Essays on Education Policy

by

Dania V. Francis

Public Policy Studies
Duke University

Date: ____________________
Approved:

____
William A. Darity, Jr., Supervisor

____
Charles T. Clotfelter

____
Jacob L. Vigdor

____
Anna Gassman-Pines

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Public Policy Studies
in the Graduate School of Duke University
2013

Abstract

Essays on Education Policy

by

Dania V. Francis

Public Policy Studies
Duke University

Date: ___________________________
Approved:

William A. Darity, Jr., Supervisor

Charles T. Clotfelter

Jacob L. Vigdor

Anna Gassman-Pines

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Public Policy Studies
in the Graduate School of Duke University
2013

Abstract

This dissertation consists of three essays on the topic of education policy. In the first essay, I evaluate the impacts of a teacher quality equity law that was enacted in California in the fall of 2006 prohibiting superintendents from transferring a teacher into a school in the bottom three performance deciles of the state's academic performance index if the principal refuses the transfer. The primary mechanism through which the policy should affect student outcomes is through the mix of the quality of teachers in the school. Using publicly available statewide administrative education data, and two quasi-experimental methodologies, I assess whether the policy had an effect on the district-wide distribution of teachers with varying levels of experience, education and licensure and on student academic performance. I extend the analysis by examining whether the policy has differential effects on subgroups of schools classified as having high-poverty or high-minority student populations. I find that, as a result of the teacher quality equity law, low-performing schools experienced a relative increase in fully-credentialed teachers and more highly educated teachers, but that did not necessarily translate to an increase in academic performance. I also find evidence that the dimension along which the policy was most effective was in improving teacher pre-service qualifications in schools with high minority student populations.

In the second essay, I estimate racial, ethnic, gender and socioeconomic differences in teacher reports of student absenteeism and tardiness while controlling for

administrative records of actual absences. Subjective perceptions that teachers form about students' classroom behaviors matter for student academic outcomes. Given this potential impact, it is important to identify any biases in these perceptions that would disadvantage subgroups of students. I use longitudinal data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 in conjunction with longitudinal, student-level data from the North Carolina Education Data Research Center to employ a variation of a two sample instrumental variables approach in which I instrument for actual eighth grade absences with simulated measures of eight grade absences. I find consistent evidence that teacher reports of the attendance of low-income students are negatively biased and that math teacher reports of male attendance are positively biased. There is mixed evidence with regard to student race and ethnicity.

The third essay is a co-authored effort in which we employ a quasi-experimental estimation strategy to examine the effects of state-level job losses on fourth- and eighth-grade test scores, using federal Mass Layoff Statistics and 1996-2009 National Assessment of Educational Progress data. Results indicate that job losses decrease scores. Effects are larger for eighth than fourth graders and for math than reading assessments, and are robust to specification checks. Job losses to 1% of a state's working-age population lead to a .076 standard deviation decrease in the state's eighth-grade math scores. This result is an order of magnitude larger than those found in previous studies that have compared students whose parents lose employment to otherwise similar students, suggesting that downturns affect all students, not just students who experience parental job loss. Our findings have important implications for accountability schemes: we calculate that a state experiencing one-year job losses to 2% of its workers (a magnitude observed in seven states) likely sees a 16% increase in the share of its schools failing to make Adequate Yearly Progress under No Child Left Behind.

In loving memory of Dr. Enid V. Blaylock

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Symbols

θ_i	A measure of the quality of teacher i
ϕ_j	A measure of the quality of school j

Abbreviations

ACS	American Community Survey
API	Academic Performance Index
AYP	Adequate Yearly Progress
BLS	Bureau of Labor Statistics
D-in-D	Difference-in-Differences
ECLS-K	Early Childhood Longitudinal Study - Kindergarten Sample 1998
ELL	English Language Learner
FRL	Free or Reduced Priced Lunch
LAUSD	Los Angeles Unified School District
NAEP	National Assessment of Educational Progress
NCERDC	North Carolina Education Research Data Center
NCLB	No Child Left Behind
RD	Regression Discontinuity
REGS	Race, Ethnicity, Gender and Socioeconomic Status
STEM	Science, Technology, Engineering and Math

TIC	Total Initial Claimants
YRBS	Youth Risk Behavior Study

Acknowledgements

I would like to thank the following funding agencies for providing support for this research: The William T. Grant Foundation, The Smith Richardson Foundation, The Spencer Foundation, The Kenan Institute for Ethics and The Society of Duke Fellows. I would also like to thank my committee members for their guidance, input and advice: William A. Darity, Jr. (Chair), Charles T. Clotfelter, Anna Gassman-Pines and Jacob L. Vigdor. I'd like to acknowledge Elizabeth O. Ananat, Anna Gassman-Pines, and Christina Gibson-Davis as co-authors of "Children Left Behind: The Effects of Statewide Job Loss on Student Achievement."

Thank you to Clara Mushkin and Kara Bonneau for their support in accessing and using data from the North Carolina Education Data Research Center (NCERDC). Thank you to the North Carolina Department of Public Instruction for making their data available to the NCERDC. Also, thank you to the National Center for Education Statistics for access to and support using the Early Childhood Childhood Longitudinal Study - Kindergarten Sample of 1998.

To my mentors, thank you for your guidance and preparation. Thank you to William A. Darity, Jr. for being an excellent mentor and teacher, and for helping to guide me back to graduate school in the first place. I'd also like to thank Elizabeth Ananat and Anna Gassman-Pines for their advice and expertise in successfully navigating the graduate school experience. I recognize that there are many demands on the time of an Assistant Professor, and that you two chose to spend some of it dedicating to helping me be successful means the world to me. Thank you to Jake Vigdor for clearly and plainly laying out the expectations and responsibilities of a career in the academy. Thank you to Helen F. Ladd for being a role model, and for helping me to hone my skills as a researcher through our thoughtful discussions.

I'd like to thank the following staff members of the Sanford School of Public Policy: Marylu Knight and Patrick Morris for helping me navigate funding applications, deadlines, and all the ins and outs of the university administration. Anne Fletcher, Astrid Gatling, and the rest of the Sanford IT department for providing data security and support. They made meeting the stringent requirements of obtaining restricted-use data seem like a walk in the park - though I know better.

To Sarah Fuller and Maeve Gearing. We made a great cohort. Thank you both for sharing both the ups and downs of graduate school with me.

Thank you to participants in the Sanford Seminar Series, the Sanford Graduate Research Workshop, and participants in the research conclaves of the Research Network on Racial and Ethnic Inequality for providing input and insight on earlier drafts of these essays. Thank you to the Social Science Research Institute for computing and programming resources and support.

Finally, I'd like to thank my family. To my husband Kato Francis, thank you for your full emotional and financial support. Your patience knows no bounds. To my parents, Bryan and Dellis Frank, thank you for providing me a solid foundation, and continuing to support my dreams without fail. To my brother and sister, Jannon and Camia Frank, thank you for your love and support. Knowing that when we're together, we will share belly laughs and great memories makes it easier to get through challenging days. To my Grandfather, Lorenzo V. Blaylock, being able to exchange letters with you throughout my graduate experience was great for my sanity. Our conversations always served to refresh and inspire me. To my grandmother, Ethel Gahee, and my father, Greg Gahee, thank you for teaching me that the only limits are those that I place on myself. To Dr. Carla Shedd, thank you for being my academic and personal role model, for letting me bounce my ideas off of you, for listening to me and commiserating with me, and for being my sister and friend. To my mother in law, Linda Francis, thank you for your advice and prayers. And to the entire Francis family, thank you for taking me into your family, and for allowing me time and energy to complete this work by watching after my daughter. Knowing she was in your loving hands allowed me peace of mind. And last, but not least, thank you to my daughter Sloan for sleeping through the night so that Mommie could write well into the wee hours of the morning.

1

Introduction

The current economic climate has led to shrinking budgets for federal, state, and local governments and an increased scrutiny of publicly provided programs including primary and secondary education. This increased attention serves to amplify long standing debates about policies that address the efficiency and efficacy of public schools, especially surrounding the issues of teacher performance and the racial and socioeconomic achievement gap. In addition, the requirements of the No Child Left Behind Act of 2001 have led to increased educational data collection - as well as increased transparency and access to that data - which allows researchers to conduct data-driven analyses of education policies at an unprecedented level. This dissertation - a compilation of three essays - seeks to follow in this tradition of data-driven analysis by providing a rigorous examination of three issues related to education policy.

In the first essay, I evaluate the impacts of a teacher quality equity law that was enacted in California in the fall of 2006 prohibiting superintendents from transferring a teacher into a school in the bottom three performance deciles of the state's academic performance index if the principal refuses the transfer. The primary mechanism through which the policy should affect student outcomes is through the mix of the quality of teachers in the school. I find evidence of heterogeneous treatment effects in which low-performing schools with high concentrations of poor and minority students benefit more from the policy than low-performing schools with high concentrations of white or more affluent students. Thus, a policy initially geared toward low-performing schools may also have the unintended consequence of improving teacher quality in schools with high poor and minority student populations.

The second essay is concerned with a policy question that has the potential to impact the racial and socioeconomic academic achievement gap. I estimate racial, ethnic, gender and socioeconomic differences in teacher reports of student absenteeism and tardiness while controlling for administrative records of actual absences. Subjective perceptions that teachers form about students' classroom behaviors, such as attendance, matter for student academic outcomes as well as later in life outcomes. If certain subgroups of students, like low-income students for example, are consistently perceived as exhibiting worse behavior, they may be disadvantaged in the classroom relative to other students - teachers may be less likely to recommend them for honors or enriched activities because of their behavior regardless of their academic potential. Given this potential impact, it is important, from a policy perspective, to identify whether these perception differences are due to differences in actual behavior on the part of students, or due to biases on the part of teachers. If low-income students, for example, simply behave worse on average, then a proposed policy solution could be a behavioral intervention aimed at improving their behavior. If, however, teachers are simply biased against low-income students, then a proposed policy solution could be sensitivity and awareness training for teachers to address biases. This essay addresses this important policy question.

Finally, the third essay reveals ways in which accountability policies like the No Child Left Behind Act of 2001 can have unintended consequences in times of economic distress. We estimate the effects of job losses on student academic performance and find strong evidence in the support of spillover effects - job losses affect all students, not just students who experience parental job loss. As it pertains to accountability policies, we calculate that a state experiencing one-year job losses to 2% of its workers likely sees a 16% increase in the share of its schools failing to make Adequate Yearly Progress under No Child Left Behind. Thus, schools are potentially sanctioned under accountability policies as a result of circumstances beyond their control. In order to adjust for this potential policy spillover, accountability policies should be flexible to economic shocks that could effect student academic performance.

Taken together, these three essays rely on the tools of the economic discipline to illuminate issues in education policy using rigorous data analysis. A consistent theme that results from the analysis is that nuances in the way policies are designed can significantly impact the effectiveness of the policies themselves. It is difficult to predict, *a priori* all of the ways in which policy implementation and policy effects may differ from the initial policy goals, but a key lesson to take from these three essays is that careful analysis of policies once they have been implemented may help improve their effectiveness and efficiency.

2

The Effects of Giving School Principals More Agency: Evidence on Policy Changes that Facilitate Discretion in Hiring Decisions

2.1 Introduction

Highly effective teachers can significantly improve student performance on standardized achievement tests (Aaronson et al., 2007; Clotfelter et al., 2007a; Hanushek et al., 2005; Lai et al., 2011; Nye et al., 2004; Rockoff, 2004).[1] They also can have benefits for students later in life, such as increased odds of college attendance and increased earning potential (Chetty et al., 2011). However, highly effective teachers are distributed unevenly across schools in districts that are segregated by race and/or income. Disadvantaged schools - those with higher concentrations of poor or minority students or low-performing students - tend to have teachers with lower pre-service qualifications (Clotfelter et al., 2007a, 2005; Lankford et al., 2002; Murnane and Steele, 2007).

[1] Most of these studies measure teacher effectiveness by estimating an education production function that includes teacher fixed effects. In these specifications, effective teachers produce larger test score gains on average for their students.

The uneven distribution of highly effective teachers is caused by forces within teacher labor markets. On the supply side, teachers typically prefer to work in schools that have fewer low-income and minority students.(Borman and Dowling, 2008; Boyd et al., 2005, 2010; Feng, 2009; Guarino et al., 2006; Hanushek et al., 2004). Scafidi et al. (2007) demonstrate that many teachers prefer to avoid teaching in schools with predominately black student populations regardless of the socioeconomic status or academic performance of the students. These race and income preference produce an excess supply of willing teacher candidates for positions in advantaged schools while disadvantaged schools - which are also plagued with higher teacher turnover - may face shortages of high quality teachers (Goldhaber, 2008; Jacob, 2007).

On the demand side, teacher transfer policies within school districts may contribute to the uneven distribution of high quality teachers (Boyd et al., 2011). Districts that allow preferential transferring to teachers with more seniority, or districts that allow superintendents to control teacher transfer decisions tend to exacerbate differences in teacher quality across advantaged and disadvantaged schools (Clotfelter et al., 2007a; Lankford et al., 2002).

This paper seeks to evaluate the impact of a teacher quality equity law that was enacted in California in the fall of 2006 in order to shed light on one mechanism within teacher labor markets that may affect the distribution of high quality teachers. California's Teacher Quality Equity Law (Senate Bill 1655) prohibits superintendents from transferring a teacher into a school in the bottom three performance deciles of the state's academic performance index (API) if the principal refuses the transfer. The primary mechanism through which the policy should affect student outcomes is through the mix of the quality of teachers in the school. While there have been many recent studies of the effects of teacher quality on student academic achievement, less is known about policies that are effective in reducing disparities in teacher quality between advantaged and disadvantaged schools. Paying higher salaries or offering salary bonuses to teachers in disadvantaged schools may decrease turnover in those schools (Clotfelter et al., 2008); however, the salary differentials that are required to even out teacher quality are often too large to be feasibly implemented. Clotfelter et al. (2010) estimate that salaries may have to be up to 50% higher in disadvantaged schools to attract teachers with strong qualifications.

Using publicly available statewide administrative education data, I assess whether California's teacher transfer policy has an effect on the district-wide distribution of teachers with varying levels of experience, education and licensure and on student academic performance.[2] I extend the analysis by examining whether the policy has differential effects on subgroups of schools classified as having high-poverty or high-minority student populations. According to the previously cited literature, many teachers have preferences over schools along the dimensions of racial and socioeconomic composition in addition to academic performance. Thus, even though the policy targets schools that are low-performing, there may be differing effects of the law by racial and socioeconomic composition since that composition tends to affect a school's ability to attract new and experienced teachers.

Results indicate that as a result of the transfer policy, low-performing schools experience a relative increase in fully-credentialed teachers and more highly educated teachers. Those increases may translate into improved academic performance for

[2] It is important to note the difference between teacher qualifications (credentialing, education, experience, etc.) and teacher effectiveness - the ability of a teacher to improve student academic performance. Teacher effectiveness is a more ideal measure to study than teacher qualifications for two reasons. First, strong qualifications do not necessarily translate into effective teaching (Aaronson et al., 2007; Harris and Sass, 2011; Jacob, 2007). Thus, measuring whether or not the policy leads to the hiring of more highly-credentialed teachers in low-performing schools would not necessarily indicate whether those schools receive more effective teachers. Second, while teacher qualifications are highly transferrable across schools, teacher effectiveness may not be transferrable. A teacher who is highly effective in a school where the majority of students are from middle-class backgrounds, for example, may not be nearly as effective in a school where the majority of students are from low-income households (Jackson, 2010). Unfortunately, due to data limitations, I am not able to measure teacher effectiveness. The inability to measure teacher effectiveness as opposed to teacher qualifications is a limitation of this study.

elementary school students, but not necessarily for middle school and high school students. Treatment group schools may also experience a decrease in pupil/teacher ratios in middle and high schools, but not in elementary schools. Finally, I find evidence that schools with high poor and minority student populations experience larger gains in teacher qualifications as a result of the law, but they are also less able to attract new teachers into their schools, resulting in smaller improvements in pupil/teacher ratios.

2.2 Conceptual Framework

The distribution of teachers that results from the mechanisms of teacher labor markets may be inefficient as well as inequitable due to imperfect information regarding the quality of individual teachers. Although there is convincing evidence that credentials such as experience, education, licensure and teacher test scores can have positive effects on student academic achievement (Clotfelter et al., 2007b; Goldhaber, 2008; Murnane and Steele, 2007), there is also evidence that these characteristics only account for a very small fraction of what makes an effective teacher (Aaronson et al., 2007; Harris and Sass, 2011; Jacob, 2007). Thus, principals may face uncertainty about the true quality or effectiveness of both new hire or transfer applicants. Some studies have found that principals are often able to hire effective teachers - where effectiveness is measured using value-added - without actually having access to those teachers' value-added measures (see for example Boyd et al. (2011) and Dobbie (2011)). These studies, however, are based on data from large urban school districts, and it is unclear whether principals in smaller or less densely populated areas would be equally as successful in hiring effective teachers based solely on information available to them at the time of the hiring decision.

This potential uncertainty about teacher quality presents a challenge because it affects the ability of hiring managers to identify the best candidates for open positions among both transfer applicants and new-hire applicants (Staiger and Rockoff, 2010). Furthermore, knowing that many teachers prefer to avoid teaching in disadvantaged schools, principals of disadvantaged schools may have reason to suspect that transfer applicants who apply to open positions in their schools may less effective teachers on average, since high-quality applicants would most likely apply to transfer to higher-performing schools. A rational, utility maximizing principal who gains from the quality of teacher hired would have an incentive to turn away a transfer applicant if the expected effectiveness of the transfer applicant is lower than the expected effectiveness of a novice teacher. Having to rely on the expected effectiveness of teachers in a pool of transfer applicants due to imperfect information may lead to inefficiency in the teacher transfer market in the sense that there are potential transfers that would benefit both the principal and the transferring teacher that do not take place.

2.2.1 A Basic Model of Adverse Selection in Teacher Transfer Markets

Consider a basic model with teachers $i = 1, ..., N$ who are currently employed in a school district and exist along a continuum of quality θ_i where each individual teacher's quality is only observed by the teacher and the teacher's current school. Assume θ_i is distributed among teachers according to a probability distribution function $F(.)$. We can imagine that θ represents the random component of teacher quality that is not accounted for by observable characteristics such as tenure, credentialing and education. Teachers with high values of θ are very productive at increasing student test scores while teachers with low values of θ are not.

Within the district there is also a continuum of schools $j = 1, ..., J$ with a quality measure ϕ_j which is common knowledge. In this case, ϕ_j represents factors that make a school attractive or unattractive to potential teachers, such as the socio-demographic composition and academic performance of the students and the condition of the facilities. Schools with high values of ϕ attract more teachers than schools with low values of ϕ.

Each school has a principal who is in charge of making hiring decisions for the school. Following Staiger and Rockoff (2010) I assume the principal's hiring objective is to maximize the average teacher effectiveness in his school. To that end, he values teachers with high values of θ. Thus the principal makes hiring decisions such that he maximizes the objective function $u(\theta^j)$ where θ^j is a vector containing the quality measures of all the teachers in school j. I assume $u(.)$ is increasing in teacher quality and exhibits decreasing marginal returns.

Teachers of each possible quality type, θ_i, have a reservation level of school quality to which they will consider transferring, represented by $r(\theta_i)$, which is assumed to be increasing in θ_i. This reflects the assumption that higher quality teachers generally have better outside options.

It is useful to think of the timing of the teacher transfer market is as follows:

1. Principals announce open positions

2. Potential transfers apply to schools that have a quality level above their reservation level.

3. Principals observe transfer applicants and decide whether to hire a transfer applicant or to reject all transfers and post the position on the new hire teacher market. On the new hire teacher market, a principal can expect to hire a new teacher with expected type $\theta_{NEW} = \alpha E(\theta)$ where $\alpha \in (0, 1)$ is a parameter that measures the average expected difference in quality between new hire teachers and current teachers. This reflects the evidence that new teachers have a learning curve in their first two or three years of teaching that make them less effective on average at producing higher test scores (Clotfelter et al.,

2007b; Goldhaber, 2008; Murnane and Steele, 2007).[3]

Given these assumptions, equilibrium in the teacher transfer market can be characterized in three steps:

1. There is a pool of transfers $T(\phi_j) = \{\theta | r(\theta) \leq \phi_j\}$ who are willing to work at each school of quality ϕ_j.

2. Given this pool of transfers, a principal's expected return to hiring a transfer is

$$\mu = E[\theta | \theta \in T(\phi_j)] = E[\theta | r(\theta) \leq \phi_j] = E[\theta | \theta \leq r^{-1}(\phi_j)]$$

3. A principal will hire a transfer if $\mu \geq \theta_{NEW} = \alpha E(\theta)$.

Equilibrium in the transfer market is a threshold value of school quality $\phi_j^* = f(\alpha)$ such that all schools with quality of at least ϕ_j^* will hire a transfer applicant and all schools with quality lower than ϕ_j^* will reject any transfers and post their opening on the new hire market. Figure 2.1 demonstrates this equilibrium condition graphically. School quality is represented on the x-axis and teacher quality is represented on the y-axis. The dashed line represents the reservation school quality for each teacher type. The line labeled θ_{NEW} represents the expected teacher quality from hiring a new teacher, while the line labeled $\mu(\phi_j)$ represents the expected teacher quality from hiring a teacher from the transfer pool available to a school with quality ϕ_j. Thus any principal in a school with quality to the right of ϕ_j^* will hire a transfer applicant, while any teacher to the left of ϕ_j^* will take her chances on the new hire teacher market.

Figure 2.2 reproduces the equilibrium depicted in Figure 2.1 in order to demonstrate the potential inefficiency of the adverse selection in the teacher transfer market

[3] θ_{NEW} can also reflect the probability that a school is not successful at hiring a teacher on the new hire market.

