Title
Essays on the Economics of Education

Author
Bergman, Peter Sturmthal

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on the Economics of Education

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Peter Sturmthal Bergman

2012

ABSTRACT OF THE DISSERTATION

Essays on the Economics of Education:

by

Peter Sturmthal Bergman

Doctor of Philosophy in Economics

University of California, Los Angeles, 2012

Professor Adriana Lleras-Muney, Chair

I study three separate questions in this dissertation. In Chapter 1, I examine how information frictions between parents and their children affect human capital investment, and how much reducing those friction can improve student effort and achievement. I find that providing additional information to parents regarding missing assignments is a potentially cost-effective strategy to increase parental investments and improve student achievement. In Chapter 2, we measure the impact of high-quality charter schools on teen fertility using admission lotteries to several Los Angeles charter schools as a natural experiment. We find evidence that admission to high-quality charter schools can substantially reduce teen pregnancies. In Chapter 3, we semi-parametrically estimate teacher effects on student test scores using data from the Los Angeles Unified School District. We document that there is significantly more within-teacher variation in teachers' effects than across teacher variation. We find that interacting the teacher indicator variables with a function of the students' lagged test scores captures most of the nonlinearities, preserves the heterogeneity of teacher effects, and provides more accurate estimates.

The dissertation of Peter Leopold Sturmthal Bergman is approved.

Dora Costa

Dayanand Manoli

Meredith Phillips

Adriana Lleras-Muney, Committee Chair

University of California, Los Angeles

2012

I dedicate this dissertation to Kiri and my parents.

Contents

List of Figures

List of Tables

Preface

This dissertation explores three separate topics in the economics of education. In the first chapter, I describe a field experiment designed to answer how information frictions between parents and their children affect human capital investment and how much reducing these frictions can improve student achievement. A random sample of parents was provided additional information about their child's missing assignments, grades and upcoming exams. As in a standard principal-agent model, more information allowed parents to induce more effort from their children, which translated into significant gains in GPA, test scores and measures of student effort. Importantly however, some parents were not fully aware of these information problems; the additional information changed parents' beliefs and spurred demand for more information directly from the school. Relative to other interventions, providing more information to parents potentially produces large gains in achievement at a low cost. I am extremely grateful to Adriana Lleras-Muney, Pascaline Dupas and Karen Quartz for their advice and support in this project. I also thank Sandra Black, Jon Guryan, Day Manoli and Sarah Reber for their detailed feedback. Seminar participants at Case Western University, the CFPB, Mathematica, RAND, UC Berkeley, the University of Chicago Becker Friedman Institute and UCLA also provided helpful comments and suggestions.

In Chapter 2, which is joint work with Karen Coller and Mitchell Wong, we use admission lotteries to several high-achieving charter schools in Los Angeles as a natural experiment to study how school quality can affect teen pregnancy. We first document that the offer for admission to these schools had a significant impact on the quality of schools students attended, as measured by the school's prior test-score performance. We find that the offer for admission reduces this by more than 50%, with a larger reduction for those who attended following the offer. We show that sexual activity decreased and there was no increase in condom use. Students did not receive any additional sex-education as a result of the admission offer, but do become more concerned about the possibility of getting a sexually-transmitted disease. Overall admitted students have superior sex-health knowledge. More fundamentally, these students spend significantly more time in school and doing school work during their junior year; a plausible explanation is because the higher-achieving charter schools have a stronger focus on college preparation. We find no evidence of significant changes to preferences for risk and patience. Thank you to Mitchell Wong and Karen Coller for the opportunity to collaborate. Thank you also to Adriana Lleras-Muney and Day Manoli for the feedback.

In Chapter 3, joint research with Matthew Baird, we estimate semi-parametric models of teachers' value added to student test scores. Estimations of a teacher's value added are founded on models of the education production function. These models typically allow achievement to be a cumulative function of student, family and school inputs. In practice, teacher effects are often identified by assuming these variables are additively separable and that unobservable inputs and endowments correlated with teacher assignment are captured by prior test scores, which imposes strong restrictions on how prior inputs and ability enter the model. This chapter presents estimations that relax the assumption of linear, additive separability. We document several new facts: First, teachers' value added varies significantly by students' initial performance, as measured by prior test scores. There is much greater within-teacher variation in value added than across teacher variation. Second, we find evidence of policy-relevant variation in teachers' value added by model specification. Our most flexible model finds at least 18% of teachers would be reclassified out of the lowest or highest quintiles. We find that a simple OLS specification that includes teacher fixed effects interacted with higher-order terms for lagged test scores approximates our most flexible semi-parametric model best. From a theoretical perspective, our results imply potential complementarities between teachers and student characteristics. From a policy standpoint, optimal teacher-assignment initiatives may want to take into account how a teacher's value added varies by student characteristics. We thank Rosa Matzkin, Jinyong Hahn and Moshe Buchinsky for research advice, as well as seminar participants for helpful advice.

Thank you to Adriana Lleras-Muney, Pascaline Dupas, Day Manoli and Sandra Black for their patience, advice and support.

Curriculum Vitae

Education

- PhD Candidate Economics, University of California Los Angeles 2009
- Masters of Science Economics, University of California Los Angeles 2009
- B.A. Political Economy, University of California Berkeley 2004

Teaching and Research Fields

- Education, Applied Econometrics, Public Finance

Teaching Experience

- Introduction to Econometrics, UCLA, TA for Professor Maria Cassanova
- Designed Markets, UCLA, TA for Professor Marek Pycia
- Microeconomic Theory, UCLA, TA for Professor Jernej Copic
- Stats for Economists, UCLA, TA for Professor Bruce Brown
- Microeconomic Theory, UCLA, TA for Professor Richard Buddin

Research Experience and Other Employment

- Partnership for Los Angeles Schools, Partnership Family and Community Advisory Board Member, 2011-Present
- California Center for Population Research, Affiliate, 2008-Present
- UCLA, Research Assistant for Sarah Reber, 2010
- UCLA, Research Assistant for Day Manoli, 2009-2010
- UCLA, Research Assistant for Frederico Finan, 2008-2009

- New York City Teaching Fellow, Special Education Teacher, 2004-2005

Awards, Fellowships, and Other Activities

- Dissertation Year Fellowship, University of California Los Angeles, 2011-2012

Current Research

- Parent-Child Information Frictions and Human Capital Investment
- School Quality and Teen Fertility (with Karen Coller and Mitchell Wong)
- Semi-Parametric Estimations of Teachers' Value Added (with Matthew Baird)

Chapter 1

Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment

1.1 Introduction

When asked what would improve education in the United States the most, Americans cite more parental involvement twice as often as anything else (*Time*, 2010). Numerous papers reinforce the importance of parental investments in their child's human capital (e.g. Cunha et al., 2006; Houtenville and Smith Conway, 2007; Todd and Wolpin, 2007). However, if children hide information about their academic progress from their parents, a principal-agent problem might arise that impedes these investments. Providing parents more information about their child's academic progress is potentially a cost-effective way to improve academic outcomes. This paper uses a field experiment to answer how information frictions impede parental investments and if reducing these frictions can improve student effort and achievement.

At the outset it is uncertain whether additional information to parents can improve outcomes. There is an association between the quality of information schools provide and school performance: In schools where most students go on to college, 83% of parents are satisfied with the school's ability to communicate information about their child's academic performance, but in schools where most students do not go on to college, 43% of the parents are satisfied with this communication (Civic Enterprises, 2004). However it is not clear what underlies this association. Parents of students who are performing well might be receiving better information or they might require less information because their children have greater academic ability. Even if more information does increase parental investment in their child's education, it is not obvious this will improve academic outcomes. More information might help parents induce more effort from their children, but complementary inputs of the education production function, such as teacher instruction, also determine if this effort translates into measurable gains in achievement.

To measure the causal effect of additional information on parents' investments in their children and student achievement, I conducted an experiment at a low-performing school near downtown Los Angeles. Out of 462 students in grades six through eleven, 242 students' parents or guardians were randomly selected to receive additional information about their child's academic progress. This information consisted of emails, text messages and phone calls listing students' missing assignments and grades several times a month over a six-month period. The information provided was detailed. Messages contained the class, assignment names, problems and page numbers of the missing work whenever possible. Course grades were sent to families every five to eight weeks.[1] To quantify the effects on student effort, achievement and parental investments, I gathered administrative data on assignment completion, work habits, cooperation, attendance and test scores. Parent and student surveys were conducted immediately after the school year ended to provide additional data about each family's response.

The results for high school students suggest there are significant information frictions between parents and their children. As in a standard principal-agent model, more information

[1]This is in addition to the report cards sent by mail to all families in the treatment and control groups.

increased the intensity of parents' incentives and improved their child's effort. Importantly however, some parents are not fully aware of these frictions. Parents in the high school treatment group were twice as likely as the control group to believe that their child does not tell them enough about their schoolwork and grades. This change in beliefs coincides with an increase in parents' demand for information from the school about their child's academic progress. Parents in the treatment group contacted the school about this information 83% more often than the control group, and parent-teacher conference attendance increased by 53%. Unfortunately, the middle-school teachers replicated the treatment for all students in their grades, contaminating the results for those families. There was no estimated effect on middle school parent or student outcomes.

In terms of achievement, additional information potentially can produce gains on par with education reforms such as high-quality charter schools. For high school students, GPA increased by .19 standard deviations. There is evidence that test scores for math increased by .21 standard deviations, though there was no gain for English scores (.04 standard deviations). These effects are driven by a 25% increase in assignment completion and a 24% and 25% decrease in the likelihood of unsatisfactory work habits and cooperation, respectively. Classes missed by students decreased by 28%. For comparison, the Harlem Children's Zone increased math scores and English scores by .23 and .05 standard deviations and KIPP Lynn charter schools increased these scores .35 and .12 standard deviations (Dobbie and Fryer, 2010; Angrist et al., 2010).

Relative to other interventions, providing information could be a cost effective way to reduce the achievement gap.[2] Interventions aimed at adolescents' achievement are often costly because they rely on financial incentives, either for teachers (Springer et al., 2010; Fryer, 2011), for students (Angrist and Lavy, 2002; Bettinger, 2008; Fryer, 2011) or for parents (Miller, Riccio and Smith, 2010). Providing financial incentives for high school

[2]Examples of other information-based interventions in education include providing families information describing student achievement at surrounding schools (Hastings and Weinstein, 2008; Andrabi, Das and Khwaja 2009), parent outreach programs (Avvisati et al., 2010), providing principals information on teacher effectiveness (Rockoff et al., 2010) and helping parents fill out financial aid forms (Bettinger et al., 2009).

students cost $538 per .10 standard-deviation increase, excluding administrative costs (Fryer, 2011). If teachers were to provide additional information to parents as in this study, the cost per student per .10 standard-deviation increase in GPA or math scores would be $156 per child per year. Better information technology could reduce this cost further.

These costs raise an important related question, which is how much parents might be willing to pay to reduce these information problems. This study does not address this question, but Bursztyn and Coffman (2011) use a lab experiment with low-income families in Brazil to show parents are willing to pay substantial amounts of money for information on their child's attendance.

While this paper shows that an intensive information-to-parents service potentially can produce gains to achievement, its policy relevance depends on how well it translates to other contexts and scales up. Large school districts such as Los Angeles, Chicago, and San Diego have purchased systems that make it easier for teachers to improve communication with parents by posting grades online, sending automated emails regarding grades, or text messaging parents regarding schoolwork. The availability of these services prompts questions about their usage, whether teachers update their grade books often enough to provide information, and parental demand for this information. This paper discusses but does not address these questions empirically.

The rest of the paper proceeds as follows. Sections II and III describe the experimental design and the estimation strategy. Sections IV and V present the results for the high school students and middle school students, respectively. Section VI concludes with a discussion of external validity and cost-effectiveness.

Section II outlines a basic framework to interpret the empirical analysis.

1.2 Framework

Typically, models of human capital investment do not incorporate information frictions between parents and their children.[3] Students may wish to hide information from their parents about their human capital investment due to higher discount rates or difficulty planning for the future (Bettinger and Slonim, 2007; Steinberg et al., 2009; Levitt, List, Neckermann and Sadoff, 2011).

The framework below posits a simple non-cooperative interaction between parents and their children to show how additional information can affect student achievement. Student achievement A is a function of student effort e and teacher quality T. Effort e is also a function of a vector z and a vector of parental investments I. z includes a student's ability, peers, discount rate, and value of education. A child takes I as given and maximizes $\max_e \{u(A(e; I, z, T)) - e\}$. Solving, the best response to I is $e(I, z, T)$.

Parents choose I at a cost of $c(I; \varepsilon, t)$. The cost function represents, in a reduced form, the costs parental monitoring, helping with schoolwork directly, or providing incentives. The vector ε captures parental heterogeneity such as their value of education, work schedule, and parenting skills. t is an indicator variable for receiving the information treatment. Parents take their child's best response as given and maximize the utility they get from their achievement (normalized to A): $A(e(I; z, T)) - c(I; \varepsilon, t)$. The first-order condition yields

$$\frac{\partial A(e(I; z, T))}{\partial e} \times \frac{\partial e(I; z, T)}{\partial I} = \frac{\partial c(I; \varepsilon, t)}{\partial I} \tag{1.1}$$

The right-hand side of equation (1) is the marginal cost of parents' investments. The information treatment t could reduce this marginal cost, for instance, by lowering monitoring costs for parents. In a standard moral hazard problem, this lower monitoring cost could increase the intensity of incentives, and in turn, improve student effort. I examine this implication using survey and administrative data.

[3] Several important exceptions in the context of human-capital investment are Weinberg (2001), Berry (2009), and Bursztyn and Coffman (2011).

Equation (1) also highlights several important complementarities. The treatment effect depends on parents' valuation of education (part of ε). Parents who place little value on education are likely to ignore information regarding their child's academic progress. Even if investment increases and student effort increases ($\partial e(I; z, T)/\partial I > 0$), this does not imply a positive effect on student achievement, which depends on the quality of teacher inputs T in the achievement function. If teachers provide students work that is either unproductive or that does not translate into higher test scores, then there will be no measured effect on achievement. The effect size of additional information on investments and achievement is uncertain *ex ante*.

1.3 Background and Experimental Design

1.3.1 Background

The experiment took place at a K-12 school during the 2010-2011 school year. This school is part of Los Angeles Unified School District (LAUSD), which is the second largest district in the United States. The district has graduation rates similar to other large urban areas and is low performing according to its own proficiency standards: 55% of LAUSD students graduate high school within four years, 25% of students graduate with the minimum requirements to attend California's public colleges, 37% of students are proficient in English-Language Arts and 17% are proficient in math.[4]

The school is in a low-income area with a high percentage of minority students. 90% of students receive free or reduced-price lunch, 74% are Hispanic and 21% are Asian. Compared to the average district scores above, the school performs less well on math and English state exams; 8% and 27% scored proficient or better in math and English respectively. 68% of teachers at the school are highly qualified, which is defined as being fully accredited

[4]This information and school-level report cards can be found online at http://getreportcard.lausd.net/reportcards/reports.jsp.

and demonstrating subject-area competence.[5] In LAUSD, the average high school is 73% Hispanic, 4% Asian and 89% of teachers are highly qualified.[6]

The school context has several features that are distinct from a typical LAUSD school. The school is located in a large building complex designed to house six schools and to serve 4,000 students living within a nine block radius. These schools are all new, and grades K-5 opened in 2009. The following year, grades six through eleven opened. Thus in the 2010-2011 school year the sixth graders had attended the school in the previous year while students in grades seven and above spent their previous year at different schools. Families living within the nine-block radius were allowed to rank their preferences for the six new schools. These schools are all pilot schools, which implies they have greater autonomy over their budget allocation, staffing, and curriculum than the typical district school.[7]

1.3.2 Experimental Design

The sample frame consisted of 462 students in grades six through eleven enrolled at the school in December of 2010. Of those, 242 students' families were randomly selected to receive the additional information treatment. This sample was stratified along indicators for being in high school, having had a least one D or F on their mid-semester grades, having a teacher think the service would helpful for that student, and having a valid phone number.[8] Students were not informed of their family's treatment status nor were they told that the treatment was being introduced. Teachers knew about the experiment but were not told which families received the additional information. Interviews with teachers and students suggest that students discussed the messages with each other. There was uncertainty about

[5] Several papers have shown that observable teacher characteristics are uncorrelated with a teacher's effect on test scores (Aaronson et al., 2008; Jacob and Lefgren, 2008; Rivken et al., 2005). Buddin (2010) shows this result applies to LAUSD as well.

[6] This information is drawn from the district-level report card mentioned in the footnote above.

[7] The smaller pilot school system in Los Angeles is similar to the system in Boston. Abdulkadiroglu et al. (2011) find that the effects of pilot schools on standardized test scores in Boston are generally small and not significantly different from traditional Boston public schools. For more information on LAUSD pilot schools, see http://publicschoolchoice.lausd.net/sites/default/files/Los%20Angeles%20Pilot%20Schools%20Agreement%20%28Signed%2

[8] The validity of the phone number was determined by the school's automated-caller records.

who was in the control group at the outset, but students and teachers most likely could infer who was regularly receiving messages and who was not as time went on.

The focus of the information treatment was missing assignments, which included homework, classwork, projects, essays and missing exams. Each message contained the assignment name or exam date and the class it was for whenever possible. For some classes, this name included page and problem numbers; for other classes it was the title of a project, worksheet or science lab. Overwhelmingly, the information provided to parents was negative—nearly all about work students did not do. The treatment rule was such that a single missing assignment in one class was sufficient to receive a message home about. All but one teacher accepted late work for at least partial credit. Parents also received current-grades information three times and a notification about upcoming final exams.

The information provided to parents came from teacher grade books gathered weekly from teachers. 14 teachers in the middle school and high school were asked to participate by sharing their grade books so that this information could be messaged to parents. The goal was to provide additional information to parents twice a month if students missed work. The primary constraint on provision was the frequency at which grade books were updated. Updated information about assignments could be gathered every two-to-four weeks from nine of the fourteen teachers. Therefore these nine teachers' courses were the source of information for the messages and the remaining teachers' courses could not be included in treatment. These nine teachers were sufficient to have grade-book level information on every student.

The control group received the default amount of information the school provided. This included grade-related information from the school and from teachers. The school mailed home four report cards per semester. One of these reports was optional—teachers did not have to submit grades for the first report card of the semester if they did not want to. The report cards contained grades, a teacher's comment for each class, and each teacher's marks for cooperation and work habits. Parent-teacher conferences were held once a semester.

Attendance for these conferences was very low for the high school (roughly 15% participation) but higher for the 7th and 8th grade (roughly 50%) and higher still for the 6th grade (100%). Teachers could also provide information to parents directly. At baseline, many teachers had not contacted any parents, and the median number of calls made to parents regarding their child's grades was one. No teacher had posted grades on the Internet though two teachers had posted assignments.

Figure 1.1 shows the timeline of the experiment and data collection. Baseline data was collected in December of 2010. That same month, contact numbers were culled from emergency cards, administrative data and the phone records of the school's automated-calling system. In January 2011, parents in the treatment group were called to inform them that the school was piloting an information service provided by a volunteer from the school for half the parents at the school. Parents were asked if they would like to participate, and all parents consented. These conversations included questions about language preference, preferred method of contact—phone call, text message or email—and parents' understanding of the A-F grading system. Most parents requested text messages (79%), followed by emails (13%) and phone calls (8%).[9]

The four mandatory grading periods after the treatment began are also shown, which includes first-semester grades. Before the last progress report in May, students took the California Standards Test (CST), which is a state-mandated test that all students are supposed to take.[10] Surveys of parents and students were conducted over the summer in July and August.

Notifications began in early January of 2011 and were sent to parents of middle school students and high school students on alternating weeks. This continued until the end of June, 2011. A bar graph above the timeline charts the frequency of contact with families over six months. The first gap in messages in mid February reflects the start of the new

[9]A voicemail message containing the assignment-related information was left if no one picked up the phone.

[10]Students with special needs can be exempted from this exam.

semester and another gap occurs in early April during spring vacation. This graph shows there was a high frequency of contact with families.

1.3.3 Contamination

The most severe, documented form of contamination occurred when middle school teachers had a school employee replicate the treatment for all students, treatment and control. This employee called parents regarding missing assignments and set up parent-teacher conferences additional to school-wide conferences. This contamination began four-to-five weeks after the treatment started and makes interpreting the results for the middle school sample difficult.

For the high-school sample, a math teacher threatened his classes (treatment and control students) with a notification via the information treatment if they did not do their assignments. These sources of contamination likely bias the results toward zero.

Due to the degree of contamination in the middle school, I analyze the results for the stratified subgroups of middle school and high school students separately.

1.4 Data and Empirical Strategy

1.4.1 Baseline Data

Baseline data include administrative records on student grades, courses, attendance, race, free-lunch status, English-language skills, language spoken at home, parents' education levels and contact information. There are two measures of GPA at baseline. For 82% of high school students, their cumulative GPA prior to entering the school is also available, but this variable is missing for the majority of middle school students. The second measure of GPA is calculated from their mid-semester report card, which was two months before the treatment began. At the time of randomization only mid-semester GPA was available. Report cards contain class-level grades and teacher-reported marks on students' work habits and cooperation. As stated above, there is an optional second-semester report card, however the data in this paper uses mandatory report cards to avoid issues of selective reporting of grades by teachers. Lastly, high school students were surveyed by the school during the first semester and were asked about whom they lived with and whether they have Internet access. 73% of students responded to the school's survey.

Teachers were surveyed about their contact with parents and which students they thought the information treatment would be helpful for. The latter is coded into an indicator for at least one teacher saying the treatment would be helpful for that student.

1.4.2 Achievement-Related Outcomes

Achievement-related outcomes are students' grades, standardized test scores and final exam or project scores from courses. Course grades and GPA are drawn from administrative data on report cards. There are four mandatory report cards available after the treatment began, but only end-of-semester GPA and grades remain on a student's transcript. Final exam and project grades come from teacher grade books and are standardized by class.

The standardized test scores are scores from the California Standards Tests. These tests are high-stakes exams for schools but are low stakes for students. The math exam is subdivided by topic: geometry, algebra I, algebra II and a separate comprehensive exam for students who have completed these courses. The English test is different for each grade. Test scores are standardized to be mean zero and standard deviation one for each different test within the sample.

1.4.3 Effort-Related Outcomes

Measures of student effort are student work habits, cooperation, attendance and assignment completion. Work habits and cooperation have three ordered outcomes: excellent, satisfactory and unsatisfactory. There is a mark for cooperation and work habits for each class and each grading period, and students typically take seven to eight classes per semester. Assignment completion is coded from the nine available teacher grade books. Missing assignments are coded into indicators for missing or not.

There are three attendance outcomes. Full-day attendance rate is how often a child attended the majority of the school day. Days absent is a class-level measure showing how many days a child missed a particular class. The class attendance rate measure divides this number by the total days enrolled in a class.

1.4.4 Parental Investments and Family Responses to Information

Telephone surveys were conducted to examine parent and student responses to the intervention not captured by administrative data. For parents, the survey asked about their communication with the school, how they motivated their child to get good grades, and their perceptions of information problems with their child about schoolwork. Parent-teacher conference attendance was obtained from the school's parent sign-in sheets. The student survey asked about their time use after school, their communication with their parents and their valuations of schooling.[11]

The parent and student surveys were conducted after the experiment ended by telephone. 52% of middle-school students' families and 61% of high-school students' families responded to the telephone survey.[12] These response rates are analyzed in further detail below.

To reduce potential social-desirability bias—respondents' desire to answer questions as they believe surveyors would want—the person who sent messages regarding missing assignments and grades did not conduct any surveys. No explicit mention about the information service was made until the very end of the survey.

[11] Students were also asked to gauge how important graduating college and high school is to their parents, but there was very little variation in the responses across students so these questions are omitted from the analysis.

[12] The school issued a paper-based survey to parents at the start of the year and the response rate was under 15%. An employee of LAUSD stated that the response rates for their paper-based surveys is 30%.

1.4.5 Attrition, Non Response, Missing CST Scores

Of the original of 462 students in the sample, 32 students left the school, 8% of whom were in the treatment group and 6% of whom were in the control group. The most frequent cause of attrition is transferring to a different school or moving away. Students who left the school are lower performing than the average student. The former have significantly lower baseline GPA and attendance as well as poorer work habits and cooperation. Table 1.11 shows these correlates in further detail for middle school and high school students separately.

Just over one third of high-school parents did not respond to the survey. Table 1.12 shows nonresponse for families of high school students. Nonresponse is uncorrelated with treatment status for both children and parents. However, if those who did not respond differ from the typical family, then results based on the surveys may not be representative of the school population. This is true, as a regression of an indicator for non response on baseline characteristics shows the latter are jointly significant (results not shown). Nonetheless, the majority of families responded and provide insight into how they responded to the additional information.

Lastly, many students did not take the California Standards Test. 8% of scores are missing for math and 7% of scores are missing for English. These tests were taken on different days. Table 1.13 in the appendix shows the correlates of missing scores for high school students. Baseline controls are added for each of the first three columns with an indicator for missing math scores as the dependent variable. The remaining three columns perform the same exercise for missing English scores. The treatment is negatively and significantly associated with missing scores. The potential bias caused by these missing scores is analyzed in the results section on test scores.

