiting the mobile students themselves (i.e., by reducing the amount of instructional time needed to help new students catch up or reducing disruptive behavior caused by new students feeling lost in new material). Finally, estimates of mobility-related peer effects from Hanushek et al. in Texas and Gibbons and Telhaj in the U.K. are quite similar, even though the U.K. has a national curriculum; this suggests that mobility-related peer effects are unlikely related to changes in curricula.

$$\begin{aligned}
A_{ist}(X_{it}, c_{st} &= \tilde{c}, M_{it} = 1, c_{\widehat{st}} = \tilde{c}, \delta_{it}, \epsilon_{it}) - A_{ist}(X_{it}, c_{st} = c_{st}, M_{it} = 1, c_{\widehat{st}} = c_{\widehat{st}}, \delta_{it}, \epsilon_{it}) \\
&= X_{it}[\beta(\tilde{c}) - \rho\beta(c_{st}) - (1-\rho)\beta(c_{\widehat{st}})] - \gamma 1[c_{\widehat{st}} \neq c_{st}] \qquad (1.4)
\end{aligned}$$

As seen in 1.3, the effect of curriculum standardization on non-movers is determined entirely by the gains (or losses) from changing the curriculum. Whether this difference is positive or negative depends on the effect of the curriculum schools would have chosen in the absence of curriculum standardization. Importantly, 1.3 could be negative if schools would choose the achievement-maximizing curriculum for their expected student populations, but standardization forced them to adopt a different one. But, 1.3 could be positive if schools had not chosen the achievement-maximizing curriculum, for example, because they were maximizing an outcome other than measured achievement.

For movers, the effect of curriculum standardization is a weighted average of the gains from changing the curriculum and an extra term that measures the extent to which standardization reduces achievement losses associated with differences between schools' curricula. By the assumptions on γ, the second term in 1.4 (-$\gamma 1[c_{\hat{st}} \neq c_{st}]$) is non-negative. Thus, even if curriculum standardization resulted in the adoption of a curriculum that was not targeted toward the characteristics of movers, it could still benefit them by eliminating the differences between schools' curricula. Ultimately, however, the effect of curriculum standardization on student achievement is theoretically ambiguous and must be resolved empirically.

The model also highlights a separate empirical issue related to estimating

$$\hat{\lambda} = \underbrace{E(\lambda_{it}|X_{it})}_{\text{causal effect of changing schools}} + \underbrace{\frac{Cov(\epsilon_{it}, M_{it}|X_{it})}{Var(M_{it}|X_{it})}}_{\text{selection bias}} \qquad (1.5)$$

To estimate the causal effect of changing schools, I propose using the conditional achievement gap between students who move after the exam (or school year) and non-mobile students as an estimate of the bias term. These students do not change schools before achievement is measured, but they likely experience similar omitted factors (e.g., parents' job loss, divorce) as those who do change schools. If students who move before the exam are more negatively selected than those who move after, this estimate provides a lower bound on the magnitude of the bias term. Moreover, since we might expect all three terms in 1.2 to be non-positive (i.e., movers do not systematically move to schools with curricula that are worse for them, changing curricula harms achievement, and non-curriculum related effects of changing schools harm achievement), the difference between the OLS estimate of λ and the estimate of the bias term provides an upper bound on the magnitude of the achievement loss related to changing curricula, γ.

1.3 Institutional Background

In January of 2003, New York City Mayor Michael Bloomberg and District Schools Chancellor Joel Klein unveiled an ambitious educational reform agenda called "Children First." In addition to a number of organizational reforms, the plan called for standardized math and reading curricula to replace the 50 different math and 30 different reading programs that were being used in New York City's 1,200 schools.[7] Notably, the Mayor and Chancellor announced that the "top 200" schools, which were revealed weeks later, would be exempt from

[7]The reforms included: replacing the 40 community and high school districts with 10 regional "instructional divisions" and 6 support centers, increasing principal authority over hiring and budget decisions, closing local school district offices in city schools, putting a parent coordinator at each school, creating 10 support centers for parents, replacing elected school boards with parent engagement boards, reorganizing special education and bilingual programs, and ending social promotion. Importantly, all of these reforms were citywide, so their implementation did not depend on the same criteria as exemptions from the standardized curricula.

the standardized curriculum requirements for two years.[8] A timeline of the events related to the announcement and implementation of the standardized curricula reforms is displayed in Figure 1.1.

Although many people believed that the exemptions would simply go to the 200 highest scoring schools, which served predominately middle-class white and Asian students, the Department of Education (DOE) based exemptions on a formula meant to ensure racial and economic diversity. To determine exemptions, schools were separated into high-, medium- and low-need categories based on the percentages of students receiving free lunch, English Language Learner, and special education services. Different cutoffs, based on the percent of students passing math and English Language Arts exams in 2002, were established for each of the three categories.[9]

209 schools initially received exemptions based on these cutoffs, but schools that did not receive an exemption were given the chance to appeal. 233 schools or special programs in schools (e.g., gifted and talented) applied for an exemption during the appeals process, but fewer than half received them; 31 schools or special programs within schools received two-year exemptions, and 88 schools received one-year exemptions.[10] My analysis is based on the initial exemption cutoffs, but the appeals process explains why not all schools that

[8]Originally, the district intended to have a yearly re-evaluation process to extend and grant new exemptions. However, Diana Lam, the Deputy Chancellor in charge of curriculum and instruction, resigned in March 2004, and it is unclear if these re-evaluations occurred.

[9]For example, a school with 40 percent of students passing math and 60 percent passing English would receive a score of 100. The exemption cutoff for low-need schools was 140 for elementary and middle schools and 160 for high schools. Cutoffs for medium-need schools and high-need schools were set 15 and 30 points lower, respectively, and were lowered an additional 10 points for schools that ranked in the top 20% of schools making one-year gains citywide.

[10]Two-year exemptions in the appeals process were based on a formula similar to the one used to determine the original 209 exemptions, except that math and reading results were analyzed separately instead of being combined. Thus, some schools were only exempted from math, while others were only exempted from reading. One-year exemptions in the appeals process were granted to schools based on special circumstances: 43 because they had received grants to carry out other curricula, 33 because they lacked testing data, 10 because they had made significant improvements in recent years, and 2 because of special needs.

fell below the cutoff adopted the standardized curricula.

The standardized curricula were mandatory for all schools that did not receive an exemption, but exempt schools could still implement them voluntarily. In addition to selecting a single curriculum per grade and prescribing an instructional approach, the standardized curricula mandated the minimum amounts of time daily that schools must dedicate to reading, writing, and math. Complete details on the selected programs and their publishers can be found in the Appendix (Table A1). Details on the minimum time limits are also provided, although there is no evidence that these time limits were binding.[11]

The standardized reading curriculum started in fall 2003 and was based on a "balanced literacy" approach that combined "whole-language" and "phonics-based" approaches to reading instruction.[12] One key component of the reading curriculum was the purchase of books for classroom libraries, so students would read from books instead of basic readers. More controversial was the DOE's choice of *Month-by-Month Phonics* as its primary reading program for grades K-3. Although proponents argued that a less scripted approach would improve students' critical thinking skills, critics charged that it "[fell] short as an effective systematic phonics program [and its] effectiveness [had] not been validated scientifically" (Wolf 2007).[13]

The implementation of the standardized math curriculum was phased in over three years:

---

[11]A number of studies on the length of the school year, weather-related school closings, and subject-specific instructional time provide evidence that increasing instructional time might improve achievement (e.g., Pischke 2007, Marcotte 2008, Lavy 2010).

[12]Described in the DOE's handbook *A Comprehensive Approach to Balanced Literacy*, "balanced literacy" emphasized practices such as providing instruction in systemic phonemic awareness, phonics, spelling, and word study; teaching students strategies for reading and writing; providing time for independent reading; and engaging in shared reading. " Whole-language" emphasizes reading for meaning and memorizing words as whole units, while phonics teaches students to break down words into smaller components (e.g., letter combinations) that can be joined together to form words. However, NYC's "balanced literacy" approach drew fire for its more whole-language approach to reading instruction.

[13]The National Reading Panel (2000) has identified systematic phonics instruction as producing significant benefits for children's reading and writing skills. In response to this criticism, the DOE supplemented *Month-by-Month Phonics* with *Voyager Passport.* I have not found any credible studies on the effectiveness of *Month-by-Month Phonics* or *Voyager Passport.*

grades K-2, 6, and 9-12 by fall 2003; grades 3-5, and 7 by fall 2004; and grade 8 by fall 2005. The " constructivist" math curricula *Everyday Mathematics* and *Impact Mathematics* were selected for grades K-5 and 6-8, respectively.[14] There is little credible evidence on the effectiveness of *Everyday Mathematics*. The best study to date (Waite, 2000) compares the math test scores of 732 students in grades 3, 4, and 5 in 6 schools using *Everyday Mathematics* to those of 2,704 students in 12 schools matched on baseline achievement and student demographics that were using a more traditional math curriculum. While the study's author concluded that *Everyday Mathematics* had a statistically significant positive effect on math achievement, there were several methodological problems with the analysis that likely lead to an overstatement of the program's effects.[15] I have been unable to find any research studies on the effectiveness of *Impact Mathematics*.

[14] "Constructivist" or " reform" math is distinguished from " traditional" math by its student-centered approach; it emphasizes students constructing math knowledge through their own processes of reasoning and teaches multiple primary approaches to solving problems. Proponents of this approach argue that it fosters a deeper understanding of math, while critics counter that multiple methods are confusing and hinder students' computational skills. The Report of the National Mathematics Advisory Panel (2008) finds that high quality research does not support either exclusively "student-centered" or "teacher-directed" approaches to math instruction. But, the Panel also reports that " explicit-instruction" (i.e., providing clear models for solving a problem type, allowing students to think aloud while solving problems, and providing feedback) is beneficial for low-performing and learning disabled students. For example, *Everyday Mathematics* teaches four methods for addition: "partial sums," "column addition," "fast method," and "opposite change rule." Consider the problem 269+83. With "partial sums," students move left to right, adding up the place values represented by the digits in the column and then adding the partial sums, so the problem becomes 200+140+12= 352. In the "column addition" method, students write the addends in columns separated by vertical lines. In the first stage, students add the digits in each column $\begin{array}{rrr} 2| & 6| & 9 \\ +\ \underline{0}| & \underline{8}| & \underline{3} \\ 2| & 14| & 12 \end{array}$ and do the carries in the second stage. The "fast method" is the traditional right to left method of addition with carries. The "opposite change rule" has students add and subtract the same numbers to simplify the math by making one addend end in zero (e.g., 270+82). For a nice description of the *Everyday Mathematics* algorithms, see http://math.nyu.edu/~braams/links/em-arith.html

[15] In addition to the endogeneity of the schools' curricula choices, Waite dropped teachers from the sample if they had not followed the *Everyday Mathematics* curriculum to his liking (i.e., they reported using other instructional materials for a majority of the instructional time, their lesson plans did not use *Everyday Mathematics* curriculum daily, or their student assessments did not reflect the variety of the curriculum). In addition, a U.S. Department of Education review concluded that the effects estimated by Waite lose statistical significance when the standard errors allow for clustering.

New York City spent an estimated $35 million for the math and phonics materials and classroom libraries associated with the standardized curricula. To prepare teachers to use the new curricula, the DOE offered voluntary training sessions during the summer — attendance could not be required under the teachers' union contract — and conducted mandatory training during days prior to the start of school. Each school was also assigned a math and a reading coach — at a cost of around $55,000 each — to provide teachers with training and assistance with the new curricula during the school year, such as sharing instructional strategies, modeling sample lessons, and helping them with lesson plans.[16]

1.4 Data

I use several sources of data from the New York City Department of Education to study elementary and middle schools in the first two years after the inception of the standardized curricula, school years 2003-2004 and 2004-2005, hereafter referred to as the years 2004 and 2005, respectively. I do not examine effects in later years because, due to the two-year time limit on exemptions, there is no longer a discontinuity in curricula adoption by 2005-2006.

Data on schools' use of the standardized curricula come from administrative records of schools' purchases of standardized curricula materials (e.g., textbooks, workbooks, teacher manuals, etc.), some of which are re-ordered each year. Data on purchases of materials for non-standardized curricula, before standardization was implemented or for exempt schools in the years afterward, are unavailable, but conversations with DOE personnel suggest that, regardless, schools often combined pieces of different programs together to form their own unique curriculum.[17] The only exception is that some schools are listed on DOE documents

[16]Schools that did not adopt the standardized curricula received $110,000 to use at their discretion, $55,000 for each coach.

[17]I have limited data on prior curricula from the 2002 school report cards, where schools could voluntarily list any curriculum programs they used. See Appendix Figure A3 and Appendix Table A3.

as using two pieces of the standardized curricula (*Everyday Mathematics* and *Month-by-Month Phonics)* prior to the citywide reform. I omit them from the math and/or reading analysis when applicable.[18]

Data on test scores cover students in the 3rd through 8th grades, who are tested annually in both math and reading. For math, due to its staggered implementation, I focus on students in the 3rd through 6th grade in 2004 and 3rd through 7th grade in 2005. For reading, I focus exclusively on students in 3rd grade, since the reading curriculum for 3rd grade (i.e., *Month-by-Month Phonics*) was better defined than that of higher grades, which were simply required to follow the balanced literacy instructional approach. Student data also include demographic characteristics and program participation (e.g., race, gender, free lunch, English Language Learner, special education, place of birth), and identifiers that allow students to be linked across years. More information on these data can be found in Kane et al. (2008).

To examine mobility, I use data on student enrollment which contain the dates on which students were admitted to or discharged from New York City schools. This is distinct from many studies on student mobility, which rely on one observation per student per year and only observe students who move schools between school years (National Research Council and Institute of Medicine 2010).[19] I use these data to distinguish between different types of mobility: (1) non-structural summer moves, which *exclude* moves related to school structure (e.g., moves that occur from elementary to middle school or from schools that closed) (2) moves during the school year that occur prior to an exam, and (3) moves during the school year occurring after an exam. My focus on students who experience non-structural moves

[18]District 15 was already following the *Everyday Mathematics* and *Month-by-Month Phonics* curriculum, and District 10 was already following *Everyday Mathematics.* District 15 schools have test scores that are around 0.1 standard deviations better than the NYC average, while District 10 schools have test scores that are about 0.2 standard deviations worse. Results including these schools are qualitatively similar, although the first stages are slightly weaker because these schools were less likely to respond to the district's mandate.

[19]Hanushek et al. (2004) observe one observation per student per six-week period.

follows previous research on student mobility and is useful because these students may be the most likely to benefit from curriculum standardization.[20] The distinction between moves before and after exams is also important for my empirical work; moves that occur after the exam cannot have a causal effect on student achievement and can be used to test the validity of interpreting estimated effects of mobility on achievement as causal. This is similar to the strategy used by Herrmann and Rockoff (2010) to examine the causal impact of teacher absence on student achievement.

Additional data include the list of the 159 elementary and middle schools that received exemptions based on test score cutoffs, information on schools' need categories (high, medium, and low), and the percentages of students who passed the math and reading exams.[21] Schools received an exemption if (i) the sum of the percentage of students who passed the math and reading exams met or exceeded the cutoff for their need category, or (ii) this sum was within 10 points of their cutoff and they placed in the top 20 percent of one-year gains citywide. Despite the efforts of district staff, the exact formula the DOE used to calculate gains cannot be located. However, using my calculations of the average gains students made between 2002 and 2001, I am able to match the first round exemption list relatively well. Thus, I use these calculations to assign the top 20 percent of schools a cutoff 10 points lower than the normal cutoff for their need category. Details on these calculations can be found in the Appendix (Figure A1).

Figure 1.2 displays a scatter plot of the probability of receiving a first round exemption on the difference between schools' scores and their cutoff, averaged in 5-point bins. There is a clear jump in the probability of receiving an exemption at the cutoff, with schools above

[20]Since moves due to school structure are usually predictable, schools may already align curricula for students that experience structural moves.

[21]50 high schools also received exemptions.

the cutoff about 60% more likely to receive an exemption.[22] This evidence helps establish the possibility of using a regression discontinuity design to evaluate the effect of curriculum standardization on student achievement, although equally important is whether receiving an initial exemption affects the probability that schools actually implement the standardized curricula.

As exemptions were based on performance on tests almost one year before the announcement of the Children First reforms, it is unlikely that schools manipulated their eligibility. As a further check on the exogeneity of the exemption cutoffs, Figure 1.3 shows that the density of schools' scores relative to the exemption cutoff is smooth, and the statistical density test proposed by McCrary (2008) does not uncover any evidence of manipulation.

Table 1.1 presents the 2002 baseline characteristics of the schools included in the analysis, separated by whether they fall above or below the exemption cutoff.[23] For both math and reading, student test scores are rescaled to have a mean of 0 and standard deviation of 1 within each grade and year. Not surprisingly, schools above the exemption cutoff have higher test scores, and fewer black, Hispanic, and free lunch students.

Table 1.2 provides summary statistics on the implementation of the standardized curricula. Implementation rates are substantially higher for schools below the cutoff than for those above. Over 90% of the schools below the cutoff implemented the 3rd grade standardized reading curriculum in 2004 and 2005, compared to about 30% and 40%, respectively, of schools above the cutoff. For math, the rates of implementation are lower in 2004 than in 2005, since schools were not required to implement the standardized curriculum for grades 3-5 until 2005. However, by 2005, nearly 90% of schools below the cutoff had implemented

[22]This is partially due to measurement error in the assignment of schools to the top 20% or bottom 80% citywide.

[23]Schools that serve students with moderate to severe disabilities and very low-performing schools managed by the district's central office are omitted from the analysis, though including these schools does not noticeably change my results.

the math curricula, compared to 40% of those above.

1.5 Empirical Approach

The DOE's policy of granting exemptions from its standardized curricula based on whether schools' pre-determined test scores met or exceeded certain cutoffs presents an opportunity to evaluate the effectiveness of these curricula using a regression discontinuity design. The main idea behind the RD design, first introduced by Thistlethwaite and Campbell (1960) and since used in many studies in education, is that the schools whose scores fell just above the cutoff, and were exempt from implementing the standardized curricula, are a good control group for schools whose scores fell just below the cutoff and were required to implement the standardized curricula.[24] As long as any systematic relationship between schools' scores and student outcomes is smooth through the cutoff, the discontinuity in the probability of treatment can be used to estimate the causal effect of the treatment. Consider the reduced form regressions:

$$A_{ist} = \alpha + \beta[Z_s < 0] + f(Z_s) + \phi_s + X_{ist}\lambda + \varepsilon_{ist} \tag{1.6}$$

$$SC_{ist} = \mu + \delta[Z_s < 0] + f(Z_s) + \theta_s + X_{ist}\pi + \eta_{ist} \tag{1.7}$$

where A_{ist} is the test score of student i in school s in year t; SC_{ist} is an indicator variable for whether the student's school implemented the standardized curricula; $1[Z_s < 0]$ is an indicator for whether the school's rescaled score Z_s is less than 0; $f(Z_s)$ is a flexible control function for the school's rescaled score (i.e., a linear model estimated separately on both sides of the cutoff); ϕ_s and θ_s are cutoff fixed effects; X_{ist} is a vector of year-grade

[24]For example, these studies include van der Klauww (2002); Jacob and Lefgren (2004); Chay, McEwan, and Urquiola (2005).

fixed effects and student and school-grade peer characteristics (e.g., race, gender, free lunch, English Language Learner, special education, foreign born, moved prior to the exam, moved prior to the school year); and ϵ_{ist} and η_{ist} are error terms, which are clustered at the school level.

The coefficient β provides the " intent-to-treat" estimate of how much a student's test score improves if her school's score is just below, rather than just above, the cutoff for its need and gains category. The coefficient δ estimates the "first-stage" relationship between the school's score falling below the cutoff and its implementation of the standardized curricula. Taking the ratio $\frac{\beta}{\delta}$ yields the treatment effect that would be estimated from a two-stage least squares (2SLS) instrumental variables regression. Under the assumptions of monotonicity (i.e., falling short of the cutoff cannot cause some schools to adopt the standardized curricula while causing others to reject it) and excludability (i.e., falling short of the cutoff cannot affect student achievement except through the implementation of the curricula), this ratio estimates the causal effect of the treatment (Lee and Lemieux 2010).

In this present case, the 2SLS approach corresponds to a "fuzzy," rather than "sharp," regression discontinuity design (RD), because neither the first round exemptions nor the adoption of the curricula are perfectly determined by falling below the cutoff. For instance, some schools below the cutoff received exemptions in the appeals process, and schools that received exemptions because they were above the cutoff could still voluntarily adopt the standardized curricula. As illustrated in the conceptual framework, the treatment effect estimate $\frac{\beta}{\delta}$ represents a composite of gains (or losses) from changing curricula and gains from reducing the losses associated with moving.

Since only a small minority of students is mobile, one would expect the RD estimate to reflect primarily effects due to the changes in curricula. However, whether standardization has differential effects on movers is still of interest. Thus, I also estimate separate specifications for different types of mobile students.

1.6 Results

Before moving to the main results, Table 1.3 displays balancing tests for student characteristics. Columns 1 and 2 in Panel A verify that there are no pre-treatment differences in students' 2002 math and reading test scores. I use the same specification, grades, and bandwidths as the main regression results; my preferred bandwidths are 39 for math and 22 for reading, which were calculated using the "leave-one out" procedure described in Imbens and Lemieux (2008) (see the Appendix for more details). The remaining columns use the data from years 2004 and 2005 and present the results from a regression of each characteristic on an indicator for being below the cutoff, controlling for cutoff and year-grade fixed effects and separate linear trends on both sides of the cutoff. These columns illustrate that the observable characteristics are balanced across the cutoff; only one characteristic, female, is even marginally statistically significant, and this difference, 1 percent, is small in magnitude.

Figure 1.4 provides a graphical analysis of the first stage results for the effect of being below the cutoff on implementing the standardized curricula. Each panel in Figure 1.4 plots the average residuals from a regression of the fraction of students for whom the school implemented the standardized curricula on cutoff and year fixed effects and the fitted values from locally weighted regressions of the residuals on schools' distances from the cutoff.[25] Panel A shows that schools just below the cutoff were about 20% more likely to implement the standardized math curricula than those just above. There is not full compliance below the cutoff because the curriculum requirement for grades 3-5 was not binding until 2005. Panel C shows that for reading, schools just below the cutoff were around 30% more likely to implement the standardized curriculum than those just above, a larger discontinuity than in math. Panels B and D display similar residual plots where the school's average test score in 2004 and 2005 replaces the use of the standardized curriculum as the dependent variable.

[25] In this figure, the unit of analysis is the school, and each school receives the same weight.

In contrast to Panels A and C, there are no sharp discontinuities in test scores at the cutoff for receiving an exemption.

A formal RD analysis is presented in Table 1.4, which reports the results from regressions using student-level data, which are analogous to the specifications in Table 1.3, Columns 1 and 2. However, in Table 1.4, I add one more panel for math results that control for cubic polynomials in prior math and reading test scores. These controls help to reduce the standard errors of the test score outcomes but are only available for grades 4-7, so I cannot add a corresponding panel for the reading sample, which only includes third graders. The regressions in Table 1.4 control for cutoff and year-grade fixed effects, separate linear trends on both sides of the cutoff, and student and school-grade peer characteristics. The first three columns report the estimated discontinuities in the implementation of the standardized curricula for three different bandwidths around the cutoff: (i) 58 points (the maximum distance on the right side of the cutoff), (ii) the cross-validation bandwidths (39 points for math and 22 points for reading), and (iii) a bandwidth 5 points within the cross-validation bandwidth (34 points for math and 17 points for reading).

These results show that there is a relatively strong first stage for the implementation of the math and reading standardized curricula. However, as expected from Figure 1.4, Columns 4-6 find no significant effect on test scores. The reduced form point estimates for the cross-validated bandwidths are all very close to zero, -0.004 standard deviations for math (when controlling for prior test scores) and -0.007 standard deviations in reading. The standard errors of these estimates are around 0.02 in math and 0.04 in reading, ruling out " intent-to-treat" improvements of more than 0.03 standard deviations in math and 0.08 standard deviations in reading at the 95% confidence level. While the estimates are somewhat imprecise, the associated 95% confidence interval for the 2SLS estimate for 4th and 5th graders in math, who used the *Everyday Mathematics* curriculum, is [-0.33, 0.24], ruling out an improvement of 0.28 standard deviations, the effect size that Waite (2000) reports for *Everyday Mathematics*.[26] The corresponding confidence interval for the 2SLS estimates for the balanced literacy reading program (i.e., *Month-by-Month Phonics*) is [-0.26, 0.22]. For comparison, in a randomized trial, Borman et al. (2007) found that a *different* literacy and school reform program *(Success for All)* increased test scores on the Woodcock Word Attack scale by a statistically significant 0.22 standard deviations in its first year and 0.25 in its second year.[27] The estimates rule out similarly large effects for NYC's standardized curricula, although it is worth noting that *Success for All* is a very intensive intervention.[28]

Like any study based on RD methodology, I can only identify the local average treatment effect (LATE) of the standardized curricula for schools that were close to the exemption threshold, which were relatively high-performing schools within each school need category. One might expect the average treatment effect (ATE) of curriculum standardization to be higher, since standardization may have the largest benefits for low-performing schools. Unfortunately, the way the curricula were implemented does not allow me to cleanly disentangle whether any improvements in the test scores of low-performing schools are related to the

[26]This estimate is not shown in the table. Evaluating the effectiveness of this curriculum requires disaggregating by grade, since *Everyday Mathematics* was implemented for grades 3-5 and *Impact Mathematics* was implemented for grades 6 and 7.

[27]They found no statistically significant effects on three other reading measures in the first year but found statistically significant effects on two other literacy measures in the second year. Effect sizes on these measures ranged from 0.21 to 0.36 standard deviations in the third year (Borman et al. 2007).

[28]Components of *Success for All* include: grouping students by performance levels for reading lessons, assessing students at 8 week intervals for regrouping, creating a Family Support Team to identify and address problems that interfere with school performance (e.g., poor attendance, problems at home), and designating a full-time Program Facilitator. See Borman et al. (2007) for more details about *Success for All*. The cost-effectiveness of curriculum standardization relative to other education interventions can also be evaluated. Assuming that the standardized curricula affected around 80% of students in 1,000 New York City schools, a rough estimate of this program's per-pupil cost is around $180. Of course, this estimate may be too high because it does not account for the cost of the curriculum materials that would have been purchased in the absence of standardized curricula. The standardized curricula were estimated to produce less than 0.17 standard deviations gain in math and less than 0.25 standard deviations gain in reading. However, since improvements of 0.20 standard deviations are usually orders of magnitude more expensive, NYC's standardized curricula do not seem particularly cost ineffective. For example, Project STAR, the famous class size experiment conduced in Tennessee, cost $16,000 per 0.2 standard deviations gain; *Success for All* was considerably cheaper, at $1,500 to $2,600 per 0.2 standard deviations (Loeb and McEwan 2010).

standardized curricula, other Children First reforms (e.g., changes to the special education and ELL programs), or mean reversion (see Chay et al. 2005).[29]

However, I can assess whether curriculum standardization has heterogeneous effects by school type by disaggregating the results by school need groups (high-, medium-, and low-), which were one determinant of schools' cutoffs. Exemptions were granted to the highest performing schools within each need group, but exempt schools in higher need groups generally had lower baseline achievement than exempt schools in lower need groups. The results are presented in Table 1.5. The main take-away from this table is that the exemption criteria primarily induced medium need schools to implement the standardized curricula. This appears to be related to general differences in the rates of implementation by need group; below the cutoff, low need schools were the least likely to implement the curricula, possibly because they may have been more likely to apply for and utilize second round exemptions, while above the cutoff, high need schools were the most likely go along with the district and implement the curricula. Nonetheless, this table also finds no significant effect of the standardized curricula on students' math or reading scores for any need group.