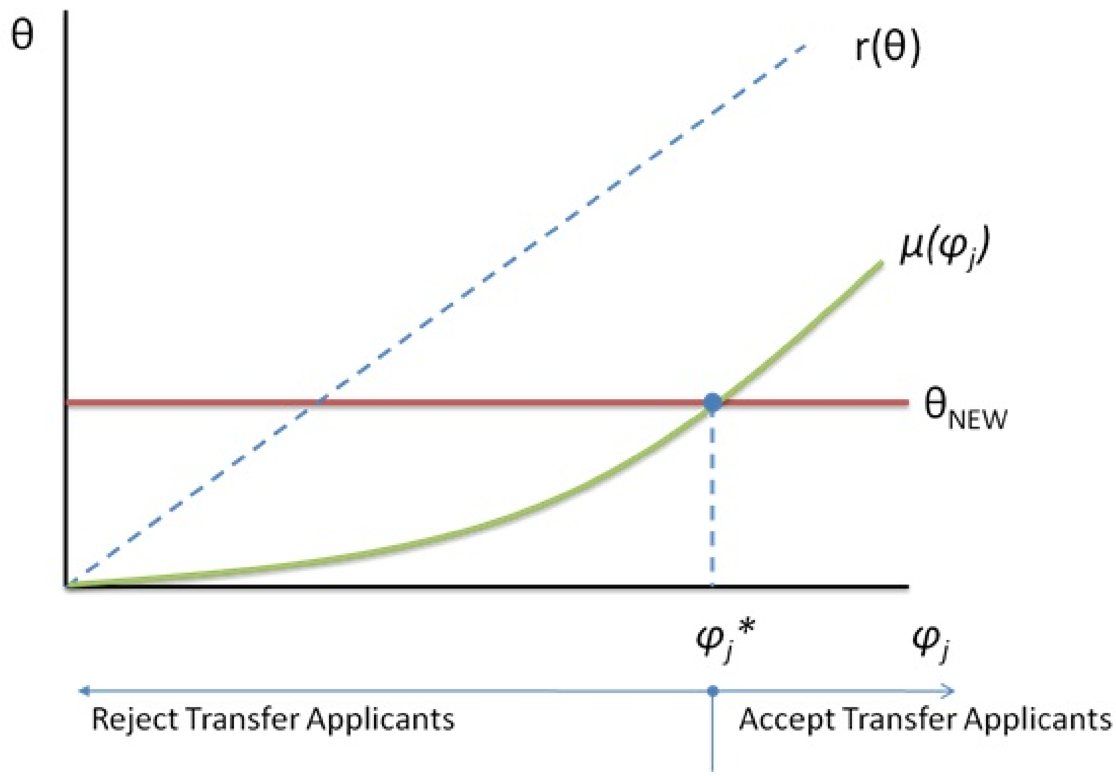

FIGURE 2.1: Equilibrium in the Teacher Transfer Market.

that is brought about by the informational asymmetry of the hiring principal not being able to observe the quality of prospective applicants. In this figure, any school with quality level between ϕ^{NEW} and ϕ_j^* would be willing to hire an applicant from their willing transfer pool if they observed the applicant's quality because there are applicants with quality above the expected new hire quality who are willing to apply to those schools. However, since the principals of those schools cannot observe an individual teacher's quality, they ascribe the average quality of the transfer pool to each transfer applicant and forgo mutually beneficial transfers.[4] The shaded region represents the deadweight loss due to this inefficiency.

[4] We can think of this behavior as akin to engaging in a form of statistical discrimination, in which a principal, lacking accurate information on a teacher's true quality, ascribes the average quality of all transfer applicants to the particular transfer applicant in question. See Arrow (1971) and Phelps (1972).

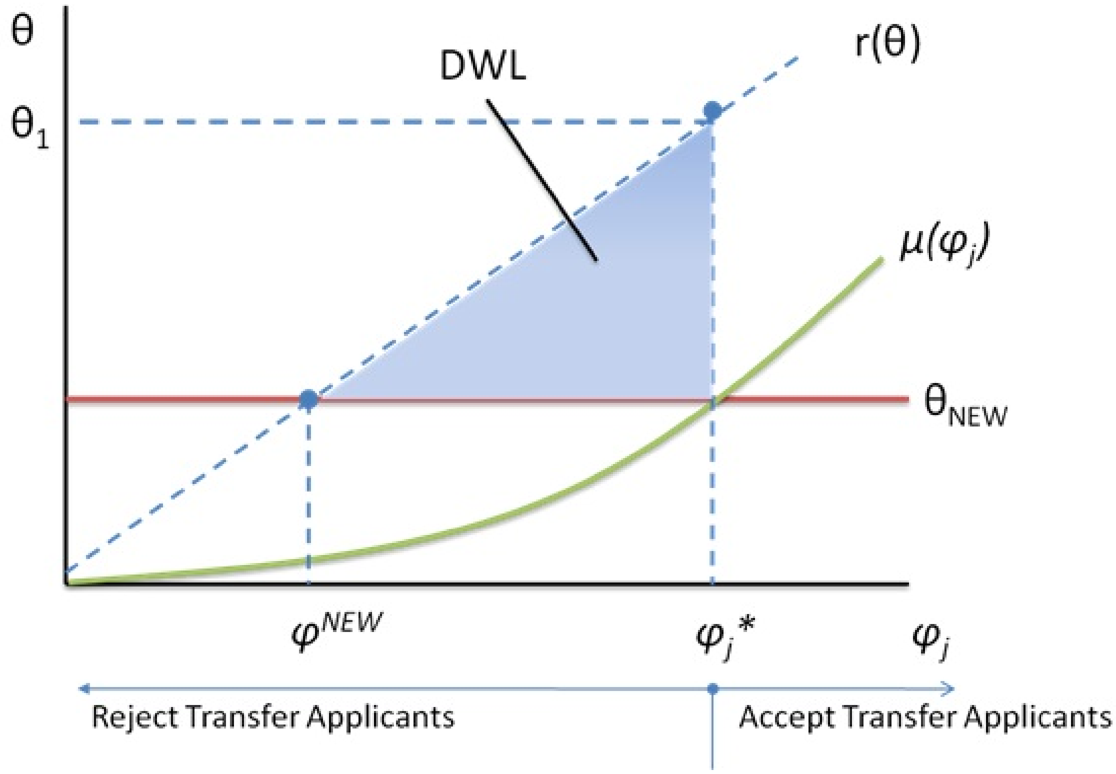

FIGURE 2.2: Inefficiency Due to Informational Asymmetry in the Teacher Transfer Market.

2.2.2 Introducing Superintendent Discretion in Transfer Hiring

There are a few possible remedies to the potential inefficiency caused by informational asymmetry in the transfer market. One solution could be to implement a signaling strategy such as national board certification where the signal is less costly for high-quality teachers to attain, and thus teachers are sorted on the market through their possession of the certification. Many districts, however, have opted for another remedy in which superintendents have the option of forcing principals to take transfer applicants they otherwise would not have hired. This type of policy may solve the problem of inefficiency due to informational asymmetry if the superintendent can observe the performance of transfers in the sending schools and force principals in receiving schools to take a transfer that is a good match for the school.

On the other hand, allowing a superintendent the discretion to force principals to accept transfer applicants could create another type of inefficiency if some principals are forced to take transfers who are of lower quality than the quality they could expect by hiring a novice teacher. Consider the incentive structure of a district superintendent. The superintendent's incentives depend partly on the academic performance of all students and schools within the district, and thus he has incentive to help improve low-performing schools. However, the superintendent's incentive structure also depends partly on appeasing well-resourced parents with political clout since he is accountable to both parents and a district school board in order to retain his position as superintendent.

On average, the complaints of affluent, well-resourced parents are more likely to be heeded by school administrators (Dauber and Epstein, 1993). This could be because complaints lodged by the parents of low-income and minority students are given less attention by teachers and administrators (Louie and Holdaway, 2009) or because low-income or minority parents have less political clout and may even become discouraged from lodging complaints. No matter the root cause, low-income or minority parents, whose children tend to be concentrated in low-performing schools have less voice and agency when it comes to complaining about ineffective teachers. Thus, a superintendent who is unable to terminate a low-quality teacher because of union contract provisions has an incentive to transfer the low-quality teacher to a low-performing school where parents are afforded less opportunities to complain. Thus it is possible that allowing superintendents to dictate transfer decisions may introduce a distributional equity issue in which low-performing schools are more likely to be assigned low-quality teachers.

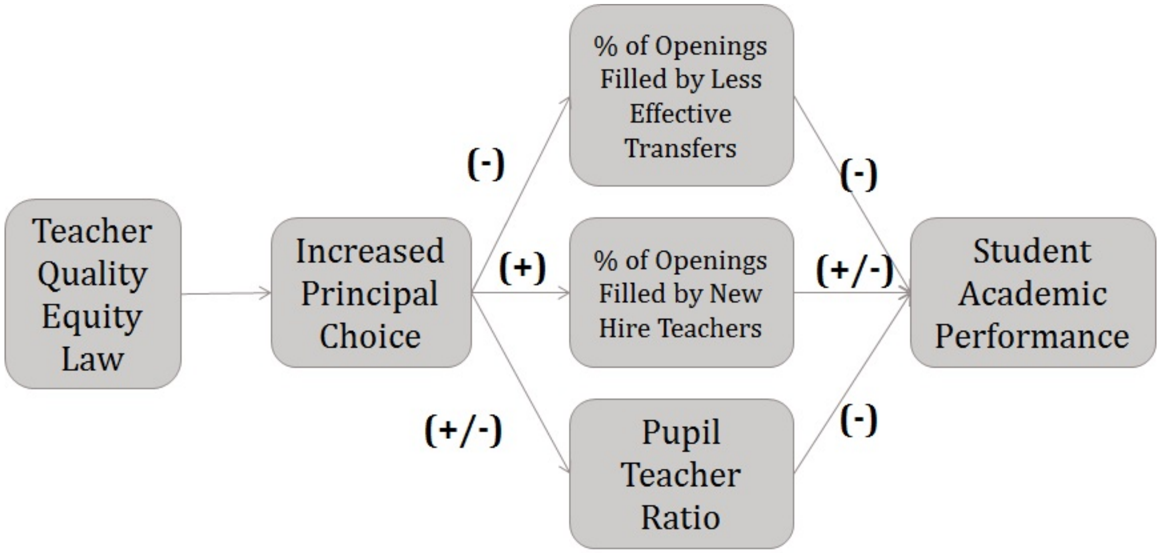

FIGURE 2.3: Predicted Effects of California's Teacher Quality Equity Law.

2.2.3 Predicted Effects of the Teacher Quality Equity Law

Figure 2.3 depicts the predicted effects of California's Teacher Quality Equity Law as operating through three potential mechanisms. First, since the policy allows principals in low-performing schools more freedom of choice in their hiring decisions, the potential for unfavorable selection in the transfer market would lead these principals to attempt to hire more novice teachers and forgo transfer applicants. This would also have the effect of decreasing the number of low-quality transfers - the second mechanism. Third, the policy could have an effect on pupil/teacher ratios. Prior to the law, principals may have had incentive to leave a position vacant rather than announce an opening and risk the superintendent forcing a low-quality transfer applicant into the position. Thus, the policy may result in lower pupil/teacher ratios as principals fill positions they otherwise may not have. However, principals of low-performing schools may have difficulty filling their positions on the new hire market, thus leaving those positions unfilled. The end effect on the pupil/teacher ratio is unknown.

The total effect of these three mechanisms on student performance is an empirical question. Decreasing the amount of low quality transfer teachers should improve student performance while increasing the number of novice teachers may have a negative effect on student performance. In expectation, the combined impact of these two mechanisms should be positive, otherwise, principals would choose to accept the transfer applicant instead of attempting to hire a novice teacher. However, the realized outcome is unclear since the realized quality of the novice teacher could be low, or the school could fail to attract a new hire teacher at all. Finally, if the pupil/teacher ratio decreases, there is some evidence that this may lead to improved student performance (Angrist and Lavy, 1999; Boozer and Rouse, 2001; Krueger, 1999), although there is also evidence that class size or pupil/teacher ratios do not have significant effects on student achievement (Hanushek, 1997; Hoxby, 2000). The conflicting results from these studies may be due to differences in sample composition and estimation techniques. The studies with the strongest internal validity, however, predict benefits to smaller class sizes, especially for the types of disadvantaged students this particular teacher transfer policy is designed to help.

2.3 Data and Methodology

I use data from two sources. The first is publicly available school-level administrative data from the California Department of Education DataQuest system for the 1999/2000 through 2008/2009 school years.[5] Schools that opened after 2006 or closed before 2006 are eliminated from the sample as they would not have observations during both the pre-treatment and post-treatment periods. Special education schools, continuing education schools, and primary centers which include only kindergarten to 2nd grade are also eliminated. These schools are not given API rankings and therefore are not subject to the provisions of the Teacher Quality Equity Law. After

[5] http://www.cde.ca.gov/dataquest/

these delimitations, the base sample includes 8,148 schools in 58 counties and over 1,000 school districts.

I also use teacher-level data from the Los Angeles Unified School District (LAUSD) that I accessed from the publicly available Los Angeles Times study on teacher-value added.[6] In the current study, I do not draw on the value-added part of the LA Times study. Rather, I use the data to track teacher mobility from the 2003/2004 school year to the 2009/2010 school year for third through fifth grade teachers. From the teacher-level data, I am able to aggregate teacher movement to the school level for 450 LAUSD elementary schools.

As causal identification strategies I employ both a difference-in differences (D-in-D) approach to identify the average treatment effect of the policy on all low-performing schools relative to their higher-performing counterparts and a regression discontinuity (RD) approach which identifies the local treatment effect of the policy by comparing schools directly above and below the academic performance cutoff specified by the law.

2.3.1 Treatment and Comparison Groups

California's Teacher Quality Equity Law applies to schools in the lowest three deciles of the state's academic performance index (API) - a measure that captures student performance in math and English language arts at the school level. Accordingly, I designate the treatment group for the D-in-D specification as those schools that were in the lowest three deciles of API scores in the fall of 2006, as reported by the state administrative records. All other schools comprise the comparison group. As a robustness check, I estimate another specification where I eliminate schools with API scores in the top three deciles and use the middle four deciles as the comparison

[6] The data are publicly available in a series of web pages on the Los Angeles Times website at http://www.latimes.com/news/local/teachers-investigation/.

Table 2.1: Statewide Sample Averages, 2005-2006 School Year.

	All	Treatment	Comparison
N	8148	2489	5659
Total Enrollment	747	799	724
Outcomes			
% Fully Credentialed	95.0	92.2	96.2
% Not Fully Credentialed	6.5	9.8	5.0
Average Experience (yrs)	12.7	11.5	13.2
% with MA	34.1	31.0	35.4
% with BA	64.7	67.5	63.4
Pupil/Teacher Ratio	20.3	20.4	20.3
API Score	747	643	793
English Language Arts Z-Score	0.01	-1.00	0.45
Math Z-Score	0.06	-0.87	0.45
Algebra Z-Score	-0.01	-0.78	0.33
Student Demographics			
% ELL	25.0	41.4	17.7
% FRL	50.5	76.5	39.1
% American Indian	1.2	1.4	1.2
% Asian	10.8	5.7	13.0
% Black	7.7	11.0	6.3
% Hispanic	44.0	67.0	33.9
% White	33.9	13.6	42.8
% Multi Ethnic	2.3	1.3	2.8
Teacher Demographics			
% Male	23.3	26.6	21.9
% Female	76.6	73.4	78.0
% American Indian	0.6	0.7	0.6
% Asian	5.9	5.9	5.9
% Black	4.1	7.6	2.6
% Hispanic	14.1	24.2	9.7
% White	73.9	59.8	80.0
% Multi Ethnic	1.4	1.8	1.2

Sample averages for the treatment and comparison groups are significantly different from each other for all variables except Pupil/Teacher Ratio and Percent of Teachers who are Asian.

Source: California DataQuest Management System (http://dq.cde.ca.gov/dataquest/)

group.

Table 2.1 presents sample averages for various pre-treatment characteristics in the immediate pre-treatment school year (2005/2006). There are roughly 2500 schools in the treatment group and nearly 5700 schools in the comparison group. Treatment group schools tend to have a higher average student enrollment than comparison group schools (799 compared to 724). This difference is statistically significant. Sample Averages for the treatment and comparison groups are significantly different from each other at the $\alpha = .05$ level for all variables except the pupil/teacher ratio and the percent of teachers who are Asian. Also of note, schools in the treatment group are much more likely to have teachers who are Hispanic and less likely to have white teachers than schools in the comparison group.

In the extended analysis, I examine the effects of the policy on subgroups of schools categorized by student demographic characteristics including race and ethnicity, free or reduced lunch eligibility (FRL) status[7] and status as an English language learner (ELL)[8]. As demonstrated in Table 2.1, treatment group schools have a higher percentage of students who are ELL - 41.4 percent compared to 17.7 percent in the comparison group. Schools in the treatment group also have a much higher percentage of students who are FRL eligible - 76.5 percent compared to 39.1 percent. Finally, students in the higher achieving comparison group schools are more likely to be white or Asian[9] while students in the treatment group schools are more likely to be black or Hispanic.

For the RD approach I treat school API scores as the variable that assigns a school to the treatment group and create a treatment indicator variable (T) using

[7] students are considered eligible for free or reduced price lunches through the National School Lunch Program based on their parents' income. Income levels at or below a poverty threshold qualify students for the program.

[8] English Language Learners are classified as students whose native language is not English.

[9] Filipinos and Pacific Islanders are categorized as Asian.

the rule:

$$T = \begin{cases} 1, & \text{if } API < API^* \\ 0, & \text{if } API \geq API^* \end{cases}$$

where *API** is the threshold value determining which schools are in the bottom three academic performance deciles.

For parts of the analysis, I will differentiate between elementary schools, middle schools, and high schools[10]. Tables A.1, A.2 and A.3 in the Appendix present the same pre-treatment sample averages as Table 2.1, for elementary, middle and high schools respectively. Table A.4 in the Appendix presents the pre-treatment sample averages for the LAUSD subset of schools which will be used to analyze teacher mobility.

2.3.2 Dependent Variables

To determine whether the teacher transfer policy has any effect on the distribution of observable teacher characteristics, I focus on two school-level measurements of teacher quality: education and credentials. *Masters* measures the percent of teachers in a school with a master's degree. Turning again to Table 2.1, schools in the treatment group are more likely to have a higher percentage of teachers with bachelor's degrees and a lower percentage of teachers with master's degrees than comparison group schools. *Full Credential* measures the percentage of teachers in a school who are fully credentialed in at least one area. *Non-Full Credential*, on the other hand, measures the percentage of teachers in a school with less than a full credential in at least one area they are currently teaching. Having less than a full credential can include teachers with an emergency credential, a waived credential, interns and those

[10] Elementary schools typically cover grades K through 5, middle schools cover grades 6 through 8 and high schools cover grades 9 through 12. In the event a school spans multiple grade categories, it is classified as belonging to the category of the highest grade level in the school.

with no credentials.[11] While most schools have a high percentage of teacher who are fully credentialed, there are still significant differences between treatment and comparison group schools (92 percent compared to 96 percent). Likewise, treatment group schools have a higher percentage of teachers who are not fully credentialed (9.8 percent) than comparison group schools (5 percent).

To assess the mechanisms that are outlined in the theoretical predictions section, I include *Pupil/Teacher Ratio* which can signify the extent to which schools are able to fill vacant positions. On average, schools in the 2005/2006 school year have pupil/teacher ratios of about 20, with no significant difference between treatment and comparison schools. I also include *Average Experience* as a crude measure of whether schools hire novice teachers or more experienced transfer teachers to fill vacant spots. For more detailed measures of teacher mobility I turn to the LAUSD teacher mobility dataset. For each school, I calculate four school-level mobility variables. The variable *#Departing* measures the number of teachers who have left the school in a given school year[12]. The variable *#Arriving* measures the total number of teachers who are new to the school. The variables *%Transfers* and *%Novice* measure what percentage of the total newly arriving teachers at the school are transfers from another school or novice teachers, respectively.

Table 2.2 presents pre-treatment means for the teacher mobility data. The first thing to notice are the sample sizes of the treatment and control groups. In the LAUSD sample, nearly half of all schools are below the API cutoff levels, since the cutoff levels are based on being in the bottom third of schools in the entire

[11] The values for *Full Credential* and *Non-Full Credential* do not sum to 100 percent because it is possible for a teacher to hold more than one type of credential. For example, a middle school teacher may hold a credential in a single subject like math or science, but not hold a full general subject credential. Thus a teacher could be both fully-credentialed and partially-credentialed at the same time. While these two variables measure similar things, they are not exactly overlapping.

[12] I am not able to differentiate between whether teachers left the school for another school in the district, for another school outside of the district, or whether they left the teaching profession.

Table 2.2: LAUSD Teacher Mobility Sample Averages, 2005-2006 School Year

	All	Treatment	Comparison
N	450	207	243
# of Grade 3-5 Teachers	14	17	12
# Teachers Leaving	4.0	6.0	3.0
# Teachers Arriving	3.0	4.0	2.0
% of Arrivers who are Novice	64.0	64.4	63.6
% of Arrivers who are Transfers	23.4	22.2	24.7

Sample averages for the treatment and comparison groups are significantly different from each other for all variables except % of Arrivers who are Novice and % of Arrivers who are Transfers

Source: California DataQuest Management System (http://dq.cde.ca.gov/dataquest/)

state, and not just in the school district. Thus the LAUSD sample is more heavily represented by low-performing treatment group schools than the statewide sample.[13] Differences between the LAUSD sample and the statewide sample may bring into question the external validity of the LAUSD movement results, however, a large enough subpopulation of students in California are educated in large urban districts such that the results may still be informative.

With respect to the teacher mobility variables, treatment group schools experience more turnover than comparison group schools. Overall, schools experience an average of four teachers leaving per year. In comparison, treatment group schools average six leavers per year while comparison group schools average three leavers per year. Similarly, treatment group schools average four teachers arriving to the school per year while comparison group schools average two arriving teachers per year. There is no significant difference, however, in the mix of arriving teachers that

[13] Comparing Table 2.1 with Table A.4 in Appendix A, the LAUSD sample is different from the statewide sample in five additional ways. There is a larger difference in total enrollment between the treatment and comparison groups (939 vs. 603 student in the LAUSD sample compared to 799 vs 724 students in the statewide sample). There are also a greater percentage of English language learners and FRL eligible students in the LAUSD sample. There are significantly more Hispanic students and significantly fewer white students in the LAUSD sample than in the statewide sample. Similarly, there are fewer white teachers and more minority teachers in the LAUSD sample when compared to the statewide sample.

are transfers or novice teaches with each group having around 64% transfers and 23% novice teachers. The remaining percentage of teachers are those who transfer from outside of the district.