1.4.6 Descriptive Statistics

In practice, the median treatment-group family was contacted 10 times over six months. Mostly mothers were contacted (62%), followed by fathers (24%) and other guardians or family members (14%). 60% of parents asked to be contacted in Spanish, 32% said English was acceptable, and 8% wanted Korean translation.

Figures 1.2 and 1.3 depict GPA and teacher-marked behavior distributions from the mandatory report card at baseline for all students.[13] For every report card, work habits and cooperation are graded as excellent, satisfactory or unsatisfactory for each student's class. Teachers describe the work habits grade as measuring how on task a student is while cooperation reflects how respectful their classroom behavior is. Figure 1.3 shows the majority of students receive satisfactory or excellent marks for class cooperation and only 10% of students receive an unsatisfactory mark. Work-habit grades are more uniformly distributed across the three possible marks.

Table 2.1 presents baseline-summary statistics across the treatment group and the control group for high school students. Panel A contains these statistics for the original sample while Panel B excludes attriters to show the balance of the sample used for estimations. Measures of works habits and cooperation are coded into indicators for unsatisfactory or not and excellent or not. Of the 13 measures, one difference—the fraction of female students—is significantly different (p-value of .078) between the treatment and control group in Panel A. All results are robust to adding gender as a control. Work habits and students' cumulative GPA from their prior grades are better (but not significantly) for the control group than the treatment group. Panel B shows that baseline GPA is .06 points higher for the control group than the treatment group in the sample used for analysis, and as shown below, results are sensitive to this control. One concern with this baseline difference is mean reversion, however students' prior GPA, which is a cumulative measure of their GPA over several years, also shows the treatment group is lower achieving than the control group. In addition, GPA for the control group is highly persistent from the end of the first semester to the end of the second semester. A regression of the latter on the former yields a coefficient near one.[14]

[13] For high school students, 22% of students have a baseline GPA of 1.00 or below, while 19% of students have 3.00 or above.

[14] Mean reversion does occur between students' prior GPA and their baseline GPA, however this reversion does not differ by treatment status (results available on request).

1.4.7 Empirical Strategy

The empirical analyses estimate intent-to-treat effects. Families in the treatment group may have received fewer or no notifications because their child has special needs (13 families); the guidance counselor requested them removed from the list due to family instability (two families); or the family speaks a language other than Spanish, English or Korean (two families). All of these families are included in the treatment group.

To measure the effect of additional information on various outcomes, I estimate the following

$$y_i = \alpha + \beta * Treatment_i + X_i'\gamma + \varepsilon_i \tag{1.2}$$

Control variables in X include baseline GPA and cumulative GPA from each student's prior school, grade indicators and strata indicators. The results are robust to various specifications so long as a baseline measure of GPA is controlled for, which most likely makes a difference due to the .06 point difference at baseline.

I estimate equation 1.2 with GPA as a dependent variable. To discern whether there were any differential effects by subject or for "targeted" classes—those classes for which a teacher shared a grade book to participate in the information treatment—I also use class grades as a dependent variable.[15] This regression uses the same controls as 1.2 above but the standard errors are clustered at the student level. End-of-semester grades are coded on a four-point scale to match GPA calculations.[16]

Similar to class grades, there is a work habit mark and a cooperation mark for each student's class as well. I estimate the effect of additional information on these marks using an ordered Probit model that pools together observations across grading periods and clusters

[15] Recall that only nine of the 14 teachers updated their grade books often enough so that assignment-related information could be provided to parents. The class-grades regression estimates whether those nine teachers' classes showed greater achievement gains than the classes of teachers who did not update grades often enough to participate.

[16] A is coded as 4, B as 3, C as 2, D as 1 and F as 0.

standard errors at the student level. The controls are the same as above with additional grading-period fixed effects. I report marginal effects at the means, but the average of the marginal effects yields similar results.

Effects on full-day attendance and attendance at the classroom level use the same specification and controls as the specifications for GPA and class grades, respectively.

1.5 Results

The Effect of the Treatment on School-to-Parent Contact

Table 1.2 assesses the effect of the treatment on survey measures of school-to-parent contact. Parents were asked how often the school contacted them during the last month of school regarding their child's grades or schoolwork. During this time all parents had been sent a progress report about their child's grades. The first column shows how much more often the treatment group in high school was contacted by the school than the control group, controlling for baseline GPA and cumulative GPA from students' prior schools.[17] The treatment increased contact from the school regarding their child's grades and schoolwork by 187% relative to the control group. The dependent variable in the second column measures the fraction of people that were contacted by the school more than once. This fraction increases by 158% relative to the control group. The treatment had large effects on both the extensive margin of contact and the intensive margin of contact from the school regarding student grades.

Recall that the experiment was contaminated when the middle school teachers had an employee call their students regarding missing assignments. Mechanically, there should be a positive effect for middle school families since parents did receive messages via the treatment. The employee who contacted families regarding missing work did so for all students—treatment and control—likely resulting in parents being contacted more than once regarding

[17] The results without controls are extremely similar.

the same missing assignment. While there is no measure of how often parents were contacted with new information, if the contamination were significant, we would expect school-contact effects to be smaller for the middle school sample. The results are considerably weaker for the middle school students. Contact increased by 106% and the fraction contacted increased by 69%.

1.5.1 Effects on GPA

Figure 1.4 tracks average GPA in the treatment and control groups over time. The red vertical line indicates when the treatment began, which is about one month before the first semester ended in mid February. There is a steady decrease in GPA for the control group after the first semester ends in February followed by a spike upward during the final grading period. The treatment group does not experience this decline and still improves in the final grading period. Teachers reported that students work less in the beginning and middle of the semester and "cram" during the last grading period to bring up their GPA, which may negatively affect learning (Donovan, Figlio and Rush, 2006).

The regressions in Table 1.3 reinforce the conclusions drawn from the graphs described above. Column (1) shows the effect on GPA with no controls. The increase is .15 and is not significant, however the treatment group had a .06 point lower GPA at baseline. Adding a control for baseline GPA raises the effect to .20 points and is significant at the 5% level (column (2)). The standard errors decrease by 35%. The third column adds controls for GPA from students' prior schools and grade level indicators. The treatment effect increases slightly to .23 points. The latter converts to a .19 standard deviation increase in GPA over the control group. Table 1.4 shows estimates of the treatment effect on class grades. Column (1) shows this effect is nearly identical to the effect on final GPA.[18] Column (2) shows the effect on targeted classes—those classes for which there was a grade book provided by teachers so that messages could be sent home regarding missing work. This analysis is underpowered,

[18]The similarity in effects between this unweighted regression and the regression on GPA is because there is small variation in the number of classes students take.

but the interaction term is positive and not significant (p-value equals .16). Columns (3) and (4) show that math classes had greater gains than English classes (p-values equal .11 and .85, respectively). This effect disparity coincides with the difference in effects shown later for standardized tests scores.

The last column shows the effects for students in which at least one teacher thought additional information would be especially helpful for them. The treatment effect is negative but not significant (p-value equals .24). Most likely there was no differential effect for these students, and if anything the effect appears negative. Teachers appear to have no additional information about whom the treatment would be most helpful for. While teachers generally take in new students every year, the teachers in this sample had known students for three months at the time this variable was measured.

Even though grading standards are school specific, the impact on GPA is potentially important. Several studies find that high school GPA is the best predictor of college performance and attainment (for instance Geiser and Santelices, 2007). GPA is also significantly correlated with earnings even after controlling for test scores (Rosenbaum, 1998; French et al., 2010). [19]

1.5.2 Effects on Final Exams and Projects and Standardized Test scores

Additional information causes exam and final project scores to improve by .16 standard deviations (significant at the 5% level, Table 1.5). However, teachers enter missing finals as zeros into the grade book. On average, 18% percent of final exams and projects were not submitted by the control group. The effect on the fraction of students turning in their final exam or project is large and significant. Additional information reduces this fraction missing by 42%, or 7.5 percentage points.

[19] One caveat, however, is that the mechanisms that generate these correlations may differ from the mechanisms underlying the impact of additional information on GPA.

Ideally, state-mandated tests are administered to all students, which would help separate out the treatment effect on participation from the effect on their score. Unfortunately, many students did not take these tests, and as shown above, missing a score is correlated with treatment status. To see whether those who did not take the test responded to the treatment differently than those who did take the test, I compare the GPA results of those who took the standardized tests with those who did not. Specifically, the indicator for treatment is interacted with an indicator for having a math test score or English test score as follows.

$$GPA_i = \beta_0 + \beta_1 * Treatment_i + \beta_2 * Treatment_i * 1(HasScore_i) + X_i'\gamma + \varepsilon_i$$

Where the variable $HasScore_i$ is an indicator for either having an English test score or having a math test score. The coefficient on the interaction term, β_2, indicates whether those who have a test score experienced different effects on GPA than those who do not have a test score. This achievement effect might correlate with the achievement effect on test scores. If β_2 is large, it suggests how the test-score results might be biased—upwards if β_2 is positive and downwards is β_2 is negative.

Table 1.14 shows the results of this analysis. The coefficients on the interaction term for having a score is insignificant (p-value equals .14) but is large and negative. Thus there is some evidence that the treatment effect is smaller for those with test scores compared to those without, which may bias the estimates on test scores downward.

To account for this potential bias, the effects on math and English test scores are shown with a varying number of controls. The first and fourth columns in Table 1.6 control only for baseline GPA. The effect on math and English scores are .08 and -.04 standard deviations respectively. Columns (2) and (5) add controls for prior test scores, race language spoken at home and free-lunch status. The treatment effect on math scores is .19 standard deviations, but remains near zero for English scores. Finally, if the treatment induces lower performing students to take the test, then those with higher baseline GPA might be less affected by this selection. This means we might see a positive coefficient on the interaction term between baseline GPA and the treatment. Columns (3) and (6) add this interaction term. While the interaction term is small for English scores, for math scores it implies that someone with the average GPA of 2.01 has a .18 standard deviations higher math score due to the additional information provided to their parents.

This disparity between math and English gains is not uncommon. Bettinger (2010) finds financial incentives increase math scores but not English scores and discusses several previous studies (Reardon, Cheadle, and Robinson, 2008; Rouse, 1998) on educational interventions that exhibit this difference as well. There are three apparent reasons the information intervention may have had a stronger effect on math than English. First, the math teachers in this sample provided more frequent information on assignments that allowed more messages to be sent to parents. Potentially, this frequency might mean students fall less behind.[20] Second, 30% of students are classified as "limited-English proficient," which means they are English-language learners and need to pass a proficiency test three years in a row to be reclassified. Looking at class grades, these students tend to actually perform *better* in English classes, though interacting the treatment with indicators for language proficiency and English classes yields a large and negative coefficient (results not shown). In contrast, this coefficient is negative but 75% smaller when the interaction term includes an indicator for math classes rather than English classes. This means that the treatment effect for students with limited English skills is associated with smaller gains for English than math, which may in part drive the disparity in effects. Lastly, math assignments might provide better preparation for the standardized tests compared to English assignments if they more closely approximate the problems on the test.

[20]This theory is difficult to test since there is no within-class variation in grade-book upkeep or message frequency conditional on missing an assignment.

1.5.3 Effects on Behaviors and Assignment Completion

The effects on work habits and cooperation are consistent with the effects on GPA. Figure 1.5 shows that the treatment group exhibits less unsatisfactory work habits than control group, on average. Figure 1.6 shows that excellent work habits increase steadily for the treatment group over time. Excellent cooperation, shown in Figure 1.8, dips in the middle of the semester but rises at the end. The average levels of uncooperative behavior exhibit a similar pattern to unsatisfactory work habits (Figures 1.7).

Table 1.7 provides the ordered-Probit estimates for work habits and cooperation (Panel A). Additional information reduces the probability of unsatisfactory work habits by 24%, or a six-percentage point reduction from the overall probability at the mean. This result mirrors the effect on excellent work habits for high school students, which increases by seven-percentage points at the mean. The probability of unsatisfactory cooperation is reduced by 25% and the probability of excellent cooperation improves by 13%.

Panel B shows OLS estimates of the effects on attendance. The effect on full-day attendance is positive though not significant, however full-day attendance rates are already above 90% and students are more likely to skip a class than a full day. Analysis at the class level shows positive and significant effects. The treatment reduces classes missed by 28%. The final column of Panel B contains the estimated probability of missing an assignment. At the mean, 20% of assignments are not turned in. Assignments include all work, classwork and homework, and the grade books do not provide enough detail to discern one from the other.[21] At the mean, the treatment decreases the probability of missing an assignment by 25%.

The behavior effects indicate that one mechanism the additional information operates through is increased productivity during school hours. Assignments may be started in class but might have to be completed at home if they are not finished during class (e.g. a lab

[21] Several teachers said that classwork is much more likely to be completed than work assigned to be done at home.

report for biology or chemistry, or worksheets and writing assignments in history and English classes). If students do not complete this work in class due to poor attendance or a slow work pace, they may not do it at home. The information treatment discourages low attendance and in-class productivity, which in turn may increase assignment completion.

1.5.4 Mechanisms

The goal of this is section is to understand how parents used the additional information provided by the treatment, how students responded outside of school, and how the information affected parents.

How Parents Used the Additional Information

Panel A of Table 1.8 shows how parents used the information. Parents were asked how many privileges they took away from their child in the last month of school, which increased by nearly 100% for the treatment group (column (1)). The most common privilege revoked by parents involved electronic devices—cell phones, television, Internet use and video games—followed by seeing friends.[22] Parents also spoke about college more often to their child, perhaps emphasizing the future returns to schooling in addition to the threat of punishment. Interestingly, parents in the treatment group asked about homework less during the last month of school, though the coefficient is not significant. The negative sign could be due to parents' interpretation of the question in that they exclude messages from the school from their count. Or, parents might substitute away from asking about this from their children if they realize the information provided via the treatment is more reliable. Lastly, children were asked how often they received help with their homework from their parents on a three point scale ("never," "sometimes," or "always," coded from zero to two). The final column of Panel A shows the coefficient is positive but not significant. Overall, more information

[22] An open-ended question also asked students how their parents respond when they receive good grades. 41% said their parents take them out or buy them something, 50% said their parents are happy, proud or congratulate them, and 9% said their parents do not do anything.

appears to facilitate parents ability to incentivize their children to do well in school.

How Students Responded

The first three columns of Panel B show how students' work habits changed outside of school. Tutoring attendance over the semester increased 42%. The coefficient is marginally insignificant at standard levels (p-value equals .11). Tutoring was offered by teachers after school for free. The positive effect on tutoring is at least partially due to several teachers' requirement that missing work be made up during their after school tutoring to prevent cheating. The second column shows the effect on whether students did their homework at the last minute, which was coded from zero to two for "never," "sometimes" or "always." Students in the treatment group were significantly less likely to do their homework at the last minute. Nonetheless, student study hours at home did not significantly increase, which implies that most of the gains in achievement are due to improved work habits at school. The remaining two columns of Panel B show students' valuations of schooling on a four-point scale. Students in the treatment group are more likely to say grades are important, but no more likely to say that college is important. One interpretation of these results is that grades are important because students will be punished if they do not do well, but their intrinsic valuation of schooling has not changed.

Information Problems and Information Demand

Panel C shows that some parents lacked awareness of the information problems with their child regarding school work. Column (1) reports the answers to the question, "Does your child not tell you enough about his or her school work or grades?" Parents in the treatment group are almost twice as likely to say yes as parents in the control group. This answer may reflect parents' about their understanding of the A–F grading system. 11% of parents responded that they did not understand it or were unsure of the meaning of the scale. 40% of parents did not graduate high school and many are not from the United States. Other countries use different grading scales, which might contribute to parents' unfamiliarity and their reliance on their children for information.[23]

Coinciding with the awareness of information problems, the second and third columns show that parents increased their demand for information regarding their child's schoolwork and progress. Over the last semester, parents in the treatment group were much more likely to contact the school regarding the latter (85% more), and this is corroborated by the school's data on parent-teacher conference attendance, which increased by 53%. The guidance counselor reported that parents arranged meetings with her because of the additional information. This increased contact could partly be due to the limited nature of the additional information in the messages home and the conflict it created in the household. The information came directly from the grade book and no further details could be provided. If a child denied missing an assignment, provided an excuse or parents otherwise needed further information, parents might have wanted to speak directly to the counselor or teacher. The treatment appears to make parents aware of communication problems between themselves and their children, which in turns spurs demand for information from the school.

Finally, the last column reports answers to whether parents agree they can help their child do their best at school. Parents in the treatment group are 16 percentage points more likely to say yes.

1.6 Middle School Results

Table 1.9 summarizes the effects on achievement and effort-related outcomes for middle school students, which are mostly small and not significantly different from zero. These results are consistent with the effects on how parents used the additional information, how children responded, and parents' awareness and demand for information (Table 1.10). Based on these results and the contamination, it is difficult to discern what effect the treatment

[23]This knowledge deficit is salient enough to LAUSD that they have begun to offer free classes to parents to teach them about the school system, graduation requirements, what to ask during conferences and other school-related information.

would have had on younger students.

There are several reasons the additional information may have had less effect on younger children. First, the middle school students have less margin to improve: Their GPA is almost a full standard deviation higher than high school students' GPA, middle school students miss 7.5% of their assignments compared to high school students who miss 20% of their assignments, and attendance and behavior are also better for middle school students. However if this were the only cause for small effects, there could still be an impact on students who were lower-performing at baseline. Unfortunately the study is underpowered to examine subgroups in the middle school, however if anything students with higher GPA respond more positively to additional information (results not shown). A second reason there might be smaller effects for middle school students is that parents might be able to control younger children better than older children. It might be less costly for parents to motivate their children or information problems arise less frequently. There is some support for this hypothesis since teacher-measured behavior at baseline is better for middle school students than high school students, which might correlate with parents' ability to control their child. Third, the repeated messages to middle school parents through the information treatment and the contamination by the school employee may have annoyed them. If they had already resolved an issue such as a missing assignment, receiving a second message regarding that work might have been confusing and frustrating. Parents could have viewed the information treatment as less reliable given the lack of coordination about school-to-parent contact and started ignoring it, which might explain the small negative coefficients on several middle school outcomes. Lastly, parents of middle-school students might already obtain information about their child's education more actively, which is reflected in the higher parent-teacher conference attendance. In addition, a comparison of the control groups in the high school and middle school shows that parents of middle school students are more likely to take away privileges from their children, be aware of information problems, contact the school about their child's grades, and feel that they can help their child try their best than parents of high school students in the control group.[24] It is possible that the contamination caused this higher level of involvement—meaning messages home did affect middle school parents—or it could be that these parents were already more involved than high school parents. If the latter, it leaves open the question of why this involvement wanes as children get older; perhaps parents perceive they have less control of their child's effort or that they no longer know how to help them. In short, the effect of additional information on younger children is inconclusive.

1.7 Conclusion and Cost effectiveness

This paper uses an experiment to answer how information asymmetries between parents and their children affect human capital investment and achievement. The results show these problems can be significant and their effect on achievement large. Additional information to parents about their child's missing assignments and grades helps parents motivate their children more effectively, but also changes parents' beliefs about the quality of information they receive from their child. Parents become more aware that their child does not tell them enough about their academic progress. These mechanisms drive an almost .20 standard deviation improvement in math standardized test scores and GPA for high school students. There is no estimated effect on middle-school family outcomes, however there was severe contamination in the middle school sample. One positive aspect of this contamination is that it reflects teachers' perceptions of the intervention. In response to this experiment, the school and a grade-book company collaborated to develop a feature that automatically text messages parents about their child's missing assignments directly from teachers' grade books.

Importantly, this paper demonstrates a potentially cost-effective way to bolster student achievement: Provide parents frequent and detailed grade-book information. Contacting parents via text message, phone call or email took approximately three minutes per student.

[24] Results available upon request.

Gathering and maintaining contact numbers adds five minutes of time per child, on average. The time to generate a missing-work report can be almost instantaneous or take several minutes depending on the grade book used and the coordination across teachers.[25] For this experiment it was roughly five minutes. Teacher overtime pay varies across districts and teacher characteristics, but a reasonable estimate prices their time at $40 per hour. If teachers were paid to coordinate and provide information, the total cost per child per .10 standard-deviation increase in GPA or math scores would be $156. Automating aspects of this process through a grade book could reduce this costs further. Relative to other effective interventions targeting adolescent achievement, this cost is quite low.

However it is important to consider how well these results extrapolate to other contexts. While the student population is fairly representative of a large, urban school district like Los Angeles Unified, there are several parameters of the education-production function that determine the effectiveness of increasing information to parents. The framework in Section II illustrates this point. Two salient factors are teacher and parent characteristics. Teacher quality affects both the capability of the school to provide information and the impact of student effort on achievement. If teachers do not frequently grade assignments, it is difficult to increase the amount of information to parents. Nine of the fourteen teachers in the sample maintained their grade books often enough to effectively participate in the experiment. It is not known whether this is a typical amount or not. In this experiment, the positive effects spilled over to classes for which there was little grade book information. Also, teachers at this school generally accepted work after the requested due date. Parents were notified about missing assignments that they could still help their child complete. This scheme might allow parents to overcome a child's high discount rate by immediately monitoring and incentivizing the work they must make up. The results may be weaker if parents are only notified about work prior to the due date or about work students can no longer turn in. Even if information can be provided, and this engenders greater effort from students,

[25] Some grade book programs can produce a missing assignment report for all of a student's classes.

the effects on achievement depend on the quality of the work teachers supply. Students may work harder but show no gains in measures of learning. For parents, the effects may differ by demographic characteristics as well. Information changes parental beliefs, and this effect may apply less to parents who know the school system well and have greater resources to invest in their children. Finally, the treatment lasted six months. The negative information about academic performance could create tension at home that might impact outcomes differently over the long run. [26]

Overall, the results support a model of human capital investment that incorporates information asymmetries between parents and their children. This experiment suggests providing information lowered monitoring costs for parents, which increased incentives and improved outcomes. More broadly, parental monitoring is positively linked to number of behavioral outcomes, such as crime and health behaviors (Kerr and Stattin, 2000). Future research could examine how parent-child information frictions affect other parental investments in their children as well.

[26] The cost-effectiveness analysis excludes this potential cost to parents and children.

1.8 Figures

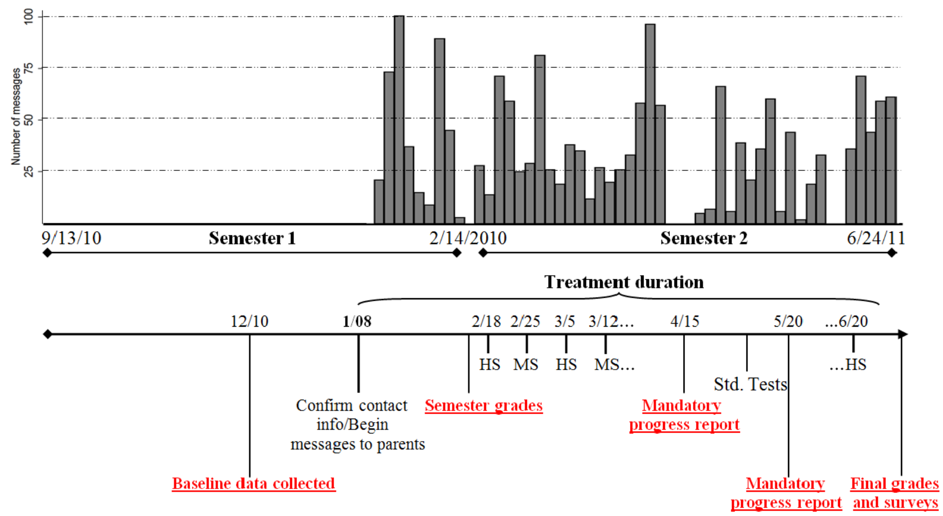


Figure 1.1: Timeline

This figure shows the timeline of the experiment. Above the timeline is a chart of the frequency of messages sent to parents. Each bar signifies the number of messages sent over a three-day period and corresponds to the timeline dates below. The abbreviations HS and MS indicate that messages were sent to families of high school (HS) students families on alternate weeks with respect to middle school (MS) students families. "Std. tests" shows when the state-mandated standardized tests took place.

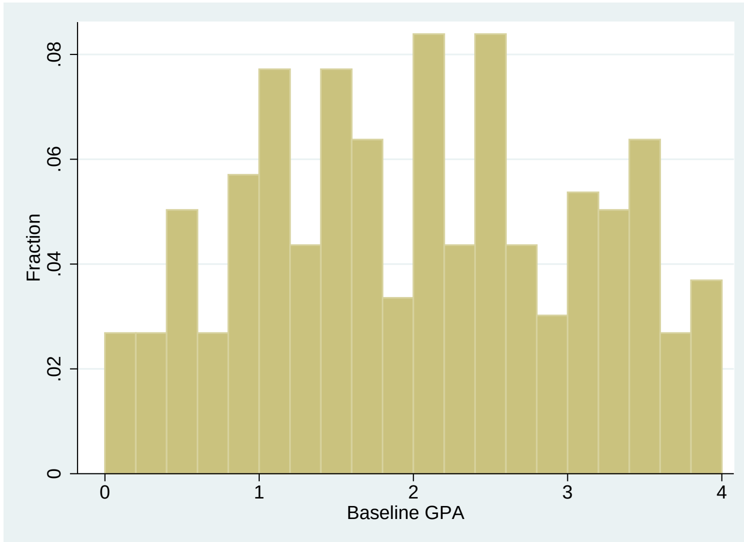


Figure 1.2: Distribution of baseline GPA

This figure shows the distribution of baseline GPA for all grades. Baseline GPA is calculated from a student's mid-semester progress report, which was two months prior to the start of the treatment.