In addition to the type of schools, the effect of curriculum standardization could also depend on the type of students (e.g., low-performing versus high-performing students). For instance, in a randomized control trial in Kenya, Glewwe et al. (2009) find that providing textbooks only increased the test scores of students with high initial achievement, likely because they were the only ones who could read the textbooks, which were written in English. Similarly, the balanced literacy and constructivist math curricula used in New York City might have been better suited for high achieving students since the less scripted approach

[29]The RD design allows the effect of the standardized curricula to be disentangled from the other Children First reforms at the cutoff, since the other reforms were citywide and student characteristics are smooth through the cutoff. Concurrent reforms are a generally a problem for identifying the effect of curriculum standardization. For example, in Los Angeles, a standardized curriculum was simultaneously implemented for the entire district at around the same time as reductions in class size, and in Pittsburgh, a standardized curriculum was implemented along with reforms in principal pay and teacher mentoring.

to literacy and student-centered approach to math might have required highly educated teachers and well-behaved students. Thus, I also estimated specifications that disaggregated the results by student characteristics (including prior test scores for math), measures of teacher quality (teacher experience, whether teachers have a master's degree), and the year of implementation.[30] These results, which are available from the author upon request, also fail to uncover differential effects of the standardized curricula for any group.

Despite finding no significant effects on overall student achievement or the achievement of various groups of students, it is still reasonable to believe that curriculum standardization could benefit mobile students. In the sample, about 6 percent of students move during the school year (5.5 percent before the math exam) and about 9 percent of students move for non-structural reasons (e.g., residential mobility) each summer. Table 1.6 provides information on the characteristics of these movers by reporting the coefficients from probit regressions of an indicator for each of type of move on prior math and reading test scores, indicators for missing test scores, controls for student characteristics, and year-grade fixed effects.[31] These regressions confirm that mobile students have lower achievement even before their moves. Of students who move during the school year, those who move before the exam appear more negatively selected on observables than those who move afterwards. Similarly, of students who experience non-structural moves during the summer, those who move before the current school year appear more negatively selected on observables than those move after the current school year.

Figure 1.5 provides more detailed information on the origin and destination schools of mobile students. The panels illustrate the composition of moves that occurred before the

[30] The separate analysis by year of implementation addresses the concern that the effects of curriculum standardization might be significant and positive once teachers learn the new curricula. It is also worth nothing that other studies (e.g. Waite (2000) and Borman et al. (2007) find effects in the first year.

[31] Test scores are missing for students from other districts and for 3rd graders.

math exam and for non-structural reasons before the school year. This figure shows that schools below the cutoff receive the majority of students who move and that a small number of students cross the cutoff when they change schools.[32] The fact that a large portion of movers come from outside the school district suggests that district-wide curriculum standardization may have little scope to improve the achievement of the average mover.

Table 1.7 reports results where the RD is separately estimated by three types of students: 1) those who experience no mobility before the school year or exam, 2) those who experience mobility during the school year before the exam, and 3) those who experience non-structural mobility before the school year. For mobile students, the distance from the cutoff is coded based on their destination schools, so the RD for mobile students compares those who moved to schools just below the cutoff to those who moved to schools just above. I only estimate effects for non-structural movers in 2005 because standardization only improves instructional continuity for students who moved in the summer of 2004. In addition, since one might expect standardization to primarily benefit students who moved between schools that used the standardized curricula, I also estimate specifications that restrict the sample of movers to those whose origin schools fell below the cutoff.

This table does not provide strong evidence that the curriculum standardization benefits mobile students in general or mobile students who were most likely to move between schools using the same curricula.[33] There is one coefficient for movers before the reading exam that is positive at the 5 percent level, but it is based on a very small sample. In addition, its first stage does not appear to have the power to identify an effect, and this result is not corroborated by the results in math. Nevertheless, this raises the question of how much

[32]There are more schools below the cutoff than above, but rates of mobility are still disproportionately higher in schools below the cutoff.

[33]The large decline in the sample size for movers between Panel A and Panel B is due to the fact that about 40% of mobile students in grades 4-7 lack prior test scores.

curriculum standardization could plausibly be expected to increase the achievement of mobile students.

This question is explored in Table 1.8, which evaluates how much scope there is for curriculum standardization to reduce the achievement gap associated with student mobility. A unique feature of the data used for this study is that they contain information on the exact timing of students' moves. This enables me to conduct a test of how much of the estimated effects of student mobility on achievement could be causally attributed to changing schools. Since moves after the exam – either during the school year or during the summer – cannot have a causal effect on student test scores, the relationship between these moves and test scores likely provides a lower bound on the amount of omitted variable bias that could be generated by unobserved factors that are correlated with student mobility (e.g., parents' job loss, divorce).[34] For this analysis, I use a sample of 3rd through 7th graders in both math and reading in 2004 and 2005; this allows me to control for prior test scores in the results for reading, which is not possible in the previous reading sample that only includes 3rd graders.

Columns 1 and 3 of Table 1.8 report the coefficients from regressions of student test scores on indicators for different types of moves, controlling for student characteristics and year-grade fixed effects. These columns demonstrate that student mobility is clearly associated with lower test scores. The achievement gap for students who move before the exam is -0.35 standard deviations in math and -0.23 in reading; for students who move before the school year, the gap is -0.13 in math and -0.10 in reading. However, both the results from Table 1.6 and the significant achievement gaps for students who move after the exam or school year suggest that this relationship could be driven by selection. Columns 2 and 4 follow previous literature on student mobility and attempt to deal with selection bias by controlling for cubic

[34]It is a lower bound if students who move before the exam are more negatively selected than those who move after the exam. Table 1.6 suggests that this is likely the case for observable characteristics, but it hard to say whether this true for unobservable factors.

polynomials in students' prior math and reading scores. Consistent with the results being driven by selection, the inclusion of controls for prior test scores significantly reduces all of the estimated effects of mobility on achievement. In math, the coefficient on moves before the exam falls from -0.35 to -0.06, and the coefficient on moves before the school year falls from -0.13 to -0.01; similar reductions occur in reading.[35] Moreover, once I control for prior test scores, the effects of mobility after the exam or after the school year are at least as large as the effects of mobility before the exam or before the school year.

Figure 1.6 extends the analysis of the relationship between the timing of moves and achievement. This figure plots the coefficients and 95 percent confidence intervals from a regression of test scores on indicators for student mobility at different times around the math or reading exam (i.e., more than 7 months before the exam, 4-7 months before the exam, 0-3 months before the exam, 0-3 months after the exam, and more than 3 months after the exam), controlling for prior test scores, student characteristics, and year-grade fixed effects. If mobility only causes low achievement by interrupting the continuity of instruction, one might expect students who move in the middle of the pre-exam period to have the largest declines in achievement; relative to students who move in the middle of pre-exam period, students who move early in the year or very close to the exam have longer periods of continuous instruction. This causal pathway would also imply that there should be no achievement losses for students who do not change schools until after the exam. If, on the other hand, achievement losses are caused by omitted variables (e.g., divorce, job loss) that are correlated with mobility, one might expect these losses to reflect the correlation between the omitted variables and achievement. Mobility after the exam could appear to have a large negative effect on achievement if it indicates the student had a difficult home life during the exam period. Consistent with the omitted variables story, there are large

[35] Some of this reduction is due to changes in the sample composition.

achievement gaps for students who move after the exam, and these are similar in magnitude to the achievement gaps for the equivalent time windows before the exam. Taken together, these findings suggest that factors correlated with mobility could potentially explain all of the achievement gap between mobile and non-mobile students, leaving little scope for improvement through curriculum standardization.

1.7 Discussion and Conclusions

A major debate over the benefits of centralization has emerged in the education sector over the issue of curriculum standardization. To shed light on this issue, I examine a substantial reform that standardized math and reading curricula for students in New York City public schools. I estimate the effect of standardization on student achievement using a regression discontinuity design. Schools that fell below a sharp exemption cutoff were more likely to adopt the standardized curricula, but these curricula changes did not have a significant effect on student achievement. The "intent-to-treat" estimates for falling below the cutoff are insignificant and very close to 0; the 2SLS estimates are fairly imprecise — [-0.22, 0.17] standard deviations in math and [-0.26, 0.22] standard deviations in reading — but exclude prior estimates from some other studies of curricula.[36] I also find no evidence that curriculum standardization significantly benefits mobile students. Importantly, this is because negative selection, not differences in schools' curricula, appears to be the main cause of the achievement gap between mobile and non-mobile students.

Since this paper examines the effect of specific math and reading curricula implemented in New York City (e.g., *Everyday Mathematics*, *Month-by-Month Phonics*), it is unknown to what extent these results may generalize to other contexts or curricular choices. However, this study does highlight a few broad facts about curriculum standardization which may

[36] These estimates for math include students in grades 4-7 and control for prior test scores.

be applicable in other settings: First, it is unlikely to improve the achievement of mobile students. Since most of the mobility achievement gap appears to be related to negative selection, schools seeking to improve the achievement of mobile students should focus on more targeted educational interventions (e.g., tutoring) rather than standardization. Moreover, even if differences in schools' curricula do harm the achievement of mobile students, standardization is limited in its ability to help mobile students who come from outside of its area of implementation (e.g., other districts, states, or countries). This paper finds that many mobile students in NYC come from outside of the school district, and Strand (2002) finds that a large fraction of mobile students in the U.K., which has a national curriculum, come from abroad.

Second, given that curriculum standardization is unlikely to benefit mobile students, the relevant debate should be over whether the proposed standardized curricula are more effective at raising students' achievement than their alternatives. Standardization may be warranted if there are curricula that raise the achievement for all types of students, and schools are not using them. On the other hand, if the effectiveness of curricula depends on the type of students, standardization is only preferable if it mandates curricula that are better matched to students' characteristics. The effectiveness of various curricula remains an important topic for future research, but this study suggests that implementation of standardized curricula may have smaller effects on achievement than some previous studies might suggest.

Finally, this study only examines student achievement and does not account for any other possible benefits of curriculum standardization, such as reductions in administrative costs or quantity discounts from ordering large amounts of curriculum materials. It is worth pointing out that even if standardization reduced test scores, it could still be justified if it resulted in sufficiently large cost-savings.

Figure 1.1: Timeline of Events Related to the Implementation of the Standardized Curricula

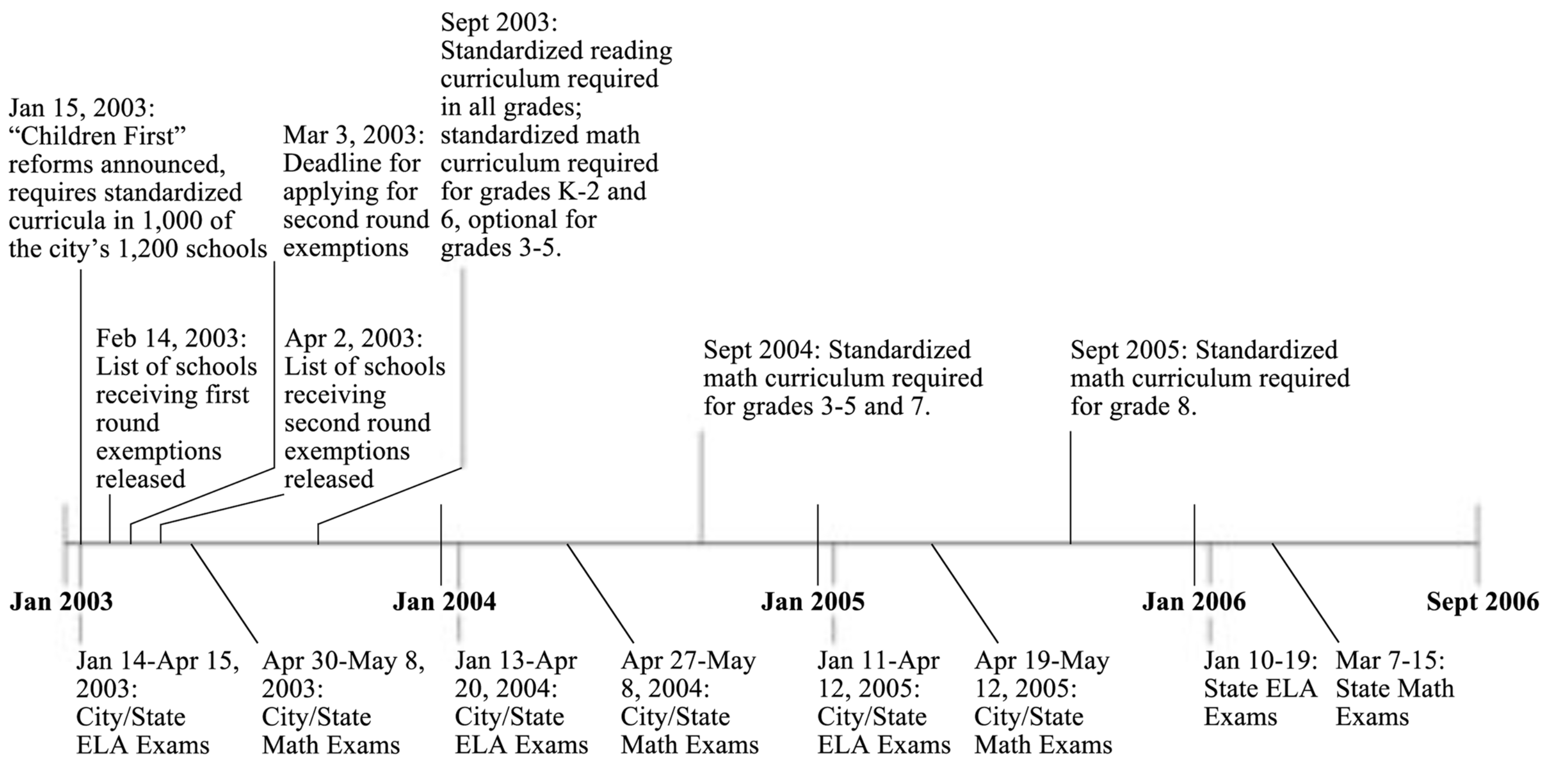
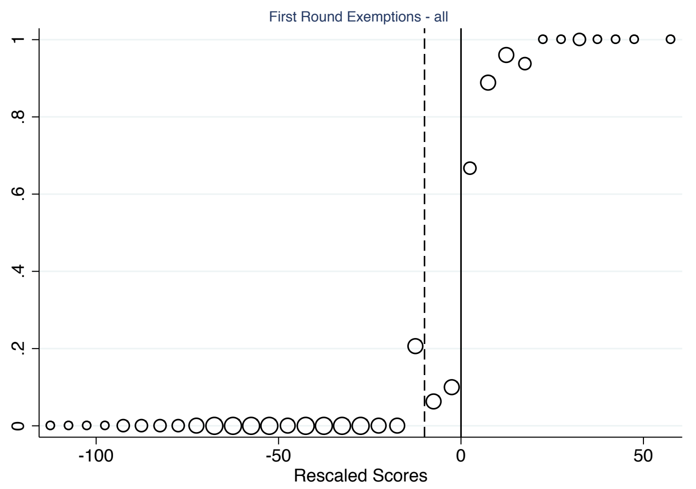


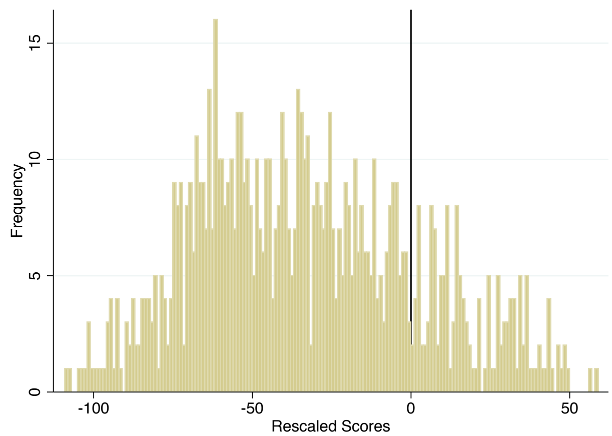
Figure 1.2: First Round Exemptions



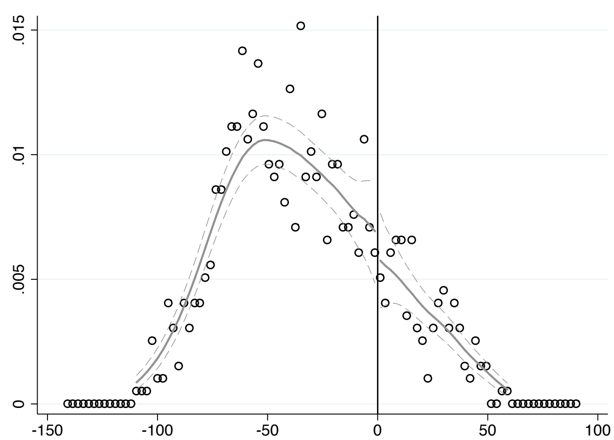
Notes: The figure plots the probability of receiving a first round exemption on the difference between schools' scores and the cutoff for their need group and improvement category.

Figure 1.3: Density of Rescaled Scores

Panel A: Histogram



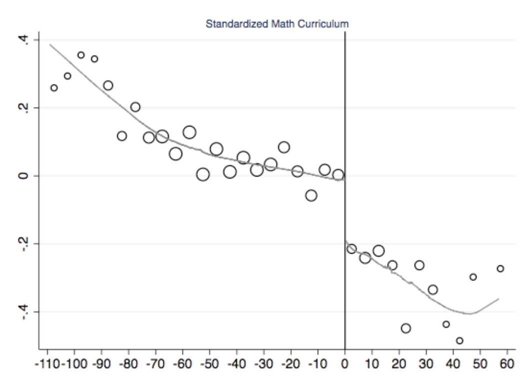
Panel B: Density Test



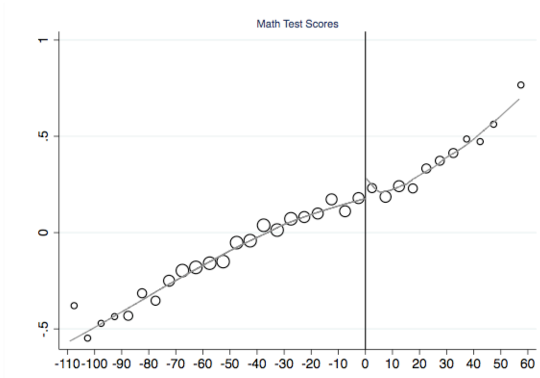
Notes: Panel A plots a histogram of rescaled scores for the schools used in the analysis.Panel B plots the density test proposed by McCrary (2008) for these schools, which has a coefficient and standard error of -0.1532 and 0.2501, respectively.

Figure 1.4: Standardized Curricula and Test Scores

Panel A: Standardized Math Curricula

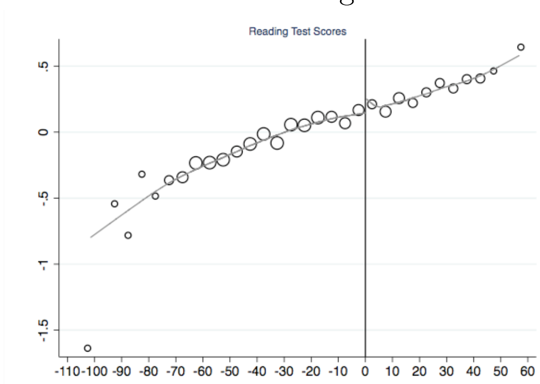


Panel B: Math Test Scores



Panel C: Standardized Reading Curricula

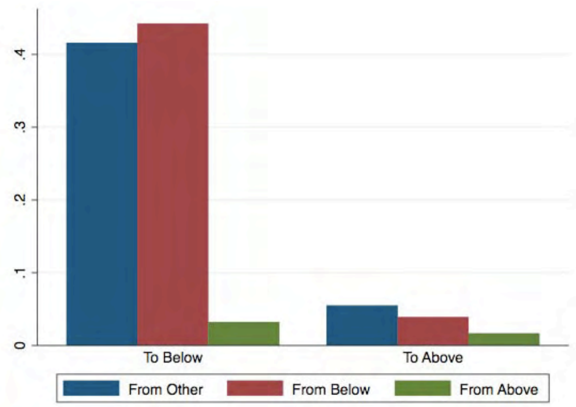
Panel D: Reading Test Scores



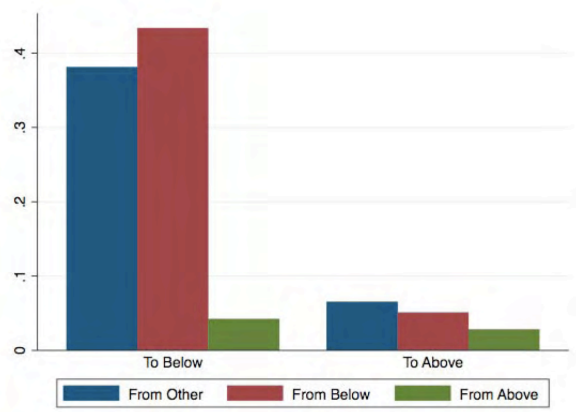
Notes: All panels plot residuals from a regression of the dependent variable on cutoff and year fixed effects and the fitted values of locally weighted regressions of these residuals on schools rescaled scores, estimated separately on each side of the cutoff using Statas *lowess* command and a bandwidth of 0.8. The dependent variable is the fraction of students for whom the school implemented the standardized curricula in Panels A and C and the average test score in Panels B and D.

Figure 1.5: Origin and Destination Schools' Positions Relative to Cutoff

Panel A: Moves During the Year Before the Math Exam



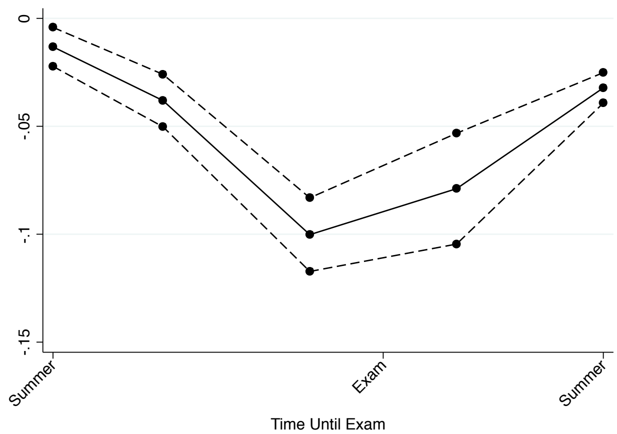
Panel B: Moves Before the Year for Non-structural Reasons



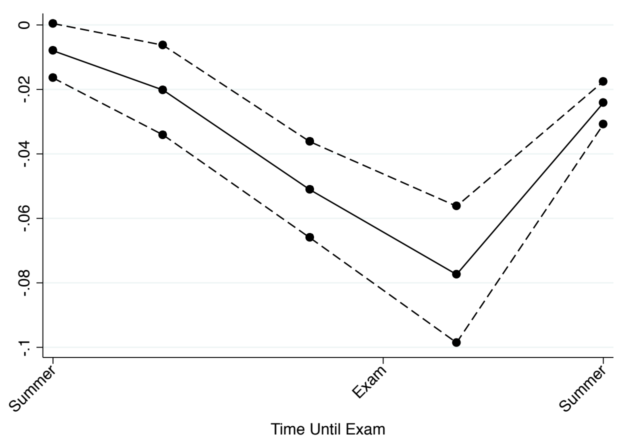
Notes: These statistics include students in the analysis sample on the date of the math exam and are averaged over the years 2004-2005. In each panel, the fraction of students sums to 1. Other schools mostly includes students from outside of the district (i.e., missing previous school information and not previously enrolled in the district) but also includes late enrollments (i.e., students who were previously enrolled in the district but who were not enrolled in a NYC school until after October 1 of the current school year) and students from schools whose position relative to the cutoff is undefined (e.g., schools were not open in 2002).

Figure 1.6: Mobility Achievement Gaps by Timing of Mobility

Panel A: Math



Panel B: Reading



Notes: These figures plot the coefficients and 95 percent confidence intervals from a regression of test scores on indicators for student mobility at different times around the math and reading exam (>7 months before the exam, 4-7 months before the exam, 0-3 months before the exam, 0-3 months after the exam, >3 months after the exam). Both the math and reading samples contain students in grades 3-7 in 2004 and 2005. All specifications include year-grade fixed effects and control for cubic polynomials in prior math and reading test scores and student characteristics (Female, Black, Hispanic, Other Race, Foreign Born, Free Lunch, English Language Learner, Special Education, Mover Prior to the Exam, Non-Structural Mover Prior to the School Year).

Table 1.1: Descriptive Statistics - Baseline 2002

Characteristic	Below	Above
Math Test Score	-0.09	0.70
	(0.33)	(0.28)
Reading Test Score	-0.10	0.57
	(0.31)	(0.30)
Female	0.49	0.49
	(0.03)	(0.05)
Black	0.41	0.13
	(0.31)	(0.17)
Hispanic	0.40	0.20
	(0.27)	(0.16)
Asian/Other Race	0.08	0.26
	(0.12)	(0.21)
Foreign Born	0.12	0.14
	(0.08)	(0.09)
Free Lunch	0.87	0.59
	(0.16)	(0.27)
English Lang. Learner	0.07	0.01
	(0.08)	(0.02)
Special Education	0.07	0.03
	(0.06)	(0.05)
Moved During Year - Before Exam	0.05	0.03
	(0.03)	(0.02)
Moved Prior to Year - Non-Structural	0.09	0.07
	(0.04)	(0.03)
Number of Schools	675	151
Number of Students	307336	50208

Notes: The sample includes students in grades 3-7 who were enrolled in schools in the analysis sample on the date of the math exam. The columns display school-level means of student characteristics.

Table 1.2: Summary Statistics on Implementation

	2004		2005	
Variables	Below	Above	Below	Above
Standardized Math Curricula	0.62	0.13	0.92	0.41
	(0.45)	(0.32)	(0.24)	(0.48)
Standardized Reading Curricula	0.94	0.26	0.96	0.36
	(0.23)	(0.44)	(0.20)	(0.48)

Notes: For math, the school-level means and standard deviations of the variables are presented for students in grades 3-6 in 2004 and grades 3-7 in 2005 who were enrolled in schools in the analysis sample on the date of the math exam. For reading, statistics are presented for students in grade 3.