The final set of dependent variables capture the extent to which the teacher transfer policy affects student performance in math and English language arts. For each school in the sample, the state compiles standardized test scores in math and English language arts. Students in grades three through twelve are administered a grade-level English language arts exam. For math exams, students in grades three through seven are administered a grade-specific math exam, while students in grades eight through twelve are given subject-specific tests. For these upper-level tests, results are estimated for algebra test scores and for a general high school math course. I standardize the test scores to have a mean of zero and a standard deviation of one.

2.3.3 Difference-In-Differences Analysis

I employ the following difference-in-differences causal identification specification using ordinary least squares regression estimation:

$$Y = \beta_0 + \beta_1 Post + \beta_2 Treat + \beta_3 (Post * Treat) + \Gamma X + \delta_d + \delta_t + \delta_{dt} + \varepsilon \quad (2.1)$$

In this specification all observations are at the school level and ε represents a school level error term that is clustered at the district level. *Post* is a dichotomous variable equal to 1 if the observation occurs after the 2006/2007 school year and equal to 0 otherwise. *Treat* is a dichotomous variable equal to 1 if the school is in the treatment group as described in the previous section and equal to 0 otherwise. The coefficient on the interaction term *Post*Treat* represents the treatment effect of the policy. A separate regression is estimated for each of the dependent variables, represented by the variable Y. School district fixed effects (δ_d) are included since the policy is operationalized at the district level. Time trends (δ_t) are included to distinguish between the effects due to decline or improvement in an outcome that happen over time and the effects due to the one-time shock of the teacher transfer policy. District-year trends (δ_{dt}) are included to distinguish between the effect of the policy, and the possibility that some school districts that may have had a more uneven distribution of highly effective teachers before the policy may be making other district-level improvements over time that lead to relative improvements in the outcome variables among treatment group schools. Finally, I include a vector of school-level characteristics (X) which includes *FRL*, *ELL*, total school enrollment, and the percent of students in the following racial and ethnic group categories: American Indian, Asian, black, Hispanic, mulit-racial, and white.

In order to examine whether there are heterogeneous treatment effects among school with different populations of students, I include a subgroup analysis where I estimate equation (2.1) separately for above and below median categories of the following school-level percent demographic variables: FRL eligible students, ELL students, Asian students, black students, Hispanic students and white students. This strategy allows comparison of the treatment effect both between high and low levels of a single demographic variable (i.e. comparing schools with high FRL eligible student populations to those with low FRL eligible student populations) as well as comparison across high or low levels of different demographic subgroups (i.e. comparing schools with high levels of Asian student populations to schools with high levels of white student populations).

2.3.4 Regression Discontinuity Analysis

In the regression discontinuity (RD) analysis, the strength of the causal identification comes from the assumption that schools on either side of the API cutoff value for assignment into the treatment group are not significantly different from each other on observable or unobservable dimensions, and they therefore serve as valid comparison groups (Shadish et al., 2002). For example, among elementary schools, the API cutoff score for inclusion in the treatment group (the bottom three performance deciles of schools) is 712. The assumption is that schools with API scores of 711 are not significantly different from schools with API scores of 713. Based on that assumption, schools directly above the cutoff serve as a valid control group for schools directly below the cutoff because there should not be unobservable differences that impact the outcome variables. Thus we can be confident that any differences in the outcome variables are due to the differential effect of being subject to the provisions of the teacher transfer policy.

An advantage of this approach is the strength of the internal validity of the estimates. A disadvantage is that the estimates are based on a small subset of observations, and therefore any possible effects of the policy must be treated as local treatment effects that are applicable to schools that are close to the API cutoff score. The effects may not be generalizable to lower-performing schools whose API scores are much farther from the cutoff value, such as schools in the lowest decile.

One way to estimate the effect of the policy on the various outcome measures using the RD framework is to fit a curve with the assignment variable (API) as the independent variable and the outcome variables as the dependent variables, allowing the curve to jump at the treatment cutoff value. The magnitude of that discontinuous jump will be the estimate of the effect size. If there is no significant jump for a particular outcome variable, then the effect of the policy on that outcome is not significantly different from zero. A potential threat to the statistical validity of RD estimates is the misspecification of the functional form used to fit the curve that estimates the relationship between the assignment variable and the outcome variable (Shadish et al., 2002). To avoid this particular pitfall, I use a nonparametric strategy to estimate the boundary points on either side of the discontinuity. The difference between these two boundary points is the estimate of the effect of the policy on the outcome variable:

$$\widehat{\beta_{RD}} = \widehat{Y^-} - \widehat{Y^+} \tag{2.2}$$

In this specification, $\widehat{Y^+}$ is the estimated boundary point to the right of the assignment cutoff and $\widehat{Y^-}$ is the estimated boundary point to the left of the assignment cutoff. The boundary points $\widehat{Y^+}$ and $\widehat{Y^-}$ are estimated using local linear regression by fitting lines in small neighborhoods (or bandwidths) around each data point up to the boundary point. Imbens and Lemieux (2008) provide a detailed discussion of the application of local linear regression techniques in RD frameworks. This technique has been used in many studies, including studying the effects of medicare on mortality rates (Card et al., 2009) and the effects of the government's Head Start program on child mortality rates and educational attainment (Ludwig and Miller, 2007). Estimates of $\widehat{\beta_{RD}}$ are computed for each of the outcomes of interest, separately for elementary schools, middle schools, and high schools using the rd command from the statistical software package *Stata* (Nichols, 2011).

The local linear regression approach also helps to provide evidence in support of the underlying assumption for the RD design that schools on either side of the cutoff do not significantly differ on unobserved characteristics. While it is not possible to examine unobserved characteristics directly, Figure 2.4 provides graphical evidence that various observable pre-treatment characteristics do not significantly differ at the cutoff point. The percentage of students who are English language learners, who are eligible for free or reduced price lunches, who are Hispanic, and who are white do not differ significantly on either side of the cutoff score (normalized to an API score of 0 in these figures). While the graph for the percentage of Hispanic students appears to have a discontinuous jump, that jump is not statistically significant. The discontinuities for total enrollment and for the percentage of students who are black, however, are marginally significant, indicating that comparison group schools

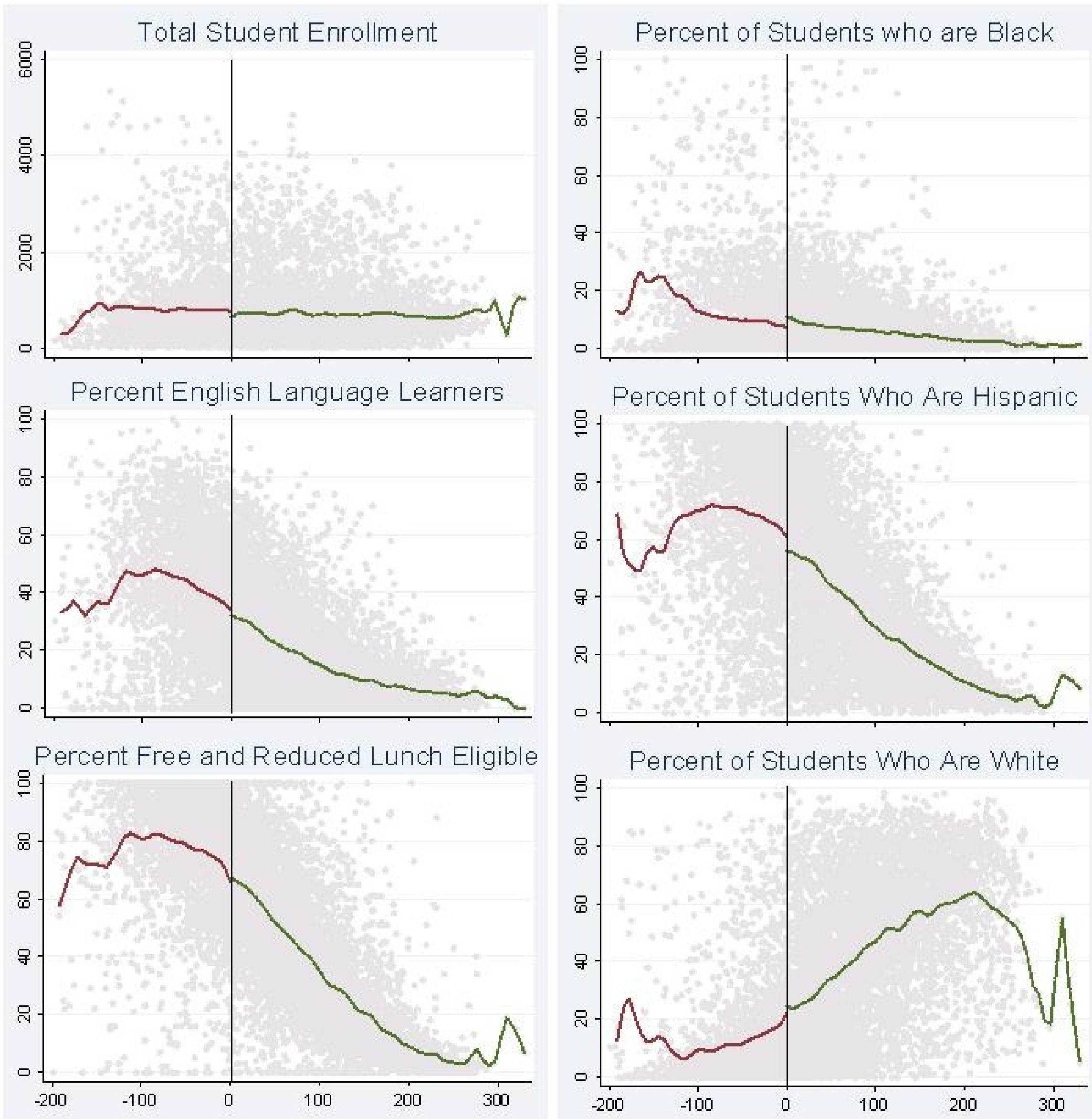

FIGURE 2.4: RD Estimates of Select Pre-Treatment Characteristics. Normalized API Score is on the X-Axis

have about 100 fewer students and about 4% fewer black students. These are the only two marginally significant results out of 28 different estimated pre-treatment characteristics.[14]

2.4 Results

In this section, I present results from the difference-in-differences analysis for both the full sample of comparison group schools, and a restricted sample of comparison group schools where I include only those schools in API deciles four through seven. I then present results from the regression discontinuity analysis followed by the analysis of possible heterogeneous treatment effects across demographic subgroups. Finally, I present the mobility analysis using the subsample of schools from the Los Angeles Unified School District.

2.4.1 Difference-In-Difference Analysis

Turning first to the outcomes that involve teacher characteristics, Table 2.3 presents the results for the D-in-D analysis. Each cell in the table is derived from a unique regression of Model (2.1) for the outcome listed in the first column and represents the point estimate and robust standard error for β_3 - the treatment effect. The percent of teachers fully credentialed in at least one area increased significantly among all school levels in the treatment group relative to the comparison group, with the largest increases found among middle schools. The point estimates range from 1.5% in elementary schools to 2.1% in middle schools. The percent of teachers without a full credential in an area they are currently teaching decreased significantly among elementary schools in the treatment group relative to the comparison group (by -

[14] I only present graphs for six pre-treatment characteristics here, however, I estimated differences for a wide range of pre-treatment characteristics including parent, student and teacher demographic characteristics. I also estimated separate figures for elementary schools, middle schools and high schools.

Table 2.3: Change in Outcomes in Treated Schools Post 2006 - Difference-in Differences

	Elementary	Middle School	High School
Teacher Characteristics			
% Fully Credentialed in at Least One Area	1.477**	2.125*	1.547**
	(0.591)	(1.122)	(0.661)
% Not Fully Credentialed in a Subject Currently Teaching	-1.763**	-0.854	-0.609
	(0.695)	(0.754)	(0.914)
Average Experience	0.0918	-0.107	0.249
	(0.0837)	(0.155)	(0.202)
% with Masters Degree	1.214***	0.571	-0.218
	(0.443)	(0.610)	(0.942)
School Outcomes			
Pupil/Teacher Ratio	0.0394	-0.793***	-0.662
	(0.161)	(0.223)	(0.410)
ELA Z-Score	0.0855***	-0.0775***	-0.0203
	(0.00945)	(0.0279)	(0.0253)
Math Z-Score	0.0684***	0.0197	-0.107**
	(0.0105)	(0.0420)	(0.0466)
Algebra Z-Score	–	-0.0640	-0.0612*
		(0.0645)	(0.0323)
Observations	52,854	12,202	10,991

Notes: Each cell is derived from a unique regression of model (2.1) for the specified outome variable

Robust standard errors clustered at the district level in parentheses

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

1.8%), but not among middle schools or high schools. This lack of significant effect among middle schools and high schools is not entirely surprising. Teachers in middle schools and high schools are often required to have single subject credentials as well as general education credentials, where many elementary school teachers are only required to have one credential. This increases the likelihood that middle schools and high schools will find teachers who have a full credential in at least one area, but may be lacking the credential in another area. The overall results with respect to teacher credentialing are consistent with the hypothesis that low-performing schools were able to be more discriminating in the teacher hiring practices after the implementation of the teacher transfer policy.

The coefficient estimates for average experience among teachers in a school are not statistically different from zero at any school level. It is possible that average experience is too aggregate a variable to pick up the effects of the movements of a few teachers into or out of a school. The mobility analysis using the LAUSD schools may provide more insight into whether or not the policy has an effect on the mix of novice or experienced teachers in a school. Among elementary schools, the policy is associated with a 1.2% increase in teachers with masters degrees relative to the comparison group. The effects are not significantly different from zero among middle schools and high schools.

With regard to the pupil/teacher ratio, the estimated coefficient is negative for middle schools (-0.79) and high schools (-0.66) although not significantly so for high schools. The coefficient for elementary schools is not significantly different from zero. The negative estimates for middle schools and high schools imply that pupil/teacher ratios decline in treatment schools relative to comparison group schools, even after controlling for total enrollment. Recall from the discussion in the conceptual framework that a decrease in the pupil/teacher ratio for treatment group schools is consistent with the hypothesis that, in the pre-treatment period, principals may prefer to forgo the announcement of an open position, rather than be forced by a superintendent to hire an ineffective transfer teacher. The implementation of the teacher transfer policy may allow those principals the freedom to fill positions they may otherwise have chosen not to fill and instead cover with existing teachers.

Finally, the results for academic performance indicate that treatment group schools, at the elementary level, experience a relative improvement in both English language arts scores and math scores. English language arts scores improve by 0.09 standard deviations and math scores improve by 0.07 standard deviations. To put these improvements in perspective, Angrist et al. (2010) estimate that the interventions involved with KIPP Charter Schools increase student academic performance by 0.35 standard deviations in math and by 0.12 standard deviations in English. Thus these estimates imply that the policy has both a statistically and economically meaningful effect on student academic performance. Among middle schools and high schools, however, the treatment effects are either significantly negative or not significantly different from zero. Thus, even though middle schools and high schools experience an increase in fully credentialed teacher and a decrease in pupil/teacher ratios, those improvements do not necessarily translate into improved academic performance, at least in the short run.

Table 2.4 presents results for the restricted D-in-D estimation where the comparison group is limited to schools that fall into deciles four through seven of the academic performance rankings. These estimates compare the low-performing schools with middle-range schools, and exclude the highest performing schools. The results are consistent with those in the general model, but with slightly smaller point estimates.

2.4.2 Regression Discontinuity Analysis

In the regression discontinuity analysis, as with the difference-in differences analysis, effects of the teacher transfer policy are estimated separately for each grade level.

Table 2.4: Change in Outcomes in Treated Schools Post 2006 - Restricted Difference-in Differences

	Elementary	**Middle School**	**High School**
Teacher Characteristics			
% Fully Credentialed in at Least One Area	1.224***	2.139*	1.161*
	(0.434)	(1.155)	(0.657)
% Not Fully Credentialed in a Subject Currently Teaching	-1.506***	-1.183	-0.358
	(0.498)	(0.812)	(0.992)
Average Experience	0.00837	-0.0760	0.266
	(0.0844)	(0.153)	(0.215)
% with Masters Degree	1.119***	0.802	-0.201
	(0.397)	(0.643)	(1.034)
School Outcomes			
Pupil/Teacher Ratio	0.0940	-0.639***	-0.472
	(0.172)	(0.226)	(0.482)
ELA Z-Score	0.0663***	-0.0503*	-0.0276
	(0.00869)	(0.0268)	(0.0295)
Math Z-Score	0.0575***	0.0176	-0.0851
	(0.0103)	(0.0416)	(0.0534)
Algebra Z-Score	–	-0.0268	-0.0181
		(0.0688)	(0.0277)
Observations	36,898	8,572	7,744

Notes: Each cell is derived from a unique regression of Equation (2.1) for the specified outcome variable using the sample of schools restricted to include only API performance deciles 1-7

Robust standard errors clustered at the district level in parentheses

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

Table 2.5 presents the results of the local treatment effects obtained from estimating Equation (2.2) for the eight different outcomes of interest at each grade level.

The signs on the coefficients for teacher credentials are consistent with the D-in-D results. The percent of fully credentialed teachers increases for the treatment group relative to the control group at all school levels while the percent of non-fully credentialed teachers decreases among elementary and middle schools, but not among high schools. These effects are statistically significant among elementary schools, but not among middle schools and high schools. It is possible that the point estimates for middle schools and high schools are not precisely estimated due to the much smaller sample sizes for those grade levels.

The estimates for average experience are positive, though not significantly different from zero for middle schools and high schools. For elementary schools, the results imply an increase in average experience of almost one additional year. This is in contrast to the D-in-D results which are negative, but not significant among elementary schools. With regard to educational attainment, the signs of the coefficients imply an increase in the percent of teachers with masters degrees, though the estimates reach statistical significance only among the high school sample (6.0%). This also is in contrast to the D-in-D results in which the coefficient on masters degree is negative but not significant. There are no other significant treatment effects among the RD estimates. The observed differences in some of the RD and D-in-D estimates could be due to the difference between a local treatment effect and an average treatment effect. The D-in-D estimates compare a wider range of schools from very low to very high achievement levels, as opposed to the RD estimates which only compare those schools on either side of the assignment cutoff. Thus, it could be that the schools that are affected the most by the teacher transfer policy are among the lowest performing schools, and not those schools that near the API cutoff level.

Table 2.5: Change in Outcomes in Treated Schools Post 2006 - Regression Discontinuity

	Elementary	Middle School	High School
Teacher Characteristics			
% Fully Credentialed in at Least One Area	1.199***	0.560	1.144
	(0.407)	(2.021)	(1.933)
% Not Fully Credentialed in a Subject Currently Teaching	-1.616**	-1.204	0.261
	(0.638)	(2.343)	(2.516)
Average Experience	0.877***	0.417	0.365
	(0.327)	(0.641)	(0.911)
% with Masters Degree	0.670	2.309	6.035*
	(1.750)	(4.352)	(3.366)
School Outcomes			
Pupil/Teacher Ratio	0.0950	0.0792	0.284
	(0.252)	(0.591)	(0.834)
ELA Z-Score	-0.00533	-0.0103	-0.00950
	(0.0198)	(0.0420)	(0.0489)
Math Z-Score	0.0347	0.0271	-0.179
	(0.0281)	(0.0561)	(0.124)
Algebra Z-Score	–	0.0330	-0.00764
		(0.119)	(0.0635)
Observations	16,352	3,795	3,690

Notes: Each cell is derived from a unique estimation of Equation (2.2) for the specified outcome variable.

Bootstrapped standard errors in parentheses

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

2.4.3 Demographic Subgroups

Since academic performance is the dimension along which the teacher transfer policy is designed to operate, the analysis thus far has relied on comparing low-performing schools to high-performing schools. However, other school characteristics, such as the composition of the student body, may influence the ability of a school to attract and retain high quality teachers, which can have an impact on the effectiveness of the policy. For this reason, I estimate treatment effects on subgroups of schools categorized as having above or below median concentrations of free or reduced lunch (FRL) eligible students, English language learners (ELL), and four racial/ethnic subgroups - Asian students, black students, Hispanic students, and white students.[15]

The treatment effects on the percent of teachers who are fully credentialed in at least one area are presented in Figure 2.5 separately by above median levels of the demographic subgroups. At all grade levels, schools with above median percentages of FRL eligible students, ELL students, and minority students see a significant increase in the percent of fully credentialed teachers. Schools with and above median percent of white students, however do not see a statistically significant treatment effect. These results indicate that, while no subgroup of schools were affected negatively by the policy, schools with above average populations of historically disadvantaged populations benefited the most.

With respect to teacher education, Figure 2.6 presents the treatment effect on the percent of elementary school teachers with a masters degree separately by above median levels of the demographic subgroups. Again, schools with above median FRL eligible students, ELL students, and minority students experience a significant positive impact on the percent of teachers with a masters degrees while schools with

[15] I also estimate a model interacting the continuous demographic percent variables (*%FRL, %ELL, %Asian, %Black, %Hispanic* and *%White*) with treatment group status. The results are in line with the demographic threshold subgroup estimates, so I choose to present the threshold estimates, since they are slightly easier to interpret.