Figure 1.3: Distribution of behaviors at baseline

This figure shows the distribution of baseline work habits and cooperation for high school students. Work habits and cooperation are measured as excellent, satisfactory and unsatisfactory. Students receive these marks for each class they take. Several teachers stated that work habits reflect how on task students are during class, while cooperation measures their interactions with the teacher and peers. These measures were drawn from students' mid-semester progress report, which was two months prior to the start of the treatment.

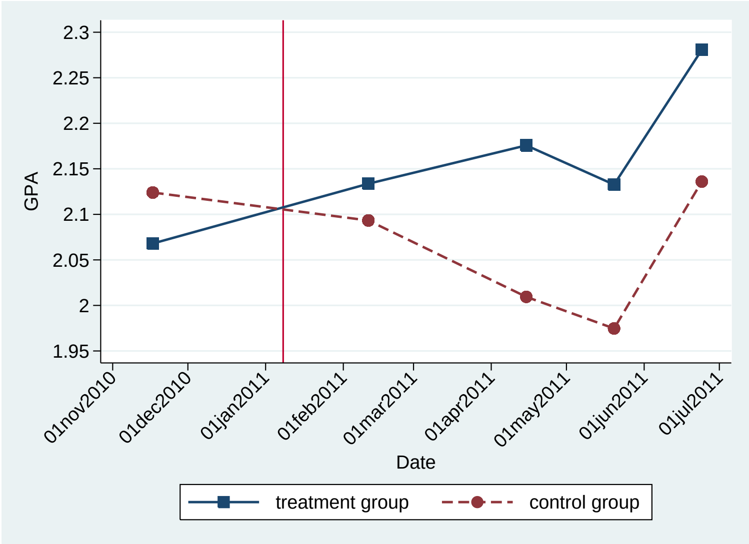


Figure 1.4: GPA over time for high school students

This graph plots the GPA of high school students in the treatment and control group over time. Each point represents the average GPA in a group calculated from progress report grades. The vertical red line indicates when the treatment began. To hold the composition of the sample constant over time, this plot excludes students who left the school prior to the end of the second semester.

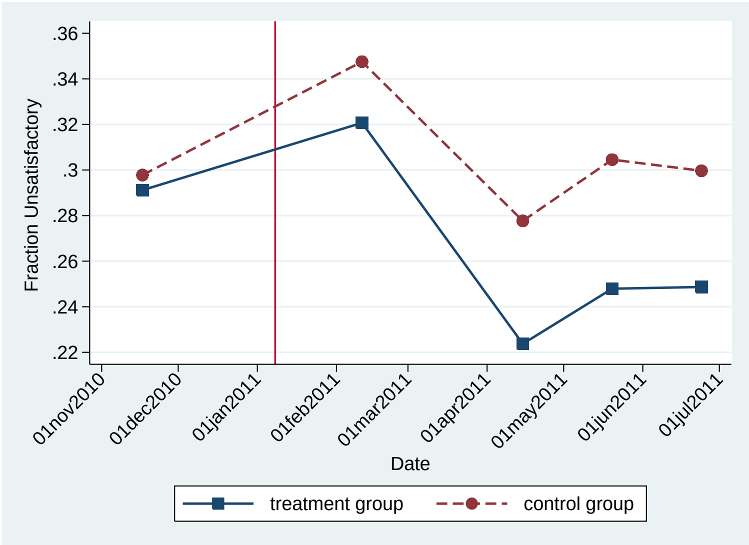


Figure 1.5: Fraction of work habits marked unsatisfactory over time

This graph plots the fraction of unsatisfactory work habit marks for the high school treatment and control groups over time. Work habits are graded as either excellent, satisfactory or unsatisfactory. Each point is calculated using progress report marks from each class. The vertical red line indicates when the treatment began. To hold the composition of the sample constant over time, this plot excludes students who left the school prior to the end of the second semester.

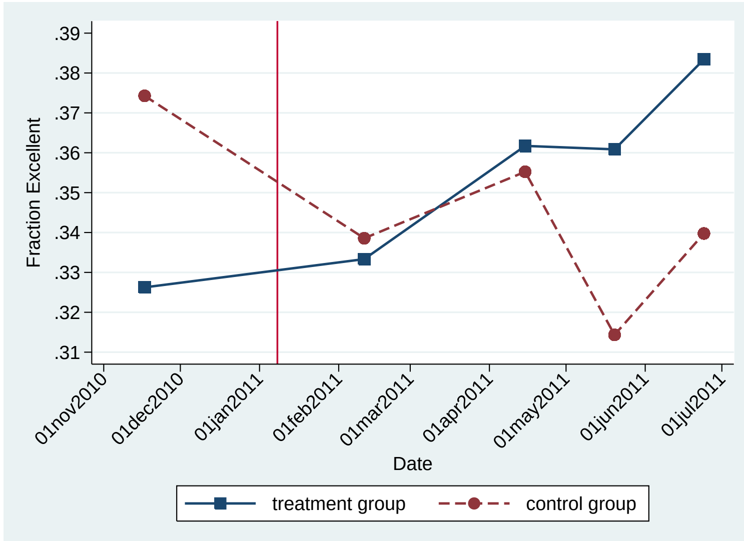


Figure 1.6: Fraction of work habits marked excellent over time

This graph plots the fraction of excellent work habit marks for the high school treatment and control groups over time. Work habits are graded as either excellent, satisfactory or unsatisfactory. Each point is calculated using progress report marks from each class. The vertical red line indicates when the treatment began. To hold the composition of the sample constant over time, this plot excludes students who left the school prior to the end of the second semester.

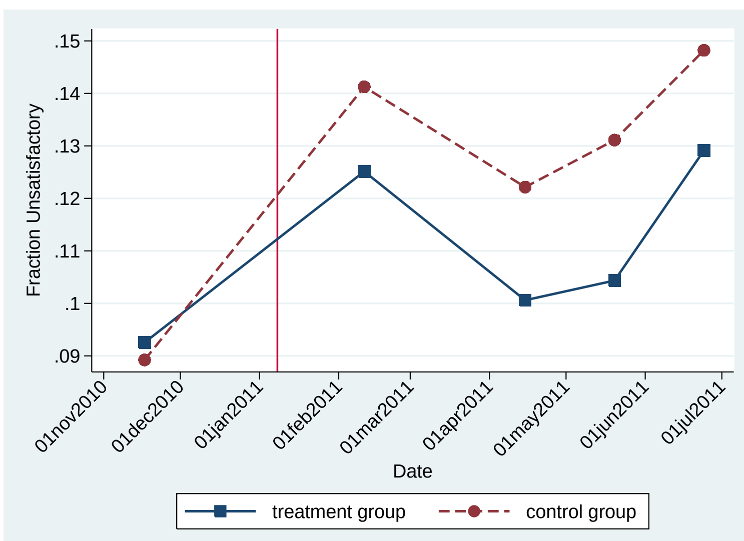


Figure 1.7: Fraction of cooperation marks rated unsatisfactory over time

This graph plots the fraction of unsatisfactory cooperation marks for the high school treatment and control groups. Cooperation is graded as either excellent, satisfactory or unsatisfactory. Each point is calculated using progress report marks from each class. The vertical red line indicates when the treatment began. To hold the composition of the sample constant over time, this plot excludes students who left the school prior to the end of the second semester.

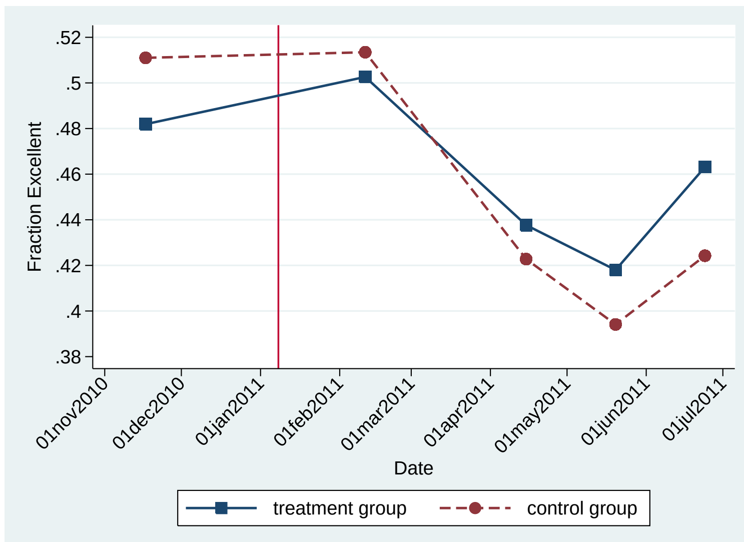


Figure 1.8: Fraction of cooperation marks rated excellent over time

This graph plots the fraction of unsatisfactory cooperation marks for the high school treatment and control groups over time. Cooperation is graded as either excellent, unsatisfactory or excellent. Each point is calculated from progress report marks from each class. The vertical red line indicates when the treatment began. To hold the composition of the sample constant over time, this plot excludes students who left the school prior to the end of the second semester.

1.9 Tables

Table 1.1: Summary Statistics and Treatment-Control Group Balance

Panel A.	Sample balance including attriters					
	Control Mean	Treatment Mean	Difference	p-value	Students	Obs.
Female	0.363	0.463	0.099	0.078	306	306
Attendance	0.928	0.942	0.014	0.278	306	306
Baseline GPA	2.019	1.995	-0.024	0.848	298	298
Prior GPA	2.173	2.043	-0.130	0.282	252	252
Asian	0.24	0.219	-0.021	0.664	306	306
Hispanic	0.699	0.725	0.026	0.612	306	306
Free/Reduced Lunch	0.89	0.869	-0.022	0.563	306	306
Work habits unsatisfactory	0.326	0.308	-0.019	0.585	297	1953
Cooperation unsatisfactory	0.111	0.105	0.006	0.751	297	1952
Work habits excellent	0.354	0.310	-0.044	0.207	297	1953
Cooperation excellent	0.487	0.458	-0.029	0.385	297	1952
Panel B.	Sample balance excluding attriters					
	Control Mean	Treatment Mean	Difference	p-value	Students	Obs.
Female	0.382	0.462	0.079	0.182	279	279
Attendance	0.949	0.952	0.003	0.797	279	279
Baseline GPA	2.124	2.068	-0.056	0.658	272	279
Prior GPA	2.267	2.137	-0.130	0.289	228	228
Asian	0.243	0.238	-0.005	0.924	279	279
Hispanic	0.691	0.706	0.015	0.784	279	279
Free/Reduced lunch	0.904	0.888	-0.016	0.657	279	279
Work habits unsatisfactory	0.298	0.291	-0.007	0.846	279	1804
Work habits excellent	0.374	0.326	-0.048	0.188	279	1804
Cooperation unsatisfactory	0.089	0.093	0.003	0.855	279	1803
Cooperation excellent	0.511	0.482	-0.029	0.388	279	1803

Note: p-values are for tests of equality of means across the treatment and control group.

Table 1.2: Contact from the School Regarding Grades

Dependent variable	(1) School contact to parent	(2) Contacted more than once	(3) School contact to parent	(4) Contacted more than once
Treatment	2.125***	0.453***	1.445***	0.308***
	(0.370)	(0.068)	(0.350)	(0.111)
Baseline GPA	0.100	-0.105**	-0.191	-0.074
	(0.411)	(0.052)	(0.320)	(0.120)
Prior GPA	-0.125	0.100*	-0.408	-0.147
	(0.341)	(0.051)	(0.420)	(0.123)
Control mean	1.134	0.286	1.360	0.448
Sample	H.S.	H.S.	M.S.	M.S.
Observations	183	183	80	80
R-squared	0.173	0.248	0.324	0.246

The dependent variable is drawn from surveys of parents. Parents were asked how many times they were contacted by the school regarding their child's grades or schoolwork during the last month of school. Columns (1) and (3) use the number of times contacted by the school as the dependent variable while columns (2) and (4) use an indicator for whether a parent was contacted more than one time. The first two columns are for the high school sample (HS), while the remaining two columns are for the middle school sample (MS), where the experiment was contaminated. Baseline GPA is calculated from students' mid-semester progress reports from two months before the experiment began. Prior GPA is students' cumulative GPA from middle school and beyond. Strata and grade-level indicators are also included in each regression. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.3: GPA effect on High School Students

	(1)	(2)	(3)
Dependent variable	GPA	GPA	GPA
Treatment	0.145	0.203**	0.229**
	(0.143)	(0.093)	(0.090)
Baseline GPA		0.931***	0.760***
		(0.060)	(0.071)
Prior GPA			0.334***
			(0.072)
Grade 10			-0.248**
			(0.119)
Grade 11			-0.164
			(0.117)
Observations	279	279	279
R-squared	0.004	0.601	0.645

The dependent variable is students' end-of-semester GPA. Data used in these regressions are from administrative records. Baseline GPA is calculated from students' mid-semester progress reports from two months before the treatment began. Prior GPA is students' cumulative GPA from middle school and beyond. Strata indicators are also included in each regression. High school in this sample includes only grades nine through eleven because the school had just opened. Robust standard errors are in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.4: Effects on Grades

	(1)	(2)	(3)	(4)	(5)
Dependent variable	Grade	Grade	Grade	Grade	Grade
Treatment	0.231***	0.188**	0.208**	0.232**	0.351***
	(0.088)	(0.095)	(0.089)	(0.090)	(0.135)
Treatment*Target		0.120			
		(0.086)			
Treatment*Math class			0.212		
			(0.132)		
Treatment*English class				0.022	
				(0.119)	
Treatment*Help					-0.204
					(0.175)
Students	279	279	279	279	279
Observations	2,224	2,224	2,224	2,224	2,224
R-squared	0.399	0.438	0.417	0.405	0.400

The dependent variable in these regressions is each students' class grade, which is coded into a four-point scale from their letter grades. Data used in these regressions are from administrative records. Each student typically takes eight classes. Grades marked incomplete are coded as missing. Additional controls in each regression are students' baseline GPA, prior GPA, grade-level indicators and strata indicators. Baseline GPA is calculated from students' mid-semester progress reports from two months before the treatment began. Prior GPA is students' cumulative GPA from middle school and beyond. Standard errors clustered by student are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.5: Effects on Final Exams and Projects

	(1)	(2)	(3)	(4)
Dependent variable:	All Scores	Math Scores	English Scores	No Score
Treatment	0.160**	0.180*	0.329**	-0.075***
	(0.081)	(0.110)	(0.106)	(0.034)
Students	279	239	100	279
Observations	639	239	100	676
R-squared	0.347	0.430	0.465	0.184

All exam and final project scores are standardized by class to have a mean equal to zero and a standard deviation equal to one. Data in these regressions are from teacher grade books. Additional controls not shown are baseline GPA, prior GPA, grade-level indicators and strata indicators. Baseline GPA is calculated from students' mid-semester progress reports from two months before the experiment began. Prior GPA is students' cumulative GPA from middle school and beyond. The final column shows the effect of the treatment on not having score, excluding excused absences. If a student does not have a score it means they did not turn in any test or project. 18% of tests or final projects were not turned in by the control group. Standard errors clustered by student are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6: Effects on Standardized Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Math Score	Math Score	Math Score	English Score	English Score	English Score
Treatment	0.077	0.190*	-0.282	-0.039	0.008	0.036
	(0.107)	(0.105)	(0.236)	(0.104)	(0.091)	(0.209)
Baseline GPA	0.311***	0.235**	0.119	0.539***	0.339***	0.346***
	(0.076)	(0.087)	(0.104)	(0.083)	(0.078)	(0.101)
Prior GPA	0.292***	0.201***	0.198***	0.140*	0.075	0.018
	(0.084)	(0.077)	(0.076)	(0.078)	(0.070)	(0.071)
Treatment*baseline GPA			0.223**			-0.013
			(0.104)			(0.090)
Additional controls	No	Yes	Yes	No	Yes	Yes
Observations	256	256	256	257	257	257
R-squared	0.306	0.457	0.468	0.337	0.605	0.605

This table reports the effect of the treatment on the state-mandated test, the California Standards Test, for high school students. Scores are standardized by test subject to have a mean of zero and a standard deviation equal to one. The additional controls not shown above are prior test scores, race, language spoken at home and free or reduced-price lunch status. Baseline GPA is GPA calculated from students' mid-semester progress reports two months before the treatment began. Prior GPA is students' cumulative GPA from middle school and beyond. Robust standard errors are in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.7: Effects on Behaviors

	In-Class Work Habits		In-Class Cooperation	
Dependent variable	Pr(Unsatisfactory)	Pr(Excellent)	Pr(Unsatisfactory)	Pr(Excellent)
Treatment	-0.061***	0.069***	-0.024**	0.056**
	(0.021)	(0.022)	(0.011)	(0.024)
Predicted probability	0.256	0.339	0.096	0.432
Students	279	279	279	279
Observations	8,795	8,795	8,795	8,795
		Attendance		Assignment Completion
Dependent Variable:	Full-day Rate	By-Class Rate	Classes Missed	Pr(Missed Asst.)
Treatment	1.675	2.879*	-1.401**	-0.049***
	(1.146)	(1.567)	(0.633)	(0.018)
Control mean	92.81	88.505	5.350	0.197
Students	278	278	278	279
Observations	278	2,252	2,252	27,297

The upper panel reports the effects of the treatment on the probability of unsatisfactory and excellent work habits or cooperation. These behaviors are measured as excellent, satisfactory and unsatisfactory. The coefficients reported are marginal effects at the means from ordered Probit models. Controls not shown are baseline GPA, cumulative GPA from prior schools, grading-period indicators, grade-level indicators and strata indicators. Baseline GPA is GPA calculated from students' mid-semester progress reports two months before the treatment began. Prior GPA is students' cumulative GPA from middle school and beyond. The number of observations differs from the number of students because each student receives a behavior mark for each class and for each of the four grading periods. In the lower panel, full-day attendance measures whether a student attended the majority of the school day, by-class measures attendance for each class, and classes missed measures how many classes a student did not attend over the semester, by course. Lastly, the probability of missing an assignment is reported as the marginal effect at the means from a Probit model. Standard errors clustered at the student level are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.8: Families' Responses to Additional Information

Panel A.	How Parents Used the Information			
Dependent variable	Privileges	Talk College	Ask HW	Help Kid
Treatment	1.660**	2.611*	-2.533	0.088
	(0.718)	(1.415)	(1.688)	(0.069)
Control Mean	1.729	7.637	18.773	0.210
Data source	Parent	Parent	Parent	Child
Observations	180	183	184	183
R-squared	0.168	0.163	0.048	0.091

Panel B.	How Students Responded				
Dependent variable	Tutoring	Homework last minute	Study hours	Grades important	College important
Treatment	5.978	-0.227*	0.146	0.234**	0.040
	(3.763)	(0.116)	(0.263)	(0.102)	(0.074)
Control Mean	14.250	1.202	0.380	3.681	3.639
Data source	Child	Child	Child	Child	Child
Observations	154	152	153	155	154
R-squared	0.086	0.087	0.160	0.181	0.133

Panel C.	Information Problems and Information Demand			
Dependent variable	Information problem?	Contacted School	Attended Conference	Can help
Treatment	0.195***	1.783***	0.079*	0.161**
	(0.070)	(0.668)	(0.046)	(0.072)
Control Mean	0.210	2.102	0.150	0.600
Data source	Parent	Parent	School	Parent
Observations	176	179	181	181
R-squared	0.160	0.147	0.105	0.101

All columns show the effects of the information treatment on parents. Treatment effects are estimated using regressions that control for baseline GPA, prior GPA, strata indicators and grade-level indicators. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.9: Middle School Student Outcomes

Dependent Variable	Treatment	Standard Error	Students	Observations
GPA	-0.108	(0.102)	149	149
Final Exams	-0.054	(0.199)	87	87
Math CST	0.034	(0.119)	139	139
English CST	-0.017	(0.13)	145	145
Pr(missed assignment)	0.005	(0.01)	87	7,692
Work habits unsatisfactory	0.019	(0.019)	149	2,635
Work habits excellent	-0.041	(0.039)	149	2,635
Cooperation unsatisfactory	0.004	(0.009)	149	2,635
Cooperation excellent	-0.021	(0.038)	149	2,635
Full-day attendance	-1.101	(0.684)	149	148
By-class attendance	-1.918	(1.24)	149	1,933
Classes missed	0.782	(0.49)	149	1,933

This table summarizes the results of the treatment effects on middle-school student outcomes, where the experiment was contaminated. The results shown are the coefficients on the treatment indicator in a regression that controls for baseline GPA, prior GPA, grade-level indicators and strata indicators. The treatment effect on missing an assignment is the marginal effect at the means from a Probit model. Work habits and cooperation treatment effects are the marginal effects at the means from an ordered Probit model. All remaining results are estimated by OLS. Where the number of observations differs from the number of students, this is because each student receives a behavior mark for each class as well as each of the four grading periods. By-class attendance is an end-of-semester measure given for each class a student takes. The data for these regressions are drawn from administrative records. Final exam scores could not be obtained for the sixth grade. Standard errors clustered by student are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.10: Middle School Family Survey Outcomes

Dependent Variable	Treatment	Standard Error	N
	How Parents Used the Information		
Privileges taken last month	0.260	(0.320)	79
Talk about college	-0.548	(1.325)	82
Ask about homework	-4.532	(2.745)	81
Help with homework	-0.186	(0.111)*	65
	How Students responded		
Tutoring	0.845	(1.500)	65
HW last minute	-0.001	(0.120)	64
Study hours	-0.267	(0.322)	60
Grades important	-0.167	(0.190)	65
College important	-0.152	(0.135)	65
	Information Problems and Information Demand		
Information problem?	0.100	(0.089)	80
Contacted School	-0.460	(0.601)	81
Can help	0.060	(0.077)	82

The dependent variables in these OLS regressions are from parent and student surveys. Additional controls in these regressions are baseline GPA, grade-level indicators and strata indicators. These results are for families of middle school students only, where the experiment was contaminated. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

1.10 Appendix Tables

Table 1.11: Attrition

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable			Left the Sample			
Treatment	-0.016	-0.012	0.008	0.032	0.031	0.031
	(0.029)	(0.026)	(0.018)	(0.032)	(0.031)	(0.026)
7th grade		0.005	-0.032			
		(0.025)	(0.026)			
8th grade		0.061	-0.017			
		(0.046)	(0.027)			
Baseline GPA			-0.058			-0.017
			(0.037)			(0.022)
Prior GPA			0.007			-0.046*
			(0.036)			(0.027)
Full-day attendance			0.496			-0.907***
			(0.306)			(0.228)
Female			-0.012			0.001
			(0.022)			(0.029)
Black			0.937***			0.175
			(0.115)			(0.122)
Hispanic			-0.160			0.085*
			(0.133)			(0.046)
Asian			-0.107			0.108**
			(0.132)			(0.047)
Free/Reduced Lunch			0.027			-0.068
			(0.018)			(0.041)
9th grade					0.000	0.000
					(0.000)	(0.000)
10th grade					0.095**	0.042
					(0.040)	(0.034)
11th grade					0.048	0.036
					(0.038)	(0.038)
Control mean	0.041			0.085		
Sample	MS	MS	MS	H.S.	H.S.	H.S.
Observations	156	156	156	306	306	306
R-squared	0.031	0.050	0.500	0.040	0.060	0.281

The dependent variable in these regressions is an indicator for having left the school. Columns (1)-(3) show the correlates of leaving for the middle school (MS). Baseline GPA is from mid-semester report cards two months before the treatment began and prior GPA is students' cumulative GPA from previous grades. Columns (4)-(6) show these correlates for the high school (HS) sample only. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.12: Survey Response Correlates

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable			Responded to Survey			
Treatment	0.043	0.033	0.036	0.026	0.013	0.015
	(0.056)	(0.055)	(0.050)	(0.057)	(0.056)	(0.053)
Baseline GPA		-0.067**	0.042		-0.062**	0.031
		(0.029)	(0.030)		(0.030)	(0.032)
9th grade		0.219***	0.224***		0.122*	0.129*
		(0.069)	(0.062)		(0.071)	(0.066)
10th grade		0.100	0.089		0.056	0.046
		(0.074)	(0.067)		(0.076)	(0.070)
Full-day attendance		0.765***	0.651***		0.914***	0.809***
		(0.265)	(0.241)		(0.274)	(0.255)
Female			-0.045			0.000
			(0.052)			(0.055)
Hispanic			0.305***			0.300***
			(0.108)			(0.114)
Asian			-0.225*			-0.175
			(0.115)			(0.121)
Free/Reduced lunch			0.075			0.108*
			(0.059)			(0.062)
Control Mean	0.582			0.493		
Sample	Parents	Parents	Parents	Children	Children	Children
Observations	306	306	306	306	306	306
R-squared	0.002	0.073	0.252	0.001	0.053	0.200

The dependent variable in these OLS regressions is an indicator for responding to the survey. Columns (1)-(3) show the correlates of response for the parent survey. Columns (4)-(6) show these correlates for the child survey. The control mean shows the percentage of control-group members who responded to the survey. These results are for families of high school students only. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.13: Missing CST Scores

Dependent variable	(1) Missing math	(2) Missing math	(3) Missing math	(4) Missing English	(5) Missing English	(6) Missing English
Treatment	-0.054	-0.060*	-0.065*	-0.047	-0.051	-0.056
	(0.033)	(0.033)	(0.035)	(0.032)	(0.032)	(0.035)
Baseline GPA		-0.050*	-0.041		-0.053*	-0.050
		(0.028)	(0.028)		(0.021)	(0.031)
Control mean	0.110			0.103		
Additional controls	No	No	Yes	No	No	Yes
Observations	279	279	279	279	279	279
R-squared	0 .010	0.010	0.122	0.008	0.090	0.173

The dependent variable in these OLS regressions is an indicator for having no test score. Additional controls include prior GPA, prior scores, test-subject indicators and demographic characteristics. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.14: Sample selection and test scores

Dependent variable	(1) GPA
Treatment	0.700**
	(0.356)
Treatment*(has score)	-0.546
	(0.369)
Observations	279
R-squared	0.653

This table shows the treatment effect on GPA, and interacts the treatment variable with an indicator for whether or not a student has a math standardized test score an English standardized test score. These effects are estimated with an OLS regression that controls for baseline GPA, GPA from a students prior school, grade-level indicators and strata indicators. Results are shown for high school students only. All data are from administrative records. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Bibliography

[1] Aaronson, Daniel, Lisa Barrow and William Sander (2007). "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, vol 25(1), 95-135.