Table 1.3: Balancing Tests

	2002 Test Scores		2004-2005 Characteristics			
Panel A	Math Score	Reading Score	Female	Black	Hispanic	Other Race
	(1)	(2)	(3)	(4)	(5)	(6)
1[score<cutoff]	-0.0203	0.0003	0.0101+	0.0355	-0.0082	0.0229
	(0.0201)	(0.0340)	(0.0058)	(0.0514)	(0.0392)	(0.0408)
Controls for Characteristics	Yes	Yes	No	No	No	No
R-squared	0.325	0.227	0.000	0.065	0.099	0.057
N	166509	23692	317162	317162	317162	317162
			2004-2005 Characteristics			
Panel B	Free Lunch	ELL	Special Ed	Moved Before Exam	Moved Before Year	Foreign Born
	(1)	(2)	(3)	(4)	(5)	(6)
1[score<cutoff]	0.0428	0.0091	-0.0061	-0.0046	-0.0044	0.0148
	(0.0337)	(0.0172)	(0.0093)	(0.0033)	(0.0046)	(0.0186)
Controls for Characteristics	No	No	No	No	No	No
R-squared	0.209	0.024	0.003	0.002	0.003	0.013
N	317162	317162	317162	317162	317162	317162

Notes: All specifications include cutoff and grade-year fixed effects and separate linear trends on both sides of the cutoff. All specifications use a bandwidth of 39 and students in grades 3-7, except for column (2), which uses a bandwidth of 22 and only includes students in grade 3. Characteristics controls include controls for student and school-grade peer characteristics (Female, Black, Hispanic, Other Race, Foreign Born, Free Lunch, English Language Learner, Special Education, Mover Prior to the Exam, Non-Structural Mover Prior to the School Year). Standard errors are clustered by school and displayed in parentheses. Significance levels are + 0.10 * 0.05 **0.01.

Table 1.4: Baseline Results

	Standardized Curriculum			Test Scores		
	Within 58 points (1)	CV Bandwidth (2)	CV Bandwidth Minus 5 points (3)	Within 58 points (4)	CV Bandwidth (5)	CV Bandwidth Minus 5 points (6)
Panel A: Math						
1[score<cutoff]	0.2186**	0.1902**	0.2085**	0.0320	0.0090	-0.0062
	(0.0609)	(0.0725)	(0.0771)	(0.0244)	(0.0285)	(0.0312)
R-squared	0.382	0.362	0.344	0.320	0.302	0.292
N	457151	317149	269144	457151	317149	269144
Panel B: Math With Controls for Prior Test Scores						
1[score<cutoff]	0.2243**	0.1845*	0.2030*	0.0123	-0.0043	-0.0103
	(0.0645)	(0.0777)	(0.0828)	(0.0171)	(0.0182)	(0.0194)
R-squared	0.381	0.355	0.332	0.667	0.652	0.644
N	293601	199242	168197	293601	199242	168197
Panel C: Reading						
1[score<cutoff]	0.4533**	0.3466**	0.3929**	0.0382	-0.0070	0.0440
	(0.0755)	(0.1211)	(0.1428)	(0.0278)	(0.0433)	(0.0466)
R-squared	0.544	0.420	0.362	0.286	0.239	0.242
N	103425	42156	32884	103425	42156	32884

Notes: All specifications include cutoff and grade-year fixed effects and control for student and school-grade peer characteristics (Female, Black, Hispanic, Other Race, Foreign Born, Free Lunch, English Language Learner, Special Education, Mover Prior to the Exam, Non-Structural Mover Prior to the School Year). Specifications in Panel B only include students in grades 4-7 and include controls for cubic polynomials in prior math and reading test scores. For math, the cross-validated bandwidth is 39, and for reading, the cross-validated bandwidth is 22. Standard errors are clustered by school and displayed in parentheses.
Significance levels are + 0.10 * 0.05 **0.01.

Table 1.5: Results by Need Group

	High		Medium		Low	
	Standard-ized Curriculum	Test Scores	Standard-ized Curriculum	Test Scores	Standard-ized Curriculum	Test Scores
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Math						
1[score<cutoff]	0.0968	0.0163	0.4360**	-0.0149	0.1211	0.0019
	(0.1313)	(0.0658)	(0.1189)	(0.0385)	(0.1217)	(0.0438)
R-squared	0.298	0.238	0.385	0.260	0.424	0.280
N	104491	104491	116978	116978	95680	95680
Panel B: Math With Controls for Prior Test Scores						
1[score<cutoff]	0.0726	-0.0329	0.4283**	-0.0135	0.1253	0.0169
	(0.1314)	(0.0433)	(0.1218)	(0.0349)	(0.1301)	(0.0224)
R-squared	0.291	0.622	0.365	0.638	0.418	0.631
N	61361	61361	70314	70314	67567	67567
Panel C: Reading						
1[score<cutoff]	0.2081	0.0864	0.4352*	0.0211	0.4091+	-0.1080
	(0.1733)	(0.0704)	(0.2087)	(0.0581)	(0.2089)	(0.0657)
R-squared	0.424	0.190	0.506	0.203	0.457	0.229
N	12031	12031	15657	15657	14468	14468

Note: All specifications use the CV bandwidths of 39 points for math and 22 points for reading. They include cutoff and year-grade fixed effects and control for student and school-grade peer characteristics (Female, Black, Hispanic, Other Race, Foreign Born, Free Lunch, English Language Learner, Special Education, Mover Prior to the Exam, Non-Structural Mover Prior to the School Year). Standard errors clustered by school are displayed in parentheses. Significance levels are + 0.10 * 0.05 **0.01.

Table 1.6: Probit Regressions of Type of Moves on Student Characteristics

	Moves During the Year		**Non-Structural Summer Moves**	
Characteristic	Before the Exam	After the Exam	Before the Year	After the Year
Prior Math Score	-0.1078**	-0.0774**	-0.0650**	-0.0438**
	(0.0050)	(0.0094)	(0.0044)	(0.0041)
Prior Reading Score	-0.0850**	-0.0395**	-0.0891**	-0.0142**
	(0.0054)	(0.0101)	(0.0046)	(0.0042)
Missing Prior	1.2493**	0.3319**	1.2735**	0.1687**
Math Score	(0.0145)	(0.0312)	(0.0127)	(0.0141)
Missing Prior	-0.0355*	0.0122	0.1536**	0.0313*
Reading Score	(0.0144)	(0.0301)	(0.0123)	(0.0129)
Female	-0.0350**	-0.0390**	-0.0125**	-0.0242**
	(0.0056)	(0.0109)	(0.0047)	(0.0045)
Black	0.2665**	-0.0249	0.2481**	0.1693**
	(0.0107)	(0.0188)	(0.0083)	(0.0080)
Hispanic	0.1880**	-0.0476*	0.1483**	0.1104**
	(0.0107)	(0.0190)	(0.0084)	(0.0080)
Other Race	0.1271**	0.0417+	0.0310**	0.0082
	(0.0121)	(0.0216)	(0.0100)	(0.0094)
Foreign Born	0.2653**	0.1055**	0.0703**	0.0228**
	(0.0075)	(0.0155)	(0.0071)	(0.0070)
Free Lunch	0.1295**	0.1653**	0.0034	0.0196**
	(0.0102)	(0.0202)	(0.0076)	(0.0074)
English Lang. Learner	0.0754**	-0.0029	0.0464**	0.0199*
	(0.0094)	(0.0196)	(0.0085)	(0.0087)
Special Education	0.1898**	0.0605**	0.1409**	0.1647**
	(0.0104)	(0.0205)	(0.0093)	(0.0088)

Notes: This table displays the coefficients from probit regressions of the type of move on student characteristics and year-grade fixed effects. N=625390. Standard errors are clustered by school and displayed in parentheses. Significance levels are +0.10 *0.05 **0.01.

Table 1.7: Heterogeneity by Student Mobility

		2004-2005		**Only 2005**	
		Movers Before the Exam		**Non-structural Movers Before School Year**	
	Non-Movers	From Any School	Only From Schools Below Cutoff	From Any School	Only From Schools Below Cutoff
	(1)	(2)	(3)	(4)	(5)
Panel A: Math					
Standardized	0.1879*	0.2511**	0.2711**	0.2027*	0.1943*
Curriculum	(0.0728)	(0.0723)	(0.0813)	(0.0865)	(0.0904)
Test Scores	0.0134	0.0223	0.0357	-0.0699	-0.0477
	(0.0285)	(0.0501)	(0.0640)	(0.0603)	(0.0799)
N	277692	13212	5080	12941	5016
Panel B: Math With Controls for Prior Test Scores					
Standardized	0.1821*	0.2810**	0.2864**	0.1822+	0.1582
Curriculum	(0.0779)	(0.0835)	(0.0897)	(0.0984)	(0.1089)
Test Scores	-0.0024	-0.0396	-0.0550	-0.0468	-0.0615
	(0.0181)	(0.0446)	(0.0575)	(0.0561)	(0.0727)
N	185747	4212	2753	4577	2849
Panel C: Reading					
Standardized	0.3515**	0.2990*	0.1730	0.3329*	0.3603**
Curriculum	(0.1216)	(0.1326)	(0.1436)	(0.1353)	(0.1355)
Test Scores	-0.0027	0.1065	0.2711*	-0.0138	0.1811
	(0.0433)	(0.0777)	(0.1364)	(0.0970)	(0.1326)
N	36939	1536	730	1717	704

Note: All specifications use the CV bandwidths of 39 points for math and 22 points for reading. They include cutoff and year-grade fixed effects and control for student and school-grade peer characteristics (Female, Black, Hispanic, Other Race, Foreign Born, Free Lunch, English Language Learner, Special Education). Standard errors clustered by school are displayed in parentheses. Significance levels are + 0.10 * 0.05 **0.01.

Table 1.8: Re-evaluating the effect of mobility on student achievement

	Math		**Reading**	
	(1)	(2)	(3)	(4)
Move During Year Before Exam	-0.3524**	-0.0631**	-0.2300**	-0.0352**
	(0.0083)	(0.0053)	(0.0083)	(0.0055)
Move During Year After Exam	-0.1786**	-0.0790**	-0.1707**	-0.0716**
	(0.0143)	(0.0131)	(0.0124)	(0.0098)
Non-structural Move Before Year	-0.1272**	-0.0130**	-0.0986**	-0.0080+
	(0.0053)	(0.0046)	(0.0058)	(0.0043)
Non-structural Move After Year	-0.0997**	-0.0322**	-0.0869**	-0.0240**
	(0.0055)	(0.0036)	(0.0056)	(0.0034)
Controls for Prior Test Score	No	Yes	No	Yes
R-squared	0.296	0.678	0.275	0.667
N	688366	478796	635152	482084

Note: Both the math and reading samples contain students in grades 3-7 in 2004 and 2005. All specifications include year-grade fixed effects and control for student characteristics (Female, Black, Hispanic, Other Race, Foreign Born, Free Lunch, English Language Learner, Special Education, Mover Prior to the Exam, Non-Structural Mover Prior to the School Year). Controls for prior test scores are cubic polynomials in prior math and reading test scores. Standard errors are clustered by school and displayed in parentheses. Significance levels are + 0.10 * 0.05 **0.01.

Chapter 2

Special Education and Peer Effects in School

2.1 Introduction

Economists have studied peer effects in a wide range of settings — retirement savings choices, efficiency at work, neighborhoods, and education — and have arrived at varying conclusions regarding their importance (e.g., Duflo and Saez 2002; Mas and Moretti 2009; Oreopoulos 2003; Kling, Liebman, and Katz 2007; Hanushek et al. 2003; Angrist and Lang 2004; Hoxby and Weingarth 2005; Imberman et al. 2009, Carrell et al. 2012). While there are a number of possible explanations for these differing results (different settings, difficulties in defining the relevant peer groups, bias due to the endogeneity of peer group formation, etc.), one possibility is that economic agents may take actions which compensate for and conceal peer effects.

This paper focuses on one such action taken by schools — the referral of disabled students to special education services — which may ameliorate negative peer effects. Studies have found that having more peers that receive special education services does not reduce, and may even increase, the achievement of regular education students (Hanushek et al. 2002; Friesen el al. 2010).[1] While these results seem somewhat surprising since disabled students are more likely to have low achievement and behavioral problems, I argue that these findings are consistent with the notion that special education mitigates negative peer effects.

I illustrate this argument by using panel data on New York City students to isolate the uncompensated peer effects of disabled students. I classify students as "disabled" if they are ever observed in special education between grades 4 and 7. This allows me to examine

[1] In Texas, Hanushek et al. (2002) find a positive relationship between a high percentage of special education peers and the academic achievement of regular education students, while, in British Columbia, Friesen et al. (2010) find a negative, small, and mostly insignificant relationship. Friesen et al. (2010) suggest that the difference in results may be due to differences between the funding mechanisms in Texas and British Columbia; although school districts in Texas receive additional funds for each special education student, school districts in British Columbia receive funds which are proportional to student enrollment (assuming a constant proportion of students have disabilities) to treat mild disabilities. British Columbia does provide additional funds for severely disabled students.

the peer effects of "disabled" students who do not receive special education services, either because they are "undiagnosed" and have not yet entered special education, or because they have been "declassified" and have exited special education. I identify the effects of these disabled students — when they are and are not receiving special education services — on their non-disabled peers by using variations in peer composition between cohorts in the same school and grade. Like Hanushek et al. (2002) and Friesen et al. (2010), I find that disabled students who receive special education do not harm the achievement of their non-disabled peers. However, these benign results mask the negative effect disabled students would have in the absence of special education services. Disabled students who do not receive special education services, particularly those with undiagnosed disabilities, have a significant negative effect on their non-disabled peers. The estimates imply that if one non-disabled student in a grade of 75 students were replaced by a student with an undiagnosed disability, the exam scores of the 74 non-disabled students would be expected to fall by 0.6% of a standard deviation in math and English.[2]

One explanation for these negative peer effects is that, along the lines of Lazear (2001), disabled students are more disruptive in class. Consistent with this notion, not only are students with undiagnosed disabilities more likely to be suspended, but they also raise the probability of suspension for their non-disabled peers. I also find that having an experienced teacher lessens the negative effects of students with undiagnosed disabilities, consistent with the notion that effective teachers may be able to limit the extent of disruptions in their classrooms. Finally, I find that segregation may be one of the main mechanisms through which special education mitigates the effects of disabled peers, although special education may also improve the classroom environment by providing other inputs, such as counseling/tutoring for disabled students or a teacher's aide (Hanushek et al. 2003; Cohen 2008).

[2] Each student has 74 peers, so each non-disabled student's score is reduced by 0.6% in math ($1/74 \times -0.4579 \times 100\%$) and 0.7% in English ($1/74 \times -0.4879 \times 100\%$).

I perform a number of robustness checks to ensure the results are not driven by selection. I find no evidence of negative selection among the peers of students with undiagnosed disabilities, and the results are relatively robust to the use of within school-year and within individual variation in peer group composition. There is also no evidence that the results are being driven by temporary negative shocks which cause some students to be classified as disabled and temporarily reduce the achievement of non-disabled peers; these shocks would generate correlations between the outcomes and timing of diagnosis which are not supported by the data.

These findings have important implications for education policy. First, they are relevant to the concern that disproportionate spending on special education drains resources from regular education; in 2000, the cost of educating the 13.2% of students who received special education services was $77.3 billion and accounted for 21% of spending on elementary and secondary education services (U.S. Office of Special Education Programs 2007; Chambers et al. 2004). Of course, disproportionate spending may be warranted if it generates positive spillovers for regular students, in addition to benefiting disabled students. An approximate dollar value of the benefit to regular students can be calculated from estimates of test score gains to non-disabled students and the labor market value of test score gains. A conservative estimate from Kane and Staiger (2002) is that a one standard deviation increase in math test scores is worth around $90,000 at age 9. This estimate implies that the receipt of special education by 13.2% of students yields math test gains of 0.06 standard deviations for each regular student, and with 40.9 million regular students, the test score gains alone are worth $220 billion, far surpassing the $77.3 billion cost of educating special education students. Second, as these gains may largely result from segregating disabled students, these findings suggest that peer effects should figure into the cost-benefit calculus of policies that mainstream special education students.

This paper continues as follows: Section 2.2 provides background information on special education and peer effects. Section 2.3 presents the empirical strategy. Section 2.4 describes the data. Section 2.5 discusses the results, Section 2.6 presents robustness checks, and Section 2.7 concludes.

2.2 Background

2.2.1 Special Education in New York City

In 1975, the United States federal government enacted the Education for All Handicapped Children Act, renamed the Individuals with Disabilities Education Act (IDEA) in 1990. The purpose of IDEA was to ensure that states and localities provided early intervention, special education, and related services to children with disabilities. Prior to the passage of IDEA, U.S. schools educated only about 20% of students with disabilities, and many states excluded children who were deaf, blind, emotionally disturbed, or mentally retarded from the education system entirely (U.S. Office of Special Education Programs 2007).

As seen in Figure 2.1, in the 30 years since the passage of IDEA, the percentage of students receiving special education services has increased substantially, from 8.3% in 1977 to 13.5% in 2007. Some explanations for this increase are expanded cultural norms about what constitutes disability, increased take-up due to reduced stigma, and increased childhood poverty (Cullen 2003). However, since much of this increase occurred for "specific learning disabilities," which have more subjective diagnostic criteria than other disabilities (e.g., mental retardation), there have been concerns that schools are increasingly misclassifying students as disabled in response to fiscal and school accountability incentives. For example, Cullen (2003) finds that fiscal incentives explain about 40% of the growth in student disability rates in Texas. Figlio and Getzler (2002) find that the introduction of high-stakes testing and test-based school accountability systems in Florida led schools to classify students as disabled and, thus, exempt from testing. Jacob (2005) and Cullen and Reback (2006) document similar phenomena in Chicago and Texas, respectively.

Although these papers suggest that schools may strategically classify students as disabled in response to incentives, there is a rigorous evaluation process which determines special education placement and services. In New York City, the process begins when a teacher, parent, or clinician refers the child for an evaluation. After the referral, a school psychologist administers a battery of tests to measure the child's "reasoning, motor skills, language, executive functions, visuo-spatial skills, social/emotional and behavioral functioning, memory, [and] academic achievement in reading, mathematics, written expression, and oral communication" (Wernikoff 2007; conversations with NYC special education teacher).[3] The results of these tests are used in conjunction with student records, classroom based assessments, observations of student behavior, and interviews with the student, school staff, and parents to determine eligibility. A student is eligible for special education services if his School Based Support Team (SBST), which includes his parents, a regular education teacher, a special education teacher, and a district representative, determines that he meets the criteria for one or more of New York State's disability classifications and requires special education services to benefit from instruction.[4]

For example, prior to the reauthorization of IDEA in 2004, the criteria for being diagnosed with a "learning disability" (LD) was that the student must "[exhibit] a discrepancy of 50 percent or more between expected achievement and actual achievement" (Board of Education of the City of New York n.d.). Due to concerns that the discrepancy requirement resulted in late identification and misidentification of LD, the reauthorization eliminates the discrepancy requirement but requires schools to document responsiveness to intervention (Cortiella 2010).

[3]For example, the Wechsler Intelligence Scale for Children (WISC) and Wechsler Individual Achievement Test (WIAT).

[4]A school psychologist, school social worker, and/or physician may also participate.

If the student is eligible for special education services, the SBST must also develop an Individualized Education Plan (IEP) that establishes goals for the student and specifies what environment and services the student needs to meet those goals. The IEP and placement are reviewed by the SBST annually and a reevaluation is conducted at least every three years. If the SBST determines the student no longer requires special education services, it may still arrange services for the student for up to one year following declassification.

IDEA mandates that special education students be educated in the "least restrictive environment," that is, "to the maximum extent appropriate, children with disabilities...are educated with children who are not disabled." Special classes, separate schooling, or other removal from the regular educational environment should only occur when education in regular classes with supplementary aids and services cannot be achieved due to nature or severity of a child's disability (Individuals with Disabilities Education Improvement Act 2004). However, the extent to which this is followed in practice is uncertain; numerous lawsuits have been filed by parents alleging that their children have not been placed in the least restrictive environment (Crockett and Kauffman 1999).

Special education students who are educated in the general education environment receive supplementary aids and services, such as curriculum modification, speech/language therapy, or a paraprofessional assigned to the classroom (e.g., teacher's aide). In New York City, regular and special education students are sometimes taught together by one general education and one special education teacher in "collaborative team teaching" classrooms.[5] Other special education students are taught in self-contained classrooms in general education schools. Students diagnosed with moderate to severe disabilities, such as autism spectrum, severely emotionally challenged, and/or multiply disabled attend schools in District 75, the Special Education District.

[5]No more than 40% of the students in collaborative team teaching classes can be special education students.

2.2.2 Peer Effects

Identifying the effects of disabled students on their regular education peers is complicated by a few well-known identification problems. The first is the selection problem: students (or students' parents) may select peer groups based on unobservables which are correlated with the outcome measures. A number of studies avoid selection bias in estimates of peer effects by exploiting knowledge of the assignment rule or quasi-experimental designs (Sacerdote 2001; Zimmerman 2003; Angrist and Lang 2004; Hoxby and Weingarth 2005; Lehrer and Ding 2007; Imberman et al. 2009).

However, since this type of variation is not always available, other methods have been used to address selection bias. Hoxby (2000) and Hanushek et al. (2003) identify peer effects using variation between cohorts within a grade within a school. This strategy, which is also used in this paper, relies upon the fact that even when parents make school choices based on the achievement or demographic composition of schools' past cohorts, the actual achievement or composition of their child's cohort may deviate from their expectations. Moreover, once parents have chosen a school, it may be costly for them to change in response to unanticipated deviations. Using variation at the grade level, rather than at the classroom level, further removes selection bias due to non-random sorting of students across classrooms within a grade (e.g., schools tracking students by achievement or highly motivated parents pressuring principals to put their children in a particular teacher's class).

The second problem is reflection or simultaneity: a student both affects and is affected by her peers (Manski 1993; Moffit 2001). While this is not a problem for studies which look at the effect of peer demographic characteristics, it is an significant obstacle for studies that want to examine the effect of peers' achievement on students' own achievement. To address this problem, studies generally use the lagged outcomes (e.g., test scores, GPA) of peers as measures of peer achievement. When peer groups are composed of students who have not previously been peers, simultaneity bias is eliminated. For example, Sacerdote (2001), Zimmerman (2003), and Lehrer and Ding (2007) look at college or high school peer groups that are composed of students from many different high schools or middle schools, and Imberman et al. (2009) study peer groups formed due to hurricane evacuations. However, when students remain peers as they progress through school, their lagged scores are also likely biased due to simultaneity.

While the literature provides substantial evidence for the existence of peer effects, understanding the mechanisms through which they operate is critical for understanding how to address them. A number of papers show that peer effects operate through behavior problems (e.g., Figlio 2007; Aizer 2008; Lavy et al. 2008; Fletcher 2009; Imberman et al. 2009; Carrell and Hoekstra 2010). Two of these papers focus on students who may have disabilities: Fletcher (2009) observes lower test scores among classmates of severely emotionally disturbed students, while Aizer (2008) finds that having a classmate with undiagnosed Attention Deficit Disorder (ADD) reduces test scores.[6] If children who disrupt the class reduce the ability of other students to learn, Lazear's (2001) theoretical model of peer effects shows that, in some cases, the segregation of more disruptive students is efficient.[7] On the other hand, Aizer (2008) finds that diagnosis and treatment of ADD mitigates negative externalities without classroom reassignment.

Peer effects may also operate through changes in teacher behavior. Duflo et al. (2010) provide evidence from a randomized evaluation in Kenya that tracking students by prior achievement benefits all students, which they attribute this to teachers being better able to tailor instruction in more homogeneous classes. Other studies of tracking provide more

[6]Emotional disturbance and other health impairment, which includes ADD when it adversely affects a student's performance, are considered disabilities under IDEA (Department of Education 1999).

[7]This assumes that students' propensity to be disruptive is independent of their peers' propensity to be disruptive. Integration may be efficient if integration can transform disruptive students into non-disruptive students.

mixed results (e.g., Betts and Shkolnik 1999; Figlio and Page 2002; Zimmer 2003; Lefgren 2004).

Either of these mechanisms could explain negative peer effects of disabled students; disabled students may cause more disruptions than non-disabled students, or they may require additional time to understand the subject material, slowing down the pace of the class. Special education may compensate for negative peer effects in a number of ways. It provides additional resources to disabled students who remain in the same class and facilitates tracking by funding smaller, separate classes for disabled students. Resources, such as tailored instruction or counseling, may improve the achievement or behavior of disabled students who remain in the same class, while disabled students who attend separate classes or schools should have no direct effect on non-disabled students.

2.3 Empirical Strategy

To examine the peer effects of disabled students, I estimate the following specification for non-disabled students:

$$Y_{igst} = X_{igst}\beta + P_{-igst}\theta + D_{gst}\gamma + \mu_{gs} + \nu_{gt} + \epsilon_{igst} \tag{2.1}$$

where Y_{igst} is the outcome variable (i.e., math test score, English test score, or probability of suspension) for student i in grade g, school s, and year t, and X_{igst} is a vector of student level covariates, including indicators for gender, race/ethnicity, free lunch status, English Language Learner status, whether the student is repeating the grade, and whether the student has switched schools in the previous year.[8] P_{-igst} is a vector of the characteristics of student i's school-grade peers, including the proportions of peers, excluding i, who are

[8]Math and English test scores have both been standardized to have a mean 0 and standard deviation 1 for each year and grade.

female, black, Hispanic, other non-white race, receiving free lunch, and English Language Learners. The vector D_{gst} contains the variables of interest: the proportion of peers who are disabled and receive special education services, the proportion of peers who have undiagnosed disabilities, and the proportion of peers who are disabled but have been declassified. Thus, the coefficients in γ test whether disabled students who do and do not receive special education services have negative peer effects. Finally, μ_{gs} is a school-grade fixed effect, ν_{gt} is a year-grade fixed effect, and ϵ_{igst} is the error term.[9] To account for intra-school correlations between students' test scores, the standard errors are clustered by school.

This specification will produce unbiased estimates if the remaining variation in the proportion of disabled students between cohorts in the same grade and school is exogenous to the sorting of students across schools. This assumption would fail if there were unobserved time-varying factors that were correlated both with the proportion of disabled students and the achievement of students' peers (e.g., if an increase in school resources increased diagnosis of disabled students and improved the achievement of non-disabled peers). To address this concern, I perform a number of robustness checks and find no evidence that the results are being driven by selection bias.

Because students' disability/special education status is unlikely to be affected by their peers, the reflection problem is less of a concern in this context, especially compared to studies which use the lagged test scores of students' peers. However, the reflection problem could still be relevant if having low achieving peers causes students to be diagnosed as disabled (e.g., if students are diagnosed as disabled because they perform poorly, and they perform poorly primarily because they have low achieving non-disabled peers). Yet, poor performance on the standardized exams is not sufficient for a positive diagnosis; as previously explained, the

[9] I include school-grade rather than school fixed effects because spurious correlations might arise if age-specific trends in student achievement and disability classification vary by schools. For example, at a high poverty school, a spurious correlation could be produced by the following unrelated trends: student achievement decreases as students approach adolescence and the number of students classified as disabled increases.

School Based Support Team bases its determination of disability on tests administered by a school psychologist and other measures of the student's behavior and ability, not solely the student's standardized exam scores.

A related concern over the subjectivity of diagnosis is that schools could misclassify disruptive students as disabled in order to reduce their negative peer effects. While this is possible, these misclassifications would still have to make it through the evaluation process; school staff can refer students for evaluation, but students would have to be evaluated and determined to have a disability by the School Based Support Team, which includes students' parents. Moreover, even if such misclassifications occur, this type of manipulation is not a reflection problem but does change the interpretation of these results. In this case, special education is an intervention which mitigates the negative peer effects of disruptive (non-disabled) students.