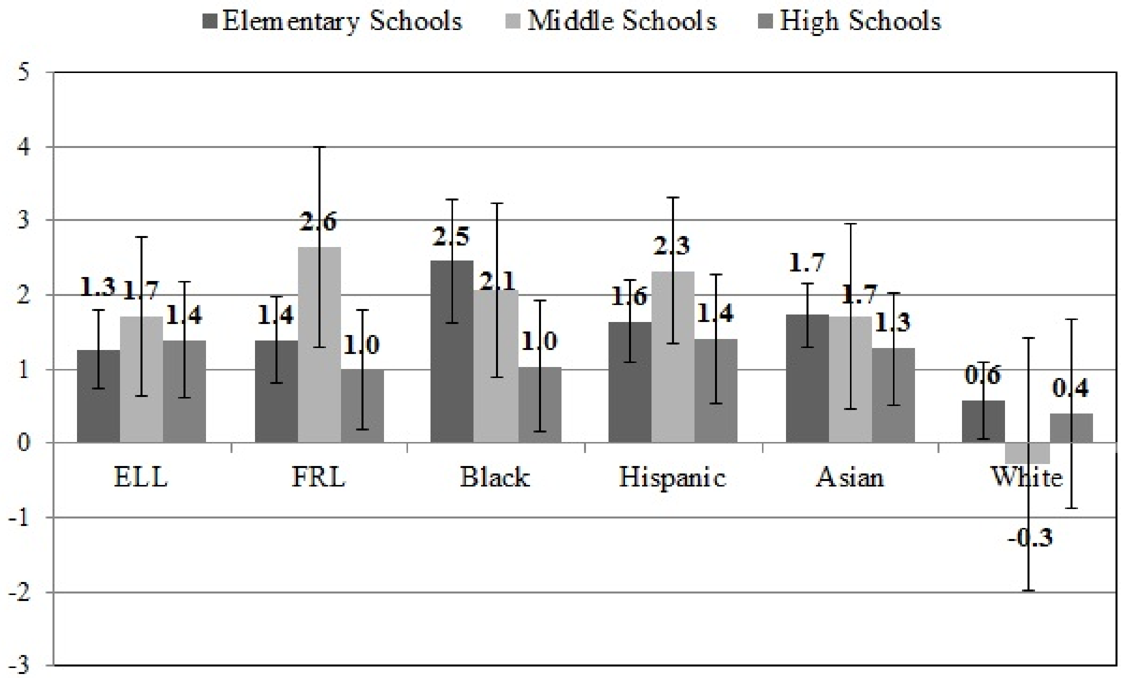

FIGURE 2.5: Treatment Effect of Teacher Transfer Policy on %Fully Credentialed in at Least One Area by Above Median % of Demographic Subgroups.

above median populations of white students do not see a significant change. Among middle schools and high schools, recall that the main effects with respect to teacher education are not significant. That pattern holds true among the demographic subgroups. Since there are no significant differences among the demographic subgroups, those results are not presented here.

The effects on pupil/teacher ratios are not significantly different across subgroups among elementary and middle schools. However, distinct patterns emerge in the high school level results, shown in Figure 2.7. Comparing across demographic subgroups, schools with below median percentages of disadvantaged students experience greater decreases in pupil/teacher ratios, while schools with below median percentages of white students do not see significant decreases in pupil/teacher ratios. Conversely, schools with above median populations of white students experience greater decreases

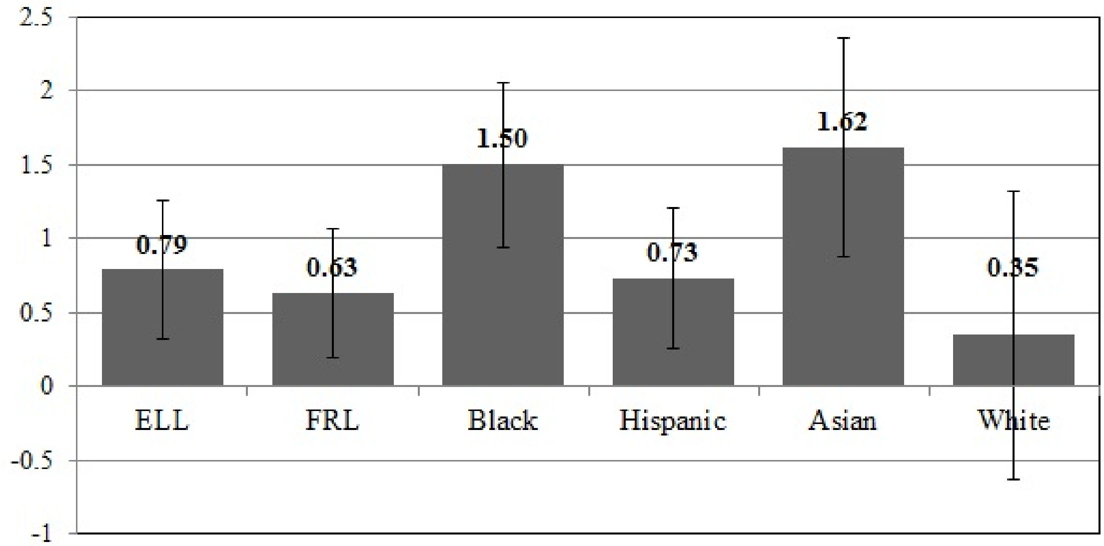

Figure 2.6: Treatment Effect of Teacher Transfer Policy on %with a Masters Degree in Elementary Schools by Above Median % of Demographic Subgroups.

in pupil/teacher ratios than those with above median populations of FRL eligible students, ELL students or minority students. These comparisons are borne out in the within-group comparisons as well. With the exception of Asian students and white students, all other subgroups experience a larger decrease in pupil/teacher ratios among the below median subgroup than the above median subgroup. This pattern of results is consistent with the hypothesis that in the post-treatment period, treatment group schools may opt to forgo transfer applicants for novice teachers, but those with lower populations of traditionally disadvantaged students are better able to fill open positions than those with higher populations.

There are no significant subgroup differences with respect to academic performance. Overall, the demographic subgroup results suggest that schools with high populations of FRL eligible students, ELL students, and minority students benefit

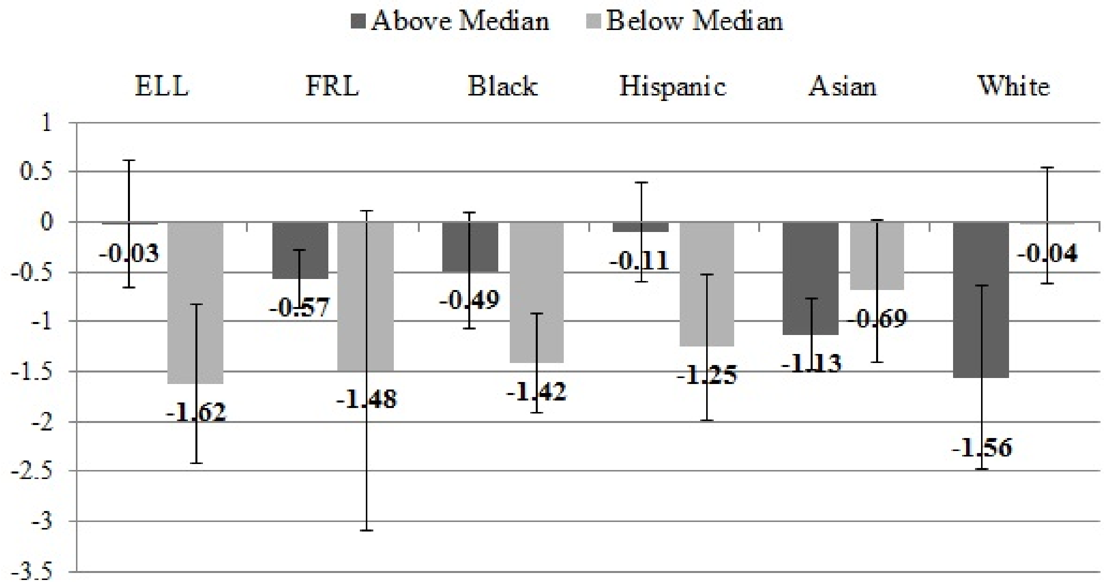

FIGURE 2.7: Treatment Effect of Teacher Transfer Policy on Pupil/Teacher Ratio in High Schools by Above and Below Median % of Demographic Subgroups.

most with respect to improved teacher credentialing and education as a result of the transfer policy. These schools, however, benefit less with respect to decreased pupil/teacher ratios. It is possible that principals of treatment group schools with relatively high populations of disadvantaged students are able to improve teacher qualifications by being more selective with respect to transfer teachers, but at the same time have more difficulty hiring and retaining new teachers when they decide to forgo transfer students.

2.4.4 Mobility Analysis

I estimate the effects of the teacher transfer policy on teacher mobility between schools using the LAUSD sample of schools.[16] Results from both the D-in-D and

[16] For the sake of comparison, I also estimate the eight main outcomes (%fully credentialed, %not fully credentialed, experience, %masters, pupil/teacher ratio, and the test score outcomes) for the

RD analyses are presented in Table 2.6. The D-in-D results indicate the transfer policy does not have a significant effect on the number of teachers leaving a school, but that it does decrease the number of teachers newly arriving at a school. While not statistically significant, the directional signs on the coefficients for the make-up of newly arriving teachers indicate that treatment group schools experience a decrease in transfer teachers and a slight increase in novice teachers. This is consistent with the hypothesis that principals in treatment group schools possibly turn away transfers and attempt to hire novice teachers as a result of the teacher transfer policy. However, since the total number of newly arriving teachers decreases for treatment group schools, the results may indicate that treatment group schools are not successful in filling openings on the new-hire teacher market. The RD results are less informative as none of them are significantly different from zero.

The results from the D-in-D subgroup analysis are presented in Table 2.7. Schools with above median percentages of FRL eligible students, ELL students, black students and hispanic students experience a significant decrease in the number of teachers newly arriving as a result of the teacher transfer policy. The mix of newly arriving teachers for these subgroups are comprised of fewer transfer teachers and more novice teachers, though these estimates are not statistically significant. Schools with above median populations of Asian students and white students, however, do not experience a significant change in the number of newly arriving teachers. The percentage of arriving teachers who are transfers decreases among all demographic subgroup types, though not significantly so, while the percentage of arrivers who are novice teachers increases among all demographic subgroup types. These results are consistent with the previous results. Treatment group schools overall may see a decrease in transfer teachers and an increase in novice teachers. However, schools with above median

LAUSD subsample. The results are consistent with the statewide sample results for elementary schools, namely, and increase in the %fully credentialed, a decrease in the %not fully credentialed, an increase in the %of teachers with masters, and increased academic performance.

Table 2.6: Change in Mobility Outcomes in Treated Schools Post 2006 - D-in-D and RD.

	D-in-D	**RD**
# of Teachers Departing	0.202	2.288
	(0.149)	(4.533)
# Teachers Newly Arriving	-0.798***	-1.337
	(0.191)	(11.66)
% of Arrivers who are Transfers	-1.980	-25.83
	(2.351)	(27.20)
% of Arrivers who are Novice	0.645	-7.697
	(2.977)	(106.7)
Observations	2,684	1,223

Notes: Each cell in column (1) is derived from a unique regression of Equation (2.1) for the specified outcome variable. Each cell in column (2) is derived from a unique estimation of Equation (2.2) for the specified outcome.

Robust standard errors in parentheses in column (1). Bootstrapped standard errors in parentheses in column (2)

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

populations of disadvantaged students attract fewer teachers overall than schools with above median populations of white students.

2.5 Discussion

In an attempt to alter the mechanisms within school districts that create an imbalance between schools in the presence of high quality teachers, California implemented a law during the 2006/2007 school year that allows principals of low performing schools the ability to refuse to accept a teacher transferring from another school. Prior to the law, principals would have had to accept those transfers at the superintendent's behest. This paper estimates the effects of this law on school-level teacher pre-service qualifications, teacher mobility, and student academic performance.

This study is subject to a few limitations. First, there are data limitations.

Table 2.7: Change in Mobility Outcomes in Treatment Group Schools Post 2006 by Above Median Demographic Subgroups.

	% FRL	% ELL	% Asian	% Black	% Hisp	% White
# of Teachers Departing	0.261	0.465*	0.314	0.399*	-0.0426	0.564**
	(0.237)	(0.261)	(0.274)	(0.221)	(0.217)	(0.250)
# of Teachers Newly Arriving	-0.810***	-0.639**	-0.204	-0.590**	-1.066***	-0.471
	(0.306)	(0.313)	(0.347)	(0.280)	(0.281)	(0.335)
% of Arrivers who are Transfers	-5.979	-6.066	-1.765	-0.307	-2.299	-1.954
	(4.064)	(4.004)	(4.236)	(3.480)	(3.360)	(4.129)
% of Arrivers who are Novice	0.212	2.517	-1.954	4.081	0.666	0.860
	(4.666)	(4.985)	(4.129)	(4.360)	(4.130)	(5.189)

Notes: Each cell is derived from a unique regression of Equation (2.1) for the specified outcome variable and demographic subgroup.

Robust standard errors in parentheses

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

Having access to student-level data to estimate teacher effectiveness in a value-added framework would shed more light on the mobility of effective teachers as opposed to simply examining pre-service qualifications. Second, with respect to external validity, California is unique in many ways, not the least of which is that the majority of public school students are Hispanic. The subgroup results that demonstrate greater effects among schools with higher minority student populations might not generalize to states with different demographic make-ups. Also, the mobility analysis which focused specifically on Los Angeles area schools, may not be applicable to smaller, more suburban districts. Third, the study could benefit from examining more long-term student outcomes such as graduation rates, and longer-term test scores. This will become a possibility as more data become available.

The most consistent finding is that treatment group schools (those in the lowest three deciles of the state's academic performance index) experience an increase in fully-credentialed teachers and a decrease in non-fully-credentialed teachers relative to comparison group schools. This pattern also persists when examining demographic subgroups. Schools with high concentrations of FRL eligible, ELL or minority students experience a relative increase in fully-credentialed teachers while schools with high concentrations of white students experience no significant change in the percent of fully-credentialed teachers. An increase in fully-credentialed teachers as a result of the policy could be an intermediate step that leads to increased academic achievement by students in the school given past evidence that fully-credentialed teachers are more effective teachers (Clotfelter et al., 2007b).

There is also evidence that California's teacher transfer policy results in a decrease in the percentage of teachers who transfer into schools and an increase in the percentage of novice teachers. This evidence would be consistent with treatment group principals exercising their discretion not to accept teacher transfers that they found to be undesirable. The total number of new teachers declines among treatment group schools, however, possibly indicating that those schools were not successfully able to hire novice teachers. As a result of the policy, treatment group schools may also have experienced a relative increase in more highly educated teachers, especially among elementary schools.

With respect to pupil/teacher ratios, the D-in-D analysis provides evidence that middle schools and high schools in the treatment group may see a relative decrease in pupil/teacher ratios, supporting the hypothesis that, during the pre-treatment period, these schools may have chosen not to announce an open position, rather than risk being forced to take an ineffective transfer teacher. Among demographic subgroups, the results suggest that schools with lower concentrations of ELL, minority or FRL eligible students experience greater decreases in pupil/teacher ratios, implying that those schools are better able to attract teachers (either transfers or new hires) than those with high concentrations of ELL, minority or FRL eligible students.

Finally, estimates from the D-in-D analysis appear to imply that elementary schools in the treatment group schools experience a relative increase in academic performance as a result of the teacher transfer policy, while middle schools and high schools in the treatment group experience a relative decrease in the academic performance. Differential treatment effects between elementary schools and middle to high schools could be due in part to structural differences between the three levels of schools. In elementary schools, students most often have one teacher who teaches them all subjects, while middle school and high school students are taught by multiple teachers for various subjects. Thus, in an elementary school, improving the average teacher quality by adding one or two new teachers may have a larger impact than in middle schools where that teacher may just be one of many that students interact with.

Perhaps most interesting are the demographic subgroup results. Given a superintendent's incentive to appease parents who may complain about ineffective teachers who cannot be easily terminated, it is likely that superintendents attempt to transfer those teachers from schools with more well-resourced parents to schools where parents may be disenfranchised, and less able to give voice to their complaints about ineffective teachers. In addressing this incentive, California's teacher transfer policy perhaps naively focuses on the dimension of academic performance. This assumes that the forced transfers are taking place from high-performing schools to low-performing schools. The subgroup analysis, however, supports the notion that these transfers are also likely taking place between schools with high populations of middle class white students and schools with high populations of poor and minority students, even if they are at the same academic performance level. In focusing only on low performing schools, instead of also focusing on other school characteristics, the policy may be missing an opportunity to further improve the distribution of high quality teachers within a school district.

In sum, there are at least two policy-relevant lessons from this analysis of the impact of California's Teacher Quality Equity Law. First, the law does not appear to have lead to any apparent increase in student academic performance among low-performing middle schools and high schools, at least in the short-run. It is possible that treatment group schools are able to attract teachers who are successful in improving student performance in non-tested subjects like science and social studies. This study would not pick up gains in those non-tested areas. It is also possible that gains may be realized in the longer term as principals in low-performing schools have more time to recruit more talented teachers and improve the overall teaching core and culture in the school. However, in the short-term and in the tested subjects of reading and math, the policy does not result in improved academic performance at the upper grade levels.

The second take-away from this study is that allowing principals at low-performing schools more discretion in declining transfer requests from potentially undesirable teachers may have the unintended consequence of leaving low-performing schools that also have high concentrations of low-income or minority students unable to fully staff their schools. There is evidence that those schools may turn down transfer teachers for a chance to hire novice teachers but may not be able to attract novice teachers to teach in their schools. This indicates that a policy that gives more agency to principals in their hiring decisions should be coupled with policies that assist schools with high minority and low-income student concentrations to recruit new teachers.

3

Bias or Behavior? Using Differences Between Teacher Reports and Administrative Records to Identify Bias in Teacher Perceptions of Student Behavior

3.1 Introduction

Subjective perceptions that teachers form about students' classroom behaviors such as effort, participation, and disruptiveness matter for student educational outcomes. These perceptions often inform academic track placement decisions (Condron, 2007; Hughes et al., 2005) that can alter a student's entire academic trajectory (Darity Jr. and Jolla, 2009; Dauber et al., 1996; Eder, 1981; Lleras and Rangel, 2009; Oakes, 2005), especially if decisions that are based on behavioral perceptions induce a mismatch between the student's actual ability and the student's track placement.[1] Given the potential impact behavioral perceptions can have on academic outcomes, and given evidence that some subgroups of students are persistently rated as having

[1] Track placement refers to the practice of tracking and ability grouping where students are placed in learning groups that are stratified by academic ability. Students in low-ability learning groups are often exposed to less rigorous curricula and tend to stay in low-ability groups throughout their entire academic trajectories.

worse behavior than others (Francis, 2012), it is important to identify any possible biases in these perceptions that would disadvantage subgroups of students.

This paper adds to the literature on teacher behavioral perceptions by estimating racial, ethnic, gender and socioeconomic differences in subjective teacher reports of a student's absenteeism while controlling for the student's actual administrative attendance records. I use longitudinal data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) which include teacher reports of individual eighth grade students' absenteeism. The data do not include administrative records for student absences in the eighth grade, but do include those measures in the two previous survey waves (fifth grade and third grade). Using the attendance measures from the two previous survey waves, along with longitudinal, student-level data from the North Carolina Education Data Research Center (NCERDC), I employ a variation of a two sample instrumental variables approach in which I instrument for actual eighth grade absences with simulated measures of eight grade absences. The simulated measures are based on parameters that are estimated from data on measured absence in previous waves of the ECLS-K data as well as from data on measured absence from fifth through eighth grade students in the NCERDC data. This empirical design has the added benefit of reducing bias from both classical measurement error and reverse causation that would be associated with using measured eighth grade absences from the ECLS-K.

I extend the analysis by computing estimates of bias for subgroups of teachers by race/ethnicity, gender, age, and education level to examine whether there are heterogeneous effects among teachers from different backgrounds. I find consistent evidence that teacher reports of the attendance behavior of low-income students are negatively biased and that math teacher reports of male attendance behavior are positively biased. There is mixed evidence with regard to student race and ethnicity. Teacher reports of the attendance behavior of Asian students appear to be in line with their actual attendance behavior, while science teachers appear to express a positive bias towards Hispanic students, relative to white students. These results indicate that some subgroups of students may be at a disadvantage if their teachers perceive their behavior to be comparatively worse than that of their peers, even when their actual behavior is the same.

3.2 Conceptual Framework

Conceptually, we can think of teacher reports of student's behavior as being made up of three components - a students actual behavior, a teacher's prior beliefs about that student's actual behavior, and classical measurement error. Thus for a student i and a teacher j we have:

$$\begin{aligned} TeacherReport_{ij} = {} & ActualBehavior_i + TeacherPrior_{ij} \\ & + ActualBehavior * TeacherPrior_{ij} + \varepsilon_{ij} \end{aligned}$$

Based on this relationship, we can imagine three possible scenarios. In the first, the teacher holds no prior beliefs about a student's behavior and his or her report of a student's behavior is reflective of that students' actual behavior, plus the possibility of measurement error. This is the case of no bias. However, even in this case, it would be difficult to empirically establish a causal relationship between actual behavior and teacher perceptions of that behavior since there is the potential for reverse causality - how a teacher perceives a student can influence that student's behavior (Hughes et al., 2008; Jussim and Harber, 2005; Steele and Aronson, 1995).

In a second scenario, a teacher may observe a student's race, ethnicity, gender, or socioeconomic status (REGS) and form a perception of that student's behavior based solely on that teacher's prior beliefs about how students from that background tend to behave. This would be the case of pure bias (either positive or negative) where the teacher's perceptions are not at all based on the student's actual behavior.

Finally, there is a third possibility, in which a teacher's prior beliefs influence how he or she views a student's actual behavior. For example, if a teacher has a prior belief that boys are generally more disruptive than girls, she may perceive the behavior of the boys in her class to be more disruptive than that of the girls, even if their actual behavior is the same. In this way, teachers' subjective perceptions are incongruous with students' actual behaviors. There is convincing evidence that teachers give higher subjective grade ratings for given measured ability levels to students from similar social backgrounds as the teacher(Goldwater and Nutt, 1999) and to students whose parents the teacher has a higher quality relationship with (Hughes et al., 2005). The relationships between teacher bias and student behavior also have the potential to be reciprocal - student behavior or teacher perceptions about student behavior may reinforce teacher biases. In this scenario, there are nonzero values for *ActualBehavior*, *TeacherPrior* and the interaction term between the two.

I define teacher bias as any difference between the teacher's report of student behavior and the student's actual behavior that cannot be explained by classical measurement error:

$$\begin{aligned} Bias_{ij} &= TeacherReport_{ij} - (ActualBehavior_i + \varepsilon_{ij}) \\ Bias_{ij} &= TeacherPrior_{ij} + ActualBehavior * TeacherPrior_{ij} \end{aligned}$$

Behaviorally based biases do not necessarily have to reflect overt prejudice or preference on the part of the teacher. The dominant culture in the United States public school setting is reflective of white, middle class cultural values (Boykin et al., 2005). Thus, it is possible that when viewed through the dominant cultural lens, a teacher's perception of the behavior of low-income or minority students is culturally misinterpreted as lacking effort, interest or discipline (Tyler et al., 2006, 2008). Alternatively, teacher bias may reflect statistical discrimination in which teachers, lacking accurate information on the actual behavior of an individual student, ascribe the average characteristics of the student's REGS grouping to that individual student as a less time-consuming means of processing information on a large number of students (Arrow, 1971; Phelps, 1972). Finally, drawing upon the anthropological concept of fictive kinship, teacher biases may reflect affinities that teachers may have towards students who are similar to them through latent, cultural social ties.