[2] Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane and Parag A. Pathak (2011). "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *The Quarterly Journal of Economics*, 126, 699-748.

[3] Andrabi, Tahir, Jishnu Das and Asim Ijaz Khwaja (2009). "Report Cards: The Impact of Providing School and Child Test-Scores on Educational Markets," Mimeo, Harvard University.

[4] Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, Christopher R. Walters (2010). "Inputs and Impacts in Charter Schools: KIPP Lynn," *American Economic Review: Papers and Proceedings*, vol 100(2), 239-243.

[5] Angrist, Joshua D. and Victor Lavy (2002). "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, vol 99(4), 1384-1414.

[6] Avvisati, Francesco, Marc Gurgand, Nina Guyon and Eric Maurin (2010). "The Influence of Parents and Peers on Pupils: A Randomized Experiment," Mimeo, Paris School of Economics.

[7] Berry, James (2011). “Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India,” Mimeo, Cornell University.

[8] Bettinger, Eric P. (2010). “Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores,” NBER Working Paper No. 16333.

[9] Bettinger, Eric, Bridget T. Long, Philip Oreopoulos and Lisa Sanbonmatsu (2009). “The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment,” NBER Working Paper No. 15361.

[10] Bettinger, Eric and Robert Slonim (2007). “Patience in Children: Evidence from Experimental Economics,” *Journal of Public Economics*, vol 91(1-2): 343-363.

[11] Bridgeland, John M., John J. Dilulio, Ryan T. Streeter, and James R. Mason (2008). *One Dream, Two Realities: Perspectives of Parents on America's High Schools*, Civic Enterprises and Peter D. Hart Research Associates.

[12] Buddin, Richard (2010). “How Effective Are Los Angeles Elementary Teachers and Schools?” Retrieved online on April 2011, from http://s3.documentcloud.org/documents/19505/how-effective-are-los-angeles-elementary-teachers-and-schools.pdf.

[13] Bursztyn, Leonardo and Lucas Coffman (2011). “The Schooling Decision: Family Preferences, Intergenerational Conflict, and Moral Hazard in the Brazilian *Favelas*,” Mimeo, UCLA Anderson.

[14] Cunha, F., James J. Heckman, Lance Lochner and Dimitriy V. Masterov (2006), “Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation,” *Handbook of the Economics of Education*, vol 1: 697-812.

[15] Dobbie, Will and Roland G. Fryer (2011). “Are High Quality Schools Enough to Close

the Achievement Gap? Evidence from the Harlem Children's Zone," *American Journal: Applied Economics*, vol 3(3), 158-187.

[16] French, Michael, Jenny Homer and Philip K. Robins (2010). "What You Do in High School Matters: The Effects of High School GPA on Educational Attainment and Labor Market Earnings in Adulthood." Mimeo, University of Miami, Department of Economics.

[17] Fryer, Roland G. (2010). "Financial Incentives and Student Achievement: Evidence from Randomized Trials," NBER Working Paper No. 15898.

[18] Fryer, Roland G. (2011). "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," NBER Working Paper No. 16850.

[19] Gelser, Saul and Marla V. Santelices (2007). "Validity of High-School Grades in Predicting Student Success Beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes," Research & Occasional Paper Series: CSHE.6.07.

[20] Hastings, Justine S. and Jeffrey M. Weinstein (2008). "Information, School Choice, and Academic Achievement: Evidence from two Experiments" *Quarterly Journal of Economics*, vol 123(4) 1373-1414.

[21] Houtenville, Andrew J. and Karen Smith Conway (2008), "Parental Effort, School Resources and Student Achievement," *Journal of Human Resources*, vol 43(2): 437-453.

[22] Jacob, Brian and Lars Lefgren (2008). "Can Principals identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics*, vol 26(1), 101-136.

[23] Kerr, Margaret and Hakan Stattin (2000). "Parental Monitoring: A Reinterpretation," *Child Development*, vol 71(4), 1072-1085.

[24] Levitt, Steven D., John A. List, Susanne Neckermann and Sally Sadoff (2011). "The Impact of Short-Term Incentives on Student Perofrmance" Mimeo, University of Chicago, Becker Friedman Institute.

[25] Los Angeles Unified School District (2009). "Memorandum of Understanding Between Los Angeles Unified School District and United Teachers Los Angeles." Retrieved on January 10, 2011, from http://publicschoolchoice.lausd.net/sites/default/files/Los%20Angeles%20Pilot%20Schools%20Agreement%20%28Signed%29.pdf.

[26] Los Angeles Unified School District (2011). *School Report Cards.* Retrieved on January 10, 2011, from http://getreportcard.lausd.net/reportcards/reports.jsp.

[27] Miller, Cynthia, James Riccio and Jared Smith (2009). "A Preliminary Look at Early Educational Results of the Opportunity NYC-Family Rewards Program" MDRC.

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

[29] Rockoff, Jonah E., Douglas O. Staiger, Thomas J. Kane and Eric S. Taylor (2010). "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools," NBER Working Paper No. 16240.

[30] Rosenbaum, James E. (1998). "College-for-All: Do Students Understand What College Demands?" *Social Pyschology of Education*, vol 2(1) 55-80.

[31] SFGate.com (2010). *Poll Finds Most Blame Parents for School Woes.* Retrieved on January 10, 2011, from http://articles.sfgate.com/2010-12-12/news/25188015_1_parents-united-responsible-education-student-discipline.

[32] Springer, Matthew G., D. Ballou, L. Hamilton, V. Le, J.R. Lockwood, D. McCaffrey, M. Pepper and B. Stecher (2010). "Teacher Pay for Performance: Experimental Evi-

dence from the Project on Incentives in Teacher," Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

[33] Steinberg, Laurence, Lia O'Brien, Elizabeth Cauffman, Sandra Graham, Jennifer Woolard and Marie Banich (2009). "Age Differences in Future Orientation and Delay Discounting" *Child Development*, vol 80(1): 28-44.

[34] Time Magazine (2010). *TIME Poll Results: Americans' Views on Teacher Tenure, Merit Pay and Other Education Reforms*. Retrieved online on January 10, 2011, from http://www.time.com/time/nation/article/0,8599,2016994,00.html#ixzz1dFb1oSwE.

[35] Todd, Petra and Kenneth I. Wolpin (2007), "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," *Journal of Human Capital*, vol 1(1): 91-136.

[36] Weinberg, Bruce (2001), "An Incentive Model of the Effect of Parental Income on Children," *Journal of Political Economy*, vol 109(2): 266-280.

Chapter 2

School Quality and Teen Fertility: Evidence from Charter-School Admission Lotteries

2.1 Introduction

The pregnancy rate for women ages 15-19 in the United States is 7% (Kearney and Levine, 2012). From a policy stand point, teen pregnancy is associated with negative outcomes for parents and their children (e.g. Geronimus and Korenman, 1992); it is important to understand the factors that drive these outcomes. Prior research suggests a relationship between education and teen fertility. McCrary and Royer (2011) summarize several theoretical arguments: Education causes permanent income to rise, which induces women to reduce fertility and invest in the quality of their children; the latter could be reinforced by positive-assortative matching. Education may also improve cognitive ability and in turn improve the ability to understand information regarding fertility decisions.

Empirically, research has focused on the relationship between the *quantity* of education received and fertility. Black, Devereux and Salvanes (2008) find increasing attainment through compulsory schooling laws cause a reduction in teen fertility; McCrary and Royer (2011) find no effect of attainment increases through school starting rules; and Geruso, Clark and Royer (2011) uncover significant effects on the timing of fertility due to compulsory schooling laws in the United Kingdom. However there is significantly less research on the role of education *quality* and fertility.

In theory, increased school quality may induce higher attainment and improve cognitive ability, resulting in similar predictions to the theory described above. Figure 2.1 plots the predicted probability (via a Probit model) of ever being pregnant or getting pregnant with respect to a composite measure of schools' standardized-test score performance. The sample for this plot is drawn from a set of students who applied to charter schools but could not attend due to losing the admission lottery. The relationship is sharply downward sloping as test score performance increases. This evidence is suggestive of a relationship between school quality and teen pregnancy, but is not causal.

To examine the causal relationship between school quality and teen pregnancy, this paper uses admission lotteries to several high-achieving charter schools ("experimental" schools) in Los Angeles as a natural experiment. We draw information from state standardized test scores and surveys of 694 charter school applicants who applied to three charter schools between 2008 and 2010. This is a unique data set, involving tracking students up to three years after applying to the experimental schools who attended 59 different schools in the Los Angeles area. We surveyed more than 75% of all applicants in the sample frame. The survey contains information about a range of sexual behaviors and the potential determinants of those behaviors.

We first document that the offer for admission to these schools had a significant impact on the quality of schools students attended, as measured by the school's prior test-score performance. We then measure the effect of being offered admission on the fraction of students who got pregnant or got someone pregnant. We find that the offer for admission reduces this by more than 50%. The effect is therefore larger for those who attended the schools (82% of those offered admission), with a reduction of more than 70%.

We then examine a number of possible proximate and fundamental mechanisms for this result. As to the former, we show that sexual activity decreased while there was no increase in condom use. Students did not receive any additional sex-education as a result of attending the experimental schools, but do become more concerned about the possibility of getting an sexually-transmitted disease. Overall the admitted group has superior sex-health knowledge. More fundamentally, students in the admitted group spend significantly more time in school and doing school work during their junior year. This could be because the experimental schools have a much stronger focus on college preparation. We find no evidence of significant changes to preferences for risk and patience or valuations of the future.

2.2 Background Information and Study Design

Families have several options for their child's high-school education. Students may attend their neighborhood public school, which is determined by their residential location, or they may apply to a magnet school, private school, or a charter school. Similar to traditional public schools, charter schools require no tuition. In contrast to many traditional public schools, charter schools in the Los Angeles area are required to accept all applications for admission irrespective of a family's residential location. If the number of applications exceeds the number of seats available at a charter school, the law mandates that admission be conducted via a lottery. Families may apply to more than one charter school.

Our sample frame is the set of applicant to three California charter schools located in Inglewood School District, Los Angeles Unified School District, and Lennox School District. Sampling was stratified by school and class. Students who accepted a seat at a charter school other than three we sampled applications from or who accepted a seat at a private school were excluded from the sample. Unfortunately, this potentially induces significant bias in the sample by excluding highly motivated parents from the control group. Overall, We obtained 2,327 applications that span the classes of 2011-2014. The class of 2014 were surveyed with the intention of being tracked through high school. For consistency, this class will be excluded from analyses as they were surveyed early in their freshman year.

We surveyed students who were offered admission (the treatment group) and those who were not offered admission (the control group). Both parental and student consent were required to participate in the study, and students were offered $40 to participate in the 90 minute survey. Sensitive questions were asked via computer and were self-administered.

2.3 Data

The final sample contains students attending 59 different schools across 15 school districts in the Los Angeles area. Figure 2.2 shows the locations of these schools across the Los Angeles area. California schools use a composite measure of various student test scores aggregated to the school level, called the Academic Performance Index (API), to measure school quality. Standardizing this measure across high schools in California, the average sample school has an API of $-.85\sigma$. The experimental schools have API scores $.5\sigma$ to $.75\sigma$ greater than other attended schools.

The survey covers a range of health outcomes and behaviors, health knowledge, peer attributes, future plans, risk aversion and patience, as well as school outcomes and characteristics. The key outcome of interest is whether a student has ever been pregnant or ever gotten someone pregnant. This question is drawn from the National Youth Behavior Survey (YRBS). The surveys also contain information on proximate causes for changes in teen fertility: frequency of various sexual activities in the last three months (intercourse, oral sex, kissing), number of sexual partners, and use of drugs and alcohol while having sex. In addition, we surveyed about condom use in the last three months, concerns about STDs and pregnancy, sex education and health knowledge, and time use. In terms of fundamental mechanisms, we collected information about students college intentions, academic performance, time and risk preferences, and perceptions of the school's focus on college preparation.

Figure 2.3 breaks down the non-response into refusal, unable to locate and ineligible. Refusals occur with similar frequency across the treatment and control groups (12%), however there are disconcertingly higher fractions of ineligible and unlocatable students in the control group (44% v. 27% and 11% v. 7%, respectively). These differences are significant at the 1% level. The bias induced from differences in ineligibility is difficult to estimate empirically, but it is likely that highly-motivated or involved parents are excluded from the control group. This is because the latter may be more likely to apply to more than one charter school, and if their child attends this school, they would be excluded from the study. The difference in ability to locate students across treatment and control may imply that the worst-off control students were the most difficult to find if they no longer attend school regularly. However, it could be that the group we could not locate also contain parents who were motivated enough or had sufficient resources to move long distances if their child was not accepted to a good charter school. Empirically, we find that the amount of phone calls and home visits required to reach a study subject is associated with worse academic and health outcomes, perhaps indicating that those we could not locate have worse outcomes than those we could locate.[1]

Table 2.1 presents comparisons of exogenous characteristics across the treatment and control groups. The two significant differences are the larger fraction of treatment students in the graduating class of 2011 and the greater difficulty in locating control group students. Finding the oldest students, who applied in 2008, was more difficult, which also drives the overall difference in how hard to reach a student was. Students in the freshman and sophomore control groups were, if anything, easier to find than treatment group students (results not shown).

[1]This analysis is available upon request.

2.4 Empirical Strategy

The equation of interest is

$$y_{it} = \alpha + \gamma' X_i + \delta' S_i + \rho Attends_{it} + \varepsilon_{it}$$

Here, $Attends_{it}$ is an indicator for whether or not the student is currently attending an experimental school. X_i are controls for race, gender, indicators for middle school attended and language spoken at home. S_i is a vector of strata indicators.[2] We instrument attendance with the treatment of being offered admission to an experimental school. The first stage we estimate is therefore

$$Attends_{it} = \alpha_1 + \alpha_2' X_i + \alpha_3' S_i + \alpha_4 Treatment_{it} + \nu_{it}$$

.

Table 2.2 shows the coefficients and F-statistics for a joint test of the regressors. Across the three classes, 81% of students offered admission to an experimental school are attending it; though not shown, this coefficient is relatively constant across classes. Columns one and two show the first-stage results without and with controls, respectively. The F-statistics are above 500 in both regressions.

2.5 Results

2.5.1 Effects on School Quality

Table 2.3 shows the effects of the treatment and attending an experimental school on a particular measure of school quality: the API of the school that students attend. Recall that API is a composite measure of student test scores aggregated to the school level. This

[2]Strata include which schools a student applied to and the year in which they applied.

measure is then standardized within the set of schools in the sample. Columns one and two show that the reduced form or effect of being offered admission to an experimental school (the treatment) results in a $.6\sigma$ increase in API without and with controls. Columns three and four likewise show an increase of $.7\sigma$ for students who attend the experimental schools. These regressions show one characteristic of these schools, high test scores; other aspects that could affect health behaviors are investigated in the mechanisms section below.

2.5.2 Effects on Teen Pregnancy

Figure 2.4 graphs the reduced-form differences in ever-pregnant or gotten-someone-pregnant rates by class. We see no significant differences for freshman, but the gap between treatment and control widens into a significant difference for the oldest class, 2011.

Table 2.4 shows the regression results pooling the 2011-2013 classes together. Column one shows the effects of the admission treatment, a three-percentage point (50%) reduction, which is significant at the 10% level. Adding controls increases the effect sizes to four-percentage points (columns two and four). This increase is possibly due to the difficulty in finding control students who might be might be more likely to have gotten someone pregnant or become pregnant. Adding controls might help account for this sampling problem.

2.5.3 Proximate Causes

This section examines the proximate causes for the reduction in pregnancies: Are students having less sex? Increasing contraceptive use? Fewer partners or substituting other behaviors for sex? Table 2.5 shows the analysis the effect of attending an experimental school (the 2SLS estimates) on these outcomes. The first and second entry of each dependent variable shows these effects without and with controls, respectively.

Columns one and two show that attending the experimental schools caused fewer students to have sex in the last three months. 35% of students who were not offered admission had sex in the last three months compared to seven to nine percentage points less for students who attended experimental schools. There are no significant changes to the number of partners or oral sexual activity. On net there's no robust reduction in kissing, and there is an increase in using drugs or alcohol while having sex. The latter might occur from a selection effect: given the reduction in sexual activity, those still having sex while attending the experimental schools exhibit riskier behavior than control group students. Including those who have not had sex in the last three months as students who also do not use alcohol during sex eliminates this effect (results not shown).

Table 2.6 shows condom usage and concerns about STDs and pregnancy. Columns one and two show no significant increase in condom use conditional on having had sex in the last three months. There are also no gains to intentions to use condoms (columns three and four) or to whether students could use condoms. Columns seven and eight ask whether students used condoms the last time they had sex, and there is no effect as well. In terms of concerns for the consequences of sexual behavior, students are more concerned about an STD but not directly concerned about pregnancy.

2.5.4 Fundamental Mechanisms

Changes in sexual behaviors might directly effect pregnancy rates, but these changes do not show why these changes took place. Behaviors might change for a number of reasons: schools improve health knowledge, increased valuation of the future (through income or non-pecuniary improvements), social networks, time constraints and time and risk preferences. We investigate several of these mechanisms below.

Sexual-Health Knowledge

A straightforward reason why students who attend the experimental schools are less likely to get pregnant is because these schools offer sex education classes. Table 2.7 examines this possibility. Columns one and two shows whether students attending the experimental schools received more sex education. This is not the case. Nonetheless, sexual-health knowledge is greater among attending students. Columns three through ten use questions from Add-Health to ascertain students' reproductive health knowledge; asking about the merits of pulling out as a means for contraception (columns three and four), when the most fertile time of month is for a woman (columns five and six), the appropriate fit for a condom (columns seven and eight), and whether Vaseline is a suitable condom lubricant (columns nine and ten). Finally, the index in columns eleven and twelve sums the correct and incorrect answers (scored +1 and −1 respectively) into an aggregate score. The latter demonstrates that the experimental school attendees have significantly greater sexual-health knowledge. This could be due to higher-quality sex-education classes, improved cognition that facilitates knowledge gains, or because of changes in valuations of the future cause students to seek out this information.

Time Use

If students perceive higher returns to their education due, for example, to an improvement in school quality or peers, students may invest more time in school-related work that crowds out time for riskier sexual behaviors. Table 2.8 shows the effects of experimental-school attendance on study time outside of school and time with a boyfriend or girlfriend. On net, there is no significant effect, as shown by columns one and three. However, there is a large increase in study time for the class of 2012, students primarily in their junior year, and a large reduction in time with a boyfriend or girlfriend for that class as well. This could be because junior year is often considered an important year for college admissions; students who intend to go to college may work particularly hard during this year as shown in the next section.

School Attitudes and Performance

Table 2.9 shows measures of school performance, attitudes and college plans. Reiterating the potential for school-related time to crowd out risky sexual behaviors, students are significantly less likely to cut class in the past month in the experimental schools (columns one and two). Students also have the perception that the experimental schools are much safer than the alternatives, which implies that the impact of school quality cannot be disentangled from improvements in school safety. Columns five through ten report whether students believe most students at the school are going to college, if teachers help students prepare for college, and if the school offers classes that help students get into college. These results are all large and significant in favor of the experimental schools improving students' college related perceptions. Columns eleven and twelve however show that students are not any more likely to feel that grades are important. The effects on these perceptions suggest that students in the experimental schools may have a higher return to studying with respect to the marginal utility of risky sexual behaviors compared to the control group.

Preferences

Finally, in Table 2.10 we examine whether the experimental schools altered risk and patience preferences. The evidence here suggests there is no affect. Columns one and two ask students to subjectively rate how much they like to take risks from one (risk avoiding) to five (risk loving). Columns three and four show the results from asking students whether they would accept a lottery with higher expected value than a certainty value. There is no effect on either. Similarly, columns five and six ask students how patient they are from one (not very patient) to five (very patient) as well whether they are willing to wait for more money later than less money now (columns seven and eight). Again, there is no effect.

2.6 Conclusion

This paper uses admissions lottery to several high-achieving charter schools to study the relationship between school quality and teen fertility. We find that the offer of admission to these schools reduces the fraction of students who have ever been pregnant or gotten someone pregnant by 50%, and by 70% for those who attended the schools. In terms of proximate causes, students reduced sexual activity but did not increase condom use. Students did not receive any additional sex-education as a result of attending the experimental schools, but concerns about the possibility of getting a sexually-transmitted disease increase. The treatment group demonstrates greater sex-health knowledge relative to students who lost the admission lottery. More fundamentally, students in the treatment group spend significantly more time in school and doing school work during their junior year. A plausible explanation is because the experimental schools have a much stronger focus on college preparation, which increases the returns to studying relative to the marginal utility of risky sexual behavior.

2.7 Figures

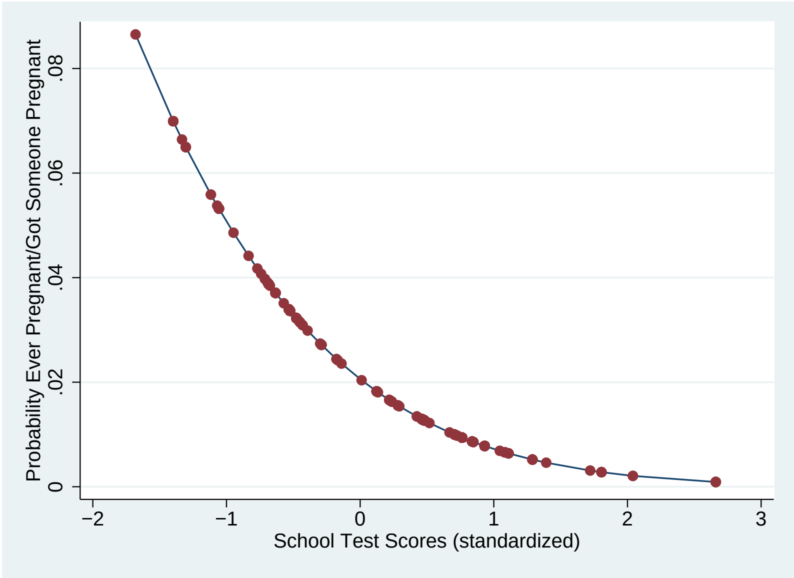


Figure 2.1: School API and Pregnancy Rate

This figure shows the relationship between a measure of school quality, API, and the probability of ever being pregnant or getting someone pregnant. The sample is drawn from a set of students who applied but were not admitted to a set of charter schools. API is a composite measure of student test scores aggregated to the school level.

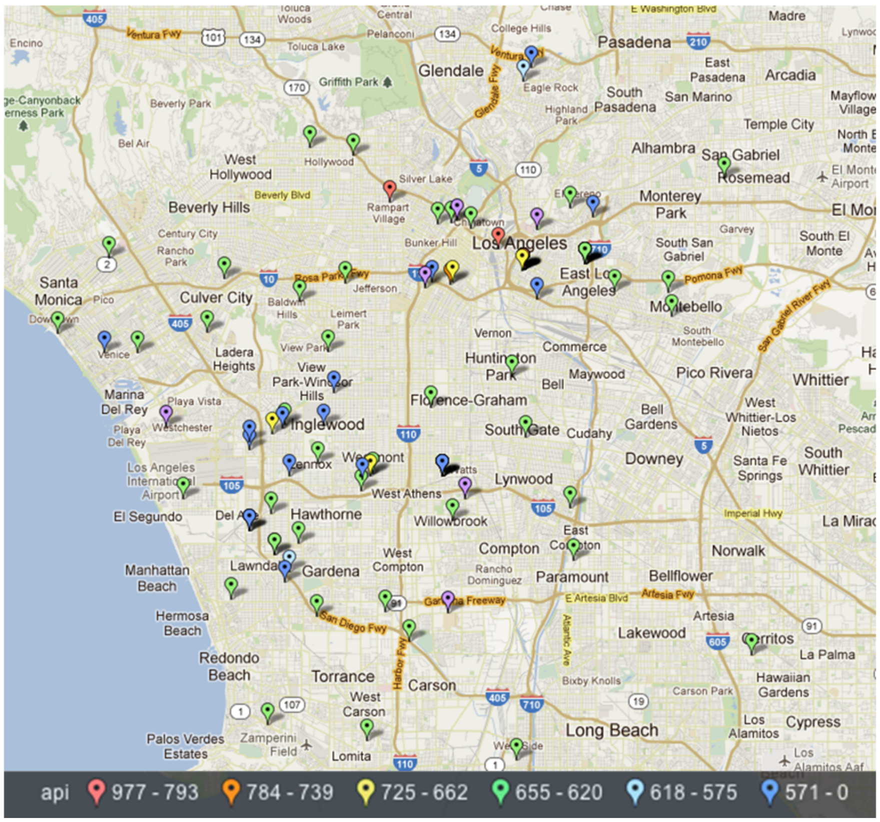


Figure 2.2: Schools in the Sample

This map shows the location of schools in the sample and their respective API scores. API is a composite measure of student test scores aggregated to the school level.