2.4 Data

This paper uses a rich panel data set from New York City to follow seven cohorts of students as they progress from 4th through 7th grade. The oldest cohort attended 4th grade in school year 1999-2000, while the youngest cohort attended 4th grade in school year 2005-2006. Each cohort contains around 80,000 students in 1,000 schools.

The data contain information on students' math and English exam scores, suspensions, gender, race/ethnicity, free lunch eligibility, and English Language Learner status. Importantly, the data contain students' special education status but not their type of disability. Students are also matched to their schools, and in most cases, to their classrooms and teachers.[10]

[10]Some schools did not use the administrative system which allows linkages between teachers and students, and older students, who are taught by separate math and ELA teachers, are less likely to be linked. Kane et al. (2008) find no statistically significant relationship between student characteristics and schools' use of

One important limitation of the data is that they contain the grade level of the exam taken by students but not students' grade level. I assume that students take the appropriate grade level exam and impute grade levels for students who were not tested using the following method. If a student took exams in previous years, I assume that the student was promoted from his grade in the previous years, and if a student never took an exam, I use his exact date of birth and New York's kindergarten entry date to place him in his age-appropriate grade.[11] Students are restricted to be in highest grade served by their school. Due to concerns about grade misclassification, I drop the school-grade-year cells with less than 28 students (the bottom 1% in school-grade enrollment) from the regression sample. I also drop a few non-disabled students who attend schools (e.g., hospital schools) in the Special Education District, District 75, from the regressions.

The panel structure of the data allows me to observe whether a student has received special education in the past or will receive it in the future. Since students must have a disability in order to receive special education services, I exploit the panel structure of the data to classify students as "disabled" if I ever observe them receiving special education services between 4th and 7th grade. This classification assumes that students' disabilities exist both before and after the student receives special education. This assumption seems realistic in most cases, such as for a student whose learning disability is not diagnosed until 5th grade, but would not properly account for individuals whose disability status changes over time, such as a student whose disability is the result of a recent car accident. However, this type of measurement error would tend to attenuate the estimates.

---

this administrative system.

[11]In New York, students must turn 5 before December 31 of their kindergarten year. For regular education students who took exams, the grade of the exam matches the age-appropriate grade more than 90% of the time for regular education students and 50% of the time for special education students. While this suggests that the grades of special education students who do not take exams may tend to be overestimated, they are a small fraction of the special education population (see Table 2.1).

Since students are classified as disabled based upon their past, present, and future receipt of special education services, some disabled students do not currently receive special education services. I classify students as "undiagnosed" if they receive special education services in the future and as " declassified" if they received special education services in the past.

Table 2.1 presents statistics by grade on the prevalence of disability, the receipt of special education services, and exam participation among the New York City students in the sample. Overall, I classify 11 percent of students as disabled, but only 9.3% of students receive special education services. As students progress through grades, the percentage of students in special education grows since more students are diagnosed than are declassified.

Panel A of Table 2.1 reports exam participation rates for regular and special education students in New York City, which are comparable or higher than those in British Columbia and Texas.[12] In Texas, about 80% of regular students have valid gain scores, compared to 90% in both British Columbia and New York City (Friesen et al. 2010; Hanushek et al. 2002).[13] Exam participation rates for special education students in New York City are noticeably higher than those in British Columbia and Texas. In British Columbia, only 67% of learning disabled students take exams, and in Texas, only 30% of special education students have valid gain scores; in New York City, about 80% of special education students take exams and 70% have valid gain scores (Friesen et al. 2008; Hanushek et al. 2002).

Panel B of Table 2.1 presents statistics on the fraction of disabled students who are undiagnosed, receiving special education services, and declassified. Most disabled students receive special education services in all years. Due to the window of observation, most of

[12]Some English Language Learners receive exemptions; thus, participation rates are lower for ELA than for math. Some special education students are exempt from these exams and may receive alternate assessments, but these exemptions seem to be limited to students with the most severe disabilities; about 95% of the special education students who do not take exams received special education services in all the years they were in the sample.

[13]In order to have a valid gain score, students must have non-missing exam scores for two consecutive exams.

the undiagnosed students are in 4th grade, and most of the declassified students are in 7th grade.[14] A quarter of disabled students do not receive special education in 4th grade, and 8% of disabled students are declassified by 7th grade. Note that I identify the effects of students with the types of disabilities that could remain undiagnosed by the 4th grade; this rules out more severe disabilities, and the results should be interpreted in this light.

Table 2.2 provides the summary statistics for the non-disabled students used in the regressions, and, for comparison purposes, the summary statistics for disabled students who also have non-missing suspension outcomes and student and peer controls. The summary statistics for disabled students are separated into three categories: special education, undiagnosed, and declassified. The table shows that, as expected, disabled students have lower academic achievement than non-disabled students. Since the most severely disabled students always receive special education services, special education students have the lowest math and English scores, almost 1.5 standard deviations below those of non-disabled students. Undiagnosed and declassified students fare better but still score about 1 and 0.5 standard deviations lower than non-disabled students, respectively. These summary statistics also suggest that disabled students may be more disruptive; while the overall rates of suspension are relatively low, less than 3% for non-disabled students, suspension rates are twice as high for special education and declassified students, and over three times higher for undiagnosed students. Disabled students also tend to be male, are disproportionately black and Hispanic minorities, and are more likely to receive free lunch and English Language Learner services.

Table 2.2 also reports the characteristics of students' peers in their school and grade. About 7% of the peers of non-disabled students receive special education services, 1.6% have undiagnosed disabilities, and 0.4% have been declassified. Compared to non-disabled students, undiagnosed and declassified students are only slightly more likely to have spe-

[14]The small percentage of students in 7th grade who are undiagnosed are due to students who have repeated 7th grade.

cial education peers; about 8% and 9% of their peers receive special education services, respectively. In contrast, 13% of the peers of special education students also receive special education services. Since receipt of special education services is correlated with other student characteristics, disabled students' peers are also more likely to be black or Hispanic and receive free lunch.

2.5 Results

Table 2.3 presents the effects of disabled peers on non-disabled students' math and English test scores (Columns 1-6) and probability of suspension (Columns 7-9). All specifications include student controls, controls for other peer characteristics, school-grade fixed effects, and year-grade fixed effects. The first specification is similar to those in Hanushek et al. (2002) and Friesen et al. (2010), who only examine the peer effects of special education students. The estimated effects of special education peers on non-disabled students' math and English scores are small and insignificant; the point estimates of -0.04 standard deviations for math (Column 1) and 0.02 standard deviations for English (Column 4), fall between the statistically significant 0.16 effect found by Hanushek et al. (2002) and the statistically insignificant -0.14 effect found by Friesen et al. (2010). However, the next specification shows that these results mask the effect that disabled students have on their peers' test scores in the absence of special education services (Columns 2 and 5). The final specification separates disabled peers not receiving special education services into two categories, those who are undiagnosed and those who are declassified, and shows that the negative effects on test scores are driven almost entirely by peers with undiagnosed disabilities (Columns 3 and 6). The results for suspension are similar: while special education students do not significantly affect the probability of suspension (Column 7), having more peers with undiagnosed disabilities significantly raises it (Column 9). Overall, the results imply that a one standard deviation increase in the proportion of peers with undiagnosed disabilities (about 2%) would reduce non-disabled students' exam scores by nearly 1% of a standard deviation in math and English and increase their probability of suspension by 0.09%. Peers who have been declassified have no significant effects on non-disabled students, probably because disabled students who no longer require special education services are positively selected.

Since schools may respond to high proportions of special education students by hiring better teachers, Table 2.4 examines whether teacher quality can explain the absence of significant peer effects for special education students. Teacher quality could also explain the negative effects of students with undiagnosed disabilities if lower quality teaching caused students to be classified as disabled and also reduced the test scores of their non-disabled peers.[15] Since I am missing classroom and teacher identifiers for some students, I restrict my analysis to the subsample of students with non-missing classroom and teacher identifiers and, using this subsample, reproduce the main results (Columns 1, 4, and 7). The next specification adds a control for teacher quality: the proportion of teachers in the school-grade-year cell with no prior teaching experience (Columns 2, 5, and 8).[16] As expected, the estimates suggest that inexperienced teachers are significantly worse than experienced ones. If one of three teachers in the grade is replaced with a novice, students' average scores will decrease by about 2% of a standard deviation. However, while teacher experience is important, it does not explain either the lack of effect of special education students or the negative effect of undiagnosed students.[17] The final specification tests the suggestion of Lazear (2001) that high quality teachers are better able to control disruptive students by

[15]However, IDEA specifies that children should not be classified as disabled due to lack of appropriate instruction in math or reading.

[16]Teacher experience has a positive effect on student learning, but these effects are mainly concentrated in the first year of teaching (Hanushek and Rivkin 2006).

[17]Hanushek et al. (2002) also find that teacher experience does not explain the positive effect of special education students on their peers.

including interactions between the teacher experience variable and the proportions of disabled peers. The interaction between undiagnosed peers and the proportion of novice teachers is large and negative for math and English test scores and positive for suspensions, although it is only statistically significant for math. Nevertheless, these results provide suggestive evidence that more experienced teachers may be better at mitigating negative peer effects than inexperienced teachers, possibly because they are better able to minimize disruptions.

Non-disabled students may also benefit if disabled students are educated in separate classrooms, as are the majority of special education students in New York City. Even excluding the special education only schools in District 75, about 78% of special education students attend special education only classes in math or reading. While students who enter or exit special education are more likely to be mainstreamed, only 32% of them attend general education classes when they receive special education services. I take two approaches to determine whether the segregation explains why students with undiagnosed disabilities have a negative effect on their peers, but special education students do not. First, I use classroom identifiers to determine the following proportions of school-grade peers: (1) the proportion who have undiagnosed disabilities and will be placed in special education only classes in the future, (2) the proportion currently attending special education only classes, and (3) the proportion who have been declassified but attended special education only classes in the past. In Table 2.5, I use the same subsample of students as Table 2.4 and include (1) through (3) as additional regressors in the main specifications (Columns 1, 3, and 5). The coefficients on (1) through (3) test whether disabled students who are ever segregated have different effects than those who are not.[18] The estimates suggest that disabled students who are ever segregated tend to have more negative peer effects than those who are not.

[18] Because I am missing classroom identifiers for some schools, primarily middle schools, I may understate the proportion of disabled students ever in special education only classes, which would tend to bias these estimates towards 0.

In math, the negative effect of students with undiagnosed disabilities comes almost entirely from students who will be placed in special education only classes, but the interaction is not statistically significant for English test scores or suspensions. Special education students who attend separate classes have more positive effects on test scores and more negative effects on suspensions, although these coefficients are imprecisely estimated. Declassified students who previously attended separate classes have negative and insignificant effects on test scores, although they have a significant negative effect on the probability of suspension.

One limitation to the approach of comparing the effects of disabled students who ever and never attended special education only classes is that it does not account for smaller changes in the distribution of students between classes. For example, even if all disabled students were mainstreamed, it might be optimal to educate all the disabled students in a grade at a school in one classroom, rather than spreading them out among many. In order to analyze smaller changes in the sorting of students between classes, I follow the racial segregation literature and construct an index to measure the isolation of the three groups of disabled students (Bell 1954; White 1986; Cutler et al. 1999). For special education students, the index of isolation is:

$$index\ of\ isolation = \frac{\sum_{i=1}^{N}(\frac{special\ ed_i}{special\ ed_{total}})(\frac{special\ ed_i}{students_i}) - \frac{special\ ed_{total}}{students_{total}}}{1 - \frac{special\ ed_{total}}{students_{total}}} \tag{2.2}$$

where $students_i$ and $special\ ed_i$ are the numbers of students and special education students in class i, $students_{total}$ and $special\ ed_{total}$ are the total numbers of students and special education students in the school-grade, and N is the number of classes in the school-grade. The term $\sum_{i=1}^{N}(\frac{special\ ed_i}{special\ ed_{total}}\frac{special\ ed_i}{students_i})$ is the proportion of special education students in the class attended by the average special education student. Note that this sum would equal 1 if all of the special education students in a grade at a school were educated in a single, segregated class. Subtracting $\frac{special\ ed_{total}}{students_{total}}$ from this sum eliminates the effect that comes from the overall proportion of special education students, and dividing by $1 - \frac{special\ ed_{total}}{students_{total}}$ scales the index to fall between 0 and 1, with 1 representing complete isolation.[19] I construct similar indices for undiagnosed and declassified students by substituting the corresponding numbers of undiagnosed and declassified students into the equation. The average isolation indices for special education, undiagnosed, and declassified students are 0.86, 0.34, and 0.65, respectively. Undiagnosed students are less isolated since their disabilities have not yet been identified, while declassified students remain somewhat isolated even after they exit special education.

In Columns 2, 4, and 6 of Table 2.5, I include these isolation indices in the main specification along with the proportion of peers in each of these categories. The isolation index for undiagnosed peers is positive and significant for math and negative and significant for suspensions, suggesting that holding the proportion of undiagnosed peers fixed, increasing the isolation of these students reduces negative peer effects. The isolation indices for both special education students and declassified students are small and insignificant. This may be because, on average, these students are already more segregated, or because schools intentionally sort students diagnosed with disabilities into classes based on their potential to be disruptive, such that removing the marginal special education student from the mainstream may yield smaller gains.

2.6 Robustness Checks

The previous section suggests that unlike disabled students who receive special education services, students with undiagnosed disabilities have negative effects on their peers. However, these estimates could be biased if there are unobserved time-varying factors which are correlated with test scores and with changes in the proportion of students with disabilities.

---

[19] The isolation index equals 1 if there are no special education students.

For example, an increase in drug use in a neighborhood could increase the number of disabled students and also cause their non-disabled peers to have lower test scores. However, if such unobservables had permanent effects on a cohort (e.g., if neighborhood drug use increases the births of children with disabilities and/or low achievement), one would also expect the proportion of special education students to be negatively correlated with the test scores of non-disabled peers, which it is not. The next three tables present specifications which address these concerns.

The first of these, Table 2.6, checks whether the proportions of disabled peers are correlated with observable characteristics of non-disabled students. It reports the coefficients from the regressions of observable student characteristics on the proportions of disabled students, peer controls, school-grade fixed effects, and year fixed effects. There are few significant imbalances, and these are relatively small and not consistently signed. Notably, the evidence does not suggest that the non-disabled peers of undiagnosed (declassified) students are systematically more (less) disadvantaged. The non-disabled peers of undiagnosed students are less likely to receive free lunch, while those of declassified students are less likely to receive ELL but more likely to repeat grades. For special education students, peers' receipt of ELL and free lunch go in the same direction, but the effects are small; a one standard deviation increase in the proportion of special education students (6%) only decreases ELL and free lunch receipt by 0.2% and 0.3%, respectively. It is also important to note that all previous specifications control for these characteristics, and more stringent specifications are reported in Table 2.7.

The final columns of Table 2.6 address concerns about selection bias due to correlations between the population of non-disabled test takers and the proportion of special education students. For example, a correlation between the exam exemptions of low achieving non-disabled students and the proportion of special education students could generate a spurious positive relationship between special education peers and the achievement of non-disabled students. Columns 9 and 10 show that the only significant correlation between the probability of being tested and the proportion of disabled peers is for undiagnosed students on the English exam. Thus, it is unlikely that selection bias accounts for the results; if anything, increased exemptions for low achieving non-disabled students should bias the English results towards zero, and the English and math results are very similar, despite the lack of evidence of selection bias in math. Moreover, the scope for such selection bias appears limited. No Child Left Behind mandates the testing of 95% of students both overall and within certain student subgroups, and while there is evidence that other jurisdictions manipulate testing populations, there is no evidence of this from New York City. For instance, Rockoff and Turner (2010) find no evidence that New York City schools decreased the testing population in response to accountability grades.

Table 2.7 employs two additional strategies to allay concerns about bias due to unobservable time-varying factors. The first uses school-year fixed effects to deal with unobservable school-specific shocks that could be correlated with peer composition and achievement. For example, an increase in school resources could increase the diagnosis of disabled students and improve the achievement of non-disabled peers. The second includes student fixed effects and identifies the effects of changes in the composition of students' peers within students over time. Overall, these results are similar.

The final table, Table 2.8, addresses the concern that these results are consistent with temporary unobservable negative shocks that reduce students' achievement and cause students to be classified as disabled.[20] Such a shock could induce a negative correlation between students with undiagnosed disabilities and the achievement of non-disabled students, and if it were temporary, the achievement of non-disabled students would recover in the next year, resulting in zero correlation between special education students and achievement. A

[20]For example, Carrell and Hoekstra (2010) find that the negative peer effects of children exposed to domestic violence are primarily driven by children whose parents had not yet reported the violence.

temporary shock would generate correlations that are strongly dependent on the timing of diagnosis; if it dissipates in one year, there should only be negative peer effects for students who were diagnosed in the following year, and there should be significant positive effects for special education students who were diagnosed in the previous year. However, the tests for these correlations in Table 2.8 do not support the idea that the results are being driven by temporary negative shocks.

2.7 Conclusion

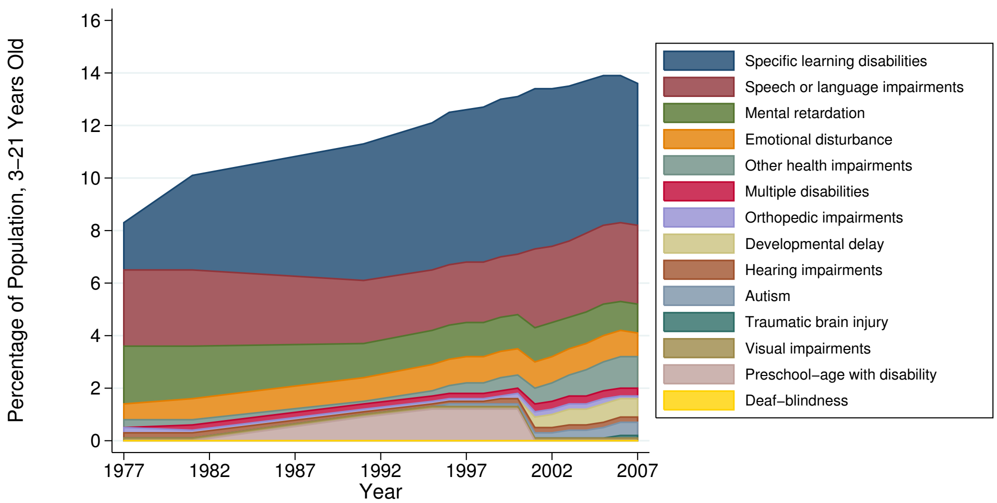
Much of the policy debate surrounding special education has centered on its cost and implementation. The cost of special education has caused worry that special education drains resources from regular education, while the policy of mainstreaming has caused worry that inclusion of special education students harms the education of regular students (Sutner 1998; Duff 1999; Coeyman 2002; Needham 2004).

Contrary to the first claim, I find that special education benefits regular students by compensating for negative peer effects of disabled students. I estimate that a one standard deviation increase in the proportion of peers with undiagnosed disabilities (about 2%) would reduce non-disabled students' exam scores by nearly 1% of a standard deviation in math and English and increase their probability of suspension by 0.09%. On the other hand, an increase in peers receiving special education services has no effect on non-disabled students. Moreover, since these effects are identified using disabled students who enter and exit special education, it is reasonable to think that students with more severe disabilities would have even larger negative peer effects in the absence of special education services.

On the other hand, the results also suggest that the second concern is not unfounded. I find that segregation may be one of the main mechanisms though which special education mitigates the effects of disabled peers.

Nevertheless, there are several reasons why these results should not be interpreted as advocating the segregation of disabled students. First, the results are based only on one school district, and their external validity is unclear. Second, it may be possible to effectively address the peer effects of disabled students by providing higher quality inputs. Indeed, I find evidence that more experienced teachers are better able to limit negative peer effects, and Aizer (2008) finds that treating students with ADD mitigates negative peer effects without classroom reassignment. Third, this paper does not study non-disabled students' social or emotional development; segregation could deny non-disabled students an important opportunity to gain tolerance and an understanding of people with disabilities. Finally, even if segregation were efficient, it might not be equitable. Disabled students already have lower achievement, and segregating them may further reduce their academic achievement and self-esteem. Thus, while this analysis sheds light on the relationship between special education and peer effects, further research is necessary to gain a more complete understanding of the costs and benefits of special education.

Figure 2.1: Special Education by Disability Type, U.S. 1977-2007



Source: U.S. Department of Education, Office of Special Education Programs (OSEP), Data Analysis System (DANS) Note: Between 1987 and 2001, data for preschool-aged children were not legally allowed to be disaggregated by disability type, so for these years, these children are categorized as "pre-school age with disability."

Table 2.1: Student Status and Exam Participation by Grade

Panel A: Special Education Status and Test Participation

		% of Students		% Taking Math Exam		% Taking English Exam	
Grade	Students	Regular	Special Ed	Regular	Special Ed	Regular	Special Ed
4	577296	91.8	8.2	95.9	82.7	90.6	77.3
5	573444	90.7	9.3	97.3	82.0	92.8	80.5
6	553291	90.3	9.7	98.0	84.1	93.7	83.1
7	551775	90.1	9.9	98.3	83.9	93.6	83.2
All	2255806	90.7	9.3	97.3	83.2	92.6	81.1

Panel B: Special Education and Disability Status

	% of Students				
	Disabled	Disabled	Disabled	Disabled Without SE	
Grade		With SE	Without SE	Undiagnosed	Declassified
4	11.2	8.2	3.0	3.0	0.0
5	11.6	9.3	2.3	2.1	0.2
6	11.4	9.7	1.7	1.1	0.6
7	10.9	9.9	1.0	0.1	0.9
All	11.3	9.3	2.0	1.6	0.4

Table 2.2: Summary Statistics of Students by Disability Status

	Non-Disabled		Disabled With SE		Undiagnosed		Declassified	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Math Score	0.15	0.89	-1.33	1.07	-0.98	0.89	-0.41	0.84
Missing Math Score	0.02	0.14	0.11	0.31	0.04	0.20	0.02	0.16
English Score	0.14	0.90	-1.32	0.99	-0.95	0.83	-0.39	0.79
Missing English Score	0.07	0.25	0.13	0.34	0.12	0.33	0.05	0.23
Probability of Suspension	0.03	0.16	0.07	0.26	0.09	0.29	0.06	0.24
Black	0.33	0.47	0.39	0.49	0.41	0.49	0.39	0.49
Hispanic	0.39	0.49	0.46	0.50	0.45	0.50	0.44	0.50
Other Race	0.14	0.35	0.05	0.21	0.05	0.22	0.06	0.23
Female	0.51	0.50	0.33	0.47	0.34	0.47	0.34	0.47
ELL	0.10	0.31	0.19	0.39	0.20	0.40	0.13	0.34
Free Lunch	0.84	0.37	0.92	0.27	0.92	0.28	0.88	0.33
Repeating Grade	0.02	0.14	0.06	0.23	0.09	0.29	0.05	0.23
Switch Schools	0.25	0.43	0.31	0.46	0.18	0.39	0.48	0.50
Percentage of School-Grade Peers								
Black	33.1	30.2	37.9	29.2	37.4	30.1	39.4	29.1
Hispanic	39.3	26.9	41.1	25.8	42.9	26.9	41.5	26.3
Other Race	13.3	17.1	9.3	13.4	9.7	14.8	9.1	13.5
Female	49.3	5.3	48.3	5.0	49.2	5.2	48.8	6.3
ELL	11.2	9.7	11.3	9.1	12.0	9.7	11.1	9.7
Free Lunch	84.5	19.7	87.1	17.2	88.7	16.4	86.2	17.0
Disabled with SE	7.2	6.3	12.7	7.7	7.8	7.0	9.2	6.6
Undiagnosed	1.6	2.0	1.6	2.0	3.2	2.9	0.8	1.4
Declassified	0.4	0.8	0.4	1.0	0.2	0.7	1.9	6.9
N	1957491		166134		34798		7608	

Note: Statistics are only presented for the non-disabled students in the regression sample, and for the purposes of comparison, disabled students who meet the same sample restrictions. For more information, see the text.