3.2.1 Are Perception Differences Due to Bias or Actual Behavioral Differences?

Estimating teacher bias has proven to be quite difficult. There is an abundance of qualitative and descriptive evidence that teachers, like everyone else, hold biases. Teachers have been shown to give higher grades for the same measured performance to students that come from a similar family background to their own (Goldwater and Nutt, 1999), or to students with whose parents they have a better relationship (Hughes et al., 2005). They have also been shown to have lower academic expectations for ethnic minorities, given the same prior academic performance (Van Ewijk, 2011).

A randomized control trial conducted among first and second graders in Sweden by Forster, Sundell, Morris, Karlberg, and Melin (2010) provides some of the most convincing evidence that teachers' expectations and priors influence their behavioral perceptions. Randomly selected students with known behavioral problems were given a behavioral intervention. Before and after the intervention, both the students' teachers and outside observers rated the behavior of both treatment group and control group students (neither knew which students were in the treatment group). After the intervention, the outside observer reported significant impacts of the behavioral intervention treatment on student behavior, while the students' teachers did not. This suggests that the new information teachers received from observing the students after the intervention was not enough to override their priors about the behavior of individual students. However, this study, as well as the bulk of the qualitative evidence on bias, suffers from external validity problems.

In an attempt to provide more broadly representative evidence on the question of teacher bias, many researchers have examined differences in perceptions of behavior between teacher-student combinations that share the same race or same gender, and those that do not. The studies consistently find that teachers rate students who share their same race or gender more favorably on subjective behaviors among high school, middle school, and elementary students (Dee, 2005; Downey and Pribesh, 2004; Ehrenberg et al., 1995; Mullola et al., 2011).[2] Despite the consistency of this evidence, even the most convincing studies in this line of research suffer from at least three limitations.

First, there is no way to distinguish between whether the observed differences in perceptions arise because students improve their actual behavior when paired with a teacher similar to them or whether teachers have expressed biases towards students who are similar to them. Second, the propensity for teachers to sort into schools so that they are racially matched with the majority of students (Feng, 2009; Hanushek et al., 2004; Scafidi et al., 2007) can overstate same-race effects by failing to take into account that the teachers who are most likely to be paired with same-race students are also more likely to hold positive biases towards those students since they've already expressed a sorting preference for working with them (Miller, 2009). A third limitation of the same-race studies of teacher perceptions involves the construct validity of using teacher-student race match as a baseline from which to measure bias. This comparison assumes *a priori* that black teachers, for example, exhibit less bias about the behavioral perceptions of black students than do white teachers, but it

[2] Despite these differences in behavioral perceptions, there is mixed evidence on whether there are same-race or same-gender associations with improved academic performance (Bishop et al., 2005; Ehrenberg et al., 1995; Harris, 2006; Miller, 2009; Ouazad, 2008) which, however, is not the focus of this paper.

could be that teachers of any race hold more positive behavior perceptions of white students than black students, in which case comparing the differential perceptions that black and white teachers hold of black and white students would understate any perceptual bias. The present study attempts to address these issues by estimating racial, ethnic, gender and socioeconomic differences in subjective teacher reports of student absence while controlling for actual administrative records of student absences. This will provide evidence as to whether systematic differences in teacher behaviorial perceptions are due to differences in actual behavior or due in part to teacher biases.

3.3 Data and Methodology

In attempting to assess whether differences in teacher reports of student behavior are due to bias on the part of the teacher or to actual behavioral differences on the part of the students, the ideal study would compare assessments of student behavior conducted by trained, independent observers with teachers' reports of that behavior. This study attempts to approach that ideal by comparing administrative records of a student's absenteeism to his or her teacher's report of that student's absenteeism as it relates to the student's race, ethnicity, gender, and socioeconomic status (REGS). In doing so, I utilize two longitudinal datasets - the Early Childhood Longitudinal Study, Kindergarten Cohort of 1998-99 (ECLS-K) and data from the North Carolina Education Data Research Center (NCERDC).

The ECLS-K provides longitudinal data for a nationally representative sample of students who attended kindergarten in the 1998/1999 school year through the end of their eighth grade year in 2006/2007.[3] Data include administrative records

[3] Approximately 86% of students in the sample who started kindergarten in 1998/1999 matriculated to the eighth grade in 2006/2007. Around 12% were still in the seventh grade, and the remaining 2% were either still in the fourth through sixth grades or beyond the eighth grade. Those students are excluded from the current analysis.

and survey data from teachers, students and parents collected in seven waves of the study. The main dependent variables of interest - teacher subjective reports of student absence - were collected in the eighth grade wave. I therefore limit the sample to students whose teachers were surveyed in the eighth grade wave.

While teacher reports of absenteeism were collected in the eighth grade, administrative data on actual absence were only collected in the third grade and fifth grade waves of the survey. To address this data limitation, I instrument for actual eighth grade absences with two simulated measures of eight grade absences. One simulated measure will be based on administrative reports of absence from the third and fifth grade waves of the ECLS-K. A second simulated measure will be based on a two sample instrumental variables approach using longitudinal administrative records of student attendance from the fifth and eighth grades of two cohorts of students in the NCERDC data - a comprehensive administrative dataset of the universe of public school students in the state of North Carolina. The first cohort of students attended the fifth grade in the 2005/2006 school year and the eighth grade in the 2008/2009 school year. The second cohort of students were a year behind the first, attending fifth grade in the 2006/2007 school year and eighth grade in the 2009/2010 school year. Similar to the ECLS-K sample, I limit the NCERDC sample to students who appear in the data for the complete fifth through eighth grade time period, and who were not retained or did not skip grade levels.[4] Instrumenting for measured eighth grade absence using these simulated measures has the added benefit of reducing potential bias from both random measurement error and simultaneity that would be associated with using actual measured eighth grade absences from the ECLS-K had those records existed.

[4] Approximately 1% of the NCERDC sample of students were either retained or skipped at some point between the fifth and eighth grades. I also eliminate approximately 80 students who switch schools during the school year. Students who switch schools between school years are retained in the sample.

Table 3.1: Sample Distribution of Race, Ethnicity, Gender and Socioeconomic Status

	ECLS-K (%)	NCERDC (%)	
		Cohort 1	Cohort 2
Gender			
Male	51.2	50.1	49.9
Female	48.8	49.9	50.1
Race/Ethnicity			
American Indian	1.4	1.5	1.5
Asian	3.7	2.1	2.3
Black	16.0	27.7	27.0
Hispanic	17.9	8.6	9.4
White	59.1	57.1	56.4
Multi-Racial	2.0	3.0	3.3
Socioeconomic Status			
Below Poverty Level	19.2	–	–
FRL Receipt	32.9	43.3	43.9
Sample Size	7270	85570	86890

All measures are taken as of the eighth grade year.

All percentages for the ECLS-K sample are weighted by the appropriate sampling weights.

American Indian category includes Native Hawaiians

Asian category includes Pacific Islanders

Multi-Racial category includes approximately 10 students for whom no race/ethnicity information was given in the ECLS-K Sample

3.3.1 Sample Demographics

Descriptors of the demographic compositions of both the ECLS-K and NCERDC samples are presented in Table 3.1. Separate statistics are presented for the each cohort of NCERDC students. The ECLS-K sample consists of 7,270 students while each cohort of the the NCERDC sample has over 85,000 students. Gender is fairly evenly split between males and females in all three samples. With regard to race and ethnicity, in all three samples, a majority of students are white - approximately 57 to 59%. In the North Carolina sample, the next largest racial group is black students at about 27%, followed by Hispanic student at between 8 and 9%. In the ECLS-K sample, however, the number of Hispanic students and black students are almost equal in size at 18% and 16% respectively. This is reflective of the proportionately larger black student population in North Carolina relative to the nation as a whole. American Indian students, Asian students, and students of multiple racial/ethnic backgrounds make up the remainder of the sample ranging from a little under 4% to about 1.5%. The North Carolina sample has a larger population of low socioeconomic status (SES) students as proxied by free or reduced lunch (FRL) receipt.[5] About 33% of students from the ECLS-K sample receive free or reduced price lunch while over 43% of the NCERDC students do.

These demographic variables make up the main independent variables of interest. Race and ethnicity are represented in a series of dichotomous variables - *American Indian*, *Asian*, *Black*, *Hispanic*, *Multi-Racial*, and *White* - which are equal to 1 if the student is from the racial or ethnic group indicated, and 0 otherwise. In the analysis that follows, students are identified as Hispanic if they indicated no other racial background, and are identified as multi-racial if they identified a race as well as Hispanic ethnicity. *Male* is a dichotomous variable that is equal to 1 if the student is male and 0 otherwise. Socioeconomic status is captured using free or reduced lunch (FRL) receipt. *FRL* is an indicator of whether the student receives free or reduced price lunch under the federal subsidized lunch program.

In the second part of the analysis, I repeat the estimation for subgroups of teachers in the ECLS-K sample based on teacher characteristics like age, race, and gender. Table 3.2 presents the sample composition of various teacher characteristics by subject taught. While the majority of teachers in all subjects are female, math and science classes have much higher male representation (30% and 37% respectively) when compared with English classes (16%). The racial/ethnic composition of teach-

[5] Students are considered eligible for free or reduced price lunches through the National School Lunch Program based on their parents' income. Income levels at or below a poverty threshold qualify students for the program. Even though students are eligible for free or reduced price lunch, parents may opt out of receipt. Both the NCERDC and ECLS-K samples measure receipt of free or reduced price lunch, as opposed to eligibility.

Table 3.2: Teacher Characteristics by Subject Taught (percent of sample)

	English	Math	Science
Gender			
Male	15.9	30.0	36.9
Female	84.1	70.0	63.1
Race/Ethnicity			
American Indian	0.5	0.6	0.4
Asian	1.1	2.1	2.1
Black	9.1	10.1	8.9
White	80.7	78.2	79.1
Hispanic	5.3	5.5	5.2
Multi	0.8	1.0	0.9
Parent Education			
Less than HS	5.3	9.9	6.9
HS Grad	35.1	35.3	34.3
Some College	10.5	8.9	11.2
College Grad	17.7	17.9	15.9
Graduate School	27.8	24.9	31.6
Age			
35 and Under	35.1	33.3	36.0
36 to 45	21.4	25.9	21.5
46 to 55	20.1	23.7	23.2
56 and Over	23.0	17.1	19.4
Certification			
Fully Certified	83.6	78.9	81.8
Less than Fully Certified	10.8	14.7	18.2
Education			
Bachelor's Degree	20.8	23.8	20.3
Some Graduate School	28.1	27.3	28.3
Graduate Degree	49.0	45.8	51.1

Categories may not sum to 100 because a small number of teachers did not answer some questions.

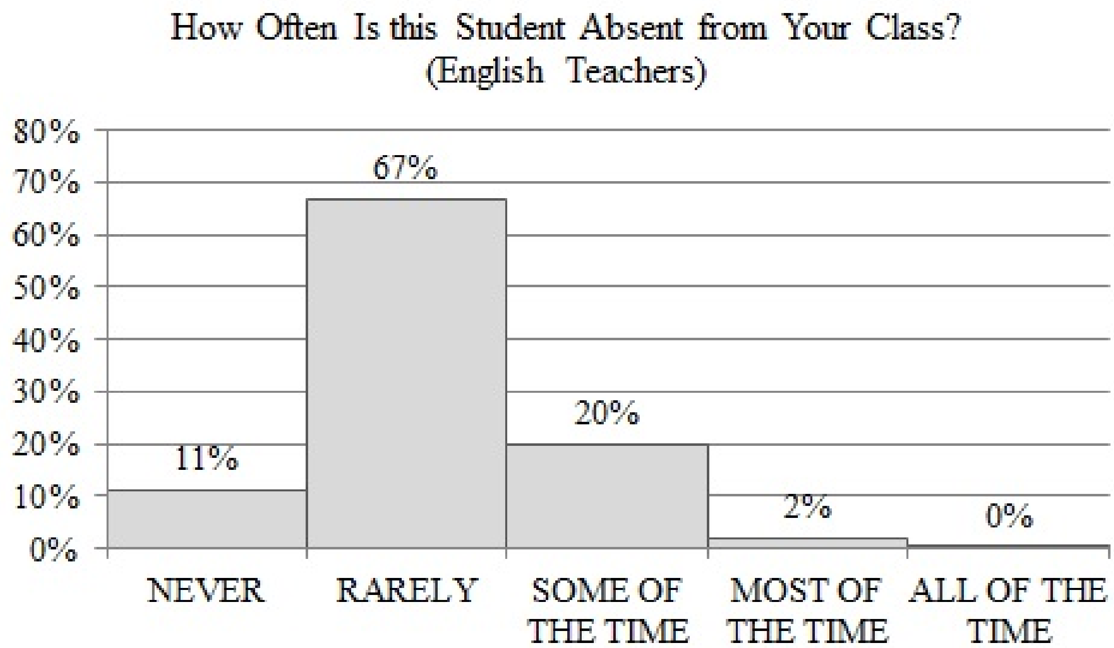

Figure 3.1: English Teacher Perceptions of Student Absenteeism

ers in the sample is comparable across subjects, with white teachers representing the majority at around 80%, followed by black teachers at about 10%, Hispanic teachers at 5% and Asian teachers between 1 and 2%. American Indian and Multi-Racial teachers make up a very small portion of the sample.

Most teachers come from homes where their parents were high school graduates (approximately 35%), followed by homes with parents who had graduate level course work (between 25 and 31%). Very few teachers have parents with less than a high school diploma. The age structure of teachers in the sample is slightly weighted towards younger teachers. Around 80% of teachers have full certification. Finally, most teachers have an advanced degree of some type (between 45 and 50%) with the remainder of the sample split relatively evenly between teachers with bachelor's degrees and teachers with some post-graduate education.

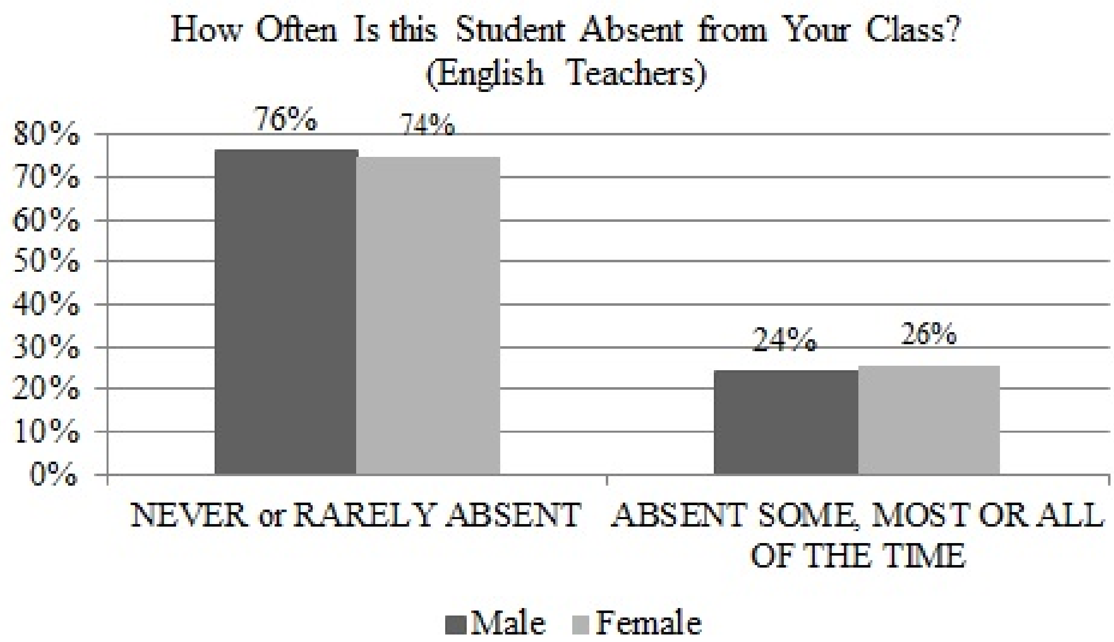

FIGURE 3.2: English Teacher Perceptions of Student Absenteeism by Gender

3.3.2 Attendance Measures

Each student in the ECLS-K sample was evaluated by two teachers - an English teacher, and either a math or a science teacher - who were asked: *"How often is this student absent from your class?"* They were given the options: *"Never," "Rarely," "Some of the time," "Most of the time,"* and *"All of the time."* These constitute the subjective teacher evaluations of a student's classroom attendance which are the main dependent variables of interest. Figure 3.1 presents histograms for English teacher reports of absences. Patterns are similar for science and math teachers. The majority of students (78%) are rated as rarely or never absent. Since much of the variation in teacher reports comes in the difference between being rated as rarely or never absent or being rated as more absent than that, I create an outcome variable to reflect the possible dichotomous nature of the perceptions question. *More Absent* takes a value of 1 if the student is reported to be absent some of the time, most of

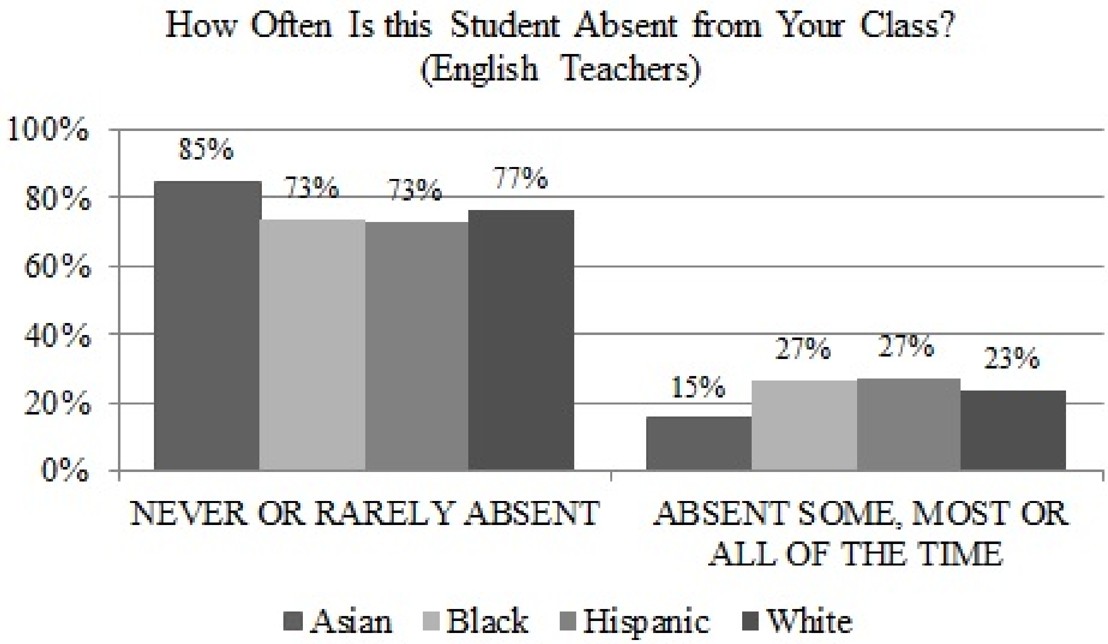

FIGURE 3.3: English Teacher Perceptions of Student Absenteeism by Race and Ethnicity

the time, or all of the time and 0 otherwise.[6]

Comparisons of English teacher reports of student attendance by gender, race and socioeconomic status are presented in Figures 3.2, 3.3, and 3.4 respectively. Again, patterns are similar for math and science teachers. Teacher reports of absence are roughly the same for male and female students. With regard to race, we see again that there are no significant difference in teacher reports of absences between student racial groups with the possible exception of Asian students who are more likely to be reported as never or rarely being absent. Finally, students who receive free or reduced price lunch are slightly more likely to be reported as having absences than those who do not.

[6] I also estimate all models using a categorical variable ranging from 1 if a student is reported as never absent and 5 if a student is reported as always absent. Since the results are qualitatively the same, I choose to present the results from the dichotomous outcome variable as they lend themselves to easier interpretation.

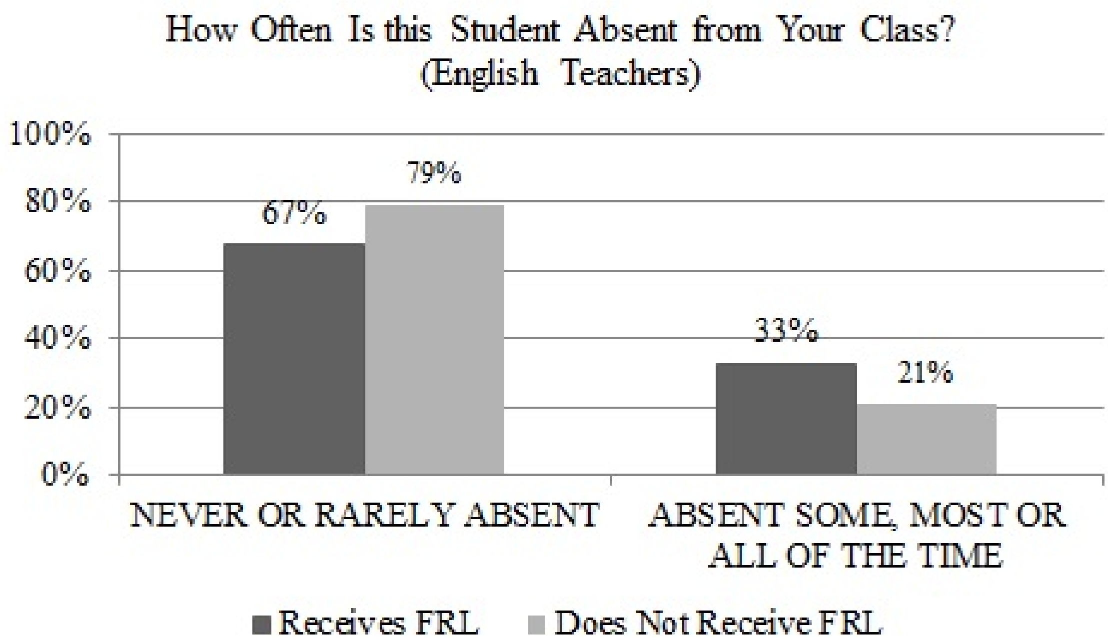

FIGURE 3.4: English Teacher Perceptions of Student Absenteeism by FRL Receipt

Descriptive statistics for actual recorded attendance measures from the ECLS-K sample for the third and fifth grades and from both cohorts of the NCERDC sample for fifth and eighth grades are presented in Table 3.3. The mean values for absences are relatively stable across third and fifth grades in the ECLS-K sample at a little over six total absences. Similar to the ECLS-K sample, students have a little over six total absences in the fifth grade. By the eighth grade, however, the average number of absences increases to between seven and eight. There is substantial variation in the number of absences, and they range from none to up to 200.