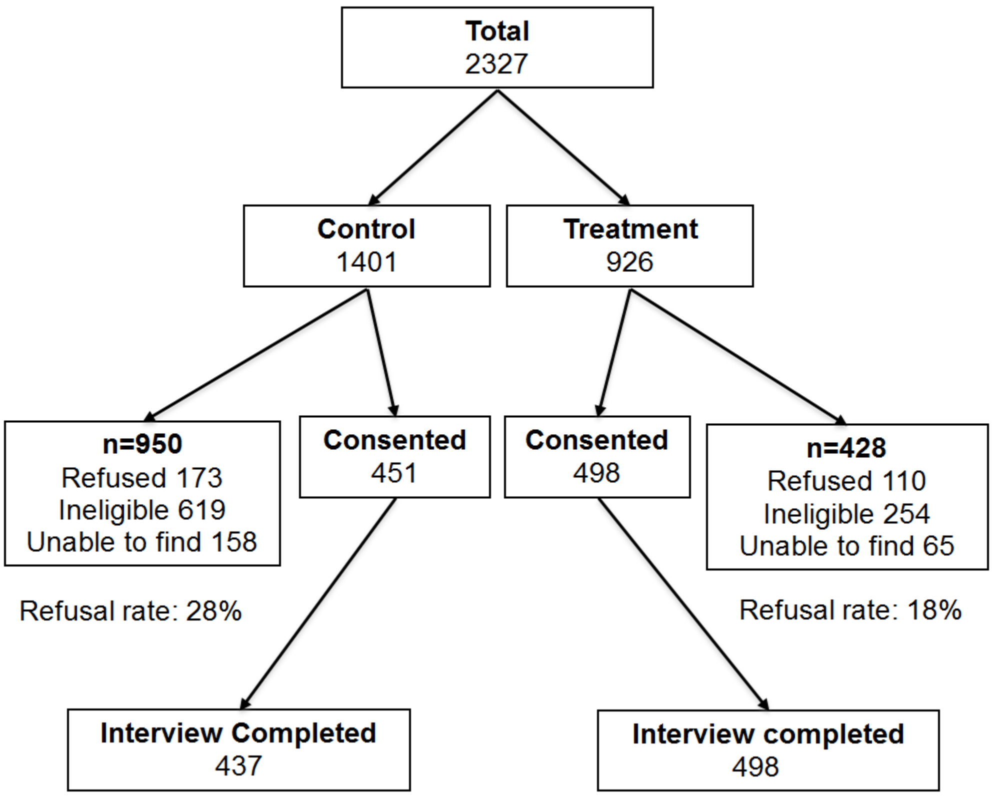


Figure 2.3: Subject Recruitment

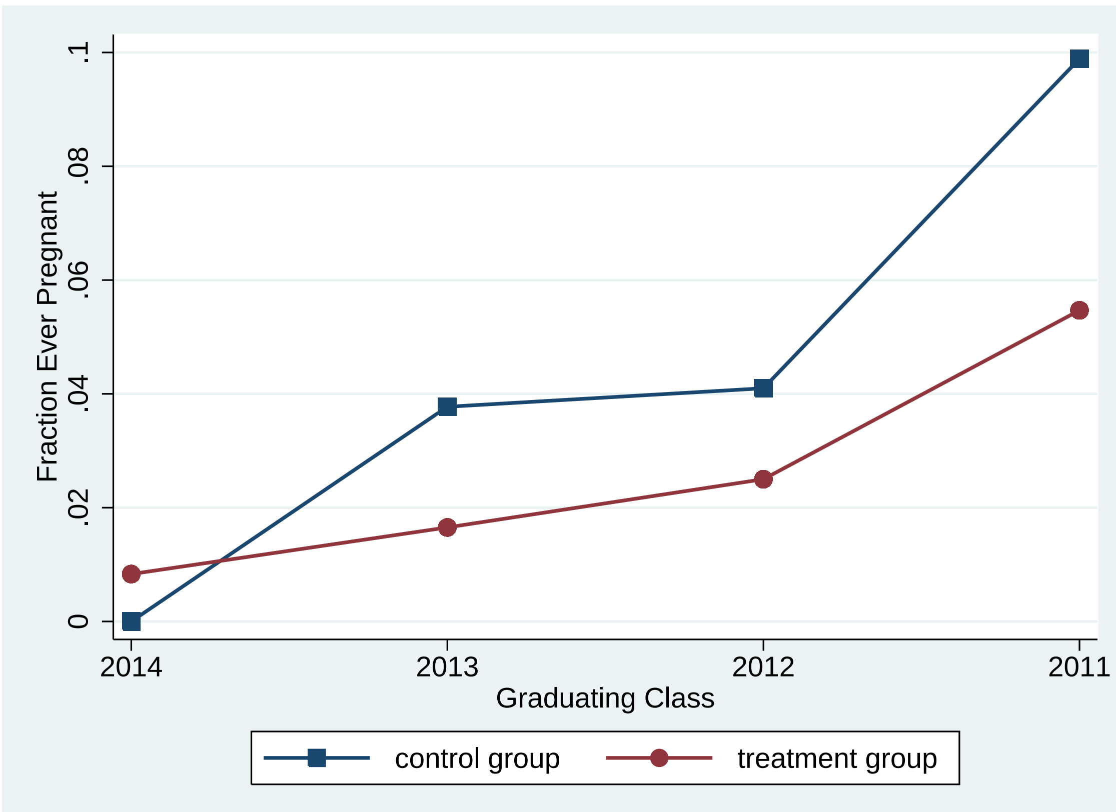


Figure 2.4: Pregnancy Rates by Graduation Class

2.8 Tables

Table 2.1: Summary Statistics and Treatment-Control Group Balance

	Control Mean	Treatment Mean	Difference	P-value	N
Age	16.431	16.390	-0.103	0.116	693
Male	0.478	0.428	-0.038	0.317	693
US born	0.875	0.857	-0.021	0.411	694
English native	0.417	0.367	-0.037	0.325	694
Parent US born	0.240	0.236	0.015	0.639	694
White	0.009	0.012	0.007	0.337	694
Latino	0.796	0.827	0.025	0.407	694
Black	0.134	0.117	-0.019	0.482	694
Mix race	0.061	0.044	-0.014	0.415	694
Parent HS grad	0.513	0.529	0.017	0.657	694
Class 2011	0.209	0.262	0.060	0.090	694
Class 2012	0.277	0.244	-0.058	0.113	694
Class 2013	0.240	0.248	-0.002	0.947	694
Hard to Reach	6.458	5.839	-1.534	0.001	694

Note: P-values are for tests of equality of means across the treatment and control group

Table 2.2: First-stage Estimation

	(1) Attends	(2) Attends
Treatment	0.814***	.818***
	(0.022)	(0.025)
F-Stat	1424	629
Controls	No	Yes
R-squared	0.658	0.735
Observations	694	691

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column one shows the effect of being offered admission to an experimental school (the treatment) on whether or not a student is currently attending an experimental school. The second column repeats this analysis with controls for race, native english speaker, language spoken at home, middle school attended and gender.

Table 2.3: Effects on School Quality

	(1)	(2)	(3)	(4)
	Reduced Form	Reduced Form	2SLS	2SLS
Dependent Variable	APIσ	APIσ	APIσ	APIσ
Treatment	0.622***	0.603***	0.750***	0.731***
	(0.049)	(0.056)	(0.058)	(0.059)
Controls	No	Yes	No	Yes
R-squared	0.213	0.410	0.252	0.459
Observations	665	665	665	665

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.4: Effects on Ever Pregnant/Gotten Someone Pregnant

	(1) Reduced Form	(2) Reduced Form	(3) 2SLS	(4) 2SLS
Treatment	-0.029*	-0.035**	-0.036*	-0.043**
	(0.016)	(0.018)	(0.020)	(0.022)
Control mean	5.6%			
Controls	No	Yes	No	Yes
P-value	0.0751	0.0473	0.0743	0.0462
Observations	688	688	688	688

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.5: Sexual Behaviors

Dependent Variable	(1) 2SLS Sex 3mo.	(2) 2SLS Sex 3mo.	(3) 2SLS Partners	(4) 2SLS Partners	(5) 2SLS O. sex	(6) 2SLS O. sex
Treatment	-0.069	-0.090*	0.043	-0.012	-0.117	-0.412
	(0.043)	(0.046)	(0.103)	(0.114)	(0.323)	(0.318)
P-value	0.112	0.0503	0.678	0.919	0.717	0.196
Observations	687	687	309	309	675	675

Dependent Variable	(7) 2SLS Kiss girl	(8) 2SLS Kiss girl	(9) 2SLS Kiss boy	(10) 2SLS Kiss boy	(11) 2SLS Alc. sex	(12) 2SLS Alc. sex
Treatment	-0.443	0.714	-1.393	-2.044*	0.313**	0.317**
	(1.131)	(1.187)	(1.091)	(1.200)	(0.140)	(0.140)
P-value	0.696	0.548	0.202	0.089	0.0251	0.0236
Observations	665	665	664	664	223	223

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values are for the test of Treatment=0.

Table 2.6: Contraception Use

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Cdm. 3mo.	Cdm. 3mo.	Cdm. Inent	Cdm. Intent	Cdm. Could	Cdm. Could
Treatment	0.143	0.326	0.044	0.022	0.046	0.014
	(0.252)	(0.267)	(0.070)	(0.075)	(0.052)	(0.058)
P-value	0.571	0.223	0.526	0.770	0.371	0.808
Observations	215	215	681	681	678	678
	(7)	(8)	(9)	(10)	(11)	(12)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Used Cdm.	Used Cdm.	HIV/STD	HIV/STD	Get Preg.	Get Preg.
Treatment	0.093	0.082	-0.085**	-0.104**	-0.033	-0.064
	(0.083)	(0.093)	(0.0413)	(0.048)	(0.055)	(0.064)
P-value	0.261	0.377	0.0480	0.0293	0.557	0.320
Observations	213	213	683	683	679	679

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. P-values are for the test of Treatment=0.

Table 2.7: Sexual-Health Knowledge

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Sex Ed.	Sex Ed.	Pull out	Pull out	Fertile	Fertile
Treatment	-0.015	0.016	0.053	0.054	0.127**	0.115*
	(0.037)	(0.039)	(0.049)	(0.053)	(0.059)	(0.065)
P-value	0.676	0.670	0.282	0.301	0.0310	0.0750
Observations	688	688	560	560	462	462

	(7)	(8)	(9)	(10)	(11)	(12)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Cdm. Fit	Cdm. Fit	Cdm. Lubr.	Cdm. Lubr.	Index	Index
Treatment	0.045	0.013	0.088*	0.158***	0.480***	0.475**
	(0.049)	(0.053)	(0.047)	(0.051)	(0.179)	(0.196)
P-value	0.352	0.805	0.0636	0.00215	0.0070	0.0150
Observations	587	587	440	440	691	691

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values are for the test of Treatment=0.

Table 2.8: Time Use

	(1)	(2)	(3)	(4)
	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Study Time/wk	Study Time/wk	BF/GF time	BF/GF time
Treatment	0.700		-0.344	
	(0.542)		(0.543)	
Treatment*2013		0.144		0.731
		(0.806)		(0.789)
Treatment*2012		2.194**		-1.807*
		(1.006)		(1.045)
Treatment*2011		-0.241		-0.011
		(1.000)		(0.984)
P-value	0.197	0.199	0.526	0.507
Observations	691	691	691	691

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values are for the test of Treatment=0.

Table 2.9: School Attitudes and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Cut Schl.	Cut Schl.	Safe Schl.	Safe Schl.	Col. Plan	Col. Plan
Treatment	-0.535***	-0.535***	0.181***	0.183***	0.372***	0.333***
	(0.078)	(0.078)	(0.059)	(0.062)	(0.071)	(0.076)
P-value	0.000	0.000	0.000	0.003	0.000	0.000
Observations	683	683	689	689	683	683
	(7)	(8)	(9)	(10)	(11)	(12)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Teach Col.	Teach Col.	Class Col.	Class Col.	Grades Imp.	Grades Imp.
Treatment	0.345***	0.298***	0.384***	0.343***	0.019	0.014
	(0.058)	(0.061)	(0.061)	(0.063)	(0.049)	(0.051)
P-value	0.000	0.000	0.000	0.000	0.694	0.776
Observations	688	688	688	688	688	688

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. P-values are for the test of Treatment=0.

Table 2.10: Preferences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Dependent Variable	Risks	Risks	Lottery	Lottery	Patient	Patient	Money later	Money later
Treatment	0.048	0.106	0.025	0.010	-0.128	-0.117	0.043	0.055
	(0.096)	(0.101)	(0.046)	(0.050)	(0.094)	(0.100)	(0.037)	(0.039)
P-value	0.620	0.294	0.583	0.845	0.175	0.241	0.239	0.154
Observations	688	688	687	687	688	688	688	688

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values are for the test of Treatment=0.

Bibliography

[1] Black, Sandra E., Paul Devereux, and Kjell Salvanes (2008). “Staying in the Classroom and out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births,” *Economic Journal*, vol 118(530), 1025-1054.

[2] Geronimus, Arline and Sanders Korenman (1992). “The Socioeconomic Consequences of Teen Childbearing Reconsidered,” *Quarterly Journal of Economics*, vol 107(4), 1187-1214.

[3] Geruso, Michael, Damon Clark, and Heather Royer (2011). “The Impact of Education on Fertility: Quasi-Experimental Evidence from the UK,” *Mimeo*, University of California, Santa Barbara.

[4] Kearney, Melissa and Phillip Levine (2012). “Why is the Teen Birth Rate so High and Why does it Matter?” NBER Working Paper 17965.

[5] McCrary, Justin and Heather Royer (2011). “The Effect of Female Education on Fertility and Health: Evidence from School Entry Policies Using Exact Date of Birth,” *American Economic Review*, vol 101(1), 158-195.

Chapter 3

Semi-Parametric Estimations of Teachers' Value Added

3.1 Introduction

Many school districts rely on subjective teacher assessments to evaluate teacher performance. However, there are reasons these measures may be inaccurate: assessments based on classroom observations can occur less than once a year and are scheduled in advance, allowing teachers to exert more effort on the announced day of the evaluation (Taylor and Tyler, 2012), and the highest rating can be given to nearly all teachers. For instance, in Los Angeles Unified School District, less than two percent of all teachers are rated as unsatisfactory and over 90% of teachers receive no negative ratings on any of the 25 ratings categories in the evaluation form (Buddin, 2011).

As an alternative, some school districts are also including in their teacher assessments objective evaluations of the teachers based on an output of the education production function: standardized test scores. Florida, Indiana, Rhode Island, Tennessee, and Colorado, as well as school districts in Houston, Denver, Dallas, Minneapolis and Washington, D.C., already make use of test scores to estimate teachers' value added (Corcoran, 2010). The New York

City Department of Education has estimated teacher effects on test scores for more than 10,000 teachers, and Los Angeles Unified School District has developed its own estimates as well.

Implicitly, these statistical models are based on estimations of an education production function. They are usually estimated by a linear regression of student test scores on previous scores, covariates from administrative data, with an additively separable teacher fixed effects variable. Additive separability of the teacher effect implies that the measured teacher value added does not change by different student characteristics. Todd and Wolpin (2003) show how the common, additively separable linear specification implies unobservable inputs and endowments must decay at a common geometric rate, one example of the restrictions assumed when using these models. Estimating the production function and the teacher value added with a linear model and additively separable teacher effect has at least two direct potential problems.

First, misspecification of the estimation model leads to biased estimation of the production function, the marginal effects of teachers, and thus the teacher rankings (the typical object of interest). The most common value-added specifications do not include flexible interactions between student and teacher characteristics. This assumption of additive separability of the teacher effect means the marginal effects of a teacher's value added is the same for all types of students. However, there may actually be a relationship between a teacher's value added and the ability of his or her students, so that one teacher works better with high performing students and another teacher performs better with low performing students. Incorrect specification of the model leads to biased estimates of the average marginal effect, and even more biased estimates of marginal effects or teacher rankings for low or high performing students.

Second, additive separability causes the teacher effect to be reduced to a single constant, which misses the rich heterogeneity of the teacher effect, providing incomplete information for specific policy questions that are interested in the performance of teachers among low or high achieving students. Not only is there higher within-teacher variation of student's characteristics than between teacher variation, we estimate significantly higher within-teacher variation of their own value added than between-teacher variation of the mean value added. Teachers have more personal variation in their ability to help students than between other teachers' average ability. The standard deviations in value added *across* teachers is measured at .34 for English and .36 for math, while the standard deviations *within* teacher is .87 for English and .83 for math.[1] While a teacher's average value added across their students might be high (low), it might be very low (high) for a subset of their students. From a theoretical perspective, this implies there are potential complementarities between teachers and student characteristics. From a policy standpoint, initiatives that seek to move high-value added teachers to low-performing schools, such as the Talent Transfer Initiative, may want to take into account how a teacher's value added varies by student performance. Also, alternative rankings of teachers can be generated from choosing different measurement criteria. For instance, evaluators might be interested in the median value added effect or including a measure of the variation in the teacher's value added into the ranking. Any additively separable teacher effect model loses information at these different points in the support of student ability.

In this chapter, we estimate several semiparametric models, among them a baseline model that uses linear regression on an additively separable teacher effect model included to represent the common estimation practices currently being used. Although we allow for slightly more flexibility in how lagged test score enters the production function than is typically employed (using a cubic expression), the baseline model is representative of the class of estimation models researchers are using. We also estimate various semiparametric additively separable econometric specifications and additively non-separable specifications. Among the non-separable specifications estimated is one estimated by linear regression that interacts the teacher indicator variables with the cubic in lagged test scores. This method is fast and

[1] The measurement unit for the dependent variable and teacher value added is standard deviations from the sample mean of the student's test score

easy to implement, but allows for student-teacher variation; the results from this chapter suggest this specification should be used in practice.

We estimate all of the models on a subsample of high-tenure elementary school teachers, and estimate the linear regression models–additively separable and additively non-separable– on the full sample of elementary school teachers with sufficient student-year observations for valid estimation of the teacher effect (over forty student-year observations). We find three major results, all of which support the use of the simple to implement and estimate additively non-separable linear regression model.

First, we find that the baseline regression results for the covariates are in line with other value added models being estimated. In particular, the marginal effects are very similar to those of Buddin (2011), who also uses data from Los Angeles Unified School District elementary schools. The methods and data sample used in this chapter are representative of the work currently being done, and the comparisons between the baseline model and the alternative models are representative of the choices facing researchers using other data.

The other two main results show that there are larger differences across the margin of the additivity assumption of the teacher effect than across the generalization of the estimation of the production function through more flexible semiparametric models.

We find that the additively separable models (including the commonly used baseline model) yield substantially different results than the additively non-separable models, evidence against using the baseline model in practice. For even the average marginal effect, where the models match the most, the most flexible additively non-separable model finds in the subsample that 18% of teachers would be reclassified out of the lowest or highest quintiles for Math test scores (18% and 27% for English) when using the baseline model, with greater movement for middle-ranked teachers. For the full sample of teachers, our most flexible model reclassifies 18% (27%) of the lowest-quintile of teachers and 18% (23%) of the highest quintile teachers for Math (English). There is even greater movement for value added at 10th and 90th percentiles of student lagged performance.

Last, the additively non-separable linear regression model matches well with the more flexible Ichimura (1993) single index model. It also still provides for heterogeneous teacher effects, overcoming the two major drawbacks of the linear additively separable model. The teacher value added methods' relative matchings are evaluated by estimating the correlation between the teacher value added between the various models at differing student lagged test score values, and by comparing which quantile teachers are ranked in using various models at differing student lagged test score values.

The additively non-separable linear regression model nests the baseline model. The only difference is that the cubic in lagged test score is interacted with the teacher indicator variables, so that teacher effects can differ with different lagged student test score. An F-test on whether these interaction terms are jointly zero is strongly rejected in all our samples.

There are several caveats to our analysis that also pertain to value-added estimations more generally. These estimates can be imprecise. Specific to our analysis, semiparametric estimations can also exacerbate imprecision. We try to mitigate this problem by using a large data set, from the second largest school district in the United States. There are also potential biases from a number of other sources, including student-to-teacher assignments based on unobservables (Rothstein 2010). In practice, value-added estimations assume these biases are small. Similarly, we make the assumption that assignment is random conditional on some function of observables in order to focus on the potential bias from model misspecification.

The additively non-separable linear regression model is easy to implement and estimate, even using a large data set of teachers such as the Los Angeles Unified School District. The small change of interacting the teacher effect with the lagged student test scores captures most of the effects given in the most flexible model we estimate which doesn't restrict the interactions to be only between teacher effect and lagged student test score. The additively non-separable linear regression model also yields more convincing estimates of the average marginal teacher effect by more flexibly estimating the underlying production function, and allows for capturing heterogeneous teacher effects by lagged student test score.

This study focuses on elementary school teachers for two main reasons: first, teacher assignment for different subjects' tests is more straightforward, and second, as Heckman and Masterov (2007) argue, the most important learning and separation of students happens early, and early interventions are the most effective.

The rest of the chapter proceeds as follows. Section 3.2 presents the empirical strategy and econometric models we will use. Section 3.3 explains the data from Los Angeles schools. Section 3.4 shows the results and discusses the implications, and Section 3.5 concludes.

3.2 Empirical Strategies

Estimation of teacher value added requires assumptions on the education production function. Similar to the model proposed by Todd and Wolpin (2003), we consider the following production function for student achievement

$$T_{it} = m_j\left[Z_{it}, \mu_{i0}, \eta_{ijt}\right]$$

T_{it} is student i's test score in academic year t. m_j is an unknown function of family and school inputs, and j is the teacher assignment. The inputs can broadly be separated into Z_{it}, the history of family and school inputs, μ_{i0}, a student's initial human capital endowment, and η_{ijt}, idiosyncratic shocks. Given family and school inputs, student ability, and random shocks, the unknown production function m_j for each teacher which translates all of these inputs to the output, the test score.

A common assumption in the literature and current practice, that Todd and Wolpin (2003) classify as the value added specification, is to use the lagged test score as a sufficient statistic for unobserved family and school inputs and mental endowment. We also make this assumption.

For each of the models used in this chapter, we assume that the idiosyncratic shocks,

η_{ijt}, are additively separable and orthogonal to all other covariates, as is also commonly done in the literature. Let X_{it} be the observable family and school characteristics, and $W_{it} = [T_{it-1}, X_{it}]$. We separate the econometric models we use into whether the teacher value added effect is additively separable or not. The assumption of an additively separable teacher effect simplifies the estimation process, but has dramatic effects on the results. This chapter establishes that there is too much information lost in the assumption of additive separability, and the results are biased. However, there is a simple additively non-separable model that can be estimated, the linear regression model.

With these assumptions, the additively non-separable models of the production function, which we call the AN models, are

$$T_{it} = m_j\left(W_{ijt}\right) + \eta_{it} \tag{AN}$$

The more restrictive additively separable models, or AS, are

$$T_{it} = m(W_{it}) + \sum_{j=1}^{J} d_{ijt}\psi_j + \eta_{it} \tag{AS}$$

$\eta_{it} = \sum_{j=1}^{J} d_{ijt}\eta_{ijt}$ is the idiosyncratic shocks, ψ_j is the additively separable contribution of teacher j, and d_{ijt} is an indicator variable for whether student i was taught by teacher j in year t. The assumption of additively separable teacher effects implies that the teacher's contribution towards a student's test outcomes does not vary by student characteristics. Teachers are restricted to have the same effect on high performing and low performing students, on male and female, on students enrolled in the free lunch program and those not, and any other factor. The advantage of additive separability is its easy estimation.

We estimate three additively separable models: linear regression, Single-Index Ichimura Model, and Artificial Neural Networks (ANN). We label these models respectively AS1, AS2, and AS3. AS1, the linear regression additively separable teacher effect model, is the baseline model and is representative of the models currently in use. AS2 and AS3 allow for more flexible estimation of the production function $m(\cdot)$ while retaining the simplicity of the additivity assumption. AS2 and AS3 are included as comparisons, to see if the problems inherent in the baseline model are because of poor approximation of the $m(\cdot)$ production function (given the additivity assumption) or result from the additivity of the teacher effect assumption.