Table 2.3: Effects of Special Ed/Disabled Students on Non-disabled Peers

	Math Scores			English Scores			Probability of Suspension		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Disabled With SE	-0.0392	-0.0513	-0.0532	0.0226	0.0085	0.0075	0.0154	0.0163	0.0167
	(0.0483)	(0.0487)	(0.0486)	(0.0401)	(0.0405)	(0.0404)	(0.0091)	(0.0091)	(0.0091)
Disabled Without SE		-0.3848**			-0.4483**			0.0294	
		(0.0968)			(0.0838)			(0.0169)	
Undiagnosed			-0.4579**			-0.4879**			0.0446*
			(0.1067)			(0.0926)			(0.0175)
Declassified			-0.0401			-0.2640			-0.0420
			(0.2154)			(0.1844)			(0.0556)
N	1920768	1920768	1920768	1821868	1821868	1821868	1957491	1957491	1957491

Note: Specifications control for student characteristics, peer characteristics, school-grade fixed effects, and year-grade fixed effects. Standard Errors are clustered by school. Significance levels are *0.05 **0.01

Table 2.4: Teacher Experience and Peer Effects

	Math Scores			English Scores			Probability of Suspension		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Disabled With SE	-0.0519	-0.0516	-0.0351	0.0071	0.0082	0.0251	0.0104	0.0103	0.0118
	(0.0497)	(0.0492)	(0.0499)	(0.0404)	(0.0400)	(0.0401)	(0.0090)	(0.0090)	(0.0097)
Undiagnosed	-0.4737**	-0.4664**	-0.3787**	-0.5085**	-0.5045**	-0.4651**	0.0483**	0.0480**	0.0363
	(0.1084)	(0.1082)	(0.1143)	(0.0964)	(0.0965)	(0.1032)	(0.0177)	(0.0177)	(0.0201)
Declassified	0.0140	0.0094	-0.0656	-0.2804	-0.2789	-0.2991	-0.0372	-0.0372	-0.0531
	(0.2567)	(0.2560)	(0.2762)	(0.2257)	(0.2245)	(0.2354)	(0.0689)	(0.0689)	(0.0743)
Proportion Novice		-0.0724**	-0.0458**		-0.0574**	-0.0379*		0.0038	0.0021
		(0.0104)	(0.0174)		(0.0090)	(0.0161)		(0.0022)	(0.0035)
Special Ed*Novice			-0.1639			-0.1604			-0.0139
			(0.1252)			(0.1131)			(0.0296)
Undiagnosed*Novice			-0.9885*			-0.4343			0.1315
			(0.4466)			(0.3994)			(0.0984)
Declassified*Novice			0.5853			0.1379			0.1331
			(0.7809)			(0.6408)			(0.2437)
N	1866220	1866220	1866220	1769117	1769117	1769117	1901608	1901608	1901608

Note: Specifications control for student characteristics, peer characteristics, school-grade fixed effects, and year-grade fixed effects. Standard Errors are clustered by school. Significance levels are *0.05 **0.01

Table 2.5: Educational Environment and Peer Effects

	Math Scores		English Scores		Probability of Suspension	
	(1)	(2)	(3)	(4)	(5)	(6)
Disabled with SE	-0.0699	-0.0521	-0.0466	0.0108	0.0245	0.0105
	(0.0812)	(0.0500)	(0.0650)	(0.0405)	(0.0159)	(0.0092)
Disabled in SE Only Class	0.0212		0.0712		-0.0197	
	(0.0857)		(0.0695)		(0.0172)	
SE Isolation Index		-0.0018		0.0037		0.0003
		(0.0104)		(0.0078)		(0.0020)
Undiagnosed	-0.0366	-0.3844**	-0.3061*	-0.4406**	0.0504	0.0316
	(0.1644)	(0.1145)	(0.1467)	(0.1004)	(0.0258)	(0.0197)
Undiagnosed and will be in SE Only Class	-0.7910**		-0.3648		-0.0035	
	(0.2355)		(0.2032)		(0.0411)	
Undiagnosed Isolation Index		0.0111*		0.0082		-0.0021*
		(0.0050)		(0.0043)		(0.0010)
Declassified	0.0803	0.1868	-0.0559	-0.0924	0.1275	0.0273
	(0.3944)	(0.3274)	(0.3624)	(0.2758)	(0.1013)	(0.0847)
Declassified and was in SE Only Class	-0.0897		-0.3592		-0.2770*	
	(0.5648)		(0.4934)		(0.1392)	
Declassified Isolation Index		0.0049		0.0052		0.0019
		(0.0058)		(0.0050)		(0.0013)
N	1866220	1866220	1769117	1769117	1901608	1901608

Note: Specifications control for student characteristics, peer characteristics, school-grade fixed effects and year-grade fixed effects. Standard Errors are clustered by school. Significance levels are *0.05 **0.01

Table 2.6: Balancing Tests for the Proportion of Special Ed/Disabled Peers

	Black	Hispanic	Other Race	Female	ELL	Free Lunch	Repeating Grade	Switch Schools	Missing Math Score	Missing English Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Disabled With SE	-0.0063	-0.0077	0.0022	-0.0273	-0.0407**	-0.0474**	0.0023	-0.0277	-0.0099	-0.0108
	(0.0124)	(0.0142)	(0.0076)	(0.0159)	(0.0097)	(0.0053)	(0.0063)	(0.0223)	(0.0076)	(0.0094)
Undiagnosed	-0.0005	0.0022	-0.0008	-0.0088	0.0335	-0.0403**	0.0165	-0.0278	0.0025	0.0843**
	(0.0296)	(0.0340)	(0.0174)	(0.0441)	(0.0204)	(0.0098)	(0.0155)	(0.0342)	(0.0254)	(0.0293)
Declassified	0.0350	-0.0180	0.0168	0.0144	-0.0915*	-0.0350	0.0752*	0.3943	0.0169	0.0316
	(0.0667)	(0.0692)	(0.0374)	(0.0783)	(0.0422)	(0.0260)	(0.0375)	(0.2021)	(0.0344)	(0.0481)
N	1957491	1957491	1957491	1957491	1957491	1957491	1957491	1957491	1957491	1957491

Specifications control for peer controls, school-grade fixed effects, and year-grade fixed effects. Standard Errors are clustered by school. Significance levels are *0.05 **0.01

Table 2.7: Robustness Checks

	Math Scores			English Scores			Probability of Suspension		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Disabled With SE	-0.0532	0.0647	-0.0279	0.0075	0.0874*	0.0075	0.0167	0.0030	0.0226*
	(0.0486)	(0.0506)	(0.0329)	(0.0404)	(0.0429)	(0.0272)	(0.0091)	(0.0076)	(0.0102)
Undiagnosed	-0.4579**	-0.3582**	-0.2658**	-0.4879**	-0.2890**	-0.2524**	0.0446*	0.0035	0.0449*
	(0.1067)	(0.1210)	(0.0844)	(0.0926)	(0.1111)	(0.0777)	(0.0175)	(0.0196)	(0.0211)
Declassified	-0.0401	-0.0209	0.1107	-0.2640	0.0090	-0.2052	-0.0420	-0.0028	-0.0074
	(0.2154)	(0.2920)	(0.1851)	(0.1844)	(0.2532)	(0.1410)	(0.0556)	(0.0460)	(0.0549)
School-Year FE	No	Yes	No	No	Yes	No	No	Yes	No
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes
N	1920768	1920768	1920768	1821868	1821868	1821868	1957491	1957491	1957491

Note: Specifications control for student characteristics, peer characteristics, school-grade fixed effects and year-grade fixed effects.Standard Errors are clustered by school. Significance levels are *0.05 **0.01

Table 2.8: Timing of Diagnosis and Peer Effects

	Math Scores		English Scores		Probability of Suspension	
	(1)	(2)	(3)	(4)	(5)	(6)
Disabled With SE	-0.0532		0.0075		0.0167	
	(0.0486)		(0.0404)		(0.0091)	
Disabled With SE - Diagnosed Before Last Year		-0.0607		0.0024		0.0107
		(0.0509)		(0.0419)		(0.0098)
Disabled With SE - Diagnosed Last Year		0.0056		0.0407		0.0721
		(0.1295)		(0.1156)		(0.0409)
Undiagnosed	-0.4579**		-0.4879**		0.0446*	
	(0.1067)		(0.0926)		(0.0175)	
Undiagnosed - Diagnosed After Next Year		-0.6347**		-0.7061**		0.0262
		(0.1463)		(0.1397)		(0.0231)
Undiagnosed - Diagnosed Next Year		-0.2944*		-0.2888*		0.0628*
		(0.1365)		(0.1136)		(0.0249)
Declassified	-0.0401	-0.0299	-0.2640	-0.2552	-0.0420	-0.0362
	(0.2154)	(0.2155)	(0.1844)	(0.1846)	(0.0556)	(0.0551)
N	1920768	1920768	1821868	1821868	1957491	1957491

Note: Specifications control for student characteristics, peer characteristics, school-grade fixed effects and year-grade fixed effects. Standard Errors are clustered by school. Significance levels are *0.05 **0.01

Chapter 3

Worker Absence and Productivity: Evidence from Teaching

with Jonah E. Rockoff

Accepted for publication by *Journal of Labor Economics* on 10/19/2011
http://www.journals.uchicago.edu/JOLE/

3.1 Introduction

There is scant evidence on the productivity losses from worker absence, despite the fact that absenteeism results in an annual loss of two percent of work time in the U.S. (Bureau of Labor Statistics, 2008). Several highly regarded studies in economics have documented drops in productivity during labor disputes (Kleiner, Leonard, and Pilarski (2002), Krueger and Mas (2004), and Mas (2008)), but labor disputes are rare — accounting for just one one-hundredth of a percent of lost work time — and it is unclear how these results generalize to more common sources of worker absence, such as illness or personal business.[1]

In this paper, we present evidence on the impact of absenteeism on productivity using detailed panel data on the timing, duration, and causes of absences among teachers and the gains in academic achievement made by their students.[2] We take advantage of this data in several ways to address the endogeneity of absenteeism. First, we base our identification on variation within teachers over time to avoid bias from the correlation of absenteeism with persistent differences in productivity across teachers. Indeed, the richness of our data allows us to identify the impact of absences using variation within the same teacher, school, and grade level. Second, we contrast estimates of the impact of absences that occur *prior* to student exams with those that occur *afterwards*; only the former can have a direct causal impact on our productivity measure. In these respects, our approach is similar to Mas and

[1] In addition, labor disputes involve more than just the replacement of full-time employees with temporary workers and are likely to have important effects on employee morale and effort. For example, Krueger and Mas (2004), who study the production of Bridgestone/Firestone tires, find that defective tires were most likely to be produced during the period before a major strike (while regular workers were still on the job) and just before a new contract was settled (when striking employees worked alongside their replacements). Statistics on the frequency of labor disputes can be found in Bureau of Labor Statistics (2009).

[2] Economists have used student achievement data extensively to study productivity in teaching, with early studies by Hanushek (1971) and Murnane (1975) and recent work by Rockoff (2004), Rivkin, Hanushek, and Kain (2005), and Aaronson, Barrow, and Sander (2007), among others. There is some debate around how student sorting affects the measurement of teacher productivity (see Kane and Staiger (2008), Rothstein (2010)). However, our identifying assumptions are much weaker than those needed to identify variation in quality between teachers, and we present direct evidence against our results being driven by student sorting.

Moretti (2009); they evaluate peer effects among supermarket cashiers using variation in productivity within workers over time and exploiting the fact that peers can only directly affect co-workers's productivity after they arrive at work. We also use a number of specifications and robustness checks to confirm that our findings are not driven by teachers taking more absences when they are assigned more difficult students, or by correlations between teacher absenteeism and student absenteeism or misbehavior.

Reductions in productivity associated with worker absence in teaching are statistically and economically significant. These negative effects occur for absences prior to student exams but not afterwards, supporting a causal interpretation. Our baseline estimates imply that the average difference in *daily* productivity between regular teachers and temporary substitutes is equivalent to replacing a teacher of average productivity with one at the 10th percentile for math instruction or the 20th percentile for English instruction.[3] We also find that productivity losses from absenteeism are greater for more experienced teachers, consistent with evidence from various studies that experienced teachers are more productive.

In addition, we provide evidence that daily losses in productivity from worker absence are decreasing in absence duration. There are several reasons why this might be so. For example, managers may engage in costly search in order to hire more productive substitute workers for longer assignments, temporary workers may learn on the job, and the supply of more productive substitutes may be greater for longer job assignments. Our estimates suggest that the daily productivity loss when a substitute is used for a single day is even greater than replacing an average teacher with one at the 1st percentile in math and equivalent to replacing an average teacher with one at the 3rd percentile in English. In other words, extremely

[3]Ours is not the first paper to estimate a negative impact of teacher absence on student achievement, but it is the first to examine variation in absence duration or cause, and the first to exploit the timing of absences relative to student exams. Miller, Murnane, and Willet (2008) and Clotfelter, Ladd, and Vigdor (2009) estimate the average effect of teacher absence on student achievement using a teacher fixed effects approach. Duflo and Hanna (2005) document the negative impact of teacher absences on student achievement using a randomized control trial in rural India, where substitutes are not used to replace absent teachers.

little production appears to take place when a teacher is absent for a single day, despite the presence of a paid temporary substitute. In contrast, the average daily productivity loss from replacing regular teachers with "long-term" substitutes is equivalent to replacing a teacher of average productivity with one at the 19th percentile in math and the 20th percentile in English.

We also investigate variation in the effects of absences with different causes. Indeed, one concern for our analysis is that shocks to worker health may lower productivity at work in addition to increasing absenteeism. Despite a large literature on the impact of health on wages, earnings, labor force participation, and education (Currie and Madrian (1999), Smith (1999), Currie (2009)), there is little research on the impact of poor health on productivity at work-what social psychologists have labeled "presenteeism."[4] If worker health shocks directly affect productivity on the job, we might expect to see outsized impacts of absences that are related to serious health conditions. However, we find that health and non-health related absences have very similar negative effects on productivity.

Last, but not least, we examine the importance of absence timing by focusing on the periods just prior to and during student examinations. We find productivity losses for absences during periods well before exams, but larger impacts for absences in the weeks and days leading up to exams. Furthermore, impacts are an order of magnitude greater for absences on the day(s) students are tested, which we show is likely mediated by the testing environment, rather than cheating. This analysis indicates that the importance of labor productivity for specific output measures can vary considerably over the production cycle. In the production of education, actions taken by teachers just prior to and during exams can have outsized effects on measured student achievement.

[4]The literature in social psychology examines cross-sectional variation in self-reported measures of health and productivity (e.g., Goetzel et al. (2004), Pauly et al. (2008)). In addition, some development economists have studied health and productivity of agricultural laborers (Strauss and Thomas (1998)).

The paper proceeds as follows. In Section 3.2 we provide a conceptual framework to motivate our empirical work. In Section 3.3 we describe the data, and in Section 3.4 we present our main empirical estimates, robustness checks, and extensions. Section 3.5 offers some conclusions and discusses the extent to which our findings might generalize to other contexts.

3.2 Conceptual Framework

We briefly present a conceptual framework that provides empirical predictions and highlights important issues for our analysis. Consider the productivity of a representative worker r on a specific day $t(q_{rt})$ as the sum of ability, work experience, and a stochastic daily component. In Equation 3.1, we write total production over days indexed from 1 to T as a function of daily labor productivity for the representative worker r, the productivity of substitute s that replaces the regular worker when absent, and other production inputs (X).

$$Q_T = f_T(q_{j1}, q_{j2}, ..., q_{jT}, X), \quad j = \begin{cases} r \text{ if present} \\ s \text{ if absent} \end{cases} \tag{3.1}$$

By assumption, production increases with labor productivity on any day. If expected productivity is lower for substitute workers than regular workers, increases in absenteeism should lower production. Also production losses from absenteeism will be greater for more productive regular workers, all else equal.

In addition, we posit that the expected *average* productivity of a substitute worker is increasing in the length of the substitute's work assignment. There are several reasons to expect the skill level (ability or experience) of substitutes to be greater for longer jobs: managers searching for better workers or allocating the best available workers to longer assignments, more highly skilled workers willing to take a longer assignment (see Gershenson, 2011), or workers learning on the job. If substitute productivity rises with job assignment length, then for any two spells of lengths M and N days, $M > N$, the expected loss from the M day spell should be less than M/N times the loss from the N day spell. We test this hypothesis explicitly in Section 3.4.

Of course, regular workers will choose when to be absent when the benefits (e.g., leisure) outweigh the costs (e.g., lower pay), and this complicates identification in a regression of productivity measured over a given period on the number of worker absences. We consider the net benefits of absence on any given day as determined by three factors: (1) worker-specific factors that do not vary over time (e.g., tastes for leisure), (2) job characteristics (including salary) which may change over time, and (3) a stochastic daily component (e.g., health) which may persist over time.

Even if substitute workers were, in expectation, equally productive as the workers they replace, one might find a spurious relationship between absenteeism and production. For example, more able workers may also derive greater enjoyment from time spent at work, creating a correlation between the value of leisure and ability, both of which are typically unobservable. To address this concern, one can compare production for the same worker across time, and examine how production varies with absenteeism.

A thornier empirical problem is that time-varying elements of productivity and the net benefits of absence may be correlated. For example, changes in production inputs will affect productivity and may also make a job less pleasant, causing workers to show up less often. A similar problem would arise if workers experience persistent negative health shocks and are less productive on the job, in addition to taking more time off from work. To address this issue, one could limit comparisons not only to the same worker over time but also to periods in which absences varied but other factors were held constant. However, there may still be bias due to factors which cannot be directly observed.[5]

To gauge the importance of a number of sources of bias, one can use a placebo test based on the idea that a worker's production over a given time period cannot be directly related to her future absences. Taking any factor that lowers productivity, makes absenteeism more attractive, and is constant within workers over a set of days 1 to T, we can see that, conditional on the number of absences between day 1 and day $T - K$, the unobservable factor will create a correlation between productivity during days 1 to $T - K$ *and* increase absences during days $T - K + 1$ to T. Thus, a relationship between current productivity and future absenteeism would be evidence of bias: we should observe no relationship between productivity measures and subsequent absenteeism if the link between productivity and absenteeism is causal.

Passing such a placebo test is, of course, not proof of causality. Unobservable factors that are imperfectly correlated across the periods from day 1 to $T - K$ and day $T - K + 1$ to T will still hold the potential for bias. While addressing all potential sources of bias in a non-experimental (or quasi-experimental) setting is quite difficult, one can assess the importance of many potential biases using detailed data. For example, one issue is that temporary negative health shocks may cause workers to take more time off and be less productive on the job. To test for this source of bias, one could compare the productivity effects of health-related absenteeism to the effects of absences for reasons such as personal business, vacation, or jury duty. If the health bias exists, one would expect health related absences to appear more detrimental to productivity.

[5] One way to address the issue of unobservable factors is to use an instrumental variable for absenteeism. In developing countries, economists have implemented field experiments which randomized introduction of financial bonuses for work attendance (Kremer and Chen (2001), Duflo and Hanna (2005)). We lack such experimental variation. We discuss one potential instrumental variable (inclement weather and commuting distance) in Section 3.4, but we find it has little power to predict absences in our setting. We therefore rely on other empirical strategies.

3.3 Data and Descriptive Statistics

Our data come from New York City, the largest school district in the U.S., and cover the school years 1999-2000 through 2008-2009. We focus on teachers of math and English in grades 4 to 8, who can be linked to students for whom we generally have math and English test scores in both the current and previous year. Students in elementary grades (4, 5, and some in grade 6) typically have the same teacher for both subjects, while older students are taught by two different teachers.[6] Over this period, the timing of exams ranged from early March to mid-May for math and from early January to mid-May for English (Appendix Table B1). Exam periods lasted from one to three days, followed by a five-day make-up exam period for students absent during all or part of the regular exam.

In addition to math and English test scores, we have information on students' absences, suspensions, demographics, and receipt of free/reduced price lunch (a measure of poverty), special education for disabled students, and English Language Learner services.[7] Data on teachers' demographics, graduate education, and experience were obtained from payroll records.

We have records of the date and reason given for all daily teacher absences over this time period. The rules governing teacher absences are set forth in a collectively bargained

[6]Students in grade 6 are taught by the same teacher in schools whose terminal grade is 6. Student-teacher links were unavailable in some schools at the start of our sample, and we only include students in school-year cells for which we match greater than 75 percent of students with teachers. Over this period, students with disabilities were typically taught in separate classrooms or schools and did not take the same standardized tests as general education students. We therefore exclude all classrooms where the portion of special education students exceeded 25 percent. We also exclude a few classrooms with less than 7 or greater than 45 students, where the teacher switches schools during the year, or where the teacher was not on active duty for more than half the year or until after the exam.

[7]We unfortunately lack daily information on student absences; we only know each student's total absences for the school year. Thus, we are unable to estimate a placebo test for whether students are affected by the absence of their regular teacher on days when they themselves do not show up at school. We leave this line of inquiry to future work. While we can test if teacher absences have smaller effects on students who themselves are absent more often, the correlation of student absenteeism with other characteristics would make the interpretation of such a test unclear.

contract between the teachers union (the United Federation of Teachers) and the school district. Teachers earn ten days of paid absence per school year (one per month). However, teachers accumulate unused absences, up to a cap of 200 days, and are paid 1/400th of their most recent salary for each unused absence when they retire. Thus, using "paid" absences poses a real financial cost for teachers unless they are certain to reach the 200 day cap.[8] These rules allow teachers to use up to ten absences each school year for "Self Treated Sickness" — sick days which do not require proof of illness from a physician — or "Personal Days." Teachers can take only three "Personal Days" each year, but there is no barrier to a teacher labeling an absence for personal business as "Self-treated Sickness."[9] Absences for "Medically Certified Sickness" (i.e., illness certified by a physician) and several other types of absences (Conferences/School Activities, Funeral/Death in Family, Jury Duty/Military Service, Injury, Graduation Attendance, Religious Holiday, and Grace Period) do not count towards the ten day cap.[10] A few absences are Unauthorized.

We also have data on the type, timing, and duration of extended work leaves and job separations, which we classify into 11 categories: Maternity Leave, Child Care Leave, Medical Leave, Sick Family Member Leave, Personal Leave, Sabbatical, Resignation or Retirement,

---

[8]This constraint is unlikely to bind for the vast majority of teachers. Among all teachers in New York (not just those teaching math and English in grades 4-8) hired in the school year 1999-2000, more than two thirds left teaching in the district by the end of our ten year sample, and only three percent of remaining teachers (1 percent of the cohort) used absences at a rate low enough to reach 200 in 25 years (i.e., 20 absences or less in over ten years).

[9]The notion that absences for Self Treated Sickness are likely to include many absences not related illness is supported by absence rates across days of the week. It is reasonable to believe that absences taken for personal reasons would be more prevalent on Mondays and Fridays, providing workers with a long weekend, and rates of absence for Self-treated Sickness and Personal Days are both nearly 50 percent higher on Mondays and Fridays than on Tuesdays through Thursdays. In contrast, absence rates on Tuesdays through Thursdays are nearly identical to rates for Mondays and Fridays if we examine illnesses certified by a doctor. Variation in absence by day of the week is not a new finding. High absence rates on Mondays have been found in studies of absence which go back many decades (e.g., Bezanson et al., 1922), and absences on Fridays are low in manufacturing jobs where workers are paid in person at the end of each week. In our setting, teachers' paychecks are mailed or directly deposited.

[10]"Grace period" typically applies to teachers who are absent prior to an extended leave (e.g., maternity). These teachers have exhausted their paid absences and are not paid, and grace period is capped at 30 days.

Involuntary Termination, Certification Termination, Death, and Other (e.g., unauthorized leave, military deployment, and leave without pay for various reasons such as working in a charter school).[11] Rules governing extended leaves are also set forth in the union contract, in accordance with applicable laws such as the Family and Medical Leave Act. Note that these events can impact students when they end as well as when they begin (e.g., women beginning their maternity leave in the summer may return several weeks or months after the school year starts).[12]

Table 3.1 shows summary statistics on the frequency and duration of spells of absence, including extended leaves and job separations. Duration is defined by the number of instructional days (i.e., work days) missed, not calendar days, though the two are highly correlated. Teachers are absent 10 days on average, or roughly 5 percent of the school year.[13] Perfect

[11]Certification Termination refers to termination of teachers who lacked required credentials; these occur primarily just before the school year 2003-2004, when state requirements were strictly enforced after a legal battle between New York City and New York State.

[12]In about 10 percent of cases, leaves are consecutive (e.g., maternity leave can turn into child care leave), and we aggregate these into a single leave, using the initial leave to classify the sequence. If daily absences are followed immediately by an extended leave (e.g., medical leaves are often preceded by absences for "Medically Certified Sickness"), we group these together and classify the spell by the extended leave of absence. In some cases, consecutive daily absences are not all labeled with the same code. In these instances, we label all absences in the spell under a single code, giving priority to more specific causes, in the following order: Injury, Medically Certified Sickness, Funeral/Death in Family, Jury Duty/Military Service, Religious Holiday, Graduation Attendance, Conferences/School Activities, Personal Day, Self-treated Sickness, Grace Period, and Unauthorized.

[13]Rates of absence for representative samples of U.S. workers are available from the Current Population Survey, which asks about time missed from full-time work during a particular week. Rates of absence were roughly 4 percent in the public sector (3 percent in the private sector) over the time period we analyze. Although this is somewhat lower than the 5 percent rate in our sample of teachers, CPS rates exclude vacation and personal days, while a non-trivial fraction of teachers' "Self Treated Sickness" absences are likely taken for personal matters. Comparable data on spells and spell length are not available in the CPS but are reported in two studies that use daily data on employee absences spanning a long time period. Ichino and Moretti (2008) report that employee absences for sickness in a large Italian bank last an average of 3.8 days; our figure for Medically Certified Sickness and Medical Leaves is 3.0 days. Barmby, Orme, and Treble (1991) examine data from a British manufacturing firm in the late 1980s and report mean absence spells of 5 days; like our data, spell length is skewed, with spells of 5 days or less in duration accounting for over 80 percent of spells but 40-45 percent of work days missed. In our data, spells of 5 days or less account for 98 percent of all spells and 78 percent of work days missed. These (admittedly limited) comparisons suggest that teachers' spells of absence may tend to be short relative to other sectors and occupations.

attendance by a teacher occurs in only three percent of cases.

"Self-treated Sickness" accounts for a large portion of all days missed, more than four days per teacher per year on average, while Medically Certified Sickness and Conferences/School Activities account for two days and one day, respectively, per teacher per year. The extended leave that accounts for the most days missed is Medical Leave, which is taken by just over one percent of teachers per year but has an average duration of almost 43 instructional days. Other types of extended leave are even less common but have similarly long durations (e.g., maternity leave is taken by 0.5 percent of teachers and has an average duration of 48.6 days).[14]

In Table 3.2 we show the mean and standard deviation of absences for the math and English teachers in our sample, both over the entire school year and broken out by timing: prior to exams, after exams, during the exam period, and during the make-up exam period.[15] As one might guess from the statistics presented in Table 3.1, the distribution of absences is right-skewed, and the standard deviation of total absences (roughly 10) is quite close to the mean. We also present standard deviations of residuals from regressions of teacher absences on teacher-school-grade fixed effects. These "within-teacher" measures are about 65 percent as large as the standard deviation based on both "between" and "within" variation. This implies that almost half of the variance in absences among teachers occurs within teachers across years, providing us considerable identifying variation.[16]

[14]To better understand teachers' potential control over the timing of extended leaves, we have examined the percentages of each type of event that begin or end during the middle of the school year. Maternity and Medical Leaves - where we do not expect much control over timing - result in missed work days 90 and 93 percent of the time, respectively, while Personal and Other Leaves - where timing may be partially under teachers' control - only result in missed work days 20 and 30 percent of the time, respectively.

[15]The unit of observation for these tables is a teacher-grade-year cell. We allow for multiple observations of teachers of multiple grades in the same year since the exam dates differ across grade levels.

[16]The within-teacher correlation in total days missed across years is just 0.18 in our sample, and there are very few teachers who do not contribute to identification. Less than 10 percent of teachers observed in adjacent years did not experience a change in their number of absences; less than 2 percent of teachers observed to three consecutive years did not experience a change their number of absences.

Before proceeding to our main analysis, we examine associations between absence frequency and the characteristics of students and teachers using negative binomial regressions. We find a marginally significant coefficient on students' prior math test scores, suggesting that teacher absence — if costly to students — may contribute slightly to inequality in educational outcomes (Table 3.3). There is no significant relationship between work days missed and free lunch receipt (our measure of poverty), special education services, or English language learner services, but we find that teachers of Hispanic students miss fewer days, relative to teachers of white students. We do find that missed work days are positively related to student absences, though the coefficient is fairly small.[17] Results from negative binomial regressions of work days missed on a set of teacher characteristics are shown in Table 3.4. Having a graduate degree is associated with fewer work days missed, as is having few years of teaching experience. Younger female teachers miss more days of work relative to teachers of different gender and age categories, and black and Asian teachers miss fewer days relative to white teachers.

Data on the individuals working as substitute teachers in New York City is unfortunately unavailable, but we can compare their employment requirements and wages to those of regular teachers.[18] Substitutes in New York do not need to pass state certification requirements (i.e., possess a degree in education and pass a series of exams) but, like regular teachers, must have a bachelor's degree and must pass a criminal background check.[19] If a substitute teacher works for more than 40 days during the school year, they must have certification

[17]One likely explanation for this finding is correlation between teacher and student illness. Since this could generate a spurious correlation of teacher absences with student achievement, we estimate specifications that control for students' current absences as robustness checks.

[18]National statistics on substitute teachers are also unavailable; the Bureau of Labor Statistics groups substitutes with other jobs (e.g., tutor, academic advisor) in the category "Teachers and Instructors, All Other."

[19]Requirements are similar in other parts of the U.S., though Henderson, Protheroe, and Porch (2002) report that 19 states do not require a bachelor's degree.

or complete additional certification coursework before the start of the following school year. Substitute teachers are currently paid just over \$150 per day of work, or about \$21 per hour given the length of a typical school day and about half of what regular teachers earn for additional hours of work (over \$40).