Figure 3.5 depicts mean values and 95% confidence intervals for recorded absences in the ECLS-K sample by student demographic characteristics. Students who receive free or reduced price lunch, on average, have significantly higher absences than students who do not. Male and female students display no significant differences in the average number of absences. While the differences do not reach significance, Asian

Table 3.3: Descriptive Statistics for Recorded Attendance

	Obs	Mean	Std.Err/ Std.Dev.	Min	Max
		ECLS-K Sample			
3rd Grade Absences	7270	6.31	0.15	0	200
5th Grade Absences	7270	6.41	0.15	0	180
		NCERDC Sample - Cohort 1			
5th Grade Absences	85570	6.06	5.75	0	170
8th Grade Absences	85570	7.35	7.86	0	140
		NCERDC Sample - Cohort 2			
5th Grade Absences	86890	6.35	5.81	0	150
8th Grade Absences	86890	7.79	8.01	0	130

Standard errors for ECLS-K sample are based on jacknife estimation using appropriate sampling weights.

Standard deviations are reported for the NCERDC sample.

Absences are measured across the entire school year.

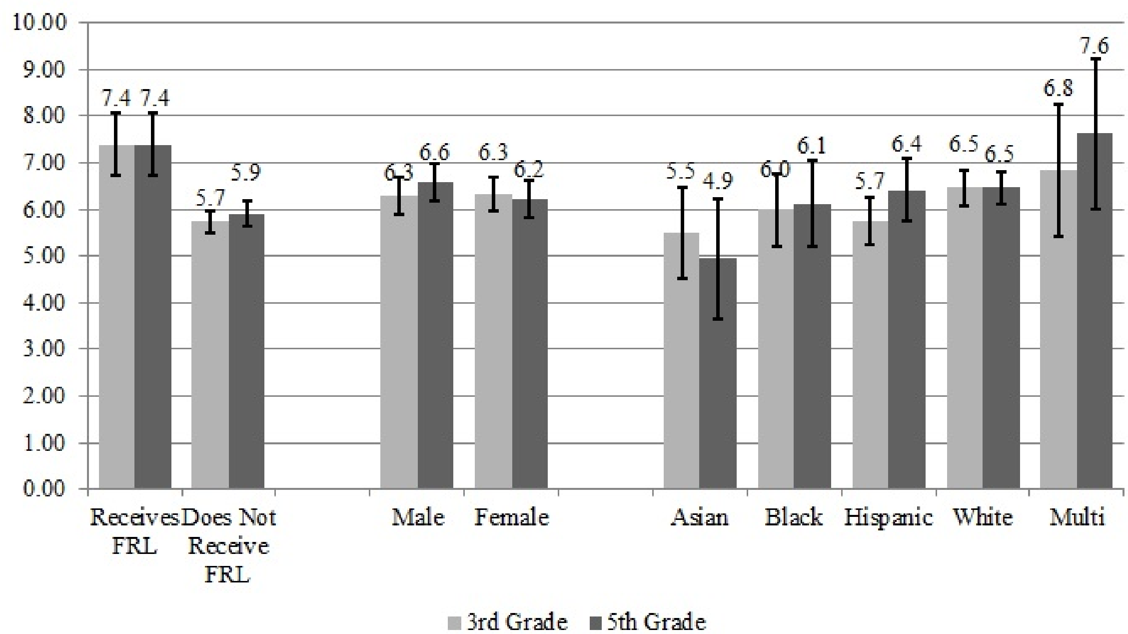

Figure 3.5: Mean Values and Confidence Intervals for Recorded School-Year Absences by Demographic Characteristics - ECLS-K Sample

students have slightly lower absences on average, while multi-racial students have slightly higher absences on average.

Mean values for administratively recorded absences for students in the NCERDC sample by race/ethnicity, gender and socioeconomic status are presented in Figure 3.6 for Cohorts 1 and 2. Again, students who receive free or reduced price lunch have significantly higher absences than non-FRL students. There appears to be no meaningful difference in mean attendance between males and females. Also similar to the ECLS-K sample, Asian students have significantly fewer absences on average than all other students, while multi-racial students have slightly higher absences.

The descriptive statistics seem to indicate that students who receive free or reduced price lunch are more likely to be reported as being absent by their teachers, and those students are also slightly more likely to actually be recorded as absent in both the ECLS-K dataset and the NCERDC dataset. Similarly, Asian students are more likely to be reported as never being absent by teachers, and they tend to have lower average administratively recorded absences than other students. These simple observations would seem to point to behavior more than bias when explaining differences in teacher perceptions of attendance behavior by race/ethnicity or socioeconomic status, however, more rigorous analysis is necessary.

3.3.3 Empirical Strategy

Given that the teacher perception variable, *Rarely Absent*, is dichotomous, I use probit estimation to examine whether student demographic characteristics are associated with better or worse teacher perceptions of student attendance while controlling for independently recorded measurements of actual student attendance. The latent dependent variable underlying the observed dichotomous variable is modeled as follows:

$$pAbsence^*_{ij} \quad = \quad \gamma_0 + \gamma_1 AbsenceRecord_i + \Gamma_2 REGS_i + \varepsilon_i \tag{3.1}$$

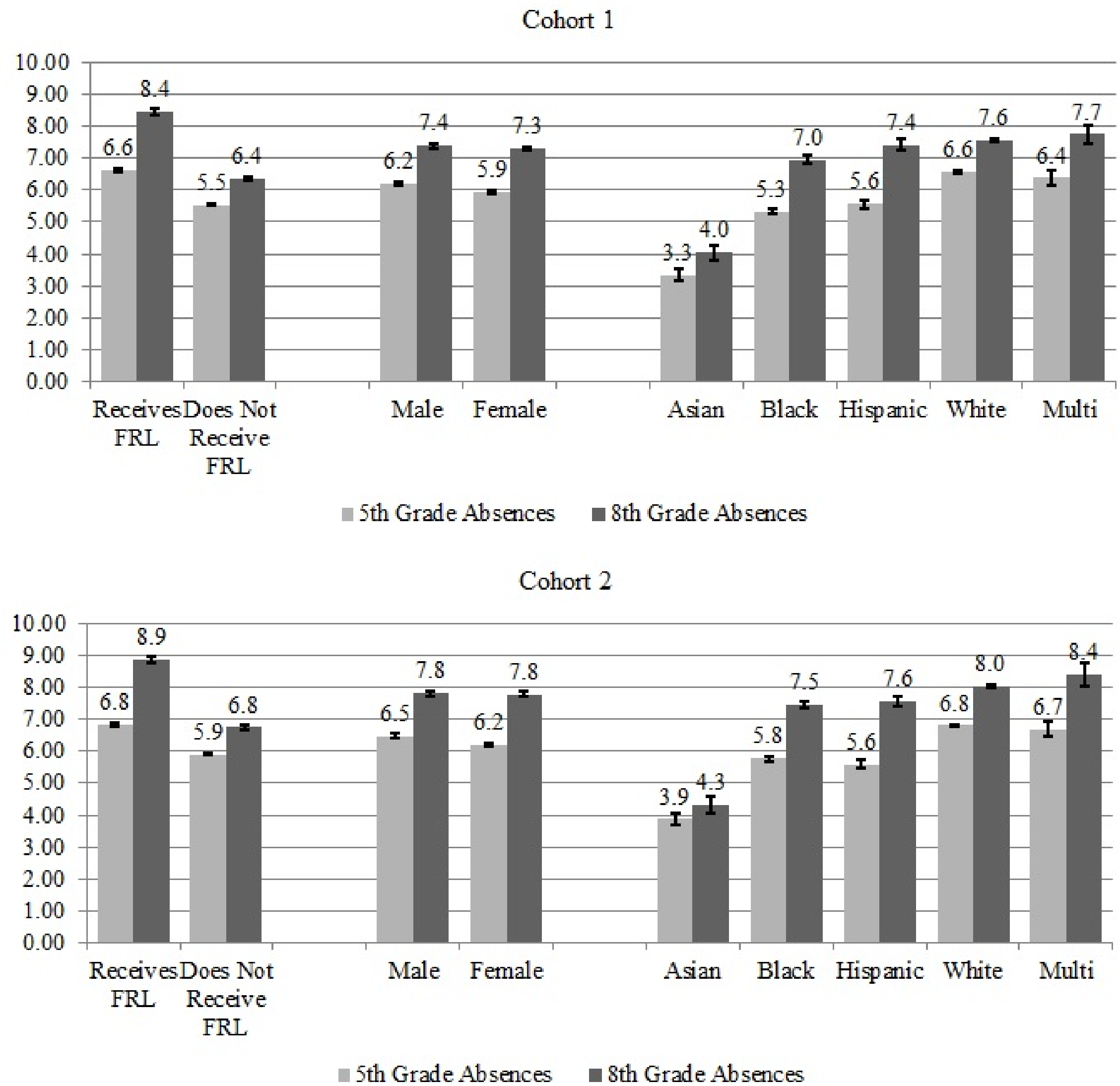

FIGURE 3.6: Mean Values and Confidence Intervals for Recorded School-Year Absences by Demographic Characteristics - NCERDC Sample.

In this specification *pAbsence** is a continuous measure of the teacher *j's* report of student *i's* absences. This measure is unobserved in practice, and instead we observe a dichotomous variable that I assume is related to this latent dependent variable through an estimable threshold value. *AbsenceRecord* is an administrative record of the actual number of recorded absences for student *i*, and *REGS* is a vector of the student specific demographic variables - *American Indian*, *Asian*, *Black*, *Hispanic*, *Multi-Racial*, *Male*, and *FRL* - as well as interaction variables for race-gender combinations, race-FRL combinations, and gender-FRL combinations. ε is an independent student level error term. The test as to whether any bias exists in teacher perceptions of student absence will be whether any of the coefficients in Γ_2 are significantly different from zero.

There are three issues with estimating this model. First, the administrative record of actual absences may suffer from random measurement error, in which case, it may result in attenuated parameter estimates. Second, actual student absenteeism in the eighth grade may be endogenously related to teacher reports of student absence in the same year. For example, a student who believes that his teacher is biased against him may feel discouraged by this belief such that he attends that teacher's class less often. Third, and perhaps most important, the ECLS-K dataset does not include administrative records of actual absence and for eighth grade students. The dataset does, however, include student-level administrative records of fifth grade and third grade absence.

In order to address these three concerns, I propose estimating a variation of a two sample instrumental variables technique that proceeds as follows:

Stage 1: Use the NCERDC attendance measures to obtain an estimate relating student attendance in one year with student attendance in another by regressing measured eighth grade absences ($absences8_i^{NC}$) on measured fifth grade absences ($absences5_i^{NC}$) and student demographics ($REGS_i^{NC}$) as follows:

$$absences8_i^{NC} = \beta_0 + \beta_1 absences5_i^{NC} + \Phi REGS_i^{NC} + \varepsilon_i \tag{3.2}$$

Stage 1 (alternative strategy:) As an alternative strategy, I will use the previous waves of the ECLS-K data and regress fifth grade absences ($absences5_i^{EC}$) as a function of third grade absences ($absences3_i^{EC}$) and student demographics

$(REGS_i^{EC})$:[7]

$$absences5_i^{EC} = \beta_0 + \beta_1 absences3_i^{EC} + \Phi REGS_i^{EC} + \varepsilon_i \tag{3.3}$$

Stage 2: Use the estimated parameters from Stage 1 to simulate eighth grade absences for the ECLS-K sample of eighth graders:

$$\widehat{absences}8_i = \widehat{\beta_0^{NC}} + \widehat{\beta_1^{NC}} absences5_i^{EC} + \widehat{\Phi^{NC}} REGS^{EC} \tag{3.4}$$

Similarly, for the alternative strategy:

$$\widehat{absences}8_i = \widehat{\beta_0^{EC}} + \widehat{\beta_1^{EC}} absences5_i^{EC} + \widehat{\Phi^{EC}} REGS^{EC} \tag{3.5}$$

Stage 3: Conduct probit estimation, modeling the latent dependent variables for absences as:

$$pAbsence_{ij}^{*} = \gamma_0 + \gamma_1 \widehat{absences}8_i + \Gamma_2 REGS_i + \varepsilon_i \tag{3.6}$$

Using simulated measures of eighth grade absences reduces the variance in the measured attendance variables that is due mainly to noise and not to true differences in the underlying behavior. This estimation strategy also addresses the concerns of random measurement error, and the potential endogeneity of measured eighth grade attendance.

A key assumption for the validity of the strategy that uses the ECLS-K data is that the attendance differences between the reference groups and comparison groups stay constant between the third and fifth grade time periods and the fifth and eighth grade time periods. As an example, when comparing males and females, the validity of the simulated estimates of eighth grade absences is based on the assumption that if males tend to have 2 more absences on average in the fifth grade than females,

[7] In the estimation that follows I also include a square term for absences in the first stage estimation, for both strategies, in order to allow for the possibility that the relationship between past and future attendance is curvilinear.

controlling for third grade absences, then they also tend to have 2 more absences on average than females in the eighth grade, controlling for fifth grade absences. If the time pattern of absences varies between males and females, any estimates of differences in teacher reports between males and females, after accounting for the simulated measures, could be due to differences in actual behavior between males and females that is not accounted for in the simulated measures because the simulated measures assume a time invariant relationship. The same assumption is not necessary for the estimates attained using the NCERDC sample, since those estimates are based on the fifth to eighth grade time period. Thus one check on the validity of the ECLS-K estimates will be if they are comparable to the estimates that are calculated using the NCERDC sample. This is one benefit of using two strategies to simulate eighth grade administrative attendance records.

I will also estimate the model for subgroups of teachers to examine heterogeneous effects among teachers from different backgrounds. While this subgroup analysis will allow me to identify potential differences in teacher reports between different subgroups, the subgroup estimates will not allow me to identify heterogeneous effects within subgroups. Thus the subgroup estimates are designed to measure average effects for each subgroup.

3.4 Results

In this section, I will first present the results from the first stage results for absences in both the ECLS-K and NCERDC samples. Recall that for the first stage in the ECLS-K sample, fifth grade absences are regressed on third grade absences while controlling for REGS in order to obtain an estimate of how absences persist over time. In the NCERDC sample, eighth grade absences are regressed on fifth grade absences while controlling for REGS. Next, I will present the main results beginning with "raw" estimates in which teacher reports of student attendance are simply estimated as a function of REGS, not controlling for the simulated measures of actual absences. These "raw" estimates will serve as a comparison for the instrumented estimates, which I present next. Finally, I will present results for subgroups of teachers based on teacher characteristics such as age, race, gender, education level, and teaching credentials.

3.4.1 First Stage Results

The first stage results are presented in Table 3.4. The first column presents estimates derived from estimating equation (3.3) which relates fifth grade absences to third grade absences from the ECLS-K sample. Similarly, in the second column, estimates are derived from estimating equation (3.2), which relates eighth grade absences to fifth grade absences from the NCERDC sample. Both regressions also include a square term for the specific attendance regressor included. For example, in the first column, a square term for third grade absences in the ECLS-K sample is included.[8] These results demonstrate evidence of strong instruments. The F-statistic for absences in the ECLS-K sample is 81 and the corresponding F-statistic in the NCERDC sample is over 3,000. Both are well above the rule of thumb threshold F-statistic of 10 suggested by Stock et al. (2002) for determining the strength of an instrument.

Turning to the magnitudes of the estimates, the ECLS-K absence estimates imply that every three absences in the third grade are associated with having about two absences in the fifth grade. The estimated relationship between fifth grade absences and eighth grade absences in the NCERDC sample is comparable, implying that roughly every three absences in the fifth grade is associated with two absences in the eighth grade. Also, the coefficient estimates on the squared absence terms imply

[8] I also estimated a cubic and fourth order polynomial functional form for the first stage. The quadratic form exhibited the best fit.

Table 3.4: First Stage Results

	ECLS-K	NCERDC
	5th Grade Absences	**8th Grade Absences**
3rd Grade Absences	0.644***	–
	(0.051)	
5th Grade Absences	–	0.677***
		(0.011)
Squared Absences	-0.004***	-0.003***
	(0.000)	(0.000)
Black	-0.807	-0.750***
	(0.760)	(0.087)
Hispanic	-0.020	0.590***
	(0.613)	(0.169)
Asian	-0.465	-1.301***
	(0.452)	(0.131)
American Indian	-1.847	1.078***
	(1.635)	(0.298)
Multi-Racial	1.406	-0.031
	(1.471)	(0.133)
Male	0.415	-0.216***
	(0.253)	(0.044)
FRL	2.271***	1.893***
	(0.741)	(0.099)
Constant	2.238**	2.885***
	(0.316)	(0.059)
Observations	7,760	163,070
F-Statistic	81.24	3325.16
R-squared	0.235	0.275

Two-way demographic interaction terms are also included in both models
Jacknifed standard errors in parentheses for ECLS-K sample
Robust standard errors in parentheses for NCERDC Sample
*** p<0.01, ** p<0.05, * p<0.1

that the relationship between prior and future absences is marginally decreasing. A student would have to have over 80 absences in the third grade for the negative marginal effect to have a larger impact on fifth grade absences than the increasing first-order effect, based on the ECLS-K estimates. Similarly, based on the NCERDC estimates, a student would have to have over 125 absences in the fifth grade for the negative marginal effect to have a larger impact on eighth grade absences than the first-order effect. Given that students have between six and seven absences on average, and that 95% of students in the distribution have fewer than 30 absences, the majority of students observe a strong positive relationship between prior absences and future absences. There are outliers who exhibit more than 80 absences in the ECLS-K sample and more than 125 absences in the NCERDC sample. However, it is reasonable to expect that for many of these students, these extreme absence values may be due to some student year-specific factors that would not carry over from year to year. Thus, we would expect that for these students, the relationship of prior absences to later absences would not be as strong. The estimates support this expectation.

3.4.2 Main Estimation Results

Marginal effects derived from probit estimations relating teacher reports of student absences to student demographic characteristics are presented in Tables 3.5 - 3.7 for English, math and science teachers respectively. The marginal effects are calculated at the mean value of the continuous variable *Simulated Absences*, and are calculated as a discrete change from 0 to 1 for the dichotomous variables. In the first column, are "raw" results which relate student demographic characteristics to teacher reports of absences without taking into account actual absences. For all three subjects, Asian students are between 5 to 12% less likely to be reported as being absent than the reference group - white students. The results for FRL receipt are also consistent

Table 3.5: Marginal Effects from Probit Estimation of MoreAbsent on Simulated Absences and Student Demographics English Teachers

	Raw	**ECLS-K**	**NCERDC**
Simulated Absences	–	0.025***	0.023***
		(0.002)	(0.002)
Black	0.004	0.046	0.001
	(0.029)	(0.031)	(0.028)
Hispanic	0.011	0.026	0.015
	(0.023)	(0.022)	(0.023)
Asian	-0.074**	-0.011	-0.023
	(0.035)	(0.038)	(0.035)
American Indian	0.182**	0.159*	0.049
	(0.090)	(0.086)	(0.069)
Multi-Racial	0.170***	0.094	0.112**
	(0.058)	(0.059)	(0.054)
Male	0.003	-0.006	-0.015
	(0.015)	(0.016)	(0.016)
FRL	0.122***	0.076***	0.077***
	(0.026)	(0.025)	(0.025)
Observations	7,270	7,270	7,270

Two-way demographic interaction terms are also included in all models
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

across all three subjects. Students who receive free or reduced price lunch are about 12 to 20% more likely to be reported as absent than non-FRL students. Math and science teachers tend to perceive male students more positively than female students, though not significantly so. Science teachers also tend to report that the attendance of Hispanic students is more favorable than that of white students on average.

These raw results do not take into account students' actual attendance behavior. Thus, Asian students, for example, may be reported by teachers as being less absent because they actually are absent less often. The second and third columns present results incorporating the simulated eighth grade absences using the ECLS-

Table 3.6: Marginal Effects from Probit Estimation of MoreAbsent on Simulated Absences and Student Demographics Math Teachers

	Raw	ECLS-K	NCERDC
Simulated Absences	–	0.029***	0.026***
		(0.003)	(0.003)
Black	-0.042	0.012	0.036
	(0.044)	(0.050)	(0.052)
Hispanic	0.062	0.052	0.062*
	(0.041)	(0.038)	(0.035)
Asian	-0.052	0.022	-0.002
	(0.048)	(0.049)	(0.048)
American Indian	0.127	0.112	-0.020
	(0.096)	(0.110)	(0.092)
Multi-Racial	0.045	-0.066	0.025
	(0.084)	(0.051)	(0.067)
Male	-0.030	-0.037*	-0.043**
	(0.021)	(0.021)	(0.019)
FRL	0.138***	0.110***	0.078**
	(0.030)	(0.030)	(0.030)
Observations	7,270	7,270	7,270

Two-way demographic interaction terms are also included in all models
Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

K and NCERDC data respectively. The first things to notice are the coefficients on the simulated eighth grade absence measures. As we would expect, each additional administratively recorded absence significantly increases the likelihood of being reported by a teacher as being absent. With regard to the coefficients on the demographic variables, a few patterns emerge. First, both the magnitude and significance levels of the coefficients on Asian students decrease in all subjects. This supports the hypothesis that the perceptual boon that Asian students have in the raw numbers is due more to behavior than to bias. Asian students are reported by teachers as being less absent because they actually tend to be less absent on average.

Conversely, the coefficients on FRL receipt remain highly significant even after controlling for the instrumented measures of recorded absences, though they decrease slightly in magnitude. Students who receive free or reduced price lunch are anywhere from 8 to 13% more likely than students who do not receive free or reduced price lunch to be reported as being absent by their teachers. This indicates that, while students who receive free or reduced price lunch have more absences on average than students who do not receive it, the difference is not great enough to explain away the less favorable teacher reports of FRL-receiving students' absenteeism. In other words, there is evidence of bias against students who receive free or reduced price lunch. This bias appears to be strongest among science teachers.