We also test two additively non-separable teacher effect models: a linear regression model where the teacher effect is interacted with the student lagged test score variables, and an Ichimura single-index model. We label these models AN1 and AN2, respectively. AN2, the Ichimura model, is the most flexible, allowing for heterogeneous teacher effects to differ by all of the inputs, including lagged student test score. However, we show that AN1 and AN2 have close results, suggesting that the simpler and quicker linear regression model that interacts student lagged test score with the teacher indicator variables is a suitable econometric model choice in application.

We explain each estimation method in detail in the Appendix in Section 3.6.1, including the specification of the estimator of the teacher value added effect for each method. An overall review is presented here.

3.2.1 AS Models: Additively Separable Teacher Effects

The additively separable teacher effect models assume that the production function is the same for every teacher (with only inputs differing), and the teacher effect is the same for any student the teacher instructs. The model is written as

$$T_{it} = m(W_{it}) + \sum_{j=1}^{J} d_{ijt}\psi_j + \eta_{it}$$

Where $\eta_{it} = \sum_{j=1}^{J} d_{ijt}\eta_{ijt}$ is the idiosyncratic shock, ψ_j is the teacher value added for teacher j, and d_{ijt} is an indicator variable for whether student i was taught by teacher j in year t. The three different models we estimate, AS1-AS3, change how $m(\cdot)$ is estimated econometrically.

AS1: Additively Separable Linear Regression

AS1, the linear regression additively separable teacher effects model is the econometric estimation method commonly in use by researchers. For this reason, AS1 will serve as the baseline model. To allow for non-linearity of student ability and heterogeneity captured in the lagged test score, we use a cubic in lagged student achievement. The other controls are assumed to enter linearly.

The intuition behind this model is that the controls on average have linear effects at the margin, and that the teacher effect can be reduced to a summary statistic (the average marginal effect), which will be captured by restricting teachers to have the same effect (ψ_j) on all students that they teach. The absolute teacher effect is not identified, because no students are observed without any teacher. Instead, we identify teacher effects relative to other teachers (the normalization used for all methods in this chapter). The comparison group is the average of all the teachers in the sample by subject, so that teachers in different schools can be compared. The interpretation of the teacher effect is how many standard deviations on average a given teacher contributes. Any normalization (to the mean, to a given baseline teacher, or any other) will create different absolute values, but the same rankings of teachers.

AS2: Additively Separable Single Index Ichimura Model

The linear regression model requires some degree of assumptions about the functional form of $m(\cdot)$. Other semiparametric additively separable teacher effect models are estimated to distinguish if the linear model fails specifically because of the additively separable assumption or because of an insufficiently flexible specification of interactions and higher order terms in the production function. We estimate an additively-separable teacher effects model using a single index Ichimura model, AS2. The model is based on the work of Ichimura (1993). AS2 allows for a much more flexible estimation of the production function by using kernel density estimation of the conditional expectation of test score on the weighted sum of the controls (the index).

AS3: Additively Separable Artificial Neural Networks

We use a model of Artificial Neural Networks (ANN), using the Ridgelet sieve, as presented in Chen, Racine, and Swanson (2001). Chen (2007) presents a review of these estimators and demonstrates that ANN performs particularly well among the class of semiparametric index model estimators as the number of indexing variables increases. The model is more flexible than the Ichimura model by allowing the weights to differ by layer, which, in essence, allows for a very flexible estimation of the production function $m(\cdot)$ where any controls can have arbitrary dependencies with other control variables for marginal effects.

3.2.2 AN Models: Additively Non-Separable Teacher Effects

The next models relax the additivity of the teacher value added effect assumption. Teacher effects are allowed to vary by different student characteristics, a more reasonable approximation of the production function that should give more consistent estimates, and retain the teacher effect heterogeneity. The econometric model is given by

$$T_{it} = m_j(W_{it}) + \eta_{ij}$$

We test two different additively non-separable models, linear regression and Ichimura index model (AN1 and AN2, respectively).

AN1: Additively Non-Separable Linear Regression

AN1, the additively non-separable teacher effect linear regression model includes interactions of the three coefficients on lagged test score (up to the cubic effect) with the teacher indicator variables, d_{ijt}.

This allows a teacher's effect to differ depending on the lagged test score, a summary statistic for the ability of the student, while retaining the fast estimation of OLS. There might be some teachers that are effective in teaching high performing students, while others might contribute more to struggling students. The goal of this model is to determine whether, if the additively separable models seem to not capture the effect well, this version of a linear regression model (AN1) is close to the Ichimura model (AN2) which allows the teacher effect to differ by other controls as well. If so, it offers a suitable, tractable model for large-sample estimation and routine use.

AN2: Additively Non-Separable Single Index Ichimura Model

The intuitive difference between AS2 and AN2 is the $m(\cdot)$ function is allowed to differ by teacher. The conditional expectation of the test score on the weighted sum of the controls is done for each teacher separately (although we require the control weights to be the same across teachers). This allows for a much more flexible estimation of the production function by teacher, but requires sufficient data and takes a lot longer to run. For this reason, we only estimate AN2 for the subsamples, when we have at least 200 student-year observations for each teacher on which to base the estimate.

3.3 LAUSD Data

The data comes from Los Angeles Unified School District (LAUSD), spanning from academic years 2002-2003 to 2009-2010. The data set includes students and teachers from all primary and secondary schools in the district. We limit our attention to teachers in grades three to five, where the teacher assignment is simpler and student learning has greater long-run effects (see Heckman and Masterov 2007). Earlier grades are not included because exams first occur in the second grade so there are no lagged test scores for use as the summary statistic before third grade.

The analysis is performed on two separate groups. First, extensive testing uses all of the estimation methods on a subsample of teachers with a substantial amount of student-year observations (over 200 student-year observations across the sample), for Math and English standardized tests. These sample restrictions help the analysis in at least three ways. First, it yields a higher number of student-year observations per teacher on which the estimates of value added are based, which is especially important for the precision of semiparametric estimation techniques. Second, it requires less parameters to be estimated for each model, as there are less teachers whose effects need to be estimated.[2] Third, the smaller sample, and in particular the smaller amount of teachers, means a much faster estimation time. Given some of the methods are slow to estimate, even programmed using multiple parallel-processor methods, we leave the full battery of tests for this subsample of high tenure teachers.

The second group we perform our analysis on is for all 3-5 grade teachers with at least 40 student-year observations, what we call the full sample. The full sample expands the number of teachers we are analyzing from just under 60 teachers (slightly different depending on whether it is Math or English scores) to over 7000 teachers. We limit the number of student-year observations to be 40 to insure a minimum threshold of accuracy estimating the teacher effect for each teacher. Given the much larger size of these data and computation constraints, we only perform two estimation techniques for the full sample, the two linear regression models AS1 and AN1. Even with these faster methods, the large sample size requires making further divisions in the sample: the analysis is performed by grade and subject, centering each grade-by-subject's teacher value added effects.

The data contain many of the standard variables in district-level teacher value added

[2] The Ichimura models and ANN all take their parameter values as the result of a simplex search; a smaller parameter space to search on decreases the likelihood of getting caught at local minima.

analysis. The most important variables are the student standardized test scores, in the current period and the lagged value. We measure this in terms of standard deviations from the mean (z scores), by year and grade. To match the teacher data to the student data, we look at which teacher gave the students their Math (Reading) marks in a given year, and assign them to be the teacher responsible for their Math (Reading) standardized test scores.

We generate three other continuous variables: the fraction of students in their class that are receiving free/reduced price lunch, the number of students in their class, and the standard deviation of the lagged test z scores in the classroom. The last provides a measure of how different abilities are in the classroom.

We also include a set of indicator variables of student characteristics: a set of race indicator variables (Black, Hispanic, Asian, and other, with White as the baseline group), an indicator variable for male, for being in the gifted program, for being in the free lunch program, and for whether one of their parents has 12 or more years of education. We also include a control indicator variable for students who either declined to report their parents' education or for whom it was missing.

Tables 3.1-3.4 contain the summary statistics for the two test subjects and the subsample and full sample.

Teachers in the subsample work with students that are substantially different than the students in the full sample. Teachers are not randomly kept in the system. On average, teachers in the subsample teach higher achieving students, with lower proportions in free lunch program and a much higher proportion in the gifted program. They teach a smaller fraction of black students, teach larger classes, and work with students almost three times as likely to have a parent that finished high school. These differences likely will bias an estimation of the population parameters, as seen in comparing the coefficients from the full sample and subsamples. However, the subsample does not bias results for the population of similar teachers in similar classrooms, and our goal is to investigate how well different estimation techniques perform within a given sample. To the extent that our subsample does not have systematic differences across the econometric models in the different samples, estimation on the entire population or on the subsample results are informative for the effects of using the different econometric specifications.

The within teacher estimates in Tables 3.1-3.4 are the average standard deviations of each teachers' students characteristics, while the between teacher estimates are the standard deviation of the means of the students characteristics by teacher. In almost every case, the between teacher variance is larger than the within teacher variance.[3] This is helpful for our analysis, implying comparability of teachers on shared supports. Also, estimators that allow for different teacher effects for different students (non-separable models) could have higher within teacher variance than between if teacher effects vary by student characteristics. This emphasizes the potential importance of using non-separable methods to evaluate teacher value added more robustly.

3.4 Results

3.4.1 Baseline Regression Results and Covariate Marginal Effects

Tables 3.5 and 3.6 present the results of the baseline linear regression estimations from model AS1 for Math and English. The first column shows the results for the high-tenure sample[4], and the remaining columns have the results for the full samples in grades three to five. The last row shows the Average Marginal Effect (AME) for lagged test scores.[5] The coefficients are similar across samples with the exception of the fraction of the class receiving free lunch and the individual-level free-lunch recipient indicator variable. The latter is positive and significant for the high tenure sample, as opposed to near zero in the full samples, while the

[3]The one exception to higher within teacher variation is the fraction of the class on the free lunch program. However, the within teacher estimate will be the same for each given class, implying zero variance for any teacher teaching one class only.

[4]The sample that only includes teachers who have at least 200 student-year observations

[5]The AME differs from the marginal effect because of the included higher-order lagged score terms. The AME for a continuous variable is calculated as $\frac{1}{\sum_{i,t} 1} \sum_{i,t} \frac{\partial E[T_{it}|X_{ijt}]}{\partial T_{it-1}}$.

coefficient on the free lunch indicator is more negative for the high-tenured sample compared to the full sample coefficients.

We can compare several coefficients to the model estimated by Buddin (2011). Looking at the Math results, the coefficients on class size are similar: Buddin's range from -.005 to -.002 compared to -.005 to -.003 in our results. Our results on gender switch signs for the third grade versus grades four and five, but the coefficients are both small. The coefficients on parents' education in Buddin's paper are similar to our results.

Broadly speaking, the baseline linear regression results suggest that the relative differences across the full sample and high-tenure samples are small, and the model and data reasonably approximates Buddin's results.

For the other econometric models, we generate the distributions of marginal effects by taking each student in the sample and estimating the marginal effect for the variable of interest with numerical derivatives for continuous variables and the discrete difference in predicted values for binary variables. The distributions of marginal effects are presented in Figures 3.4-3.16. The distributions are consistent with estimates of the same variables in the previous literature, supporting the external validity of our sample and overall methodology and demonstrating the different levels of flexibility don't generally affect the estimated marginal effects much for these control variables.

3.4.2 Correlations Across Models

We estimate four values of the teacher value added effect for each teacher: the average marginal effect and the value added at the 10th, 50th, and 90th percentiles of the lagged student test score. For the percentile estimated teacher effects, the other control variables are evaluated at the modes of the binary variables and means of the other variables. Figure 3.1 shows the distributions across teachers for these four measures by estimation method. The teacher effects are normalized at each different percentile, which is why they are centered at zero. Generally speaking, the distributions of effects are similar across models.

However, the rankings of teachers within the distribution vary by model. Tables 3.7 and 3.8 report the correlation between the estimated teacher effects for the various econometric models. The average marginal effect comparisons give insight into how well the econometric models match for the typical measurement of teacher value added. The correlations of the model at different percentiles of the lagged test score emphasize the shortcomings of using an additively separable teacher value added effect econometric model. For the AS models, the teacher effects will not vary by lagged student test score, and will be the same for the AME and the marginal effect at the 10th, 50th, and 90th percentile of lagged student test score.

Table 3.7 shows very high correlations across all percentiles for the AS models, usually around .99. For comparison, Johnson, Lipscomb and Gill (2012) find correlations from .92 to .99 (an average of .973) across models that vary the number of covariates. Under the assumption of additively separable teacher effects, various estimation methods of the education production function do not affect the value added estimates very much. Researchers should not be concerned over using more flexible semiparametric models; if the additively separable assumption is applied, the results will not vary by much.

However, the results may not be accurate, as suggested by how much they vary from the additively non-separable models, between which the correlations are much lower. For the median lagged math test score, correlations range from .92 to .97. However at the 10th and 90th percentiles, correlations are particularly low: from .52 to .77 and .72 to .75 for the 10th and 90th percentiles respectively. Correlations of effects across models for English scores range higher at the 10th and 90th percentiles. Overall, they range from .76 to .89. The correlations in AME across models are similar to the results at the median.

The correlation coefficients are higher within the additive separability assumptions than across it. For example, AN1 and AN2 have higher correlation coefficients with each other than with AS models. This is the case across all percentiles and subjects, with the exception of the 10th percentile for math scores, which has a correlation of .39 (interestingly, the correlation at the 10th percentile of English scores is .93). It is unclear to us why this correlation is so low in the high tenure sample where there should be sufficient density around the 10th percentile of scores to rely on higher-order lagged score terms. However, the overall results for Math and English suggest that using AN1, which is fast and easily implemented, will be more accurate than the additively separable models and preserves the teacher effect heterogeneity.

The Ichimura and Neural Network models show how more flexible models affect teacher effect evaluation. However, these models are often intractable and too slow for larger samples, and certainly would be difficult to apply to a data set of teachers as large as LAUSD. AN1, on the other hand, is estimable even on a larger data set of over 2500 teachers for each estimation. Given that AN1 approximates AN2 relatively well, we test the correlation of the teacher effects for the two linear regression models, AS1 and AN1, for the full sample. Doing so gives additional insight into whether the subsample results are informative for the full sample by contrasting the relative differences between AS1 and AN1 across both samples. With many more teachers involved, it also gives a wider view at the benefits of allowing for different teacher effects by lagged test scores in using the non-separable model AN1. Table 3.9 shows that the correlations between the non-separable OLS model and the baseline model are overall low, suggesting the importance of interacting teacher effect with the lagged test score. At the 10th percentile of Math and English scores, the correlations are .48 and .52, respectively. The correlation at the median fares better, .94 for math and .90 for English, but suffers again at 90th percentile (.57 for math and .31 for English). The AME also suggests that the separable OLS model misses important within-teacher heterogeneity in value added, with correlations of .81 and .65 for Math and English.

The correlations are so low between the separable and non-separable models away from the median of lagged student test score because of their failure to allow for effects to differ by student ability. We estimate the within teacher distribution of teacher effects in the subsample by evaluating the teacher effect for each teacher and each given student in the sample. Figure 3.2 presents four examples from the English value added estimated teacher distributions for a single teacher. These plots are typical of the remaining distribution estimates. Mechanically for the separable models, the distribution collapses to a vertical line because there can be no within-teacher heterogeneity by construction. The non-separable models document significant variation of the teacher effect. These plots demonstrate three important lessons. First, teachers with above (below) average value added according to the baseline OLS model can be below (above) average for a significant fraction of students, and the reduction to the single point loses quite a bit of information. Second, the distributions of the two non-separable methods AN1 and AN2 tend to be remarkably close to each other, reinforcing the justification for using the easy to implement AN1. The four teachers presented are typical of the results for all teachers. Third, sometimes even the average marginal effect estimates from the additively separable model can be substantially off base.

Overall, the semiparametric non-separable models show that teacher effects can differ substantially by student characteristics. As a summary measure, Table 3.15 documents that the within teacher variation in value added is significantly greater than the across-teacher variation in value added. This heterogeneity is not permitted in typical value-added models. The correlation results show the differences are larger between the groups AS and AN than within them.

3.4.3 Teacher Reclassification by Model

One of the primary goals of evaluating teacher value added effects is to provide a ranking of the teachers. However, the biases that come from using an additively separable model–even at the AME, and even more so for groups of students with outlying lagged test scores–cause the teachers to be incorrectly ranked. This section demonstrates the extent to which the teacher sorting can be wrong.

We look at the distribution in changes in the percentile rankings of teachers between the baseline AS1 model and the more flexible AN1 model. Figure 3.3 shows these distributions according to various percentiles of lagged test scores as well as for the AME. Overall ranking changes appear normally distributed around zero. Changes are greatest at the 10th and 90th percentiles for Math, reaching changes of around 40 percentile points in the tails. Changes at the median and AME tail off at around 20 percentile points. For English scores, there is significantly more variation in the size of the ranking changes, tailing off at around 60 percentile points at the 10th and 90th percentile of lagged scores and around 40 percentile points for the median and AME. The baseline AS1 model can widely misclassify teachers relative to the more flexible AN1 model that nests the baseline model within it.

The correlation tables previously examined show that all of the additively separable methods yield very similar results in the teacher effects, so we limit the attention of the additively separable models to the baseline linear regression model. We compare how well AS1 and AN1 match with the most flexible AN2 for the subsample. Similar to Johnson, Limpscomb and Gill (2012), we examine the policy relevance of our results by separating the distribution of teacher effects into quintiles, and comparing how closely two estimation methods' quintiles match. The results for the subsample are in Tables 3.10 and 3.11. The elements of the tables are proportions in each row conditioned on the column; for example, in the Math subsample in Table 3.10, 45.5 percent of those ranked in the 2nd quantile of teacher effects on students in the 10th percentile lagged test score by AN2 are also ranked in the second quantile by AN1.

Generally, a greater fraction of teachers are ranked differently from the preferred method by the baseline AS1 than by AN1. For English scores, across the AME, 10th, and 90th percentiles of lagged scores, AN1 generally places teachers in the same quintiles as the Ichimura model. For Math scores, at various points in the lagged-score distribution, AN1 performs better, for instance at the 90th percentile; however this is not the case at the 10th percentile. Also, there is a much larger drop-off in the matching for the 10th and 90th percentiles as opposed to the AME for the baseline model than for the additively non-separable linear regression model, showing the heightened limitations of additively separable models away from average lagged test score.

In Tables 3.12 and 3.13, we use the full sample to compare the agreement of AS1 and AN1. Similar to the subsamples, agreement between the models for Math in the highest and lowest quintiles is 82% for the AME, with greater movement in the middle quintiles. However in the full sample there is greater reclassification to quintiles different than the immediately adjacent quintile. 2% (4%) of teachers placed in the lowest (highest) quintiles by the baseline OLS models are placed in non-adjacent quintiles by the the non-separable model, as opposed to zero percent in the subsample. For English scores, the agreement is less and the magnitude of the reclassification is greater: 73% to 77% of teachers at the upper and lower quintiles agree for the AME, while lower at extreme quantiles. At the 10th and 90th percentile of lagged test score range from 32%-63%, which is very low.

The separable model AS1 is a special case of the non-separable model AN1 in the OLS case where the interaction terms between the teacher effect and the lagged test score variables have coefficients are equal to zero. We conduct two sets of hypothesis tests for the full samples with null hypotheses of no difference between the additively separable model and the additively non-separable model for each grade and subject. The results are reported in Table 3.14. The first performs separate hypothesis tests by teacher on the teacher's interacted terms. The fraction of teachers for whom the null hypothesis is rejected is reported. Between 2 to 64 percent of the teachers have different effects by different student ability than the average, according to this hypothesis test. The proportions are indicative of the proportion of teachers for which there is statistical evidence that their production functions (but not necessarily their value added) are differently shaped than the average production function. The second tests all of the interacted terms together. The p-values from these tests show the null hypotheses are strongly rejected in each instance (p-value$<$0.0000). This is strong evidence that the higher-order interaction terms are an important inclusion to the model.

3.5 Conclusion

Value added models are common in empirical investigations into teacher quality. Almost universally, the literature uses a linear OLS model with additively separable teacher value added effects to estimate the teacher value added. In this chapter, we test the additive separability assumption by using various semiparametric methods. Our results show through correlations of the estimated teacher effects at different percentiles and AME, through rankings at different quantiles, and through hypothesis testing between the two OLS models that there is a high degree of within-teacher heterogeneity in the teacher effect, and that not accounting for this through an additively non-separable model, such as one interacting the lagged student test score with teacher assignment, will bias the results.

Our research suggests that estimators that don't allow for the within-teacher value added heterogeneity will provide a very limited view into the contributions the various teachers make, and incorrectly rank the teachers even using the metric of average marginal effect. AN1, the OLS non-separable model, provides a model that is easily estimable, but still captures most of the non-linearities in the Ichimura non-separable model, and retains the teacher effect heterogeneity that is strongly displayed in the estimation and needed to evaluate for policy involved with teacher placement or rankings for low or high performing students. Within-teacher variation is much higher than between teacher variation, so a metric that reduces the distribution to one point, such as AME, even if it could be done accurately, will not provide a full view on the teachers' rankings. Misrankings of the teachers have effects in many school districts using value added methods to assess teacher performance. In Tennessee, measures of the teacher value added account for 35% of the teachers' evaluations, for example (Butrymowicz and Garland 2012). New policy in New York will allow school districts to base up to 40 percent of their teacher evaluation on standardized test performance, of which half must be based off a very simple value added model (Santos and Hu 2012).

The full sample estimations demonstrate that our conclusions are not unique to high observation teachers. In the full sample, we perform the analysis for over 7000 teachers, much larger than most school districts in the United States. Researchers and administrators can easily apply our model to other school districts. We also suggest looking at a combination of measures of the teacher's value-added distribution in their overall ranking. Doing so will yield better evaluations of teachers' varied contributions to student achievement. It will also provide for better estimations of the teacher's contribution for low or high performing students, as is of interest in many practical applications.

3.6 Appendix

3.6.1 Econometric Specifications

AS1: Additively Separable Linear Regression

The specification is given by

$$T_{it} = \beta_0 + \sum_{\ell=1}^{3} \lambda_\ell T_{it-1}^{\ell} + X_{it}\beta + \sum_{j=1}^{J} d_{ijt}\psi_j + \eta_{it}$$

The normalized teacher effect is then given by

$$\widehat{VAM}_{ijt}^{AS1} = \widehat{\psi}_j - \frac{\sum_{q,r,s} \widehat{\psi}_r}{\sum_{q,r,s} d_{qrs}}$$

AS2: Additively Separable Single Index Ichimura Model

The estimation of Ichimura's (1993) model from the assumption $E[\widetilde{\eta}_{it}|T_{it-1}, X_{it}] = 0$, the same assumption made for all of the models. This implies that

$$E\left[T_{it}|W_{it}\right] = E\left[m\left(W_{it}\beta\right)|W_{it}\right] + \sum_{j=1}^{J} d_{ijt}\psi_j$$

The left hand side is observed, and estimated using kernel density estimation:

$$E\left[T_{it}|W_{it}\right] = \frac{\sum_{q\neq i\cap s\neq t} T_{qs} K\left(\frac{W_{qs}\beta - W_{it}\beta}{h}\right)}{\sum_{q\neq i\cap s\neq t} K\left(\frac{W_{qs}\beta - W_{it}\beta)}{h}\right)}$$

As suggested by Li and Racine (2007), we jointly estimate the bandwidths and the index coefficients by minimizing the sum of squared residuals (SSR). The SSR come from the demeaned data, which serves to eliminate all ψ_j from the estimation equation.

$$
\begin{aligned}
\widehat{\eta}_{it} &= T_{it} - \frac{\sum_{q\neq i\cap s\neq t} T_{qs} K\left(\frac{W_{qs}\beta - W_{it}\beta}{h}\right)}{\sum_{q\neq i\cap s\neq t} K\left(\frac{W_{qs}\beta - W_{it}\beta)}{h}\right)} - \sum_{j=1}^{J} d_{ijt}\psi_j \\
\widetilde{\eta}_{ijt} &= \widehat{\eta}_{it} - \frac{\sum_{i,t} d_{ijt}\widehat{\eta}_{it}}{\sum_{i,t} d_{ijt}} \\
SSR &= \sum_{i,j,t} \widetilde{\eta}_{ijt}^2
\end{aligned}
$$

We estimate the teacher effects after we have gotten estimates for β and h by backing out the averaged difference between the student's exam outcomes and the predictions from $\widehat{m}(\cdot)$:

$$
\widehat{VAM}_{ijt}^{AS2} = \frac{\sum_{i,t} d_{ijt}(T_{it} - \widehat{m}(W_{it}\widehat{\beta}))}{\sum_{i,t} d_{ijt}} - \frac{\sum_{q,r,s} \frac{\sum_{i,t} d_{irt}(T_{it} - \widehat{m}(W_{it}\widehat{\beta}))}{\sum_{i,t} d_{irt}}}{\sum_{q,r,s} d_{qrs}}
$$

AS3: Additively Separable Artificial Neural Networks

The model depends on how many hidden neuron layers, or number of sieve terms, are included. Let r_N be the number of hidden neuron layers. The basic form is given by

$$
\widehat{m}(W_{it}) = \alpha_0 + \sum_{r=1}^{r_N} \frac{\alpha_r}{\sqrt{a_r}} \phi\left(\frac{W_{it}\beta_r - b_r}{a_r}\right)
$$

where the ridge function $\phi(\cdot)$ is given by

$$
\phi(\mu) = -0.8311297508e^{-2}(-105 + 105\mu^2 - 21\mu^4 + \mu^6)e^{-.5\mu^2}
$$

Conditional on the number of hidden layers, the parameters are estimated using nonlinear least squares on the demeaned data:

$$\widetilde{T}_{ijt} = T_{it} - \frac{\sum_{i,t} d_{ijt} T_{it}}{\sum_{i,t} d_{ijt}}$$

and similarly for $m(\cdot)$ and η. The number of hidden neuron layers is chosen by which number of hidden layer gives nonlinear least squares estimators with the smallest Bayesian Information Criterion, given by

$$BIC(r) = \ln(SSR_r) + (r * (k+3)) \ln(n)/n$$

where $r * (k+3)$ is the number of parameters estimated, and n is the sample size.