Another important source of substitute teachers in New York City is the Absent Teacher Reserve (ATR), which consists of certified teachers who lost their jobs due to grade reconfiguration, reduction in student enrollment, programmatic change, or phase out or closing of their school. ATR teachers have been unable to find another job, but, in accordance with the union contract, the school district pays their full salary and they work as substitute teachers, either on a per-diem or long-term basis. Individual schools using ATR teachers as substitutes pay 50 percent of the daily wage to the school district, and thus have a financial incentive relative to using other substitute teachers. A consistent series of statistics on the size of the ATR is unavailable, but recent reports put the number at around 500 teachers. Thus, with absence rates of roughly 5 percent and a teaching population of roughly 75,000, ATR teachers likely cover about 10 to 15 percent of substitute teacher assignments.

3.4 Regression Specifications and Empirical Estimates

We begin by estimating a regression specification of the following form:

$$Y_{ijkgst} = \delta A_{jt} + \beta X_{it} + \mu Z_{kt} + \lambda W_{jt} + \rho V_{sgt} + \pi_{gt} + \epsilon_{ijkgst} \tag{3.2}$$

where Y_{ijkgst} is the exam score of student i, taught by teacher j in classroom k, grade g, school s and year t. A_{jt} is the number of work day absences for the student's teacher, X_{it}, Z_{kt}, and W_{jt} are vectors of, respectively, student, class, and teacher characteristics, V_{sgt} is a vector of school-grade-year characteristics, π_{gt} is a grade-year fixed effect, and ϵ_{ijkgst}

is an error term.[20] Standard errors are clustered at the school level, which produces more conservative estimates relative to clustering at the classroom or teacher. Estimates from this specification suggest that an additional day of work missed by a regular teacher is associated with a decrease in student test scores of 0.0017 and 0.0006 standard deviations in math and English, respectively (Table 3.5, Columns 1 and 5).

Our conceptual framework motivates the concern that teachers who frequently miss work also provide lower quality instruction while on the job. We employ two strategies to address this issue. First, we separate absences by their timing-before, during, or after student exams. Since absences after exams cannot have a direct causal relationship with student exam performance, any observed relationship must be due to endogeneity.[21] When we allow the coefficient on work days missed to differ by their timing relative to student exams (Table 3.5, Columns 2 and 6), we find much larger negative effects prior to the exam than afterwards.[22] The estimated effect of absences prior to the exam is four to five times greater than absences after the exam, though absences after the exam are marginally significant, suggesting some bias in our estimates.

We then include teacher-school-grade fixed effects (π_{jsg} in the notation of Equation 3.2.

[20]Student characteristics include a cubic polynomial in prior year math and English scores, the number of absences and suspensions in the previous year, and indicators for gender, race and ethnicity, free/reduced price lunch, special education, and English Language Learner. We also interact all of these variables with the student's grade level. Teacher characteristics include indicators for the number of years of teaching experience (1, 2, 3, 4, 5, 6, 7+), gender, race, and possession of a graduate degree. School-grade-year and classroom characteristics include averages of student characteristics and class size.

[21]A more direct solution to the endogeneity problem would be an instrumental variables approach. We explored this using an instrument suggested by Miller et al. (2008), the interaction of bad weather with a teacher's commuting distance. Unfortunately, the instrument does not have a statistically significant first stage. Although living more than ten miles away from work has significant power to predict absences on the *actual days* of extreme winter weather, it has no power to predict teachers' total absences prior to exams. This suggests that teachers who have a long commute do miss work due to bad weather but "make up" that day some other time. Equivalently, teachers who live close to work and show up in bad weather may "make up" for it by taking a day off some other time. Using different distance cutoffs (e.g., less than 5 miles or less than 15 miles) does not change these results.

[22]The coefficient estimates on absences during the regular exam and make-up exam periods are also negative and statistically significant. We focus on these results in greater detail in Section 3.4.

When we control for these time-invariant dimensions of instructional quality (Table 3.5, Columns 3 and 7), the effects of absences prior to the exam become smaller (-0.0012 and -0.0006 standard deviations for math and English, respectively) but remain highly significant, while estimates for absences after the exam are statistically insignificant in addition to being quite small (roughly -0.0001 standard deviations in both subjects). These results are in line with a negative causal impact on productivity of replacing a regular teacher with a temporary substitute. They also indicate that absences are negatively correlated with the time invariant dimensions of instructional quality captured by the teacher-school-grade fixed effects.

We test the robustness of these baseline estimates in several ways. First, we drop prior test scores from our control variable (X_{it} and Z_{kt}) and put students' prior test scores as the dependent variable in our regression. In other words, we test whether teachers are absent more often in years when they are assigned students with lower prior test scores. Such a relationship would raise the concern that student sorting might bias our estimates of the impact of absences. However, we find no significant relationship between absences prior the exam and students' prior test scores (Table 3.5, Columns 4 and 8), in contrast to our baseline results.

As an additional robustness check, we take advantage of the fact that over 90 percent of middle school students in New York City take math and English with the same classmates, even though they have different teachers in each subject. If student composition caused achievement to fall and teacher absences to rise, we might expect the absences of math teachers prior to the English exam to be correlated with English achievement, and vice versa.[23] In fact, *if we omit teacher-school-grade fixed effects*, there is indeed a significant

[23]This result is also evidence against our results being driven by the correlation between teacher and student absences shown in Table 3.3. If student absences and teacher absences were related due to illness, we would expect to find effects of English teacher absences on math test scores and vice versa.

coefficient (-0.00031 standard deviations) for the "effect" of English teachers' absences prior to the math exam on math achievement (Table 3.6, Column 1). However, once the fixed effects are included, this estimate becomes much smaller (-0.00007 standard deviations) and insignificant (Table 3.6, Column 2). Math teachers' absences prior to the English exam bear no relation to English achievement, regardless of the omission or inclusion of fixed effects (Table 3.6, Columns 3 and 4).[24]

In further support of the idea that we are estimating causal effects, we have also examined whether our estimates are sensitive to the inclusion of control variables for student absences and suspensions in the *current* school year. Teacher illness could (causally) lead to student illness (and lower achievement), or vice versa, generating a spurious correlation of absences with achievement. Students might also misbehave if they think their teacher will be going away on an extended leave. However, including these control variables has no noticeable impact on our estimates, although students' own absences and suspensions are both negatively related to their level of achievement. These results are available upon request.

Having established a strong case for a causal effect of absences on productivity, it is helpful to consider the magnitude of these effects. We present a back-of-the-envelope calculation to give a better sense of the magnitude of the *daily* productivity loss from having to replace an absent teacher with a temporary substitute. To do so, we make the simplifying assumption that annual productivity differences across teachers — which are well documented by economists — are driven by a linear accumulation of differences in daily productivity. This assumption allows us to estimate the average annual productivity difference between regular teachers and substitutes by summing the daily difference in productivity (-0.0012 standard

[24]The estimate for English teachers' absences on English test scores is smaller here than in our baseline estimates because our sample is limited to middle school. While the point estimates from our baseline specification are larger for elementary grades (-0.08) than middle school (-0.03), we cannot reject that they are the same with a high degree of confidence. For math, estimates for elementary and middle school grades are quite similar to one another (-0.12).

deviations in math test scores) over the roughly 130 instructional days prior to the math exam. Doing so, we arrive at a reduction in math scores of -0.156 standard deviations. We can then compare this effect to the impact of replacing a regular teacher of average productivity with one of lower productivity for the entire school year. Given estimates in the literature, one would have to replace an average teacher with one at the 10th percentile of the teacher productivity distribution to get a similar reduction in math scores. In English, our estimated coefficient on absences (-0.0006 standard deviations) together with a pre-exam period of 110 instructional days (English exams were typically given prior to math exams) suggest that replacing a regular teacher with a substitute is, on average, equivalent to replacing an average teacher with one at the 20th percentile.[25] Thus, our analysis suggests that temporary replacements have drastically lower productivity than regular full-time teachers.[26]

3.4.1 Heterogeneity in Productivity Losses

Our baseline estimates and robustness checks strongly support the notion that productivity in teaching is significantly lower on days when regular teachers are replaced with temporary substitutes. However, it is reasonable to think that the impact of absences may be heterogeneous. Productivity losses may be greater for absences of highly productive teachers, or,

[25] To reach this estimate, we take the results from a study by Kane et al. (2008) of teachers in New York City, though their estimates are similar to other studies in this literature (see Hanushek and Rivkin (2010)). Kane et al. estimate that math test scores fall by -0.12 standard deviations for a one standard deviation decrease in teacher productivity. This implies that replacing an average teacher with one at the 10th percentile (1.3 standard deviations below the mean) would reduce scores by -0.156 standard deviations. Extrapolating our absence coefficient in English (-0.0006) over 110 instructional days implies a reduction in test scores of -0.066 standard deviations. Kane et al. (2008) find students' English test scores fall by -0.08 standard deviations for a one standard deviation decrease in teacher productivity. Given this estimate, to reduce scores by -0.066 standard deviations one would need to replace an average teacher with one at the 20th percentile (0.82 standard deviations below the mean).

[26] Note that our results are not necessarily informative about what the productivity of individuals working as substitute teachers might be if they were employed full-time. This is analogous to how studies of labor unrest (e.g., Krueger and Mas, 2004) examine the productivity of replacement workers under the temporary conditions in which they are hired, not the productivity they these "scab" workers would have if they received the same training and support as regular employees.

alternatively, these teachers may help substitutes provide effective instruction by developing easy-to-use lesson plans. While we cannot observe productivity directly, several studies find that teacher productivity rises quickly over the first few years of their careers (Rockoff (2004), Rivkin et al. (2005), Kane et al. (2008)). We therefore estimate regressions that allow the impact of teacher absences to differ by whether teachers had less than three years or three or more years of prior teaching experience.

We find evidence that absences by experienced teachers cause a greater reduction in student test scores than absences by inexperienced teachers (Table 3.7). The estimated difference in the impact of absence across the two groups of teachers is highly statistically significant in math and marginally significant in English (p-value 0.14). Although point estimates for the impact of absences on student achievement among inexperienced teachers are still negative, we can no longer reject that they are zero. This provides further support to the notion that the losses associated with the use of substitute teachers are caused by their relatively low productivity. In addition to heterogeneity across teachers, the effect of absences may vary across schools and students. Schools may differ in their abilities to find good substitutes, and some may provide substitutes with high quality instructional materials to help reduce the impact of teacher absence. Additionally, Todd and Wolpin (2003) stress that students and parents may respond to lower instructional quality by shifting household resources towards education. We do not have measures of how responsive schools and students are to changes in teacher productivity, but it is not unreasonable to think that high performing schools and high performing students may be better equipped to deal with these issues. We therefore estimated regressions that allow the effect of work days missed to differ across (a) schools with average test scores below and above the citywide median and (b) students with prior test scores below and above the citywide median. In the latter case, since students will vary in prior achievement within classrooms, we also estimated specifications that included *classroom* fixed effects. We find that the negative effects of work days missed are similar across these groups of schools and students in both math and English. These results are available upon request.

As discussed in Section 3.2, several factors suggest that daily productivity losses may decline with the duration of a spell of worker absence. In teaching, this could be due to school principals engaging in costly search for better long-term substitutes, the labor supply decisions of more highly productive substitute teachers, or temporary substitutes learning on the job (e.g., learning children's names and learning styles). To test this hypothesis, we construct variables that allow us to estimate the daily productivity losses associated with absences of different durations: 1 day, 2-3 days, 4-5 days, 6-10 days, 11-30 days, and 31 days or more.[27]

The results are in line with our hypothesis that daily productivity losses are smaller for longer duration absences (Table 3.8). In math, the coefficients decline steadily as we move from single day spells of absence (-0.0036) to spells lasting 31 days or more (-0.0008). In English, the daily productivity loss from single day spells is again the largest in magnitude (-0.0017) and then drops off precipitously. The coefficient estimates in English rise slightly as we move to the longest durations, but we cannot reject that daily productivity losses are the same for all spells of duration two days or longer.

The variation in magnitude between the estimates for single day absences and those with long durations is economically important. To illustrate this point, we again use our back-of-

[27]Let S_{itd} denote the number of spells of absence of duration d for teacher i in school year t, and define the number of work days missed during spells lasting d days as $A_{itd} = dS_{itd}$. For example, if a teacher has two five-day absence spells during the school year, then S_{it5} would equal 2 and A_{it5} would equal 10. Total work days missed over the school year (A_{it}) is the sum of the work days missed from spells of a particular duration over all possible durations (i.e., $A_{it} = A_{it1} + A_{it2} + \cdots + A_{itd}$). Our baseline estimating equation contains an implicit restriction that the daily productivity loss from worker absence is invariant to absence duration, and we relax this constraint and allow coefficients on work days missed to vary across several categories of duration. We report results on effects of absences prior to student exams; we do not find that absences after exams are related to student achievement, regardless of their duration. In cases where a spell of absence begins but does not end prior to an exam, the work days missed prior to the exam are grouped according the duration of the entire spell.

the-envelope calculation, based on a comparison with variation in productivity across regular teachers. For absences lasting just a single day, our estimates suggest that the difference in daily productivity between substitutes and the regular teachers they replace is *greater* than the difference between the daily productivity of an average teacher and a teacher at the 1st percentile in math, and on par with the difference in daily productivity between an average teacher and one at the 3rd percentile in English. Put differently, it appears that very little educational production takes place when a regular teacher misses a single day of work. In contrast, the estimates for the longest spells imply a difference in daily productivity equivalent to replacing an average teacher with one at the 19th percentile for math and the 20th percentile for English — still an important loss in productivity, but far less severe.

3.4.2 Health and Productivity at Work

In our baseline analysis, we restricted the impact of work days missed to be invariant with respect to the reason for the teacher's absence. In many cases, we believe this restriction is probably correct and, under a strict causal interpretation, is probably warranted: conditional on duration, the relative productivity of a substitute should be independent of whether the regular teacher is absent for, say, a funeral or a child's illness.[28] However, teachers may have health conditions that cause them to be less productive on the job, in addition to any impact of health on absence from work. This could potentially make health-related absences appear more detrimental to student achievement than non-health related absences; essentially, estimates of the impact of health-related absences could suffer from omitted

[28]Whether the likelihood of absence was known in advance is outside the scope of our analysis, but it is reasonable to believe that predictable absences might enable teachers or administrators to prepare and therefore be less costly. While we do not have information on predictability in most cases, we have compared the impact of maternity leaves — which are clearly known in advance — to medical leaves — which may be sudden. We find very similar negative impacts of both types of leaves prior to exams and no significant impacts of either type after the exam, suggesting the negative impact of absenteeism in this setting does not derive solely from unpredictability.

variables bias.

To investigate this possibility, we separately examine absences by type, and ask whether absences that we are confident were due to health conditions — Medically Certified Sickness, Medical Leave, and Maternity Leave — have outsized effects relative to other absences.[29]

We find no evidence that health related absences by teachers cause a greater loss in student achievement than other absences (see Table 3.9). When we estimate separate coefficients on the number of days missed prior to student exams, we actually find smaller point estimates for health-related absences, particularly for math. However, one problem with this specification is that health related absences have longer durations, and our previous results suggest that this would cause them to appear less detrimental. When we allow the coefficients for health and other absences to differ by duration, we find they both have very similar magnitudes, and in no case can we reject that they are the same. Thus, we find no evidence that teachers absent for serious health conditions are also less productive while at work. While our test for a link between health and on-the-job productivity is admittedly indirect, it is important to recognize that much of the existing literature on this issue — very little of it by economists — relies on cross-sectional variation and self-reported health and productivity measures.

3.4.3 Worker Absences and the Timing of Productivity Measurement

In the empirical results above, we focus on the significant negative impact of absences prior to student exams, and contrast them with small and insignificant estimates for absences after the exam period. However, the specifications from which these estimates were taken also

[29]Absences for Self-treated Sickness may be related to health, but our results are not sensitive to including them in the non-health-related category or including them as a separate category all to themselves. Our results are also insensitive to placing absences for maternity leave with the “other” category.

included controls for teachers' absences during the time when students were actually taking exams. In Table 3.10, Columns 1 and 3, we redisplay the results from our baseline regressions, including the coefficients on the number of absences during the main exam period — which can last between one and three days — and the five-day make-up period which directly follows it. In both math and English, absences during the main exam period have significant negative impacts on achievement (-0.0244 and -0.0128 standard deviations) that are an order of magnitude greater than the estimated impact of absences in the pre-exam period (-0.0012 and -0.0006 standard deviations). The coefficient for absences during the make-up exam period is negative and significant in math, but in English is it positive, insignificant, and quite close to zero.[30]

The striking results on absences during the main exam period have several possible interpretations. Teachers may improve student performance on the day of the exam through purposeful and permissible actions, such as reminding students of test-taking strategies or making sure that all students understand exam instructions. Teachers might also take actions which are not permissible, such as overtly (or covertly) supplying students with correct answers. Instances of teacher cheating are well-documented (e.g., Jacob and Levitt (2003), New York Times (2010)), and substitute teachers — who typically proctor exams in a teacher's absence — might have little incentive to engage in this type of malfeasance.[31] Another

[30]The negative effect of make-up period absences in math but not English is somewhat puzzling. We speculate that the result is driven from differences between the testing schedule information and the actual dates students were tested in math. Over this period, New York City was permitted to test students within a short window (usually 3 to 5 days) set forth by the state. If some math tests were administered after the originally scheduled date, then a much larger fraction of students may have been tested during what we classify as the make-up period. For example, we discovered that during the school year 2008-2009, extreme winter weather caused the DOE to cancel classes on March 2, 2009 and postpone the start of 3rd, 4th, and 5th grade math exams (New York Times, March 3, 2009).

[31]To gauge whether cheating could explain our findings, we use results from Jacob and Levitt (2003), who estimate that roughly 5 percent of teachers cheat and that cheating increases scores by 0.5 standard deviations (10 additional standard score points on the Iowa Test of Basic Skills) on average. If the probability of absence during the exam is independent of a teacher's intention to cheat, we could expect a coefficient of -0.025. This is larger than our estimate for English (-0.0128) but quite close to our estimate for math

plausible explanation is that students perform worse on high-stakes tests when their regular teacher is absent because of increased anxiety or discomfort. A meta-analysis of two dozen small-scale experimental studies on student familiarity with test examiners finds effect sizes on the order of 0.3 standard deviations (Fuchs and Fuchs (1986)), and there are also many studies demonstrating how anxiety in various forms can impact test performance (e.g., Steele and Aronson (1995)). Finally, a recent experiment by Levitt, List, and Sadoff (2011) finds that students' effort on tests can be very sensitive to small short-term incentives, and it is possible that students exert less effort when the test is administered by a temporary substitute.

Looking at absence frequency, we find some indication that teachers do not wish to be absent on the day of the exam and shift work absences in order to do so; absence rates average 5.8% on days before exams, 5.3% on days after exams, 2.6% on days during the exam, and 8.5% on days during the make-up period. This also raises the possibility that a teacher's absence at so crucial a moment in the school year is a signal about her productivity on the job. To address this issue, we look at teachers in grades 4 and 5, who provide instruction in both subjects, and repeat our regressions while controlling for a student's *current* score in the *other* subject. In other words, we ask whether students score relatively worse in math (or English) when a teacher is absent for the math (or English) exam. Our original findings are quite robust to this much more stringent test, suggesting that, whatever the interpretation, a teacher's absence during high stakes exams has an important negative causal effect on exam performance.

In addition to the effect teachers have on student performance on the day of the test, it is often noted anecdotally that teachers engage in test preparation activities in the days

(-0.0244). However, if teachers who care enough about scores to risk cheating also care enough to show up at work while ill, then teachers absent on the test day would be more honest than average, and these estimates likely overstate the impact of cheating.

and weeks prior to the exam. For example, they might focus on the material and types of questions most likely to be on the exam. We investigate this by allowing for different impacts of absences occurring 1-5 instructional days, 6-20 instructional days, and at least 21 instructional days prior to the exam. Though all absences have negative effects, we find clear evidence that absences in the weeks and days leading up to exams have greater impacts on exam performance than those occurring earlier in the year (Table 3.10, Columns 2 and 4). For math, the coefficient estimate for absences 21+ days prior is -0.00096 standard deviations, similar to our baseline, but for absences 6-20 days and 1-5 days prior, the point estimates are, respectively, double (-0.00185) and nine times (-0.0085) as large. The relative magnitude of the coefficients in English are similar, suggesting that actions taken by regular teachers just before exams are more important for exam performance than those taken earlier in the year.[32]

3.4.4 Persistent Effects of Absence

It is natural to ask whether the impact of teacher absences on students' test scores persists into the following school year. Recent studies of teacher productivity have documented that teachers' effects on scores one year later are between 20 and 50 percent as large as their effects on current test scores (Kane and Staiger (2008) and Jacob, Lefgren, and Sims (2008)).[33] Another motivation to examine persistence is the possibility that the outsized effects of absences close to the exam period reflect "teaching to the test" or that the impact

[32]Allowing for separate coefficients for absences close to the exam does dampen the estimated effects of absence during the exam period, but these coefficients (-0.0162 for math and -0.0081 for English) remain quite large and statistically significant.

[33]The issue of "fade-out" has been raised for other educational interventions, though it may be caused by differences in future resources or belie improvements in other outcomes (see Currie and Thomas (1995), Garces, Thomas, and Currie (2002), Chetty et al. (2010)). Lang (2010) makes the point that rescaling of annual tests to have a mean of zero and standard deviation one could also lead to a perception that the effects of educational interventions fade out.

of absences during the exam reflect cheating or effects on the test taking environment, rather than changes in students' knowledge of the content being tested. If so, then we should see lower persistence in the effects of absences that occur just prior to and during student exams.

Because we lack data on students in grade 9 or those that leave the school district, we first show that our baseline results are similar when we drop all students in grade 8 and students for whom we do not observe test scores the following year (Table 3.11, Columns 1 and 5). When we replace the current year's exam with the following year's exam as the dependent variable, we find that the negative impact of absences prior to the exam exhibit a similar level of fade-out as in previous studies (Columns 2 and 6). For math, the coefficient on work days missed prior to exams in the following year is about 35 percent of the coefficient in the current year; for English, the fade-out is similar, with following year effects roughly 45 percent of current effects. The coefficient on following year test scores is significant at conventional levels in math and marginally significant in English (p-value 0.11).[34]

Importantly, the impact of absences during the exam period exhibits much greater fade-out. The coefficient on following year math scores is only about 10 percent as large as the coefficient on current year scores and is not statistically significant. In English, the coefficient on following year scores is *positive*, albeit not statistically significant. This suggests that, whatever is driving the outsized effects of absence during the exam on current year scores (e.g., cheating, test anxiety), it likely does not reflect real differences in student content knowledge. Also, it is worth noting that the coefficients on absences *after* exams, while never statistically significant, suggest larger negative effects on following year test scores than current year scores, in line with our causal interpretation.

---

[34]If a student does poorly in the current year, it may trigger policies in the following year designed to remediate or improve their performance. In line with this idea, the coefficients grow slightly in magnitude and are somewhat more precisely estimated when we control for future policies (i.e., grade retention, special education, English language learner services, and assignment to a more experienced teacher). For example, the English coefficient grows from -0.027 percent of a standard deviation to -0.028 and is significant at the 8 percent level.

If we break out absences prior to exams by timing (i.e., more than 20 days prior, 6-20 days prior, and 1-5 days prior), we find suggestive evidence that the effects of absences closest to the exams fade out most quickly. Most of the coefficients on following year scores are between 25 and 65 percent of the magnitudes for current year scores. However, for math scores, absences 1-5 days prior to the exam show an unusually large amount of fade-out, with a coefficient in the following year that is only 6 percent of the current year coefficient (Table 3.11, Column 4). This provides some indication, though far from conclusive, that a significant portion of teachers engage in "test prep" activities just before an exam.

3.5 Conclusion

Worker absence is an important phenomenon across all countries, industries, and occupations. Among OECD nations, absence frequency is noticeably higher in northern European countries with generous national sick leave policies (e.g., Barmby, Ercolani, and Treble (2002), Bergendorff (2003)). Absenteeism is also a major concern in developing countries, particularly in the public sector where oversight may be very weak (Chaudhury et al. (2006)).

Despite its ubiquity, there is a paucity of empirical work which convincingly estimates the causal impact of absenteeism on labor productivity. The major hurdle in this line of research is addressing the endogeneity of work absence. To do so, we take advantage of extremely detailed data on the absences of teachers in New York City public schools. We present evidence that missed work days have an economically important negative impact on productivity in teaching. To be confident that our estimates are causal, we focus on variation within teachers over time and contrast the significant effects of absences occurring prior to exams with the lack of any effect for absences occurring afterwards. We find similar impacts of absences across different students and schools, but greater impacts for more experienced (and productive) teachers than for newly hired teachers.

Our estimates of daily productivity losses are smaller for longer spells of absence. This pattern is likely caused by several factors: managers searching for more productive substitutes on longer job assignments, more productive workers applying for longer job assignments, or substitute workers becoming more productive on the job. We also find very large negative effects of work absences just prior to and during student examinations, suggesting that actions taken by the teacher at certain crucial moments in the school year have outsized impacts on student exam performance. Finally, we find no evidence that teachers show up to work when they are too ill to be productive ("presenteeism" in the parlance of social psychologists), though an analysis based on direct observations of health and productivity on-the-job would be better suited to addressing this issue.

Our study focuses on absenteeism in a significant part of the U.S. economy and one which plays a key role in fostering growth (e.g., Mankiw, Romer, and Weil (1992), Hanushek and Woessman (2008)). However, it is natural to ask how the impact of absenteeism in education might generalize to other settings and which features of the educational process may or may not be shared by other industries. First, labor substitution may be more difficult in occupations like teaching that require skilled workers and involve personal relationships with clients (e.g., healthcare practitioners, social workers, marketing and sales managers, etc.). Second, employees may be more likely to be ill (and unproductive) while on the job when paid sick leave is not available, and rates of paid sick leave are somewhat higher for public school employees (90 percent) than for employees in private firms (70 percent).[35] Third, the production schedule in education is somewhat inflexible (e.g., classes are not be rescheduled) and losses from absenteeism cannot be addressed through overtime work (Ehrenberg (1970)) or flexible hours.

[35]One problem with these statistics, taken from the Bureau of Labor Statistics Employee Benefits Survey, is that the presence of paid sick leave may not accurately reflect the financial incentives for work attendance. As we note above, teachers in New York City get paid when they are absent, but face financial costs because they are paid for unused absences upon retirement.

What can be done to limit the production losses from worker absenteeism? One possibility is to address the root cause of absences, such as negative shocks to worker health. Indeed, absence prevention is one of the main drivers of recent growth in employer sponsored "health promotion" programs (Linnan et al. (2008)), though the evidence on the impact of these programs on absenteeism is quite mixed (Aldana and Pronk (2001)).

Alternatively, governments and firms could offer stronger incentives for workers to show up. Empirical evidence strongly suggests that financial incentives affect worker absence (e.g., Winkler (1980), Jacobson (1989), Ehrenberg et al. (1991), Barmby et al. (1991), Brown and Sessions (1996), and Lindeboom and Kerkhofs (2000)). However, only one study, a field experiment in rural India (Duflo and Hanna (2005)), presents clear evidence that incentives for workers to show up can raise productivity. Financial incentives for work attendance could, in principle, decrease productivity by inducing workers to show up while seriously ill. Though it is reasonable to think that workers would be less responsive to financial incentives when in poor health, this is ultimately an empirical question.