The results also indicate that math teachers report that male students are less absent than female students even after controlling for simulated eighth grade absences. Male students are about 4% less likely to be reported as having absences than female students by math teachers in the both the ECLS-K and NCERDC samples.

In order to explore heterogeneous treatment effects, two-way interaction variables for race-gender pairs, FRL-gender pairs and FRL-race pairs are included in each model. Given that math teachers exhibit a positive bias towards male students, it might be interesting to know if there are differences in this effect by race. Figure 3.7 presents the marginal effects of math teacher reports of male absence relative to female absence by race. The data underlying the figure are presented in Table B.1 in Appendix B. The largest, and most statistically significant results are for black males relative to black females. Black males are roughly 14 to 15% less likely to be viewed as absent by math teachers than black females. No other results in the figure are statistically different from zero.

Marginal effects from probit estimation of teacher reports of absences for students who receive free or reduced price lunch relative to those who do not are presented in Figure 3.8 by race. The data underlying the figure are presented in Table B.2 in

Table 3.7: Marginal Effects from Probit Estimation of MoreAbsent on Simulated Absences and Student Demographics Science Teachers

	Raw	**ECLS-K**	**NCERDC**
Simulated Absences	–	0.023***	0.021***
		(0.003)	(0.003)
Black	0.014	0.033	0.018
	(0.063)	(0.062)	(0.061)
Hispanic	-0.077**	-0.036	-0.065**
	(0.033)	(0.032)	(0.030)
Asian	-0.122***	-0.051	-0.063*
	(0.038)	(0.039)	(0.037)
American Indian	0.037	-0.006	-0.037
	(0.112)	(0.078)	(0.075)
Multi-Racial	0.195*	0.176*	0.154*
	(0.104)	(0.104)	(0.093)
Male	-0.005	-0.021	-0.006
	(0.025)	(0.023)	(0.024)
FRL	0.197***	0.133***	0.123***
	(0.039)	(0.038)	(0.036)
Observations	7,270	7,270	7,270

Two-way demographic interaction terms are also included in all models
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix B. One clear pattern is that low-income white students are consistently and significantly viewed less favorably than more affluent white students with respect to their absenteeism by English, math and science teachers even after controlling for simulated absences. They are anywhere from 7% to 21% more likely to be viewed as absent. Similarly, low-income Asian students are viewed significantly less favorably than more affluent Asian students by English and science teachers, but not by math teachers, with marginal effects ranging from 12% to 23%. There are no clear patterns when looking at low-income black and Hispanic students. With respect to race and ethnicity, math teachers tend to view low-income Hispanic students as being

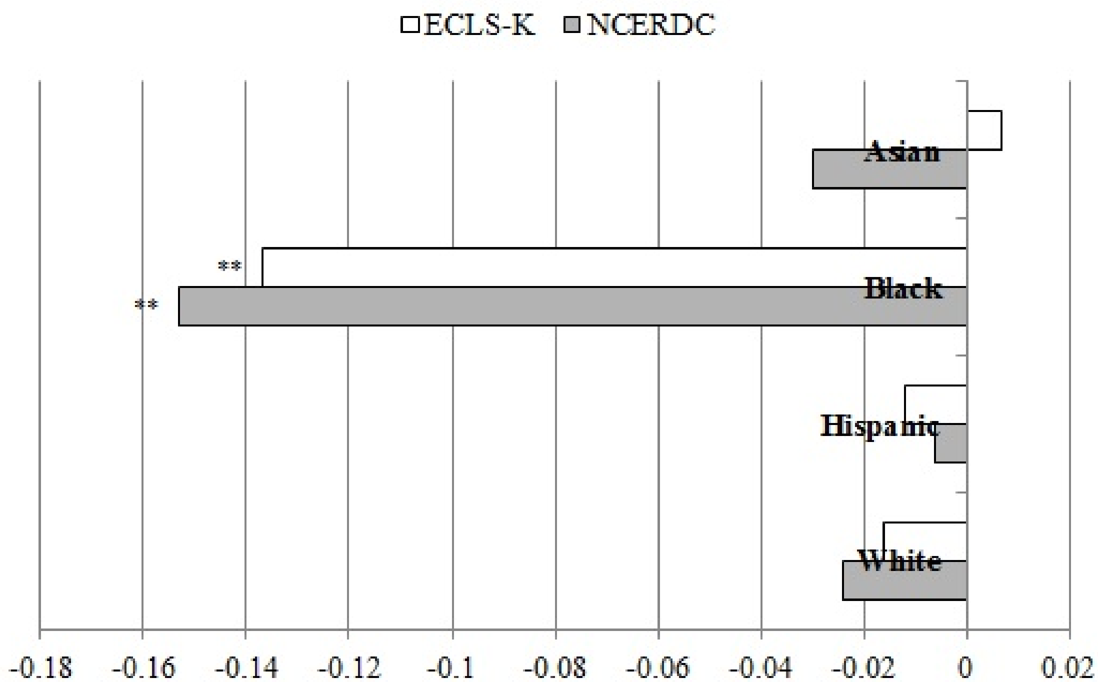

Figure 3.7: Marginal Effects of Probit Estimation of Math Teacher Reports of Male Absence Relative to Female Absence, by Race (**p<0.05)

significantly more absent than more affluent Hispanic students with an increased likelihood of between 10 and 15%. Based on the ECLS-K estimates, math teachers also tend to report low-income black students as about 13% more likely to be absent than more affluent black students. The corresponding NCERDC estimate is not significantly significant.

3.4.3 Teacher Subgroup Results

Given the potential evidence of bias found in the main results, it would be interesting to examine whether there is heterogeneity among the results for different subgroups of teachers based on their own characteristics. For example, are male teachers more likely to express a positive bias towards male students? Or are Hispanic teachers likely to express a positive bias towards Hispanic students? In order to explore

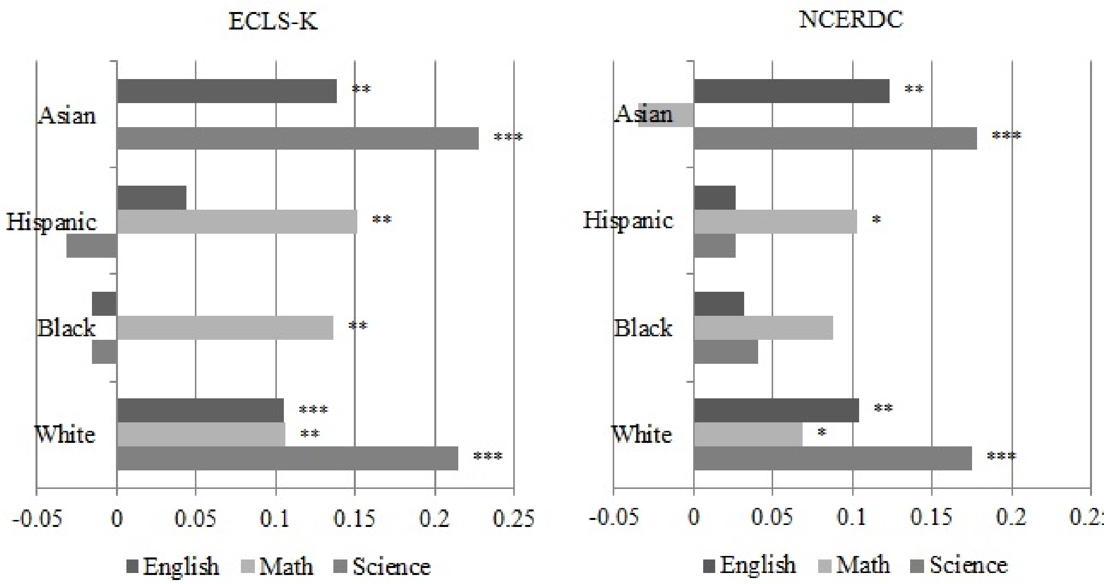

FIGURE 3.8: Marginal Effects of Probit Estimation of Teacher Reports of FRL Student Absence Relative to Non-FRL Student Absence, by Race (*** $p<0.01$, ** $p<0.05$, ** $p<0.10$)

these questions, I estimated the same equations as in the main results section, while restricting the sample to 19 different subgroups of teachers. The results between the two samples are so similar that I only present the ECLS-K results here.

With 19 different subgroups of teachers, seven demographic variables of interest for students, and the three different subjects taught, presenting tables for all of the results would be cumbersome and impractical. Instead, I present the results as a series of summary figures - Figures 3.9, 3.10, and 3.11 - for English teachers, math teachers and science teachers respectively. In the figures, each row represents a separate estimation of equation (3.6), with the sample restricted to the subgroup of teachers specified at the head of the row. A plus (+) sign indicates that the coefficient on the variable at the top of the column is positive, meaning there is an increased likelihood that a teacher reports that group as absent even after controlling for instrumented records of actual absence. Conversely a minus (-) sign indicates a

	Sample Size	Male	American Indian	Asian	Black	Hispanic	Multi-Racial	FRL
Gender								
Male	1240	+	+	-	-	-	-	+
Female	6490	-	-	-	+	+	+	+
Race/Ethnicity								
Black	540	+	-	-	+	-	-	-
White	6570	-	-	-	-	+	+	+
Hispanic	370	-	-	-	-	-	**N/A**	-
Parent Education								
Less than HS	420	-	+	-	-	-	**N/A**	+
HS Grad	2930	-	+	+	-	+	-	+
Some College	810	-	+	-	+	-	+	+
College Grad	1370	+	-	-	+	-	+	+
Graduate School	2120	+	-	-	+	+	-	+
Age								
35 and Under	2700	-	+	-	+	+	+	+
36 to 45	1720	-	+	-	+	+	+	+
46 to 55	1640	-	-	+	+	-	-	+
56 and Over	2390	+	+	-	-	-	+	+
Certification								
Fully Certified	6570	-	-	-	-	+	+	+
Less than Fully Certified	880	-	+	-	+	-	+	-
Education								
Bachelor's Degree	1650	-	+	-	+	+	+	-
Some Graduate School	2170	-	-	-	-	-	-	+
Graduate Degree	3900	-	-	-	-	-	+	+

Each row represents a separate probit estimation of equation 3.6 for the subgroup of teachers specified at the head of the row. A plus (+) sign indicates that the coefficient on the variable at the top of the column is positive, meaning there is an increased likelihood that a teacher reports that group as being absent. Conversely a minus (-) sign indicates a decreased likelihood that the group is reported as absent. A cell is shaded if the coefficient is significantly different than 0. The darkest shading represents p<0.01, the medium shading represents p<0.05 and the lightest shading represents p<0.10.

Figure 3.9: Directional Effect of Probit Estimation of MoreAbsent on Student Demographic Characteristics - English Teacher Subgroups

decreased likelihood that the group is reported as absent even after controlling for instrumented record of actual absence. Thus, a positive sign can be interpreted as teachers in that subgroup reporting a group as more absent without a basis for that report in actual behavior. Finally, a cell is shaded if the coefficient is significantly different than zero. The darkest shading represents a p-value<0.01, the medium shading represents $p<0.05$ and the lightest shading represents $p<0.10$ in a test of significance.

Turning first to the results for English teachers in Figure 3.9, males are consistently reported as being less absent than females, except by male teachers. The bias is significant among female teachers, white teachers, teachers whose parents have some college education, and teachers between the ages of 36 and 45. Thus, teachers do not appear to hold positive biases towards students of the same gender, and in the case of female teachers they actually hold a significantly positive bias towards males students. Asian students are also consistently reported as being less absent by teachers, significantly so among 5 of the 19 teacher subgroup categories. No other clear patterns emerge with respect to the other race/ethnicity variables, but it is of interest to note that black students are viewed significantly less favorably than white students by black teachers. Finally, students who receive free or reduced price lunch are consistently reported as being more absent than non-FRL students with the exception of four teacher subgroups - black and Hispanic teachers, teachers with only a Bachelor's degree, and teachers who are less than fully certified. The teacher categories in which the bias estimates against FRL receiving students reach statistical significance are some of the categories with the largest populations of teachers - female teachers, white teachers, fully certified teachers, teachers with higher than a bachelor's degree, and teachers whose parents have a high school degree.

Among math teachers (Figure 3.10), male students are reported to have less absences than female students across the board, again significantly so among female

	Sample Size	Male	American Indian	Asian	Black	Hispanic	Multi-Racial	FRL
Gender								
Male	1380	-	-	-	+	+	-	+
Female	5480	-	-	+	+	+	-	+
Race/Ethnicity								
Black	500	-	N/A	N/A	+	+	-	-
White	7070	-	-	+	+	-	-	+
Hispanic	400	-	N/A	-	+	+	N/A	+
Parent Education								
Less than HS	600	-	-	+	+	+	N/A	+
HS Grad	3020	-	-	-	+	+	-	+
Some College	840	-	-	-	+	-	+	+
College Grad	1420	-	-	N/A	-	-	+	+
Graduate School	2350	-	+	+	+	-	-	+
Age								
35 and Under	2720	-	-	-	+	+	-	+
36 to 45	2130	-	-	-	+	+	-	+
46 to 55	2110	-	+	+	-	-	-	+
56 and Over	2120	-	+	-	+	+	+	+
Certification								
Fully Certified	6790	-	-	-	+	+	-	+
Less than Fully Certified	1130	-	N/A	+	+	+	-	+
Education								
Bachelor's Degree	1690	-	-	-	-	+	-	+
Some Graduate School	2370	-	-	+	+	+	-	+
Graduate Degree	4260	-	-	-	+	+	+	+

Each row represents a separate probit estimation of equation 3.6 for the subgroup of teachers specified at the head of the row. A plus (+) sign indicates that the coefficient on the variable at the top of the column is positive, meaning there is an increased likelihood that a teacher reports that group as being absent. Conversely a minus (-) sign indicates a decreased likelihood that the group is reported as absent. A cell is shaded if the coefficient is significantly different than 0. The darkest shading represents $p<0.01$, the medium shading represents $p<0.05$ and the lightest shading represents $p<0.10$.

FIGURE 3.10: Directional Effect of Probit Estimation of MoreAbsent on Student Demographic Characteristics - Math Teacher Subgroups

	Sample Size	Male	American Indian	Asian	Black	Hispanic	Multi-Racial	FRL
Gender								
Male	1390	-	+	-	-	-	+	+
Female	5480	-	-	-	+	-	+	+
Race/Ethnicity								
Black	500	+	N/A	N/A	-	-	N/A	+
White	7080	-	-	-	+	-	+	+
Hispanic	400	+	N/A	-	-	-	N/A	+
Parent Education								
Less than HS	640	-	+	N/A	+	-	N/A	+
HS Grad	2990	-	-	-	-	-	-	+
Some College	840	+	-	-	-	-	+	+
College Grad	1470	+	-	-	+	+	+	+
Graduate School	2320	-	+	+	+	-	+	+
Age								
35 and Under	2720	+	-	-	-	-	+	+
36 to 45	2110	-	+	-	+	+	+	+
46 to 55	2110	+	-	-	-	-	+	+
56 and Over	2150	-	-	+	+	-	+	+
Certification								
Fully Certified	6790	-	-	-	-	-	+	+
Less than Fully Certified	1160	+	-	-	+	-	-	+
Education								
Bachelor's Degree	1710	-	-	-	+	-	+	+
Some Graduate School	2390	+	-	-	+	-	+	+
Graduate Degree	4240	-	-	-	-	-	-	+

Each row represents a separate probit estimation of equation 3.6 for the subgroup of teachers specified at the head of the row. A plus (+) sign indicates that the coefficient on the variable at the top of the column is positive, meaning there is an increased likelihood that a teacher reports that group as being absent. Conversely a minus (-) sign indicates a decreased likelihood that the group is reported as absent. A cell is shaded if the coefficient is significantly different than 0. The darkest shading represents p<0.01, the medium shading represents p<0.05 and the lightest shading represents p<0.10.

FIGURE 3.11: Directional Effect of Probit Estimation of MoreAbsent on Student Demographic Characteristics - Science Teacher Subgroups

teachers. American Indian students are also consistently reported as having fewer absences by the teacher subgroups, with the exception of older teachers. Similar to the English teacher results, Asian students are still viewed more positively on the whole than white students, but the results indicate less positive bias and fewer significant results than among English teachers. Black and Hispanic students are reported as having more absences among more categories of math teachers than of English teachers. Also of note is that Hispanic teachers tend to report that Hispanic students have significantly more absences than white students and black teachers report that black students have more absences than white students, but not significantly so. Finally, FRL-receiving students are still reported as having more absences across the board, with perhaps the exception of black teachers.

With respect to science teachers (Figure 3.11), males are not as likely to be reported as having fewer absences than females as they are among English and math teachers. Asian students and Hispanic students, however, are consistently more likely to be reported as being less absent, and Hispanic teachers tend to report that Hispanic students are significantly less likely to be absent than white students. As with the math and English teacher results, students who receive free or reduced price lunch are viewed as having more absences than non-FRL students across all teacher subgroups. Thus, as it pertains to absences, teachers do not tend to view students who are of the same race or same gender more positively than other students, with the exception of Hispanic science teachers.

3.5 Robustness Checks

The results presented in this paper are robust to many different specifications. First, including or omitting the squared attendance term in the first stage functional form does not alter the subsequent results, however, since the hypothesis tests indicate that including the square term explains more of the variation in attendance behavior, that is the specification I choose to present in the paper.

Second, the ECLS-K sample results are robust to annualizing the coefficients achieved from the first stage in order to calculate the simulated instruments in the second stage. The first stage estimates are based on the attendance differences between third and fifth grade, a two year time span, while the simulated measure of eighth grade absences is based on extrapolating out three years from fifth grade. To account for the additional time period, annualized coefficients were used to simulate the eighth grade absences, and results were not significantly altered.

Third, estimates are robust to estimating separate first stage equations and subsequent simulations for comparison groups. For example, I regress the first stage separately for males and females, allowing the coefficients on all first stage regressors to vary by gender. I then calculate the simulated measures for males and female separately, using the estimated coefficients from the separate first stages. This would be akin to a fully interacted model. The results are robust to this specification as well.

As a fourth check of robustness, instead of including the continuous simulated 8th grade absence variable, I create simulated absence threshold variables for threshold values of 2, 4, 10, 15, and 20 absences which are equal to one if the simulated absences are above the threshold and 0 otherwise. I then control for these simulated thresholds in the third stage of the estimation instead of the continuous variable. The idea behind this strategy is that teachers may not be thinking about a student's actual number of total absences when asked to answer the question of how absent a student is, but instead, they may have threshold values in mind, above which, they consider a student as being more absent. While the magnitudes of the point estimates change slightly, the direction and significance of the main effects are consistent. Thus, the results are also robust to this threshold specification.

A key robustness check on the ECLS-K specification is how similar the ECLS-

K results are to the NCERDC results. The concern that the ECLS-K estimates could be attributed to time-varying differences in the behavior of comparison groups between third to fifth grade and fifth to eighth grade does not apply to the NCERDC estimates, since those are based on the fifth to eighth grade time period. That the ECLS-K estimates are similar in magnitude and significance to the NCERDC estimates should be further evidence in support of their validity.

Finally, I perform a sensitivity analysis to address concerns that the results may be biased by sample attrition in both the ECLS-K and NCERDC samples. The sampling weights in the ECLS-K survey adjust for attrition due to lost follow-up of students between the third and eighth grades; however, about 12% of students are excluded from my estimation sample because they did not matriculate to the eighth grade due to grade retention. A concern might be that those students are disproportionately likely to be absent, and therefore, excluding them from the estimation may lead to biased estimates in the first stage regression. For example, if students who receive free or reduced price lunch are more likely to be retained, and therefore not included in my initial estimation, the simulated measure for FRL-student absences in the eighth grade could be lower than the true measure of their absences if the retained students are also more likely to be absent on average. Similarly in the NCERDC sample, attrition may be due to both grade retention, and to students leaving the state or leaving the public school system.

Table 3.8 presents the sample demographics of students who do not matriculate to the eighth grade in both the ECLS-K and NCERDC samples. For comparison, I also reproduce the sample demographics of students in the main estimation sample with the two NCERDC cohorts averaged together. The students who leave the sample before the eighth grade are slightly more likely to be male than female. There are no meaningful differences by race or ethnicity. However, students who leave the sample tend to be significantly and meaningfully different with respect to socioeconomic

Table 3.8: Distribution of Race, Ethnicity, Gender and Socioeconomic Status of Sample Attriters

	ECLS-K (%)		NCERDC (%)	
	Attriters	Estimation Sample	Attriters	Estimation Sample
Gender				
Male	55.0	51.2	54.6	50.0
Female	45.0	48.8	45.3	50.0
Race/Ethnicity				
American Indian	0.5	1.4	1.1	1.5
Asian	2.8	3.7	2.7	2.2
Black	16.0	16.0	26.3	27.3
Hispanic	18.8	17.9	11.8	9.0
White	58.2	59.1	54	56.7
Multi	2.0	2.0	4.0	3.2
Socioeconomic Status				
Below Poverty Level	24.9	19.2	–	–
FRL	46.1	32.9	47.9	43.6
Sample Size	860	7270	29890	172460

All percentages for the ECLS-K sample are weighted by the appropriate sampling weights.

American Indian category includes Native Hawaiians

Asian category includes Pacific Islanders

Multi-Racial category includes 10 students for whom no race/ethnicity information was given in the ECLS-K Sample

status, with low-income students more likely to leave. Given that students who leave the sample also tend to have higher initial absences (around 7.7 compared to 6.3 for non-attriters), there may be concern that the differential attrition of low-income students may be driving the main result, namely, that teachers exhibit bias against low-income students with respect to absence reporting.

In order to address this concern, I conduct a sensitivity analysis in which I assign all students with missing values for eighth grade absences due to sample attrition to have 25 absences if they receive free or reduced price lunch in the initial year, and to have 0 absences if they do not receive free or reduced price lunch in the initial year. Having more than 25 absences places a student outside of the 95th percentile of the absence distribution. Therefore assuming that all missing low-income students have such a high value for absences decreases the odds of finding bias in teacher reports of their absence behavior. I also repeat the exercise using 50 absences. Results are presented in Table 3.9.