Although very technical, this method is in the end just a highly flexible estimator evaluated using nonlinear least squares. We use it because of its good characteristics for multi-dimensional covariate spaces (Chen 2007). Once the model is estimated, estimations of the teacher value added parameters can be backed out through estimating

$$\widehat{VAM}^{AS3}_{ijt} = \frac{\sum_{i,t} d_{ijt}(T_{it} - \widehat{m}(W_{it}))}{\sum_{i,t} d_{ijt}} - \frac{\sum_{q,r,s} \left[\frac{\sum_{i,t} d_{irt}(T_{it} - \widehat{m}(W_{it}))}{\sum_{i,t} d_{irt}} \right]}{\sum_{q,r,s} d_{qrs}}$$

AN1: Additively Non-Separable Linear Regression

AN1, the additively non-separable teacher effect linear regression model includes interactions of the three coefficients on lagged test score (up to the cubic effect) with the teacher effects d_{ijt}:

$$T_{it} = \beta_0 + \sum_{\ell=1}^{3} \lambda_\ell T_{it-1}^{\ell} + \sum_{j=1}^{J} \sum_{\ell=0}^{3} \gamma_{\ell j} T_{it-1}^{\ell} d_{ijt} + X_{it}\beta + \eta_{it}$$

The value added is given by

$$\widehat{VAM}_{ijt}^{AN1} = \sum_{\ell=0}^{3} \gamma_{\ell j} T_{it-1}^{\ell} - \sum_{q,r,s} \frac{\sum_{\ell=0}^{3} \gamma_{\ell r} T_{qs-1}^{\ell}}{\sum_{q,r,s} d_{qrs}}$$

AN2: Additively Non-Separable Single Index Ichimura Model

Ichimura's (1993) index model comes from the assumption $E[\eta_{it}|W_{it}] = 0$. This implies that

$$E[T_{it}|W_{it}) = E[m_j(W_{it}\beta)|W_{it}]$$

We solve jointly for β and the bandwidths (now one for each teacher) by minimizing the SSR, given by

$$\widehat{\eta}_{ijt} = T_{it} - \frac{\sum_{\ell \neq i \cap s \neq t} d_{\ell js} T_{\ell s} K\left(\frac{W_{\ell s}\beta - W_{it}\beta}{h_j}\right)}{\sum_{\ell \neq i \cap s \neq t} d_{\ell js} K\left(\frac{W_{\ell s}\beta - W_{it}\beta}{h_j}\right)}$$

$$SSR = \sum_{i,j,t} \widehat{\eta}_{ijt}^2$$

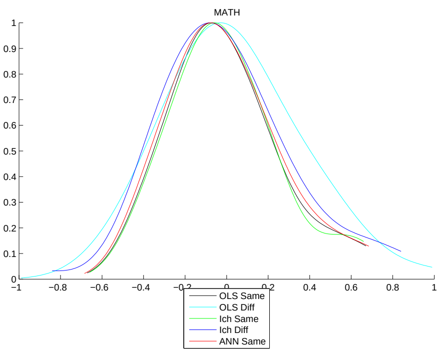
Again, the average teacher value added effects are normalized to average to zero. The teacher value added effect is given by

$$\widehat{VAM}_{ijt}^{AN2} = \widehat{m}_j(W_{it}\widehat{\beta}) - \frac{\sum_{q,r,s} \widehat{m}_r(W_{qs}\widehat{\beta})}{\sum_{q,r,s} d_{qrs}}$$

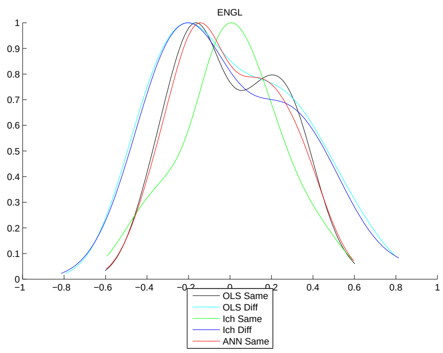
3.7 Figures

Figure 3.1: Subsample: Kernel Estimates of Density of Teacher Effects by Different Lagged Student Score Percentiles

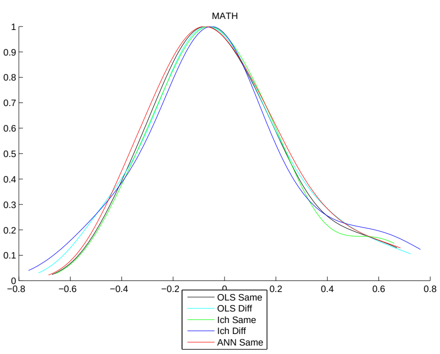
(a) Math: 10th Percentile



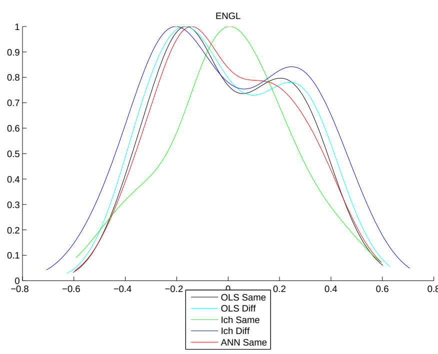
(b) English: 10th Percentile



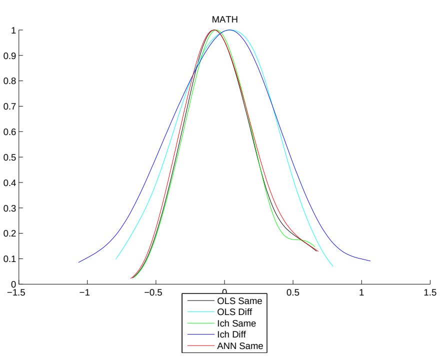
(c) Math: 50th Percentile



(d) English: 50th Percentile



(e) Math: 90th Percentile



(f) English: 90th Percentile

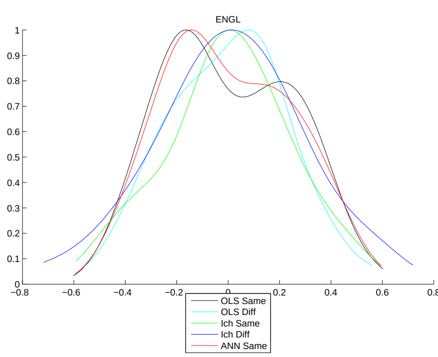


Figure 3.2: Kernel Estimates of the Density of Within Teacher Effects for 4 Teachers, English Subsample

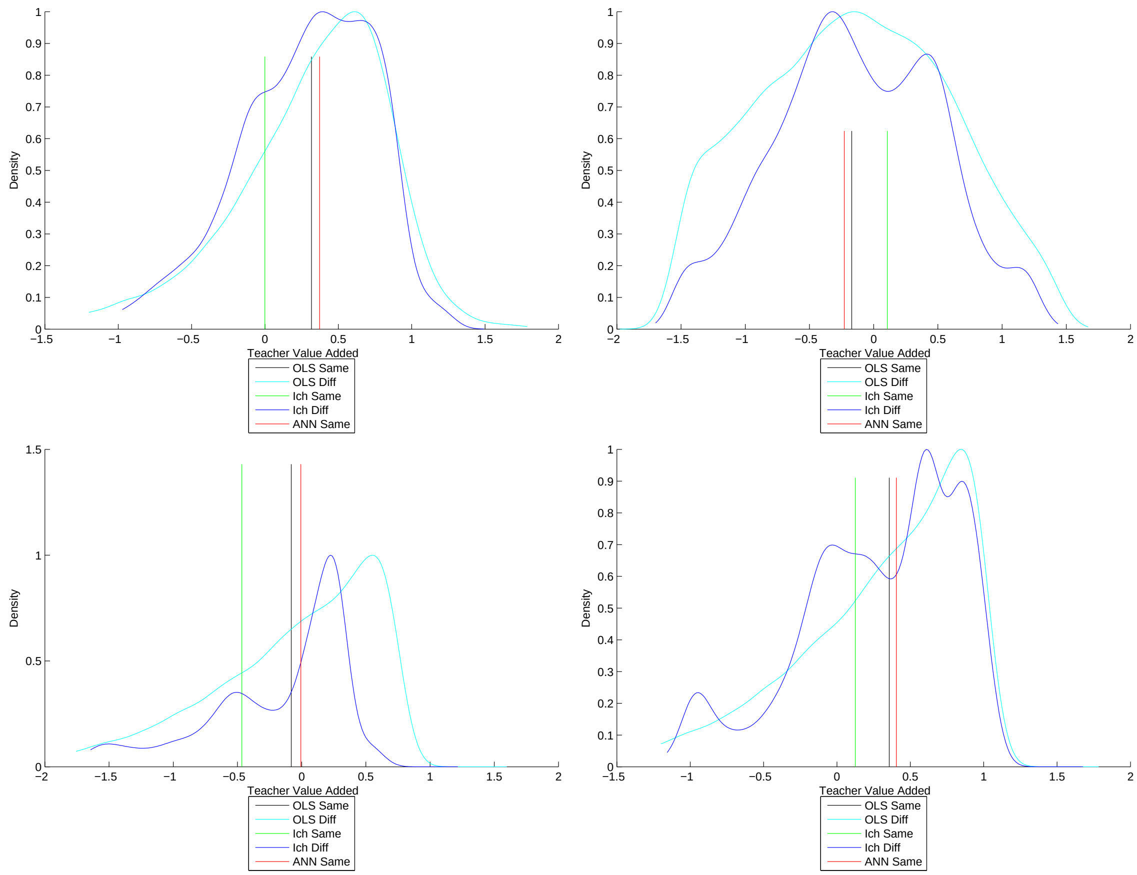
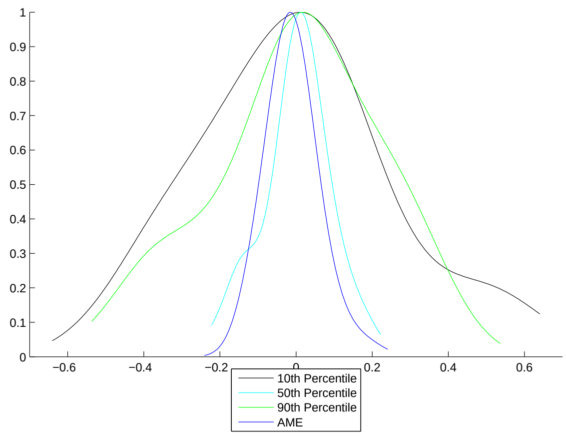
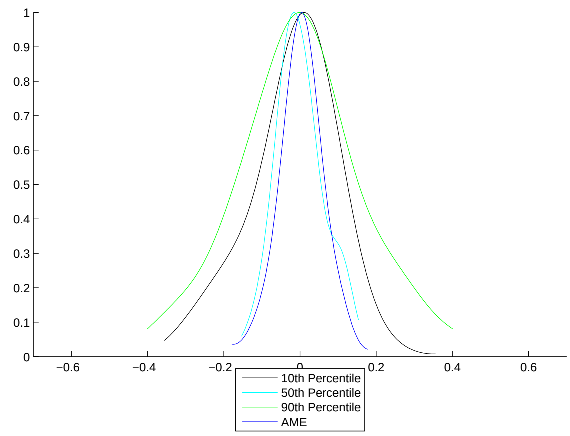


Figure 3.3: Difference in Rankings by Econometric Models

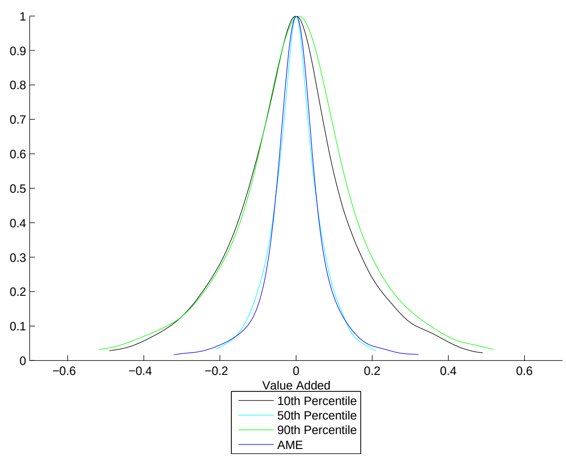
(a) Math Subsample



(b) English Subsample



(c) Math Full Sample



(d) English Full Sample

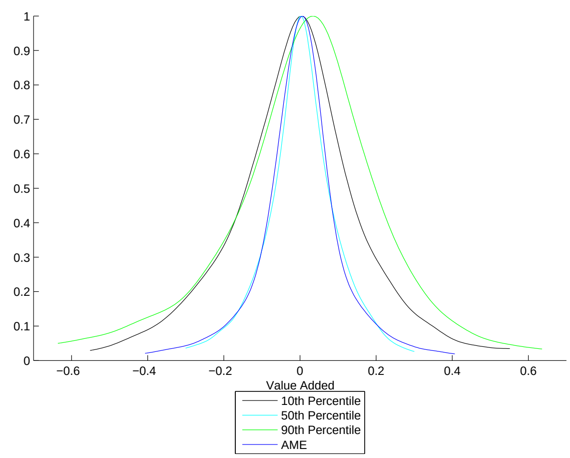
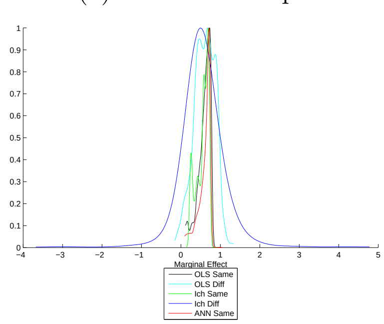


Figure 3.4: Distribution of Estimated Marginal Effects: Lagged Test Score

(a) Math Subsample



(b) English Subsample

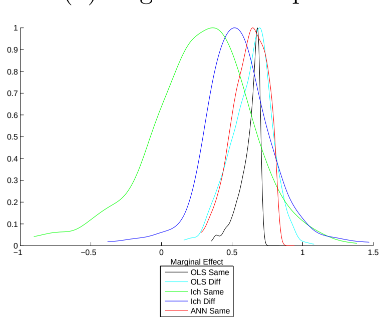
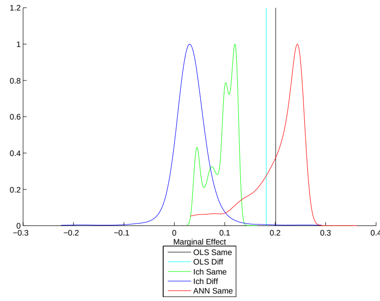


Figure 3.5: Distribution of Estimated Marginal Effects: Fraction Free Lunch

(a) Math Subsample



(b) English Subsample

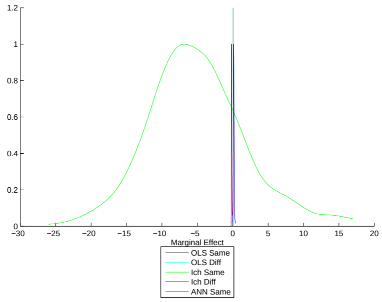
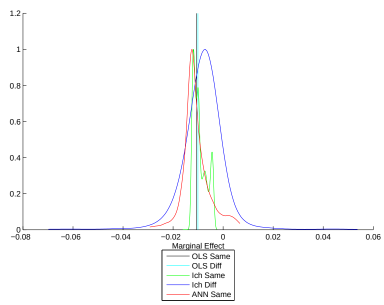


Figure 3.6: Distribution of Estimated Marginal Effects: Class Size

(a) Math Subsample



(b) English Subsample

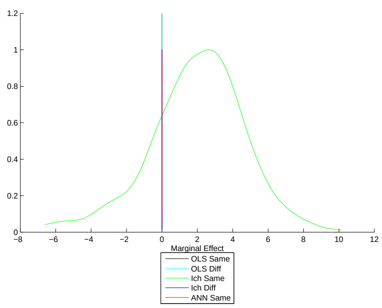
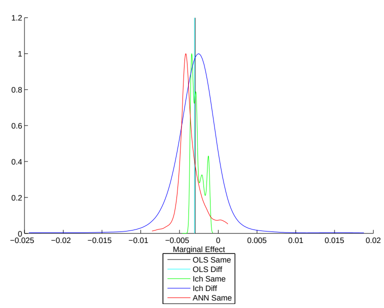


Figure 3.7: Distribution of Estimated Marginal Effects: Standard Deviation of Class Lagged Test Score

(a) Math Subsample



(b) English Subsample

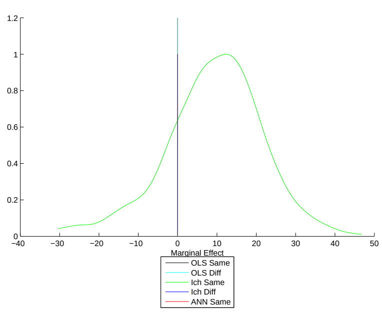


Figure 3.8: Distribution of Estimated Marginal Effects: Hispanic

(a) Math Subsample (b) English Subsample

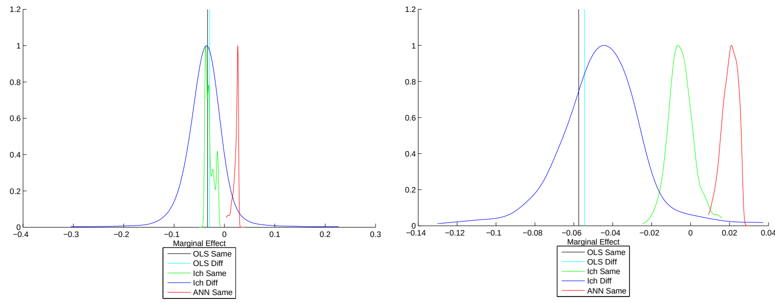


Figure 3.9: Distribution of Estimated Marginal Effects: Asian

(a) Math Subsample (b) English Subsample

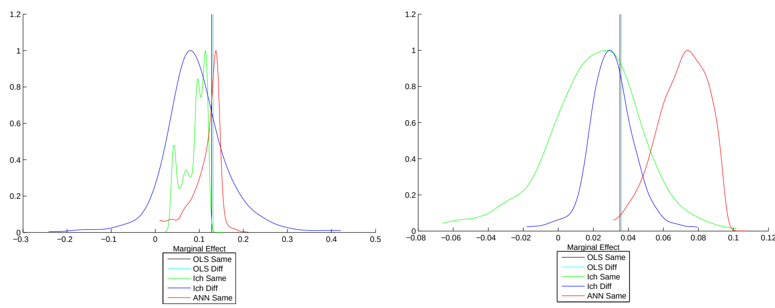


Figure 3.10: Distribution of Estimated Marginal Effects: Black

(a) Math Subsample (b) English Subsample

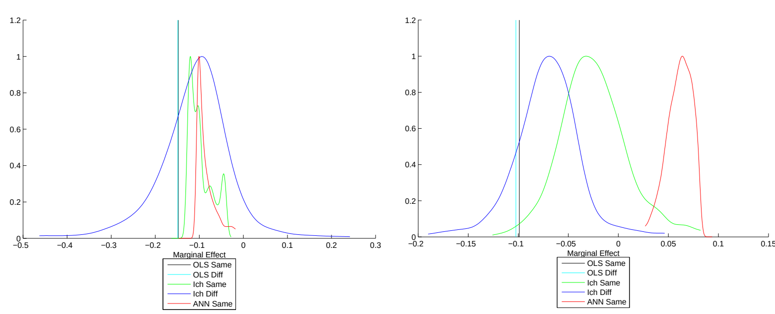


Figure 3.11: Distribution of Estimated Marginal Effects: Other Race

(a) Math Subsample (b) English Subsample

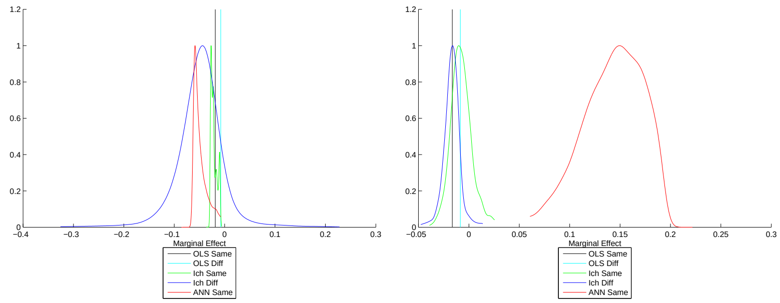


Figure 3.12: Distribution of Estimated Marginal Effects: Male

(a) Math Subsample (b) English Subsample

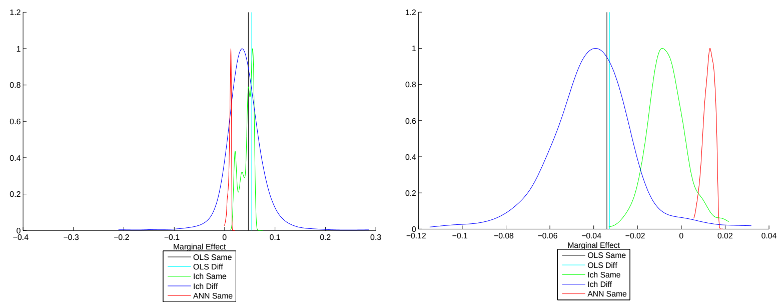


Figure 3.13: Distribution of Estimated Marginal Effects: Participation in the Gifted Program

(a) Math Subsample (b) English Subsample

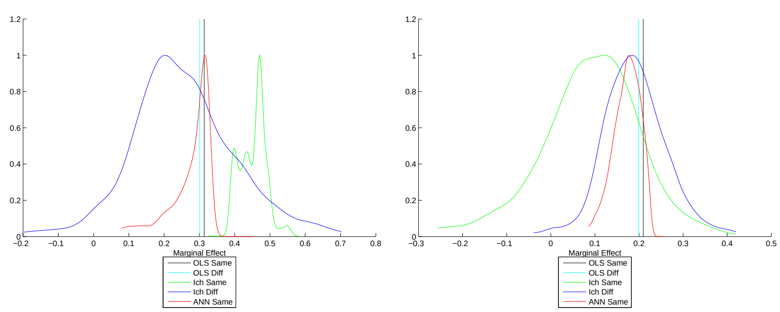


Figure 3.14: Distribution of Estimated Marginal Effects: On Free Lunch Program

(a) Math Subsample (b) English Subsample

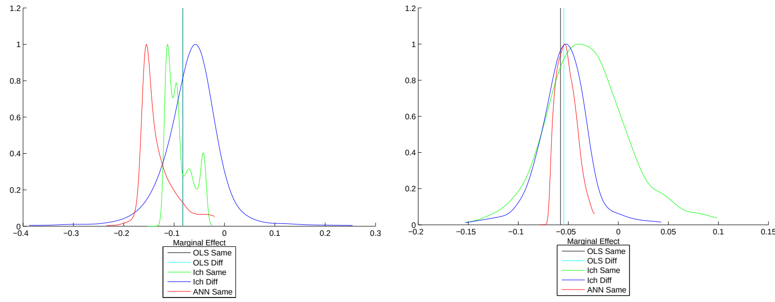


Figure 3.15: Distribution of Estimated Marginal Effects: Parents Finished High School

(a) Math Subsample (b) English Subsample

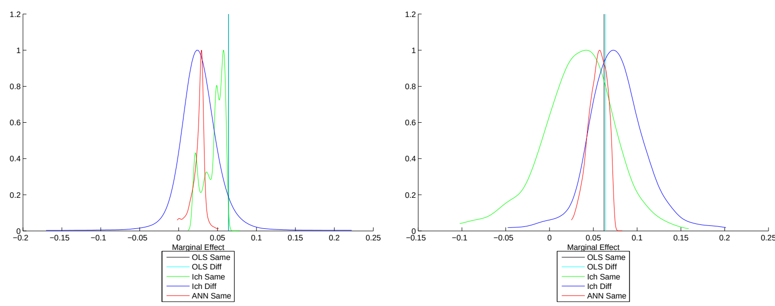
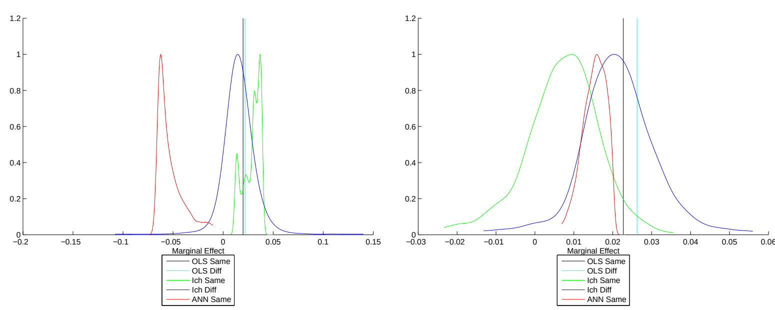