Table 3.1: Summary Statistics for Spells of Teacher Absence

	Avg. Days Missed per Teacher-Year	Average Spell Duration	Teacher-Year Observations with 1+Spells (%)	Total Spell Frequency
Total of All Types	9.98	1.56	96.9	622,843
Self-Treated Sickness	4.23	1.12	90.6	370,207
Medically Certified Sickness	2.02	2.39	41.2	82,482
Conference/School Activities	1.12	1.30	32.8	84,331
Medical Leave	0.54	42.78	1.3	1,241
Personal Days	0.47	1.32	26.3	34,476
Funeral/Death in Family	0.33	2.24	12.9	14,212
Jury Duty/Military Service	0.26	2.07	9.8	12,334
Maternity Leave	0.25	48.63	0.5	506
Child Care Leave	0.13	47.95	0.3	274
Injury	0.11	3.82	2.4	2,910
Resignation or Retirement	0.11	42.56	0.3	249
Religious Holiday	0.10	1.17	8.0	8,226
Graduation	0.10	1.36	4.0	7,217
Other Leave	0.06	44.82	0.1	134
Legislative Hearing	0.04	1.32	1.4	2,595
Grace Period	0.02	11.79	0.2	161
Personal Leave	0.02	25.56	0.1	75
Sick Family Member Leave	0.02	31.38	0.1	77
Death	0.01	37.67	0.0	15
Termination, Certification	0.01	34.15	0.0	33
Involuntary Termination	0.01	40.75	0.0	12
Unauthorized	0.01	1.55	0.4	741
Late More than Half Day	0.00	1.01	0.3	335

Note: Based on teachers in New York City teaching math and/or English to students in grades 4-8 during the school years 1999-2000 to 2008-2009. Additional information on sample restrictions is provided in the text.

Table 3.2: Between and Within Variation in Teacher Absence

	Math Teachers	English Teachers
Total Absences		
Mean	9.81	10.09
Standard Deviation	9.76	9.94
Within-Teacher S.D.	6.35	6.49
Absences Prior to Exam		
Mean	6.50	5.07
Standard Deviation	7.12	6.34
Within-Teacher S.D.	4.68	4.10
Absences After Exam		
Mean	2.78	4.58
Standard Deviation	4.91	6.76
Within-Teacher S.D.	3.39	4.62
Absences During Exam Period		
Mean	0.05	0.05
Standard Deviation	0.31	0.28
Within-Teacher S.D.	0.22	0.21
Absences During Make-up Exam Period		
Mean	0.47	0.39
Standard Deviation	1.00	0.90
Within-Teacher S.D.	0.71	0.65

Note: Based on teachers in New York City teaching math and/or English to students in grades 4-8 during the school years 1999-2000 to 2008-2009. The unit of observation is a teacher-grade-year; we create separate observations for teachers of multiple grades in the same year because exam dates differ across grades. Additional information on sample restrictions is provided in the text.

Table 3.3: Absence from Work and Students' Characteristics

	Work Days Missed
Average Prior Math Test Score	0.9819+
	(-1.9548)
Percent English Language Learner	0.9710
	(-1.3235)
Percent Receiving Free Lunch	0.9625
	(-1.4126)
Percent Special Education	0.8314
	(-1.2942)
Percent Hispanic	0.9304*
	(-2.2958)
Percent Black	0.9676
	(-1.0632)
Percent Asian	1.0163
	(0.3251)
Average Student Days Absent	1.0043*
	(3.6761)

Note: This table presents coefficients from negative binomial regressions, transformed into odds ratios. Lines separate the results of different regressions within each column. All regressions have 97,540 teacher-year observations. Robust t-statistics are shown in parentheses.
+ significant at 10% * significant at 5%

Table 3.4: Absence from Work and Teachers' Characteristics

	Work Days Missed
Master's Degree	0.9774*
	(-2.9671)
Experience (Relative to Teachers with 7+ Years)	
No Experience	0.7251*
	(-20.7370)
1 Year of Experience	0.8769*
	(-9.0892)
2 Years of Experience	0.9275*
	(-5.2895)
3 Years of Experience	0.9686*
	(-2.2772)
4 Years of Experience	0.9952
	(-0.3369)
5 Years of Experience	1.0018
	(0.1201)
6 Years of Experience	1.0015
	(0.1032)
Males' Age (Relative to Younger than 30)	
Between 30 and 44 Years Old	0.9948
	(-0.2768)
Between 45 and 54 Years Old	0.9626
	(-1.4610)
Over 55 Years Old	1.0326
	(1.0558)
Female	1.1179*
	(6.5950)
Females' Age (Relative to Younger than 30)	
Female Between 30 and 44 Years Old	1.1330*
	(6.2312)
Female Between 45 and 54 Years Old	0.9511+
	(-1.9367)
Female Over 55 Years Old	0.9332*
	(-2.2193)
Ethnicity (Relative to White)	
Asian	0.9358*
	(-2.8316)
Black	0.9624*
	(-3.1145)
Hispanic	1.0085
	(0.6439)

Note: This table presents coefficients from negative binomial regressions, transformed into odds ratios. Lines separate the results of different regressions within each column. All regressions have 97,540 teacher-year observations. Robust t-statistics are shown in parentheses.
+ significant at 10% * significant at 5%

Table 3.5: Workday Absences and Productivity, Baseline Estimates and Placebo Test on Prior Year Score

	Math Exam				English Exam			
	Current Year			Prior Year	Current Year			Prior Year
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total Absences	-0.169*				-0.063*			
	(0.009)				(0.008)			
Absences Prior to Exam		-0.201*	-0.120*	-0.021		-0.084*	-0.061*	-0.014
		(0.012)	(0.013)	(0.017)		(0.012)	(0.015)	(0.022)
Absences After Exam		-0.035+	-0.013	-0.007		-0.018+	-0.008	-0.011
		(0.018)	(0.021)	(0.025)		(0.011)	(0.013)	(0.018)
Teacher-School-Grade Fixed Effects			Yes	Yes			Yes	Yes
Dropped Controls for Prior Scores				Yes				Yes
R-squared	0.664	0.664	0.702	0.454	0.611	0.611	0.636	0.428
Observations	2,471,668	2,471,668	2,471,668	2,471,668	2,363,619	2,363,619	2,363,619	2,363,619

Note: Coefficients are expressed in percentage points of a standard deviation. All specifications control for student characteristics, teacher experience, school-grade characteristics, classroom characteristics, and grade-year fixed effects. Specifications separating absences prior to and after exams also control for absences during the exam and make-up exam period. Specifications without teacher-school-grade fixed effects also control for time-invariant teacher characteristics. For more information, see the text. Standard errors (in parentheses) are clustered by school.
+ significant at 10% * significant at 5%

Table 3.6: Absences of "Other Subject" Teachers in Middle School

	Math Exam		English Exam	
	(1)	(2)	(3)	(4)
Math Teacher's Absences Prior to Exam	-0.190*	-0.119*	-0.001	0.001
	(0.019)	(0.024)	(0.016)	(0.018)
English Teacher's Absences Prior to Exam	-0.031*	-0.007	-0.050*	-0.031
	(0.012)	(0.011)	(0.017)	(0.025)
Math Teacher-School-Grade Fixed Effects	No	Yes	No	No
English Teacher-School-Grade Fixed Effects	No	No	No	Yes
R-squared	0.692	0.717	0.625	0.642
Number of Observations	1,199,002	1,199,002	1,095,078	1,095,078

Note: Coefficients are expressed in percentage points of a standard deviation. All specifications are limited to students with different teachers for math and English. Regressions include controls for student characteristics, teacher experience, school-grade characteristics, classroom characteristics, grade-year fixed effects, and absences during the exam and make-up exam periods and after the exam. Specifications without teacher-school-grade fixed effects also control for time-invariant teacher characteristics. For more information, see the text. Standard errors (in parentheses) are clustered by school.
+ significant at 10% * significant at 5%

Table 3.7: Effects of Absence and Work Experience

	Math Exam		English Exam	
	(1)	(2)	(3)	(4)
Number of Absences	-0.120*	-0.131*	-0.061*	-0.070*
Prior to Exam	(0.013)	(0.015)	(0.015)	(0.016)
Teacher w/ < 3 Years ExperienceX		0.072*		0.048
Number of Absences Prior to Exam		(0.033)		(0.035)
R-squared	0.702	0.702	0.636	0.636
Number of Observations	2,471,668	2,471,668	2,363,619	2,363,619

Note: Coefficients are expressed in percentage points of a standard deviation. All specifications control for student characteristics, teacher experience, school-grade characteristics, classroom characteristics, grade-year fixed effects, teacher-school-grade fixed effects, and absences during the exam and make-up exam period. For more information, see the text. Standard errors (in parentheses) are clustered by school.
+significant at 10% *significant at 5%

Table 3.8: Absence Duration (in Workdays) and Productivity Loss

	Math Exam (1)	English Exam (2)
Absences Prior to Exam, 1 Day Spells	-0.356*	-0.173*
	(0.045)	(0.054)
Absences Prior to Exam, 2-3 Day Spells	-0.290*	-0.038
	(0.049)	(0.052)
Absences Prior to Exam, 4-5 Day Spells	-0.222*	-0.022
	(0.058)	(0.067)
Absences Prior to Exam, 6-10 Day Spells	-0.171*	-0.010
	(0.053)	(0.060)
Absences Prior to Exam, 11-30 Day Spells	-0.075*	-0.076*
	(0.030)	(0.037)
Absences Prior to Exam, 31+ Day Spells	-0.084*	-0.058*
	(0.017)	(0.018)
R-squared	0.702	0.636
Observations	2,471,668	2,363,619

Note: Coefficients are expressed in percentage points of a standard deviation. All specifications control for student characteristics, teacher experience, school-grade characteristics, classroom characteristics, grade-year fixed effects, teacher-school-grade fixed effects, and teacher absences during the exam and make-up exam period. For more information, see the text. Absence spells are categorized by the number of consecutive workdays missed (i.e., weekends, holidays, etc. are not counted). Standard errors (in parentheses) are clustered by school. + significant at 10% * significant at 5%

Table 3.9: Health vs. Non-Health Related Absences

	Math Exam (1)	Math Exam (2)	English Exam (3)	English Exam (4)
Health Related Absences Prior to Exam	-0.089*		-0.057*	
	(0.015)		(0.017)	
Non-Health Related Absences Prior to Exam	-0.190*		-0.069*	
	(0.024)		(0.025)	
Absences Prior to Exam in 1 Day Spells				
Health Related		-0.442*		-0.247+
		(0.119)		(0.142)
Non-Health Related		-0.346*		-0.167*
		(0.047)		(0.056)
Absences Prior to Exam in 2-3 Day Spells				
Health Related		-0.247*		-0.097
		(0.077)		(0.098)
Non-Health Related		-0.309*		-0.007
		(0.057)		(0.061)
Absences Prior to Exam in 4-5 Day Spells				
Health Related		-0.289*		-0.018
		(0.078)		(0.099)
Non-Health Related		-0.151+		-0.020
		(0.078)		(0.097)
Absences Prior to Exam in 6-10 Day Spells				
Health Related		-0.161*		0.039
		(0.069)		(0.082)
Non-Health Related		-0.178*		-0.072
		(0.083)		(0.090)
Absences Prior to Exam in 11-30 Day Spells				
Health Related		-0.058+		-0.069+
		(0.033)		(0.040)
Non-Health Related		-0.128+		-0.102
		(0.069)		(0.095)
Absences Prior to Exam in 31+ Day Spells				
Health Related		-0.076*		-0.059*
		(0.018)		(0.022)
Non-Health Related		-0.111*		-0.055+
		(0.037)		(0.033)
R-squared	0.702	0.702	0.636	0.636
Observations	2,471,668	2,471,668	2,363,619	2,363,619

Note: Coefficients are expressed in percentage points of a standard deviation. All specifications control for student characteristics, teacher experience, school-grade characteristics, classroom characteristics, grade-year fixed effects, teacher-school-grade fixed effects, and teacher absences during the exam and make-up exam period. For more information, see the text. Absence spells are categorized by the number of consecutive workdays missed (i.e., weekends, holidays, etc. are not counted). Standard errors (in parentheses) are clustered by school.
+ significant at 10% * significant at 5%

Table 3.10: Absences and the Timing of Student Exams

	Math Exam		English Exam	
	(1)	(2)	(3)	(4)
Absences Prior to Exam	-0.120*		-0.061*	
	(0.013)		(0.015)	
Absences 21+ Workdays Prior to Exam		-0.096*		-0.030+
		(0.015)		(0.017)
Absences 6-20 Workdays Prior to Exam		-0.185*		-0.208*
		(0.058)		(0.061)
Absences 1-5 Workdays Prior to Exam		-0.850*		-0.398*
		(0.144)		(0.139)
Absences During Exam Period	-2.440*	-1.653*	-1.283*	-0.853*
	(0.327)	(0.351)	(0.311)	(0.329)
Absences During Make-up Exam Period	-0.359*	-0.274*	0.030	0.097
	(0.101)	(0.103)	(0.099)	(0.100)
Absences After Exam	-0.013	-0.001	-0.008	-0.003
	(0.021)	(0.021)	(0.013)	(0.013)
R-squared	0.702	0.702	0.636	0.636
Observations	2,471,668	2,471,668	2,363,619	2363619

Note: Coefficients are expressed in percentage points of a standard deviation. All specifications control for student characteristics, classroom characteristics, school-grade characteristics, teacher experience, grade-year fixed effects, and teacher-school-grade fixed effects. Standard errors (in parentheses) are clustered by school. + significant at 10% * significant at 5%

Table 3.11: Persistence in the Effects of Workday Absences

	Math Exam					English Exam		
	Year t	Year t+1	Year t	Year t+1	Year t	Year t+1	Year t	Year t+1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Absences Prior to Exam (Year t)	-0.106* (0.014)	-0.034* (0.015)				-0.062* (0.016)	-0.027 (0.017)	
Absences 21+ Workdays Prior to Exam			-0.084* (0.015)	-0.029+ (0.017)			-0.037* (0.018)	-0.024 (0.019)
Absences 6-20 Workdays Prior to Exam			-0.159* (0.065)	-0.075 (0.066)			-0.177* (0.068)	-0.040 (0.063)
Absences 1-5 Workdays Prior to Exam			-0.805* (0.161)	-0.046 (0.161)			-0.350* (0.160)	-0.116 (0.173)
Number of Absences During Exam	-2.895* (0.371)	-0.319 (0.342)	-2.153* (0.395)	-0.253 (0.374)	-1.134* (0.352)	0.540 (0.373)	-0.782* (0.378)	0.613 (0.394)
Number of Absences During Make-Up Period	-0.317* (0.112)	-0.068 (0.113)	-0.222+ (0.114)	-0.058 (0.115)	0.031 (0.119)	0.022 (0.120)	0.091 (0.120)	0.035 (0.120)
Number of Absences After Exam (Year t)	-0.009 (0.023)	-0.011 (0.023)	0.004 (0.023)	-0.010 (0.023)	0.005 (0.016)	-0.024 (0.017)	0.010 (0.017)	-0.023 (0.017)
R-squared	0.699	0.653	0.699	0.653	0.644	0.594	0.644	0.596
Number of Observations	1,713,561	1,713,561	1,713,561	1,713,561	1,625,038	1,625,038	1,625,038	1,625,038

Note: Coefficients are expressed in percentage points of a standard deviation. Specifications are limited to students who are not in the 8th grade and who have valid test scores in the following year. All specifications control for student characteristics, teacher experience, school-grade characteristics, classroom characteristics, grade-year fixed effects, and teacher-school-grade fixed effects. For more information, see the text. Standard errors (in parentheses) are clustered by school.
+ significant at 10% * significant at 5%

Bibliography

[1] Aaronson, D., Barrow, L. and Sander, W. (2007). "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 25(1): 95-135.

[2] Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J. and Zilibotti, F. (2007). "Technology, Information, and the Decentralization of the Firm" *Quarterly Journal of Economics*, 122(4):1759-1799.

[3] Aizer, A. (2008). "Peer Effects and Human Capital Accumulation: the Externalities of ADD." *NBER Working Paper 14354*. http://www.nber.org/papers/w14354

[4] Aldana, S.G. and Pronk, N.P. (2001). "Health Promotion Programs, Modifiable Health Risks, and Employee Absenteeism." *Journal of Occupational and Environmental Medicine*, 43(1): 36-46.

[5] Angrist, J. and Lang, K. (2004). "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review*, 94(5): 1613-1643.

[6] Barmby, T.A., Orme, C.D., and Treble, J.G. (1991). "Worker Absenteeism: An Analysis Using Microdata." *The Economic Journal*, 101(405): 214-229.

[7] Barmby, T.A., Ercolani, M.G., and Treble, J.G. (2002). "Sickness Absence: An International Comparison." *The Economic Journal*, 112(480): 315-331.

[8] Bell, W. (1954). "A Probability Model for the Measurement of Ecological Segregation." *Social Forces*, 32: 357-64.

[9] Bergendorff, S. (2003). "Sickness Absence in Europe: A Comparative Study." International Social Security Association Working Paper.

[10] Betts, J.R. and Shkolnik, J.L. (1999). "Key Difficulties in Identifying the Effects of Ability Grouping on Student Achievement." *Economics of Education Review*, 19(1):21-26.

[11] Bezanson, A., Chalufour, F., Willits, J.H. and White, L.F. (1922). "Attendance in Four Textile Mills in Philadelphia." *Annals of the American Academy of Political and Social Science*, 104: 187-222.

[12] Borman, G.D., Slavin, R. E., Cheung, A.C.K., Chamberlain, A.M., Madden, N.A., and Chambers, B. (2007). "Final Reading Outcomes of the National Randomized Field Trial of Success for All" *American Educational Research Journal*, 44(3): 701-31.

[13] Board of Education of the City of New York. "Special Education Services As Part of a Unified Service Delivery System: The Continuum of Services for Students with Disabilities." Retrieved on May 27, 2009 from http://schools.nyc.gov/NR/rdonlyres/2BCCCF14-9EAE-4506-BD3E-42E9789BCE99/ContinuumofServices.pdf

[14] Brown, S. and Sessions, J.G. (1996). "The Economics of Absence: Theory and Evidence." *Journal of Economic Surveys*,10(1):23-53.

[15] Bureau of Labor Statistics, U.S. Department of Labor (2008). Current Population Survey, Annual Averages, Household Data, Table 46, http://www.bls.gov/cps/tables.htm.

[16] Bureau of Labor Statistics, U.S. Department of Labor (2009). Major Work Stoppages (Annual) News Release, Table 1, http://www.bls.gov/news.release/wkstp.htm.

[17] Carrell, S.E. and Hoekstra, M.L. (2010). "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics*, 2(1): 211-228.

[18] Carrell, S.E., Sacerdote, B.I., and West, J.E. (2012). "From Natural Variation to Optimal Policy? An Unsuccessful Experiment in Using Peer Effects Estimates to Improve Student Outcomes." Unpublished Manuscript.

[19] Caroli, E. and Van Reenen, J. (2001). "Skill Biased Organizational Change," *Quarterly Journal of Economics*, 116: 1448-1492.

[20] Chambers, J.G., Parrish, T.B., and Harr, J.J. (2004). "What Are We Spending on Special Education Services in the United States, 1999-2000?" Special Education Expenditure Project, Center for Special Education Finance, U.S. Department of Education, Office of Special Education Programs. http://csef.air.org/publications/seep/national/AdvRpt1.PDF

[21] Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. and Rogers, F.H. (2006). "Missing in Action: Teacher and Health Worker Absence in Developing Countries" . *Journal of Economic Perspectives*, 20(1): 91-116.

[22] Chay, K.Y., McEwan, P.J., and Urquiola, M. (2005). "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools" *American Economic Review*, 95(4): 1237-1258.

[23] Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., and Yagan, D. (2010). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." Working Paper no. 16381, National Bureau of Economic Research.

[24] "Closing the Door on Innovation: Why One National Curriculum is Bad for America" (2011, May 6). A Critical Response to the Shanker Institute Manifesto and the U.S. Department of Education's Initiative to Develop a National Curriculum and National Assessments Based on National Standards. Last Accessed: Sept 7, 2011 http://www.k12innovation.com/Manifesto/_V2_Home.html

[25] Clotfelter, C.T., Ladd, H.F., and Vigdor, J.L. (2009). "Are Teacher Absences Worth Worrying About in the U.S.?" *Education Finance and Policy*, 4(2): 115-149.

[26] Coeyman, M. (2002) "Special ed: Take 2." Christian Science Monitor, March 26.

[27] Cohen, J.L. (2008). "Causes and Consequences of Special Education Placement: Evidence from Chicago Public Schools." Essays on the economics of education and health, Doctoral dissertation, Massachusetts Institute of Technology,Retrieved from http://dspace.mit.edu/bitstream/handle/1721.1/42394/236206206.pdf?sequence=1 on May 27, 2009.

[28] Cortiella, C. (2010). "IDEA 2004 Close Up: Evaluation and Eligibility for Specific Learning Disabilities." Retrieved on March 3, 2010 from http://www.greatschools.org/LD/school-learning/evaluation-and-eligibility-for-specific-learning-disabilities.gs?content=943

[29] Crockett, J.B. and Kauffman, J.M. (1999). The Least Restrictive Environment: Its Origins and Interpretations in Special Education. Mahwah, NJ: Erlbaum.

[30] Cullen, J.B. (2003) "The impact of fiscal incentives on student disability rates." *Journal of Public Economics*, 87: 1557-89.

[31] Cullen, J.B. and Reback, R. (2006) "Tinkering Towards Accolades: School Gaming under a Performance Accountability System." Chapter 1, *Advances in Applied Microeconomics*, vol. 14, edited by Timothy J. Gronberg and Dennis W. Jansen, Elsevier.

[32] Culter, D.M., Glaeser, E.L., and Vigdor, J.L. (1999). "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, 107(3): 455-506.

[33] Currie, J. and Thomas, D. (1995). "Does Head Start Make a Difference?" *American Economic Review*, 85(3): 341-364.

[34] Currie, J. and Madrian, B.C. (1999). "Health, Health Insurance and the Labor Market." in: O. Ashenfelter & D. Card (ed.), *Handbook of Labor Economics*, pp. 3309-3416.

[35] Currie, J. (2009). "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood and Human Capital Development." *Journal of Economic Literature*, 47(1): 87-122.

[36] Dessein, W. (2002). "Authority and Communication in Organizations," *Review of Economic Studies,* 69:811-838.

[37] Department of Education (1999). "34 CFR Parts 300 and 303: Assistance to the States for the Education of Children With Disabilities and the Early Intervention Program for Infants and Toddlers with Disabilities; Final Regulations, Part II." Federal Register. 64(48). http://www.ed.gov/legislation/FedRegister/finrule/1999-1/031299a.pdf.

[38] Duff, A.B. (1999) "The Special Costs of Special Ed." Investor's Business Daily, March 26.

[39] Duflo, E., Dupas, P., and Kremer, M. (2010). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." Unpublished manuscript.

[40] Duflo, E. and Hanna, R. (2005). "Monitoring Works: Getting Teachers to Come to School." Working Paper no. 11880, National Bureau of Economic Research.

[41] Duflo, E., and Saez, E. (2002). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics*, 118(3): 815-842.

[42] Ehrenberg, R.G. (1970). "Absenteeism and the Overtime Decision." *American Economic Review,* 60(3): 352-357.

[43] Ehrenberg, R.G., Ehrenberg, R.A., Rees, D.I., and Ehrenberg, E.L. (1991). "School District Leave Policies, Teacher Absenteeism, and Student Achievement." *The Journal of Human Resources.* 26(1): 72-105.

[44] Fernandez, J. (2009). "At a Harlem Charter, the Cost of Getting Ahead Means School on a Snow Day." *New York Times*, March 3.

[45] Figlio, D.N. (2007) "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy*, 2(4): 376-394.

[46] Figlio, D.N. and Getzler, L.S. (2002). "Accountability, Ability, and Disability: Gaming the System." *NBER Working Paper* 9307.

[47] Figlio, D.N. and Page, M. (2002). "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?" *Journal of Urban Economics*, 51: 497-514.

[48] Fletcher, J. (2009). "The Effects of Inclusion on Classmates of Students with Disabilities: The Case of Serious Emotional Problems." *Education Finance and Policy*, 4(3): 278-299.

[49] Friesen, J., Hickey, R., and Krauth, B. (2008). "Disabled Peers and Academic Achievement." Simon Frasier University Working Paper. client.norc.org/jole/SOLEweb/865.pdf

[50] Friesen, J., Hickey, R., and Krauth, B. (2010). "Disabled Peers and Academic Achievement." *Education Finance and Policy*, 5(3): 317-348.

[51] Fuchs, D. and Fuchs, L.S. (1986). "Test Procedure Bias: A Meta-Analysis of Examiner Familiarity Effects." *Review of Educational Research*, 56(2):243-262.

[52] Gabriel, T. (2010). "Under Pressure, Teachers Tamper with Tests." *New York Times*, June 10.

[53] Garces, E., Thomas, D., and Currie, J. (2002). "Longer-Term Effects of Head Start." *American Economic Review.* 92(4): 999-1012.

[54] Gershenson, S. (2011). "How do Substitute Teachers Substitute? An Empirical Analysis of Substitute Teacher Labor Supply." Unpublished Manuscript, Michigan State University.

[55] Glewwe, P., Kremer, M., and Moulin, S. (2009). "Many Children Left Behind? Textbooks and Test Scores in Kenya" *American Economic Journal: Applied Economics*, 1:1: 112-135.

[56] Gibbons, S. and Telhaj, S. (2011). "Pupil Mobility and School Disruption" *Journal of Public Economics*, 95: 1156-1167.

[57] Goetzel, R.Z., Long, S.R., Ozminkowski, R.J., Hawkins, K., Wang, S., and Lynch,W. (2004). "Health, Absence, Disability, and Presenteeism Cost Estimates of Certain Physical and Mental Health Conditions Affecting U.S. Employers." *Journal of Occupational and Environmental Medicine*, 46(4): 398-412.

[58] Hanushek, E.A. (1971). "Teacher Characteristics and Gains in Student Achievement: Estimation using Micro Data." *American Economic Review*, 61(2): 280-288.

[59] Hanushek, E.A., Kain, J.F., and Rivkin, S.G. (2002). "Inferring program effects for special populations: Does Special Education raise achievement for students with disabilities?" *Review of Economics and Statistics*, 84: 589-599.

[60] Hanushek, E.A., Kain, J.F., and Rivkin,S.G. (2003). "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18(5):527-544.

[61] Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2004). "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economics*, 88:1721–1746.

[62] Hanushek, E.A., and Rivkin, S.G. (2006). "Teacher Quality." *Handbook of the Economics of Education* ed. E.A. Hanushek and F. Welch, p. 865-908.

[63] Hanushek, E.A. and Woessmann, L. (2008). "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature*, 46(3):607-668.

[64] Hanushek, E.A. and Rivkin, S.G. (2010). "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review*, 100(2):267-271.

[65] Henderson, E., Protheroe, N. and Porch, S. (2002). *Developing an Effective Substitute Teacher Program.* Arlington,VA: Educational Research Service.

[66] Herrmann, M.A. and Rockoff, J.E. (2010). "Worker Absence and Productivity: Evidence From Teaching" *NBER Working Paper* No. w16524.