Figure 3.16: Distribution of Estimated Marginal Effects: Missing Data on Parents' Education

(a) Math Subsample (b) English Subsample



3.8 Tables

Table 3.1: Summary Statistics, Math High Tenure Subsample; Number of Students=11,484, Number of Teachers=56

	Mean	Std. Dev.	Between Std.	Within Std.	Min	Max
T_t	0.969	1.007	0.504	0.872	-2.049	3.514
T_{t-1}	0.976	0.982	0.464	0.867	-2.182	3.607
Frac Free Lunch	0.450	0.333	0.320	0.088	0	1
Class Size	31.121	3.229	1.414	2.234	16	60
Std(T_{t-1})	63.516	12.112	6.475	9.813	33.319	105.392
Hispanic	0.061	0.240	0.066	0.213	0	1
Asian	0.157	0.363	0.135	0.313	0	1
Black	0.381	0.486	0.282	0.386	0	1
Other Race	0.043	0.204	0.029	0.186	0	1
Gifted Prog.	0.609	0.488	0.217	0.433	0	1
Male	0.501	0.500	0.032	0.500	0	1
Free Lunch Prog.	0.450	0.497	0.321	0.366	0	1
Parents College	0.426	0.494	0.239	0.426	0	1
Missing Parents Ed.	0.150	0.357	0.135	0.312	0	1

Table 3.2: Summary Statistics, Math Full Sample; Number of Students=657,406, Number of Teachers=7,072

	Mean	Std. Dev.	Between Std.	Within Std.	Min	Max
T_t	0.090	0.997	0.554	0.809	-4.527	3.514
T_{t-1}	0.096	0.989	0.491	0.843	-4.655	3.607
Frac Free Lunch	0.784	0.264	0.219	0.103	0	1
Class Size	24.167	5.150	4.484	2.251	1	61
Std(T_{t-1})	61.000	14.419	9.290	10.086	0.000	163.342
Hispanic	0.089	0.284	0.166	0.183	0	1
Asian	0.046	0.209	0.097	0.113	0	1
Black	0.748	0.434	0.272	0.290	0	1
Other Race	0.029	0.167	0.053	0.100	0	1
Gifted Prog.	0.158	0.365	0.173	0.255	0	1
Male	0.490	0.500	0.052	0.500	0	1
Free Lunch Prog.	0.785	0.411	0.220	0.311	0	1
Parents College	0.141	0.348	0.164	0.264	0	1
Missing Parents Ed.	0.281	0.449	0.218	0.382	0	1

Table 3.3: Summary Statistics, English High Tenure Subsample; Number of Students=11,685, Number of Teachers=57

	Mean	Std. Dev.	Between Std.	Within Std.	Min	Max
T_t	1.136	0.921	0.488	0.785	-2.205	5.384
T_{t-1}	1.073	0.970	0.505	0.829	-2.753	5.408
Frac Free Lunch	0.459	0.337	0.325	0.087	0	1
Class Size	31.059	3.155	1.375	2.233	17	58
Std(T_{t-1})	42.862	9.050	4.968	7.264	24.639	77.633
Hispanic	0.061	0.239	0.066	0.211	0	1
Asian	0.154	0.361	0.135	0.310	0	1
Black	0.389	0.488	0.287	0.385	0	1
Other Race	0.044	0.204	0.029	0.187	0	1
Gifted Prog.	0.605	0.489	0.219	0.433	0	1
Male	0.502	0.500	0.031	0.500	0	1
Free Lunch Prog.	0.458	0.498	0.326	0.362	0	1
Parents College	0.420	0.494	0.241	0.423	0	1
Missing Parents Ed.	0.150	0.357	0.134	0.313	0	1

Table 3.4: Summary Statistics, English Full Sample; Number of Students=658,561, Number of Teachers=7,081

	Mean	Std. Dev.	Between Std.	Within Std.	Min	Max
T_t	0.090	0.959	0.557	0.767	-6.261	5.384
T_{t-1}	0.103	0.988	0.556	0.799	-6.132	5.408
Frac Free Lunch	0.784	0.264	0.219	0.103	0	1
Class Size	24.122	5.147	4.475	2.258	1	62
Std(T_{t-1})	41.276	10.080	6.623	6.899	0.000	140.007
Hispanic	0.089	0.285	0.166	0.183	0	1
Asian	0.046	0.209	0.097	0.113	0	1
Black	0.747	0.434	0.272	0.290	0	1
Other Race	0.029	0.167	0.053	0.100	0	1
Gifted Prog.	0.158	0.365	0.173	0.254	0	1
Male	0.490	0.500	0.052	0.500	0	1
Free Lunch Prog.	0.785	0.411	0.220	0.310	0	1
Parents College	0.141	0.348	0.164	0.264	0	1
Missing Parents Ed.	0.281	0.449	0.218	0.382	0	1

Table 3.5: Math OLS Regression Results, Additively Separable Teacher Effect Model (Baseline)

	High Tenure Subsample	Full, Grade 3	Full, Grade 4	Full, Grade 5
T_{t-1}	0.742***	0.718***	0.636***	0.838***
	(0.0133)	(0.00237)	(0.00195)	(0.00223)
T_{t-1}^2	-0.0380***	-0.0275***	-0.0249***	0.0113***
	(0.0110)	(0.00134)	(0.00116)	(0.00144)
T_{t-1}^3	-0.0162***	-0.0241***	-0.0165***	-0.0362***
	(0.00328)	(0.000623)	(0.000533)	(0.000641)
Frac Free Lunch	0.201***	-0.0323***	-0.00555	-0.0577***
	(0.0606)	(0.0101)	(0.00906)	(0.0101)
Class Size	-0.0105***	-0.00343***	-0.00445***	-0.00397***
	(0.00204)	(0.000707)	(0.000411)	(0.000419)
Std(T_{t-1})	-0.00298***	-0.00115***	0.000116	-0.00120***
	(0.000584)	(0.000110)	(0.000101)	(0.000115)
Black	-0.148***	-0.173***	-0.129***	-0.115***
	(0.0276)	(0.00806)	(0.00634)	(0.00661)
Asian	0.128***	0.172***	0.138***	0.176***
	(0.0194)	(0.00858)	(0.00689)	(0.00718)
Hispanic	-0.0336*	-0.0669***	-0.0527***	-0.0475***
	(0.0180)	(0.00645)	(0.00510)	(0.00533)
Other Race	-0.0186	0.0411***	0.0445***	0.0746***
	(0.0308)	(0.0101)	(0.00793)	(0.00837)
In Gifted Program	0.314***	0.411***	0.263***	0.381***
	(0.0157)	(0.00535)	(0.00389)	(0.00389)
Male	0.0471***	0.0462***	-0.0320***	-0.0101***
	(0.0119)	(0.00272)	(0.00218)	(0.00228)
Free Lunch Prog.	-0.0829***	-0.0442***	-0.0271***	-0.0258***
	(0.0168)	(0.00441)	(0.00349)	(0.00367)
Parents College	0.0641***	0.0641***	0.0429***	0.0534***
	(0.0151)	(0.00477)	(0.00377)	(0.00393)
Missing Parent's Educ.	0.0199	-0.00117	-0.00347	-0.000249
	(0.0190)	(0.00348)	(0.00281)	(0.00292)
AME(T_{t-1})	0.5741***	0.6353***	0.5760***	0.7470***
	(0.0080)	(0.0017)	(0.0014)	(0.0017)
Observations	11,484	199,557	221,118	236,731
R-squared	0.604	0.508	0.546	0.594
Number of Teachers	56	2,623	2,681	2,621

Table 3.6: English OLS Regression Results, Additively Separable Teacher Effect Model (Baseline)

	High Tenure Subsample	Full, Grade 3	Full, Grade 4	Full, Grade 5
T_{t-1}	0.694***	0.705***	0.716***	0.773***
	(0.0111)	(0.00188)	(0.00189)	(0.00168)
T_{t-1}^2	-0.00113	0.0136***	-0.0152***	0.0246***
	(0.00719)	(0.00127)	(0.000956)	(0.00125)
T_{t-1}^3	-0.0130***	-0.0161***	-0.0158***	-0.0212***
	(0.00167)	(0.000393)	(0.000370)	(0.000457)
Frac Free Lunch	0.0648	0.0945***	-0.115***	0.0695***
	(0.0522)	(0.00918)	(0.00875)	(0.00829)
Class Size	0.00122	-0.00684***	-0.00289***	0.00170***
	(0.00178)	(0.000645)	(0.000397)	(0.000343)
Std(T_{t-1})	-0.00207***	7.60e-05	0.000448***	-0.00166***
	(0.000677)	(0.000137)	(0.000148)	(0.000139)
Black	-0.0989***	-0.117***	-0.119***	-0.0853***
	(0.0236)	(0.00732)	(0.00614)	(0.00542)
Asian	0.0352**	0.00361	0.0663***	0.0494***
	(0.0165)	(0.00781)	(0.00667)	(0.00589)
Hispanic	-0.0575***	-0.0862***	-0.0515***	-0.0516***
	(0.0154)	(0.00587)	(0.00494)	(0.00438)
Other Race	-0.0165	-0.0493***	0.0125	-0.00184
	(0.0262)	(0.00922)	(0.00768)	(0.00688)
In Gifted Program	0.210***	0.263***	0.279***	0.178***
	(0.0135)	(0.00499)	(0.00375)	(0.00323)
Male	-0.0340***	-0.0496***	-0.0578***	-0.0361***
	(0.0102)	(0.00247)	(0.00211)	(0.00188)
Free Lunch Prog.	-0.0578***	-0.0567***	-0.0469***	-0.0313***
	(0.0144)	(0.00402)	(0.00339)	(0.00302)
Parents College	0.0623***	0.0602***	0.0554***	0.0385***
	(0.0129)	(0.00435)	(0.00365)	(0.00324)
Missing Parent's Educ.	0.0227	0.0107***	-0.00182	-0.00270
	(0.0161)	(0.00317)	(0.00272)	(0.00240)
AME(T_{t-1})	0.6101***	0.6566***	0.6727***	0.7275***
	(0.0073)	(0.0016)	(0.0014)	(0.0014)
Observations	11,685	200,586	221,535	236,440
R-squared	0.653	0.544	0.597	0.627
Number of Teachers	57	2,635	2,683	2,614

Table 3.7: Math Subsample: Correlation of Teacher Effects Between Models, at 10th, 50th, and 90th Percentiles of Lagged Test Score and Average Marginal Effect

		OLS Sep.	OLS Non. Sep.	Ich. Sep.	Ich. Non. Sep.	ANN Sep.
10th Perc.	OLS Sep.	1.0000	0.5203	0.9913	0.7660	0.9970
	OLS Non. Sep.	0.5203	1.0000	0.5529	0.3940	0.5222
	Ich. Sep.	0.9913	0.5529	1.0000	0.7769	0.9850
	Ich. Non. Sep.	0.7660	0.3940	0.7769	1.0000	0.7530
	ANN Sep.	0.9970	0.5222	0.9850	0.7530	1.0000
50th Perc.	OLS Sep.	1.0000	0.9684	0.9913	0.9163	0.9970
	OLS Non. Sep.	0.9684	1.0000	0.9568	0.9232	0.9587
	Ich. Sep.	0.9913	0.9568	1.0000	0.9150	0.9850
	Ich. Non. Sep.	0.9163	0.9232	0.9150	1.0000	0.9049
	ANN Sep.	0.9970	0.9587	0.9850	0.9049	1.0000
90th Perc.	OLS Sep.	1.0000	0.7536	0.9913	0.7171	0.9970
	OLS Non. Sep.	0.7536	1.0000	0.7359	0.8749	0.7511
	Ich. Sep.	0.9913	0.7359	1.0000	0.7118	0.9850
	Ich. Non. Sep.	0.7171	0.8749	0.7118	1.0000	0.7163
	ANN Sep.	0.9970	0.7511	0.9850	0.7163	1.0000
AME	OLS Sep.	1.0000	0.9350	0.9913	0.9619	0.9970
	OLS Non. Sep.	0.9350	1.0000	0.9380	0.8754	0.9305
	Ich. Sep.	0.9913	0.9380	1.0000	0.9596	0.9850
	Ich. Non. Sep.	0.9619	0.8754	0.9596	1.0000	0.9550
	ANN Sep.	0.9970	0.9305	0.9850	0.9550	1.0000

Table 3.8: English Subsample: Correlation of Teacher Effects Between Models, at 10th, 50th, and 90th Percentiles of Lagged Test Score and Average Marginal Effect

		OLS Sep.	OLS Non. Sep.	Ich. Sep.	Ich. Non. Sep.	ANN Sep.
10th Perc.	OLS Sep.	1.0000	0.8719	0.5468	0.8936	0.9754
	OLS Non. Sep.	0.8719	1.0000	0.4823	0.9300	0.8620
	Ich. Sep.	0.5468	0.4823	1.0000	0.5864	0.4354
	Ich. Non. Sep.	0.8936	0.9300	0.5864	1.0000	0.8484
	ANN Sep.	0.9754	0.8620	0.4354	0.8484	1.0000
50th Perc.	OLS Sep.	1.0000	0.9814	0.5468	0.9452	0.9754
	OLS Non. Sep.	0.9814	1.0000	0.5109	0.9528	0.9662
	Ich. Sep.	0.5468	0.5109	1.0000	0.5794	0.4354
	Ich. Non. Sep.	0.9452	0.9528	0.5794	1.0000	0.9063
	ANN Sep.	0.9754	0.9662	0.4354	0.9063	1.0000
90th Perc.	OLS Sep.	1.0000	0.8651	0.5468	0.7568	0.9754
	OLS Non. Sep.	0.8651	1.0000	0.5640	0.8900	0.8185
	Ich. Sep.	0.5468	0.5640	1.0000	0.5432	0.4354
	Ich. Non. Sep.	0.7568	0.8900	0.5432	1.0000	0.6778
	ANN Sep.	0.9754	0.8185	0.4354	0.6778	1.0000
AME	OLS Sep.	1.0000	0.9760	0.5468	0.9568	0.9754
	OLS Non. Sep.	0.9760	1.0000	0.5532	0.9677	0.9526
	Ich. Sep.	0.5468	0.5532	1.0000	0.6675	0.4354
	Ich. Non. Sep.	0.9568	0.9677	0.6675	1.0000	0.8976
	ANN Sep.	0.9754	0.9526	0.4354	0.8976	1.0000

Table 3.9: Full Sample: Correlation of Teacher Effects Between OLS Additively Separable and Non-Separable Models

	Math	English
10th Perc.	0.4815	0.5167
50th Perc.	0.9426	0.8967
90th Perc.	0.5718	0.3090
AME	0.8112	0.6476

Table 3.10: Math Subsample: Proportion of Teacher Effects Ranked in Quantiles by Different Percentiles of Lagged Test Score, OLS Separable and Ichimura Non-Separable vs. OLS Non-Separable

			Ich Non-Sep.				
			1st	2nd	3rd	4th	5th
AME	OLS Sep.	1st	0.818	0.091	0.083	0.000	0.000
		2nd	0.182	0.636	0.167	0.000	0.000
		3rd	0.000	0.273	0.750	0.000	0.000
		4th	0.000	0.000	0.000	0.636	0.364
		5th	0.000	0.000	0.000	0.364	0.636
	OLS Non Sep.	1st	0.909	0.091	0.000	0.000	0.000
		2nd	0.091	0.818	0.083	0.000	0.000
		3rd	0.000	0.091	0.833	0.091	0.000
		4th	0.000	0.000	0.083	0.727	0.182
		5th	0.000	0.000	0.000	0.182	0.818
10th Percentile	OLS Sep.	1st	0.727	0.273	0.000	0.000	0.000
		2nd	0.091	0.455	0.417	0.000	0.000
		3rd	0.182	0.091	0.333	0.364	0.091
		4th	0.000	0.000	0.250	0.364	0.364
		5th	0.000	0.182	0.000	0.273	0.545
	OLS Non Sep.	1st	0.545	0.273	0.000	0.091	0.091
		2nd	0.364	0.273	0.167	0.091	0.091
		3rd	0.091	0.273	0.500	0.091	0.091
		4th	0.000	0.182	0.333	0.364	0.091
		5th	0.000	0.000	0.000	0.364	0.636
90th Percentile	OLS Sep.	1st	0.455	0.273	0.250	0.000	0.000
		2nd	0.364	0.364	0.083	0.182	0.000
		3rd	0.182	0.273	0.250	0.273	0.091
		4th	0.000	0.091	0.333	0.273	0.273
		5th	0.000	0.000	0.083	0.273	0.636
	OLS Non Sep.	1st	0.727	0.273	0.000	0.000	0.000
		2nd	0.182	0.727	0.083	0.000	0.000
		3rd	0.000	0.000	0.833	0.182	0.000
		4th	0.091	0.000	0.083	0.636	0.182
		5th	0.000	0.000	0.000	0.182	0.818

Table 3.11: English Subsample: Proportion of Teacher Effects Ranked in Quantiles by Different Percentiles of Lagged Test Score, OLS Separable and Ichimura Non-Separable vs. OLS Non-Separable

			Ich Non-Sep.				
			1st	2nd	3rd	4th	5th
AME	OLS Sep.	1st	0.818	0.167	0.000	0.000	0.000
		2nd	0.182	0.667	0.182	0.000	0.000
		3rd	0.000	0.167	0.727	0.083	0.000
		4th	0.000	0.000	0.091	0.667	0.273
		5th	0.000	0.000	0.000	0.250	0.727
	OLS Non Sep.	1st	0.727	0.250	0.000	0.000	0.000
		2nd	0.273	0.500	0.273	0.000	0.000
		3rd	0.000	0.250	0.545	0.167	0.000
		4th	0.000	0.000	0.182	0.667	0.182
		5th	0.000	0.000	0.000	0.167	0.818
10th Percentile	OLS Sep.	1st	0.727	0.167	0.091	0.000	0.000
		2nd	0.182	0.667	0.182	0.000	0.000
		3rd	0.091	0.167	0.455	0.250	0.000
		4th	0.000	0.000	0.273	0.417	0.364
		5th	0.000	0.000	0.000	0.333	0.636
	OLS Non Sep.	1st	0.909	0.000	0.000	0.083	0.000
		2nd	0.091	0.750	0.182	0.000	0.000
		3rd	0.000	0.250	0.727	0.000	0.000
		4th	0.000	0.000	0.091	0.750	0.182
		5th	0.000	0.000	0.000	0.167	0.818
90th Percentile	OLS Sep.	1st	0.636	0.167	0.091	0.000	0.091
		2nd	0.273	0.500	0.273	0.000	0.000
		3rd	0.091	0.333	0.364	0.167	0.000
		4th	0.000	0.000	0.091	0.667	0.273
		5th	0.000	0.000	0.182	0.167	0.636
	OLS Non Sep.	1st	0.818	0.167	0.000	0.000	0.000
		2nd	0.182	0.583	0.273	0.000	0.000
		3rd	0.000	0.250	0.545	0.083	0.091
		4th	0.000	0.000	0.182	0.667	0.182
		5th	0.000	0.000	0.000	0.250	0.727

Table 3.12: Math Full Sample: Proportion of Teacher Effects Ranked in Quantiles by Different Percentiles of Lagged Test Score, OLS Separable vs. OLS Non-Separable

			OLS Non-Sep.				
			1st	2nd	3rd	4th	5th
AME	OLS Sep.	1st	0.820	0.130	0.024	0.015	0.011
		2nd	0.140	0.679	0.156	0.018	0.006
		3rd	0.017	0.151	0.652	0.162	0.017
		4th	0.008	0.026	0.138	0.679	0.149
		5th	0.015	0.013	0.030	0.126	0.817
10th Perc.	OLS Sep.	1st	0.654	0.200	0.066	0.042	0.038
		2nd	0.250	0.442	0.220	0.073	0.016
		3rd	0.064	0.247	0.404	0.232	0.054
		4th	0.021	0.086	0.241	0.435	0.216
		5th	0.010	0.026	0.069	0.219	0.676
90th Perc.	OLS Sep.	1st	0.648	0.210	0.088	0.036	0.018
		2nd	0.248	0.445	0.208	0.080	0.020
		3rd	0.048	0.216	0.396	0.275	0.066
		4th	0.020	0.076	0.215	0.420	0.269
		5th	0.036	0.054	0.095	0.189	0.627

Table 3.13: English Full Sample: Proportion of Teacher Effects Ranked in Quantiles by Different Percentiles of Lagged Test Score, OLS Separable vs. OLS Non-Separable

			OLS Non-Sep.				
			1st	2nd	3rd	4th	5th
AME	OLS Sep.	1st	0.733	0.180	0.044	0.026	0.017
		2nd	0.190	0.567	0.194	0.038	0.011
		3rd	0.041	0.181	0.566	0.189	0.022
		4th	0.015	0.049	0.153	0.601	0.183
		5th	0.021	0.023	0.043	0.146	0.767
10th Perc.	OLS Sep.	1st	0.620	0.189	0.073	0.065	0.052
		2nd	0.257	0.416	0.218	0.077	0.033
		3rd	0.080	0.262	0.373	0.229	0.055
		4th	0.030	0.092	0.238	0.406	0.234
		5th	0.012	0.041	0.098	0.222	0.626
90th Perc.	OLS Sep.	1st	0.542	0.256	0.112	0.056	0.033
		2nd	0.253	0.353	0.262	0.101	0.031
		3rd	0.077	0.198	0.327	0.322	0.075
		4th	0.047	0.101	0.198	0.350	0.305
		5th	0.081	0.091	0.100	0.171	0.556

Table 3.14: Hypothesis Testing Interaction Terms Between Lagged Student Test Score Cubic and Teacher Effect

	Full F-Test p-value			Proportion Rejecting Null for by-Teacher Joint Tests		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
Math	0.0000	0.0000	0.0000	0.0835	0.1146	0.6748
English	0.0000	0.0000	0.0000	0.2289	0.1249	0.0241

Table 3.15: Teacher Value Added, Within vs. Between Standard Deviations

	Subsample		Full Sample	
	Between	Within	Between	Within
Math	0.2642	0.7618	0.3609	0.8257
English	0.2455	0.6758	0.3366	0.8725

Bibliography

[1] Altonji, Joseph and Rosa L. Matzkin (2005) “Cross section and panel data estimators for nonseparable models with endogenous regressors,” *Econometrica* 73, 10531102.

[2] Ben-Porath, Yoram (1967) “The Production of Human Capital and the Lifecycle of Earnings,” *Journal of Political Economy*, 75, 352-365.

[3] Buddin, Richard. (2010) “How effective are Los Angeles elementary teachers and schools?” MPRA Paper No. 27366, posted 10. December 2010.

[4] Buddin, Richard (2011) “Measuring teacher effectiveness at improving student achievement in Los Angeles elementary schools.” unpublished paper, retrieved August 11, 1011 from `http://www.gse.uci.edu/person/buddin_r/measuring_teacher_effectiveness_march_2011.pdf`

[5] Butrymowicz, S. and S. Garland (2012). “New York City Teacher Ratings: How Its Value-Added Model Compares To Other Districts.” *The Hechinger Report,* posted 3/2/12. Retrieved 5/11/2012, from: `http://www.huffingtonpost.com/2012/03/02/new-york-city-teacher-rat_n_1316755.html`.

[6] Chen, Xiaohong (2007). “Large Sample Sieve Estimatino of Semi-Nonparametric Models.” chapter in J.J. Heckman & E.E. Leamer (ed.), 2007. *Handbook of Econometrics*, Elsevier, edition 1, volume 6, number 6b.

[7] Chen,X., J. Racine, and N. Swanson (2001). “Semiparametric ARX Neural-Network

Models with an Application to Forecasting Inflation." *IEEE Transactions on Neural Networks*,12(4),674-683.

[8] Chen,X. and H. White (1999). "Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators." *IEEE Transactions on Information Theory*,45(2),682-691.

[9] Corcoran, Sean. P. (2010). "Can teachers be evaluated by their students test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice." Providence, Rhode Island: Annenberg Institute for School Reform at Brown University. Retrieved August 11, 2011, from: `http://www.annenberginstitute.org/pdf/valueAddedReport.pdf`.

[10] Heckman, J. and Masterov, D. (2007). "The Productivity Argument for Investing in Small Children." *Applied Economic Perspectives and Policy*, 29(3).

[11] Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models", *Journal of Econometrics*, 58, 71-120.

[12] Kane, Thomas J., and Douglas 0. Staiger, (2008) "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," National Bureau of Economic Research Working Paper No. 14607.

[13] Li, Qi and Jeffrey Racine (2007). *Nonparametric Econometrics*. Princeton University Press.

[14] Rockoff, Jonah E., and Cecilia Speroni. (2010). "Subjective and objective evaluations of teacher effectiveness." *American Economic Review* 100, 261-66.

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

[16] Santos, F. and W. Hu (2012). "A Last-Minute Deal on Teacher Evaluations." 2/16/12. Retrieved 5/11/2012 from: `http://www.nytimes.com/schoolbook/2012/02/16/as -deadline-nears-a-compromise-on-teacher-evaluations`.

[17] Todd, Petra, and Kenneth Wolpin (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 1 13(485): F3-F33.