[67] Hoxby, C.M. (2000). "Peer Effects in the Classroom: Learning from Gender and Race Variation." *NBER Working Paper* 7867. http://www.nber.org/papers/w7867

[68] Hoxby, C.M. and Weingarth, G. (2005). "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Working Paper, Harvard University.

[69] Ichino, A. and Moretti, E. (2009). "Biological Gender Differences, Absenteeism and the Earnings Gap." *American Economic Journal: Applied Economics*, 1(1):183-218.

[70] Individuals with Disabilities Education Improvement Act of 2004 (2004). Pub. L. No. 108-446, §302, 118 Stat. 2803.

[71] Imbens, G. and Lemieux, T. (2008). "Regression Discontinuity Designs: A Guide to Practice" *Journal of Econometrics*, 142(2): 615-35.

[72] Imberman, S., Kugler, A., and Sacerdote, B. (2009). "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." NBER Working Paper No. 15291. http://www.nber.org/papers/w15291

[73] Jacob, B.A. and Levitt, S.D. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics*, 118(3): 843-877.

[74] Jacob, B.A. and Lefgren, L. (2004). "Remedial Education and Student Achievement: A Regression Discontinuity Analysis" *Review of Economics and Statistics*, 86(1):s 226-244.

[75] Jacob, B. (2005). "Accountability, incentives, and behavior: Evidence from school reform in Chicago." *Journal of Public Economics*, 89: 761-769.

[76] Jacob, B.A., Lefgren, L., and Sims, D. (2008). "The Persistence of Teacher-Induced Learning Gains." Working Paper no. 14065, National Bureau of Economic Research.

[77] Jacobson, S.L. (1989). "The Effects of Pay Incentives on Teacher Absenteeism." *The Journal of Human Resources*, 24(2): 280-286.

[78] Janod, V. and Saint-Martin, A. (2004). "Measuring Work Reorganization and Its Impact on Firm Performance: An Estimate on French Manufacturing Firms over 1995-1999." *Labour Economics*, 11: 785-798.

[79] Kane, T.J., Rockoff, J.E., and Staiger, D.O. (2008). "What Does Certification Tell Us About Teacher Effectiveness? Evidence From New York City." *Economics of Education Review*, 27(6): 615-631.

[80] Kane, T.J. and Staiger,D.O. (2002). "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 16(4): 91-114

[81] Kane, T.J., and Staiger, D.O. (2008). "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." Working Paper 14607, National Bureau of Economic Research.

[82] Kleiner, M.M., Leonard, J.S. and Pilarski, A.M. (2002). "How Industrial Relations Affects Plant Performance: The Case of Commercial Aircraft Manufacturing." *Industrial and Labor Relations Review*, 55(2): 195-218.

[83] Kling, J.R., Liebman, J.B., and Katz, L.F. (2007). "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83-119.

[84] Kremer, M., and Chen, D. (2001). "An Interim Report on a Teacher Attendance Incentive Program in Kenya." Unpublished Manuscript, Harvard University.

[85] Krueger, A.B. and Mas, A. (2004). "Strikes, Scabs and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires." *Journal of Political Economy*, 112(2): 253-289.

[86] Lang, K. (2010). "Measure Matters: Perspectives on Education Policy from an Economist and School Board Member." *Journal of Economic Perspectives*, 24(3):167-182.

[87] Lavy, V. (2010). "Do Differences in School's Instructional Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries" *NBER Working Paper* No. w16227.

[88] Lavy, V., Passerman, D., and Schlosser, A. (2008). "Inside the Black Box of Ability Peer Effects." NBER Working Paper No. 14415. http://www.nber.org/papers/w14415

[89] Lavy, V., Silva, O., and Weinhardt, F. (2009). "The Good, the Bad, and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools." NBER Working Paper No. 15600. http://www.nber.org/papers/w15600

[90] Lazear, E. (2001). "Education Production." *Quarterly Journal of Economics*, 116(3): 777-803.

[91] Lee, D.S. and Lemieux, T. (2010). "Regression Discontinuity Designs in Economics" *Journal of Economic Literature*, 48(2): 281-355.

[92] Lefgren, Lars (2004). "Educational Peer Effects and the Chicago public schools." *Journal of Urban Economics*, 56:169-191.

[93] Lehrer, S. and Ding, W. (2007). "Do Peers Affect Student Achievement in China's Secondary Schools?" *Review of Economics and Statistics*, 89:300-312.

[94] Levitt, S.D., List, J.A., and Sadoff, S. (2011). "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment." Unpublished Manuscript, University of Chicago.

[95] Lindeboom, M. and Kerkhofs, M. (2000). "Multistate Models for Clustered Duration Data-An Application to Workplace Effects on Individual Sickness Absenteeism." *The Review of Economics and Statistics*, 82(4): 668-684.

[96] Linnan, L., Bowling, M., Childress, J., Lindsay, G., Blakey, C., Pronk, S., Wieker, S., and Royall, P. (2008). "Results of the 2004 National Worksite Health Promotion Survey." *American Journal of Public Health*. 98(8):1503-1509.

[97] Loeb, S. and McEwan, P.J. (2010). "Educational Reforms" in *Targeting Investments in Children: Fighting Poverty When Resources Are Limited*. Edited by. P.B. Levine and D. J. Zimmerman. National Bureau of Economic Research.

[98] Lyle, D.S. (2006). "Using Military Deployments and Job Assignments to Estimate the Effect of Parental Absences and Household Reallocations on Children's Academic Achievement" *Journal of Labor Economics*, 24(2): 319-350.

[99] Mankiw, N. G., Romer, D., Weil, D.N. (1992). "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics*, 107(2): 407-437.

[100] Manski, C. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60: 531-542.

[101] Marcotte, D.E. (2007). "Schooling and Test Scores: A Mother-natural Experiment" *Economics of Education Review*, 26(5): 629-640.

[102] Mas, A. (2008). “Labor Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market.” *Review of Economic Studies*, 75(1): 229-258.

[103] Mas, A. and Moretti, E. (2009) “Peers at Work.” *American Economic Review*, 99(1): 112-145.

[104] McCrary, J. (2008). “Manipulation of the Running Variable in the Regression Discontinuity Design. A Density Test” *Journal of Econometrics*, 142: 698-714.

[105] Miller, R.T., Murnane, R.J., and Willet, J.B. (2008). “Do Worker Absences Affect Productivity? The Case of Teachers.” *International Labour Review*, 147(1): 71-89.

[106] Moffitt, R. A. (2001). “Policy Interventions, Low-Level Equilibria, and Social Interactions.” In *Social Dynamics*, Durlauf S, Peyton Young H (eds). MIT Press: Cambridge, MA: 45-82.

[107] Murnane, R.J. (1975). *The Impact of School Resources on the Learning of Inner City Children.* Cambridge, MA: Balinger.

[108] National Mathematics Advisory Panel. (2008). *Foundations for Success: Report of the National Mathematics Advisory Panel.* Washington, D.C.: U.S. Department of Education.

[109] National Reading Panel. (2000).*Teaching Children to Read: an Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction: Reports of the Subgroups.*Washington, D.C.: U.S. Government Printing Office.

[110] National Research Council and Institute of Medicine (2010). “Student Mobility: Exploring the Impact of Frequent Moves on Achievement: Summary of a Workshop.” Committee on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighborhoods. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[111] New York City Department of Education. (2003). *A Comprehensive Approach to Balanced Literacy: A Handbook for Educators.*

[112] Needham, C. (2004). “ Special-needs policy provokes debate on mainstreaming.” *The Providence Journal*, p. C-01

[113] Oreopolous, P. (2003) “The Long-Run Consequences of Living in a Poor Neighborhood.” *The Quarterly Journal of Economics*, 118(4): 1533-1575.

[114] Pauly, M.V., Nicholson, S., Polsky, D., Berger, M.L., and Sharda, C. (2008). "Valuing Reductions in On-The-Job Illness: 'Presenteeism' from Managerial and Economic Perspectives," *Health Economics*, 17(4): 469-485.

[115] Pischke, J. (2007). "The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years" *The Economic Journal* , 117 (523): 1216-1242.

[116] Reynolds, A.J., Chen, C.-C., and Herbers, J.E. (2009). "School Mobility and Educational Success: A Research Synthesis and Evidence on Prevention" Paper prepared for the Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighborhoods, June 29-20, 2009. The National Academies, Washington, D.C. Available: http://www.bocyf.org/children_who_move_reynolds_paper.pdf

[117] Rivkin, S.G., Hanushek, E.A., and Kain, J.F., (2005). "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417-458.

[118] Rockoff, J.E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2): 247-252.

[119] Rockoff, J. and Turner, L.J. (2010). "Short Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy*, 2(4):119-147.

[120] Rothstein, J. (2010). "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics*, 125(1): 175-214

[121] Sacerdote, B.L. (2001). "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116: 681-704.

[122] Schacter, J. (2001a). "Geographic Mobility: March 1999 to March 2000" Current Population Reports. US Census Bureau, Washington, DC.

[123] Schacter, J. (2001b). "Why People Move: Exploring the March 2000 Current Population Series" Current Population Reports. US Census Bureau, Washington, DC.

[124] Smith, J.P. (1999). "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status." *Journal of Economic Perspectives*, 13(2):145-166.

[125] Steele, C.M. and Aronson, J. (1995). "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology*, 69(5): 797-811.

[126] Strand, S. (2002). "Pupil Mobility, Attainment, and Progress During Key Stage 1: A Study in Cautious Interpretation" *British Educational Research Journal*, 28(1): 63-78.

[127] Strauss, J. and Thomas, D. (1998). "Health, Nutrition, and Economic Development." *Journal of Economic Literature*, 36(2): 766-817.

[128] Sutner, S. (1998) "Reform won't come easy." *Telegram and Gazette*, October 6.

[129] Temple, J.A. and Reynolds, A.J. (1999). "School Mobility and Achievement: Longitudinal Findings from An Urban Cohort" *Journal of School Psychology*, 37(4): 355-377.

[130] Thistlethwaite, D.L. and Campbell, D.T. (1960). "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology*, 51(6): 309-17.

[131] Thomas, C. (2011). "Too Many Products: A Study of the Interaction of Multinational Corporations with Heterogeneous Consumers." *American Economic Journal: Microeconomics*, 3(1): 280-306.

[132] *TIMSS 2007 Encyclopedia: A Guide to Mathematics and Science Education Around the World*, Volumes 1 and 2. Edited by Mullis, I.V.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., & Stanco, G.M. (2008). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

[133] Todd, P.E. and Wolpin, K.I. (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal*, 113(1): 3-33.

[134] United States General Accounting Office. (1994). *Elementary School Children: Many Change Schools Frequently, Harming Their Education.* Washington DC: U.S. Government Printing Office.

[135] U.S. Office of Special Education Programs (2007). "History: Twenty-Five Years of Progress in Educating Children with Disabilities Through IDEA." Retrieved from http://www.ed.gov/policy/speced/leg/idea/history.pdf on May 27, 2009

[136] van der Klauww, W. (2002). "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach," *International Economic Review*, 43(4): 1249-1287.

[137] Waite, R.D. (2000). "A Study of the Effects of Everyday Mathematics on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district" *Dissertation Abstracts International*, 61(1), 3933A. (UMI No. 9992659).

[138] Wernikoff, L. (2007). "Memorandum: Children First Reforms in Special Education effective July 1, 2007." Office of Special Education Initatives, New York City Department of Education. September 12. Retrieved from http://schools.nyc.gov/NR/rdonlyres/2B33FBE7-D9EC-4417-88F0-1A6450E01DF9/0/SpecialEducationEvaluationProcessAugust2007.pdf on May 27, 2009

[139] White, M.J. (1986) “Segregation and Diversity Measures in Population Distribution.” *Population Index*, 52: 198-221.

[140] Winkler, D.R. (1980). “The Effects of Sick-Leave Policy on Teacher Absenteeism.” *Industrial and Labor Relations Review*, 33(2): 232-240.

[141] Wolf, A. (2007)). “Let the Chips Begin to Fall.” *The New York Sun*, Jan 26.

[142] Zimmer, R. (2003). “A New Twist in the Educational Tracking Debate.” *Economics of Education Review*, 22: 307-315.

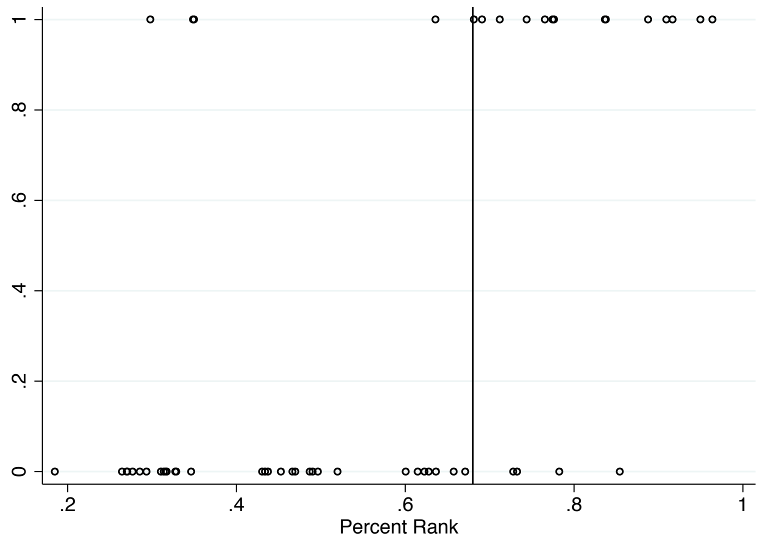
[143] Zimmerman, D.J. (2003). “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment.” *The Review of Economics and Statistics*, 85(1):9-23.

Appendix A

One Size Fits All?

Figure A.1: Top 20 Percent Calculations

First Round Exemptions by Calculated Rank – Schools Within 10 points of Cutoff

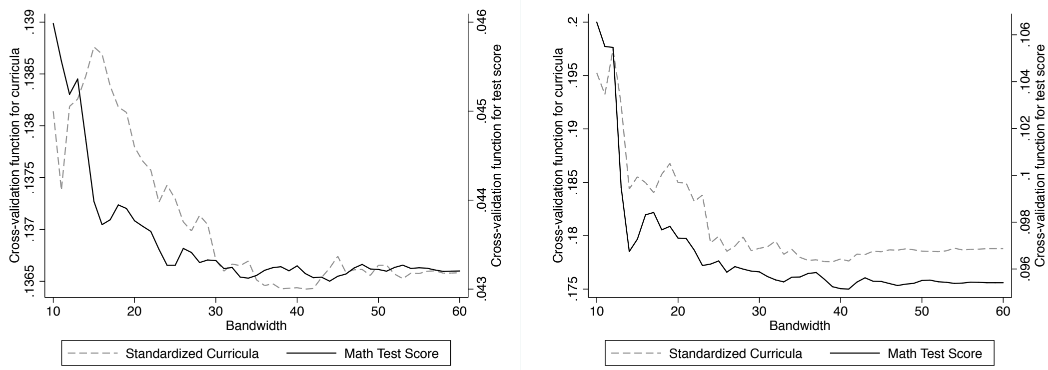


Notes: Schools received first round exemptions from implementing the standardized curriculum if their score - the sum of the percents of students who passed the math and reading exams in 2002 - met or exceeded the cutoff for their need category, or if their score was within 10 points of the cutoff for their need category and they were in the top 20 percent of improvement citywide in the previous year. I calculated schools ranks in improvement as follows: For 4th-8th grade students, I calculated each students gain for each subject as the difference between her raw scaled test scores in that subject in 2002 and 2001. For 3rd grade students, I calculated each students gain for each subject as the difference between her raw scaled test score in that subject in 2002 and the schools average raw scaled test score for 3rd graders in that subject in 2001. I then summed the schools average gain in math and reading and ranked schools by these gains. Note that reading test scores for 7th graders in 2002 are missing due to an issue with the exam. Since I only had test data for 3rd through 8th graders, I could only calculate ranks for elementary and middle schools. Thus, I use data from the first round exemptions to determine the rank cutoff that would place an elementary or middle school among the top 20 percent of all city schools (including high schools). Note that among schools within 10 points of the cutoff for their need category, only those that placed in the top 20 percent citywide should have received an exemption. The figure plots the probability of receiving a first round exemption on schools ranks in improvement (calculated by the author) for only schools with scores within 10 points of the normal cutoff for their need group. I set the break point for the top 20 percent at 0.68, the rank that maximizes the goodness-of-fit from a model of exemption receipt as a function of an indicator equal to one if the schools rank is above a particular threshold, the same procedure used in Chay et al. (2005).

Figure A.2: Cross-Validation Functions

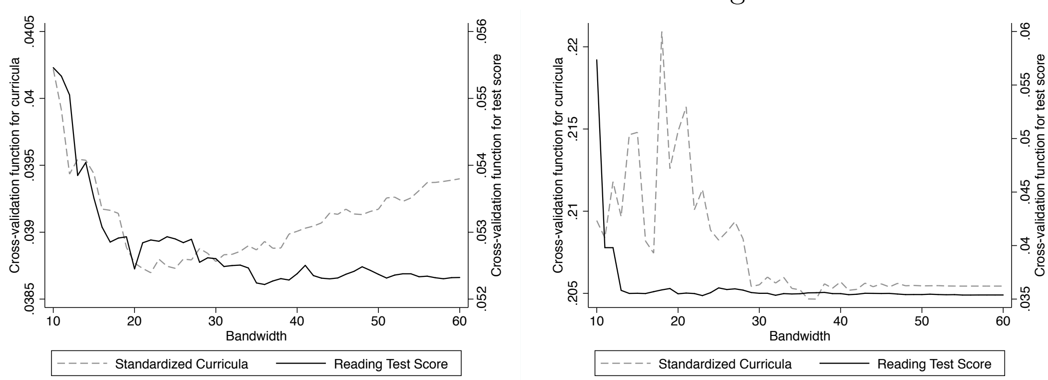
Math

Panel A: Left Side Panel B: Right Side



Reading

Panel C: Left Side Panel D: Right Side

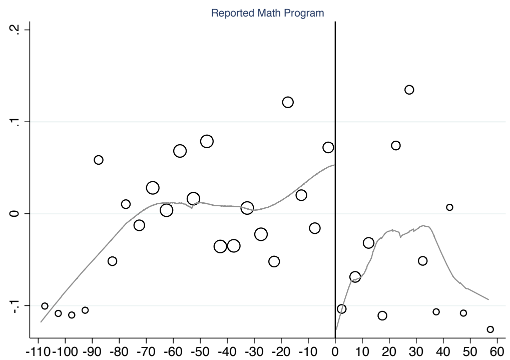


Notes: These panels present cross-validation functions for each outcome and side using the “leave one out” procedure described in Lee and Lemieux (2010). More details can be found in Appendix Table A2.

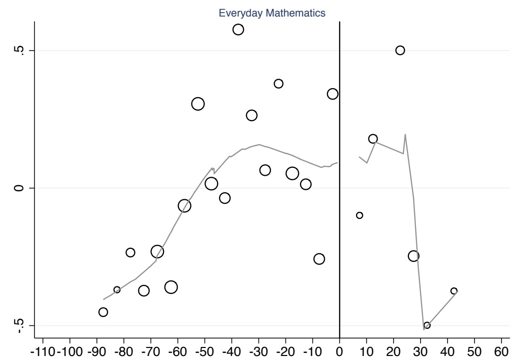
Figure A.3: Previous Math and Reading Programs

Math

Panel A: Reporting Any Math Program

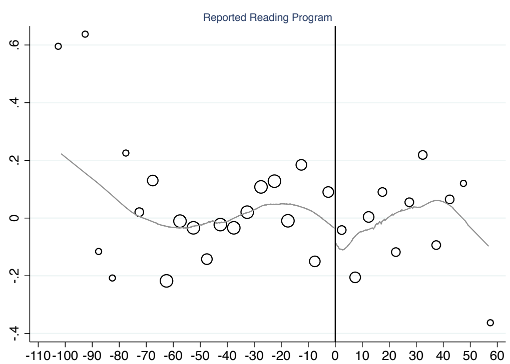


Panel B: Reporting Everyday Mathematics, Conditional on Reporting

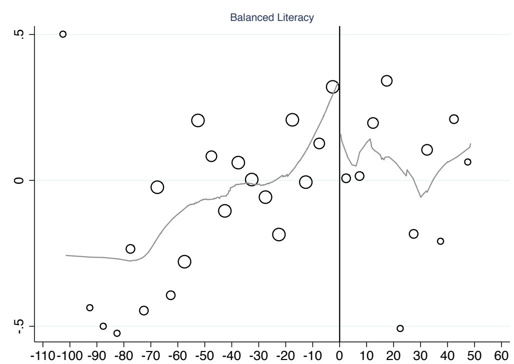


Reading

Panel C: Reporting Any Reading Program



Panel D: Reporting Balanced Literacy, Conditional on Reporting



Notes: Figures plot the residuals from regressions of the dependent variables reporting any program or reporting standardized curricula programs on cutoff fixed effects as well as the fitted values from locally weighted regressions of these residuals on schools rescaled scores. These data come from optional sections of schools 2002 report cards. Report cards were filled out in 2003, after the announcement of the standardized curricula but before they were required to be implemented.

Table A.1: Curricular Materials, Publishers, and Minimum Time Limits

Program	Publisher	Grades
Math		
Everyday Mathematics	SRA/McGraw-Hill	K-5
Math Steps*	Houghton-Mifflin	K-5
Impact Mathematics	Glencoe/McGraw-Hill	6-8
Hot Words, Hot Topics*	Glencoe/McGraw-Hill	6-8
New York Math A: An Integrated Approach	Prentice Hall	9-12
Reading		
Month-by-Month Phonics	Carson-Dellosa	K-3
Voyager Passport*	Voyager Expanded Learning	K-3
Ramp-Up to Literacy*	America's Choice	6,9
Minimum Time (Hours/Day)		Grades
Math		
1 hour/day		K-2
1.25 hours/day		3-8
0.75 hours/day, +0.75 hours/day for struggling students		9-12
Reading		
2 hours/day		K-3
1.5 hours/day		4-8
1.5 hours/day for struggling students		9-12

Note: * Signifies supplementary materials

Table A.2: Cross-validated Bandwidths

	Left Side	Right Side
Standardized Math Curricula	41	39
Math Test Scores	44	41
Standardized Reading Curricula	22	37
Reading Test Scores	36	23

Note: Cross-validated bandwidths are calculated using the "leave one out" procedure described in Lee and Lemieux (2010); for observations to the left (right) of the cutoff, for each observation i, a predicted outcome is obtained from a regression of the outcome using only observations to the left (right) of observation i within the bandwidth on cutoff fixed effects and the distance from the cutoff. The cross-validation function for the bandwidth is defined as the sum of the squared differences between the outcome for observation i and the predicted outcome from the regression using this bandwidth, divided by the number of observations. The cross-validation choice is the bandwidth that minimizes the cross-validation function. Lee and Lemieux advise choosing (i) the minimum cross-validated bandwidths, (ii) a bandwidth that minimizes the weighted average of the cross-validation function (to obtain a symmetric bandwidth), or (iii) a bandwidths that correspond to the outcome. I chose the minimum bandwidth between the standardized curriculum and the outcome and use this on both sides.

Table A.3: Previous Math and Reading Programs in 2003

Math	Percent	Reading	Percent
Listing Math Programs on Report Card	11%	Listing Reading Programs on Report Card	41%
Conditional on Listing, % Listing ...		Conditional on Listing, % Listing ...	
Everyday Math	34%	Balanced Literacy	50%
Math A	23%	Project Read	34%
TERC/Investigations	19%	Reading Recovery	21%
Math In Context	11%	Success for All	6%
Everyday Counts	5%	AUSSIE	5%
Larsen's Math	3%	Breakthrough to Literacy	5%
Scott Foresman	1%	Teacher's College Reading and Writing Project	3%
Math Steps	3%	Voyager	3%
Math Their Way	3%	Making Connections	2%
Comprehensive Instructional Management System	2%	Mondo/Building Essential Literacy	2%
Basic Skills Math	1%	Strategies for Success in Literacy	2%
Math Trailblazers	1%	CUNY Literacy	1%
		Guided Reading	1%

Note: Math and reading programs, including supplementary programs, which were listed on the 2002 report cards of schools in the analysis. Reporting of programs was optional, and percentages may add up to more than 100 since schools could list multiple programs. These data come from the following sections of the school report cards: Special Academic Programs, Extracurricular Activities, Community Support, and Parent/School Support. The 2002 report cards were completed in 2003 after the announcement of the standardized curricula but before they were required to be implemented.

Appendix B

Worker Absence and Productivity

Table B.1: New York City Math and English Testing Dates, 2000-2009

			English Exams		
School Year	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1999-2000	2/1-2/3/2000	4/12/00	4/12/00	4/12/00	5/16-5/17/2000
2000-2001	1/29-2/2/2001	4/19/01	4/19/01	4/19/01	5/8-5/9/2001
2001-2002	1/29-1/31/2002	4/16/02	4/16/02	4/16/02	3/5-3/6/2002
2002-2003	2/4-2/6/2003	4/15/03	4/15/03	4/15/03	1/14-1/15/2003
2003-2004	2/3-2/5/2004	4/20/04	4/20/04	4/20/04	1/13-1/14/2004
2004-2005	2/1-2/3/2005	4/12/05	4/12/05	4/12/05	1/11-1/13/2005
2005-2006	1/10-1/12/2006	1/17-1/18/2006	1/17-1/19/2006	1/17-1/18/2006	1/17-1/18/2006
2006-2007	1/9-1/11/2007	1/16-1/17/2007	1/16-1/18/2007	1/16-1/17/2007	1/16-1/17/2007
2007-2008	1/8-1/10/2008	1/8-1/9/2008	1/15-1/17/2008	1/15-1/16/2008	1/15-1/16/2008
2008-2009	1/13-1/15/2009	1/13-1/14/2009	1/21-1/23/2009	1/21-1/22/2009	1/21-1/22/2009

			Math Exams		
School Year	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1999-2000	5/17-5/19/2000	5/4/00	5/4/00	5/4/00	5/18-5/19/2000
2000-2001	5/6-5/8/2001	4/25/01	4/25/01	4/25/01	5/15-5/16/2001
2001-2002	5/7-5/9/2002	4/23/02	4/23/02	4/23/02	5/7-5/8/2002
2002-2003	5/6-5/8/2003	4/30/03	4/30/03	4/30/03	5/6-5/7/2003
2003-2004	5/4-5/6/2004	4/27/04	4/27/04	4/27/04	5/4-5/5/2004
2004-2005	5/10-5/12/2005	4/19/05	4/19/05	4/19/05	5/10-5/11/2005
2005-2006	3/7-3/9/2006	3/7-3/8/2006	3/14-3/15/2006	3/14-3/15/2006	3/14-3/15/2006
2006-2007	3/6-3/8/2007	3/6-3/7/2007	3/13-3/14/2007	3/13-3/14/2007	3/13-3/14/2007
2007-2008	3/4-3/6/2008	3/4-3/5/2008	3/10-3/11/2008	3/10-3/11/2008	3/10-3/11/2008
2008-2009	3/4-3/6/2009	3/4-3/5/2009	3/10-3/11/2009	3/10-3/11/2009	3/10-3/11/2009