Table 3.9: Marginal Effects from Probit Estimation of MoreAbsent on Total Absences and FRL Status by Varying Imputations of Absences for Missing Values

Specification	ECLS-K			NCERDC		
	English	Math	Science	English	Math	Science
Main Estimates	0.076***	0.110***	0.133***	0.077***	0.078**	0.123***
	(0.025)	(0.030)	(0.038)	(0.025)	(0.030)	(0.036)
25 Absences	0.055**	0.065**	0.102**	0.076***	0.077**	0.123***
	(0.025)	(0.030)	(0.040)	(0.025)	(0.030)	(0.036)
50 Absences	0.040	0.050*	0.094**	0.076***	0.077**	0.123***
	(0.024)	(0.030)	(0.040)	(0.025)	(0.030)	(0.036)
Sample Size	8130	8130	8130	8130	8130	8130

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results from the NCERDC sample estimation are robust to both of the missing value strategies. Differential attrition in the NCERDC sample is not large enough to effect the point estimates of teacher bias with respect to low-income students. Turning to the ECLS-K strategy, the results are somewhat sensitive to differential attrition. Among English teachers, the marginal effects decrease from about 7.6% to 5.5% when missing values for FRL student absences are replaced with 25 absences and to about 4.0% when missing values are replaced with 50 absences. Similar decreases in magnitude are observed for the marginal effects among math and science teachers, with math teacher effects decreasing from 11.0% to 6.5% to 5.0% and science teacher effects decreasing from 13.3% to 10.2% to 9.4%. Even though the point estimates decrease in magnitude, they are still meaningfully large, and statistically significant (with the exception of English teacher effects using the 50 absence strategy). The fact that we still observe significant bias estimates when replacing the missing values for low-income students with absence values that are well above the 95th percentile of the distribution indicates that differential attrition does not explain away the estimates of teacher bias.

3.6 Discussion

The most salient result in this paper is that low-income students - as proxied by FRL status - are consistently viewed as being more absent than their more affluent peers, even after controlling for simulated measures of actual absences. Perhaps even more to their detriment, the result is strongest among subgroups of teachers that make up a large share of the population of teachers - white, female, fully certified teachers with higher than a bachelor's degree. Thus low-income students are more likely to be taught by teachers that potentially hold negative behavioral biases against them. If these behavioral biases translate into decreased likelihood of recommending the students for honors, etc., this could unfairly disadvantage these students.

Another important result is that math teachers tend to view males more favorably than females, even after controlling for actual behavior. The same argument for low-income children above holds for females in math courses. These behavioral biases could result in fewer female students being recommended for honors or advanced math courses, which are gateway courses into science, technology, engineering and math (STEM) fields. Could these subtle biases be contributing to the fact that fewer females than males enter STEM fields? Pope and Sydnor (2010) find support for the notion that gender gaps in math test scores are due more to nurture than nature -school environments may contribute to gender differences in performance. This paper suggests that teacher perceptions may be one important aspect of the school environment that contributes to gender disparities.

There is no consistent evidence of the existence of bias for or against various racial/ethnic categories with respect to absences. In the case of black and Hispanic students, teachers don't generally view them as being more likely to be absent than white students whether controlling for simulated absences or not. In the case of Asian students, even though teachers tend to view them more positively than white students when not controlling for simulated absences, that positive difference disappears when simulated absences are controlled for. Thus the difference in behavioral perceptions for Asian students appears to be driven by differences in actual behavior.

With regard to the intersection of race and gender and the intersection of race and poverty, I find that black male students are viewed significantly more positively than black female students by math teachers with respect to their absenteeism. Also, low-income white and Asian students tend to be viewed less favorably than their more affluent counterparts. While there is evidence that math teachers are more likely to report low-income black and Hispanic students as being absent than their more affluent counterparts, there are no clear patterns, overall, that suggest low-income black and Hispanic students are subject to bias relative to more affluent black and Hispanic students.

The results that examine teacher reports by subgroups of teachers support the main results and also provide further insight. For example, female teachers do not significantly favor female students, in fact, female math teachers significantly favor male students. With the exception of Hispanic science teachers, there is also very little evidence to suggest that teachers significantly favor students from the same racial/ethnic background either. These results perhaps shed more light on the results in the previous literature that indicate that students of the same race and/or same gender of their teacher are rated more favorably with respect to their behavior. The results in this paper suggest that it may not be favorable bias on the part of teachers that produce those results, but rather, students may simply behave more favorably when paired with a teacher of the same race or gender.

How does attendance behavior relate to other behaviors in the classroom, such as disruptiveness and attentiveness? Certainly attendance behavior is a more benign form of behavior when thinking about what behaviors teachers tend to notice, and incorporate into their evaluations of students. However, even though attendance behavior may not be as forefront as other behaviors when teachers make placement recommendation decisions, the observation that there may be perceptual biases in something as latent at attendance could indicate that there are even stronger perceptual biases in other behaviors.

To examine the importance of these results with regard to student outcomes, the next step should be to see to what extent biases influence placement decisions. If teachers hold positive biases towards a specific subgroup of students, are they more likely to recommend those students for academically enriched services or honors classes even given the same academic performance as other students? Also, another avenue of investigation is how teacher biases vary across class and school contexts. Are biases against low-income students stronger in schools with higher concentrations of low-income students?

The results in this essay help build upon our knowledge of teacher biases with regard to student behavior. It may be quite likely that teachers who hold biases towards or against a specific group are not purposefully setting out to be discriminatory, and are quite unaware that they hold these biases. The fact that teachers don't consistently hold positive bias towards students who share their same race or same gender may be evidence that the biases are not purposeful. If that is the case, then a possible solution is to incorporate bias sensitivity training into teacher education or professional development courses.

In summary, this essay has examined the extent to which differences in teacher perceptions of student absence between different demographic groups are due to differences in actual behavior, or due to bias on the part of teachers. I find that low-income students are more likely to be judged as having higher levels of absenteeism than more affluent students even after controlling for actual absences. Also, math teachers tend to perceive male students as being significantly less absent than female students even after controlling for actual attendance. These biases on the part of teachers have the potential to disadvantage these subgroups of students, relative to their peers, with respect to their academic and life outcomes.

4

Children Left Behind: The Effects of Statewide Job Loss on Student Achievement

4.1 Introduction[1]

Increased emphasis on student test scores in recent years has come during a time of significant economic turmoil. Since the No Child Left Behind Act (NCLB) was passed in 2001, the United States has experienced two recessions, including the largest recession since the Great Depression, as well as high rates of job displacement as traditionally U.S.-based industries have moved overseas. Given evidence on the effects of both parental job loss and local-area job loss on youth academic success (Ananat et al., 2011; Stevens and Schaller, 2011), it is important to understand how economic downturns may affect aggregate student test scores. Such an understanding is vital to clarifying what affects student test performance and, therefore, to forming appropriate school accountability policies.

Studies that attempt to examine the effects of economic losses on academic achievement face two major challenges to validity. First, in most instances, there

[1] Co-authored with Elizabeth O. Ananat, Anna Gassman-Pines and Christina Gibson-Davis

are likely to be unmeasured or unobserved characteristics that affect both a given family's financial status and the family's well-being. For example, in families facing health or substance-use problems, parents may be less likely to maintain employment and children may also have less school success than in other families. While an instrumental variables approach can be used to address the endogeneity of parental job loss, such an identification strategy still faces a second challenge to validity. That is, it may miss effects of local economic crises on children that come through channels other than parental unemployment. Such channels could include, for example, increased stress among even continuously-employed parents and teachers, or spillover effects in the classroom from peers whose parents lose jobs. Using children whose parents have not lost employment as a control group for those whose parents are displaced, thereby assuming that these other channels are negligible, may understate the effects of job loss on both groups of children. In fact, we find in this paper that research addressing the first but not the second challenge understates the aggregate achievement effects of economic downturns by as much as an order of magnitude.

In this paper, we address these two empirical challenges by examining the impact of state-level job losses caused by business closings and layoffs on states' student achievement test scores. Using plausibly exogenous variation in business closings permits us to identify the causal effect of an economic downturn on all students and on vulnerable subgroups of students in particular. We find that job losses to 1% of a state's working-age population decrease that state's eighth-grade math scores the next year by .076 standard deviations, a large effect size commensurate with (although opposite in sign of) interventions that are designed to impact test scores. We investigate potential mechanisms for this decline, and find that it cannot be accounted for by decreased school budgets or by migration. We also present evidence against the drop in test scores representing the outcome of a "downward spiral" of youth behavior after local job losses; if anything, students appear to have better behaviors during downturns. By contrast, our results are consistent with the hypothesis that increased stress during downturns interferes with students' attainment, and that lowered income can account for some but not all of the decline in scores.

Our findings can provide insight to researchers investigating the determinants of student achievement and to educators seeking to understand the effects of the recent economic crisis. Further, our results have implications for accountability schemes: in a back-of-the-envelope calculation, we find that states experiencing one-year job losses to 1% of their workers likely see an 8% increase in the share of their schools failing to make Adequate Yearly Progress under No Child Left Behind. These results suggest that local economic conditions are an important factor in students' test performance, and so are a relevant consideration for policymakers attempting to fairly and accurately evaluate school performance.

4.2 Background

A broad consensus now exists that business layoffs and closings can be viewed as exogenous shocks to workers and communities when conditioning on prior characteristics (Jacobson et al., 1993; Stevens, 1997; Sullivan and Von Wachter, 2009) and that effects on workers and communities subsequent to layoffs and closings can therefore be interpreted as causal effects of job loss. The explanation behind this consensus is as follows. When an individual is fired or quits, it may reflect negative unobservable characteristics of that individual, or of the individual's community. In contrast, however, closings and downsizings occur because of larger macroeconomic and international trade forces. Although firms might close or relocate due to declining worker productivity in an area (which, again, might reflect unobservable community characteristics), empirically it has been repeatedly found that once fixed effects for the area are included, firm decisions are not predictable using changes in community characteristics. In this paper we provide further evidence of this empirical regularity.

One strand of literature has used this empirical strategy to examine the effects of an individual-level job loss (whether a household head loses a job because of a closing, regardless of how many others in the community are affected) on family-level outcomes such as income, parenting practices, or children's grade retention. Another strand has concentrated on the effects of community-level job losses (the total number of jobs lost in a community) on community-wide outcomes such as levels of physical health, suicide, or welfare receipt. However, few papers in this latter strand have looked at children's outcomes. We complement previous work in these two literatures by using a community-level empirical strategy while focusing on children's achievement as the outcome of interest. This approach allows us both to identify causal effects of area job loss and to identify effects on children that do not come solely through their parents' employment status. Below, we discuss the previous individual-level and community-level literatures on the effects of job loss and use them to generate hypotheses on why community-wide job losses might affect aggregate levels of child academic performance.

4.2.1 Effects of Individual-Level Job Loss

Parental job loss can affect child development in two ways. First, parental job loss can reduce families' material resources. Second, parental job loss can lead to changes in families' physical health, mental health, and behaviors, including parenting behaviors.

Job loss lowers earnings both in the short term, while parents look for new employment, and over the longer term, because people who lose their jobs due to industry downturns often must start over in new firms and new industries (Jacobson et al., 1993; Stevens, 1997). Family income affects children's outcomes. Studies have documented that changes in parental income and material resources lead to changes in children's well-being and, in particular, their achievement test scores (Morris and

Gennetian, 2004; Dahl and Lochner, 2012).

Job loss can also affect children's outcomes by affecting parents' mental health and thereby altering family functioning. Individuals who have lost employment have worse psychological (McKee-Ryan et al., 2005) and physical (Sullivan and Von Wachter, 2009) health than those who have not lost employment. Longitudinal studies that observe families before and after a parental job loss have found that job loss leads to decreased family functioning and impaired parent-child interactions (Jones, 1988; Conger and Elder, 1994; McLoyd et al., 1994; Kalil and Wightman, 2010). Parental mental health problems and impaired parent-child interactions have both been strongly linked to worse child adjustment and lower levels of school achievement (Elder Jr et al., 1995; McLoyd, 1998). While it is possible that job loss could lead parents to spend more time with their children, which could have beneficial effects on child school achievement, research has in fact shown that, compared to employed parents, unemployed parents do not spend more time with their children, either in general (Edwards, 2008; Kalil and Ziol-Guest, 2011) or specifically on education-related activities that could lead to greater academic achievement (Levine, 2011).

Researchers have also documented that parental job loss harms children's school-related outcomes. Longitudinal studies using child fixed effects have shown that parental job loss increases grade repetition (Kalil and DeLeire, 2002; Stevens and Schaller, 2011) decreases GPAs (Rege et al., 2011), and increases school-related behavior problems (Hill et al., 2011). Finally, parental job loss also appears to have long-lasting effects on children into adulthood, such as lower earnings, greater receipt of public assistance, and lower college attendance (Oreopoulos et al., 2008; Coelli, 2011).

4.2.2 Effects of Community-Level Job Losses

In addition to evidence that job loss worsens outcomes for job losers and their children, there is also evidence that firm layoffs and shutdowns affect those who live in the impacted community, whether they lose employment or not. Several researchers have measured the causal effects of job loss on community-level employment, earnings, and public-assistance receipt. A set of studies by Black et al. (2003, 2005a,b) examining booms and busts in the steel and coal industries in the 1970s and 1980s found that industry downturns lowered employment not only within but also outside of the initially affected sector. Additionally, those who remain employed in an area that has experienced large job losses also experience decreased earnings (Blanchflower and Oswald, 1995). Further, new entrants into the labor market during a downturn experience a lifelong decrease in earnings (Oreopoulos et al., 2012).

In addition to reduced employment and earnings, those who live in an area that has experienced job losses may also experience increased stress and decreased well-being, even when they do not personally experience job loss. Longitudinal research with individual fixed effects has shown that increases in the regional unemployment rate decrease employed individuals' reported life satisfaction (Clark et al., 2010; Luechinger et al., 2010). Similarly, longitudinal cross-national studies have shown that increases in countries' unemployment rates are also associated with decreases in their employed citizens' life satisfaction (Ochsen and H., 2006; Ochsen, 2008; Clark et al., 2010; Luechinger et al., 2010). Time-series analyses have shown that increases in the local unemployment rate are associated with increases in psychological distress for those who were employed (Dooley and Catalano, 1984; Dooley et al., 1988). Using two waves of data, Fenwick and Tausig (1994) also found that increases in the local unemployment rate were associated with increases in individuals' psychological distress.

Taken together, the evidence indicates that deteriorating local economic conditions are associated with deteriorating mental health, for those who lose jobs but also for those who remain employed. Well-being could decrease among those who do not lose employment because of increased feelings of job insecurity and anxiety about economic well-being or because of distress for friends and neighbors who have lost work. These changes in adults' mental health could have implications for their children, as parental mental health has been strongly linked with altered family interactions and, in turn, children's developmental outcomes, including school achievement (Downey and Coyne, 1990).

These individual changes resulting from community-level job losses could also have large effects on the school setting and on students' experience in schools. For example, given the findings reviewed above, teachers who remain employed may also experience increases in stress. Higher levels of teacher stress are related to lower levels of student academic achievement, mainly through changes in teacher-student classroom interactions (Wiley, 2000). Relatedly, if students are in classrooms with peers whose parents have lost jobs, the interactions among students within the classroom may be altered, potentially affecting all students' levels of achievement. Less positive classroom interactions are related to lower growth in children's academic achievement over time (Hamre and Pianta, 2003; Pianta et al., 2008). For example, increases in one student's behavior problems can disrupt learning by other students in the same classroom (Figlio, 2007), and such increases have been found among students who experience parental job loss (Hill et al., 2011).

In sum, the evidence consistently indicates that those who maintain their jobs in the wake of local job losses experience lower earnings and worse mental health, effects similar to, although less intense than, those experienced by individuals who lose employment. Evidence also strongly suggests that lower earnings and worse mental health among parents lead to lower academic achievement among children.

Moreover, changes within schools in teacher stress and in other students' behavior, which can also negatively affect student achievement, are by definition experienced by both those whose parents do and those whose parents do not lose employment. We hypothesize, therefore, that parents who maintain employment in the wake of local job losses, like parents who lose employment, see their children's academic performance decline, albeit by a smaller amount.

Statewide job losses likely affect test scores both through lower achievement among children whose parents lose jobs and through additional area-level mechanisms that affect all children. We do not, therefore, expect that the relationship between statewide job losses and state average test scores will be simply the relationship between individual-level parental job loss and measures of children's academic achievement identified in earlier papers (Kalil and DeLeire, 2002; Rege et al., 2011), scaled by the size of the total job loss in relation to the size of the community. Rather, we expect that our estimate of the total statewide effect will be larger than such a scaled estimate, for two reasons. First, even in a large downturn, most children do not experience parental job loss. Small effects on the majority of children whose parents do not lose employment may, in aggregate, contribute as much or more to the total relationship between statewide job loss and test scores as does the large effect on the minority of children whose parents lose employment. Second, earlier papers have used children who do not experience parental job loss as a control for those who do. If, instead, children who are unaffected by parental job loss experience academic achievement effects in the same direction as those whose parents are affected by job loss, standard "treatment minus control" effect estimates will tend to understate the true effect of parental job loss on child test scores. Based on the literature reviewed above, we believe that effects of statewide job losses will be in the same direction for children whose parents lose jobs and for children whose parents do not, although the magnitude of the effects likely differs. Thus, we hypothesize that our estimates of aggregate effects of state-level job loss on test scores will be considerably larger than would be implied by extrapolations from previous research. We note that it is unlikely that state-level job losses reflect uniform losses in all communities within a state, and hence unlikely that state-level job losses cause uniform changes in test scores across the state. Nonetheless, the relationship between losses averaged across the state and test score changes averaged across the state is interpretable as the aggregate effect of job losses within communities in that state on test scores in communities in that state.

4.3 Data

We use two main data sources, one for test score information and one for job loss information. Student academic performance data are from the National Center for Education Statistics' National Assessment of Educational Progress (NAEP), which has administered standardized tests to a nationally representative sample of students in roughly two year intervals since 1964 (NCE, 2010b). We focus our analysis on mathematics and reading assessments administered to fourth and eighth graders from 1996 to 2009, which NAEP reports, when available, as state-level average scores and state-level percentile distribution scores for all fifty states and the District of Columbia.[2] The state-level NAEP assessments are given to a representative sample of public school students in each participating state. Scores are reported for students overall as well as for subgroups of students by gender and race.[3] Math assessments

[2] NAEP also conducts assessments of twelfth-grade students' academic performance, but those data are only available at the state level beginning in 2009, the last year of our panel.

[3] For eighth graders, scores are also reported by student-reported parental education. However, the distribution of reported parent education in our sample is skewed towards educational attainment higher than is plausible given national estimates (on average, over 45% of students in the sample report that at least one of their parents has a college degree, while in the 2000 Census only 28% of households with comparably-aged children reported that at least one parent has a college degree) (calculated from IPUMS 2000 5% sample (Ruggles et al., 2004)).Thus we do not report analysis by reported parental education. Results for reported-parent-education subgroups are similar to those reported here (available upon request).

were administered in 1996, 2000, 2003, 2005, 2007 and 2009. Reading assessments were administered in 1998, 2002, 2003, 2005, 2007 and 2009. The tests are always administered in the first quarter of the year, between January and March.

Table 4.1: Descriptive Statistics - Student Demographics by Year and Subject

Year	1996	1998	2000	2002	2003	2005	2007	2009
Subject	Math	Reading	Math	Reading	Both	Both	Both	Both
A. 4th Grade								
% Male	51	49	50	51	51	51	51	51
% Black	16	17	18	19	17	16	15	15
% Hispanic	7	7	9	10	12	15	16	16
% White	71	67	65	64	64	61	61	61
B. 8th Grade								
% Male	50	49	50	50	51	51	50	51
% Black	15	16	16	17	15	15	15	15
% Hispanic	7	8	7	9	10	13	14	14
% White	72	67	69	67	68	64	63	63

Source: National Center for Education Statistics - http://nces.ed.gov/nationsreportcard/

Table 4.1 presents sample descriptive statistics of demographics of students who took the NAEP assessments. The sample is fairly evenly split between male and female students. White students make up the majority of students on average. The national trend over time shows a growth in the share of Hispanic students from 7% to between 14 and 16% and a decline in the share of white students of about 10 percentage points. This pattern is consistent among both fourth and eighth graders. The national share of black students remains stable at about 15%. These demographics are consistent with the demographics of children in the United States.

Mean assessment scores and standard deviations for all students and for subgroups of students separately by grade level and test subject are presented in Table 4.2. The NAEP assessments are designed to have a possible score range of 0 to 500 for individual students. The first two columns in each subject-year grouping rep-

Table 4.2: Descriptive Statistics - Test Scores by Grade Level, Subject and Subgroup

A. 4th Grade						
	Math			**Reading**		
	Mean	**St.Dev.**	**Indiv. Std.Dev.**	**Mean**	**St.Dev.**	**Indiv. Std.Dev.**
All Students	234	9.4	29.1	218	7.7	36.4
Female	233	9.2	28.3	221	7.5	35.7
Male	235	9.7	29.8	214	8.1	36.8
Black	214	10.7	26.8	198	8.0	34.4
Hispanic	223	9.3	27.6	202	8.1	36.3
White	241	8.1	25.9	227	5.1	32.7
B. 8th Grade						
	Math			**Reading**		
	Mean	**St.Dev.**	**Indiv. Std.Dev.**	**Mean**	**St.Dev.**	**Indiv. Std.Dev.**
All Students	277	9.6	36.4	262	6.8	34.8
Female	276	9.4	35.3	267	6.7	33.7
Male	278	9.8	37.5	257	7.0	35.2
Black	252	9.9	33.3	243	5.2	33.1
Hispanic	261	8.6	34.2	246	5.8	35.1
White	286	7.5	32.7	270	4.3	31.5

Mean is computed by taking the average across states and years of the reported state-level averages of individual student scores. The mean is weighted at the state level by the number of students in each state.

St. Dev. Is computed by taking the standard deviation across states and years of the reported state-level averages of individual student scores. The standard deviation is weighted at the state level by the number of student in each state.

Indiv. Std. Dev. Is computed by taking an average across years of the national student-level standard deviations reported by NAEP for a given assessment and year

Source: National Center for Education Statistics - http://nces.ed.gov/nationsreportcard/

resent the mean and standard deviation of the state average assessment scores. The third column contains the averages across years of the national student-level standard deviation for a given assessment and year, which are of course much larger than those for the state averages. Large differences, in the expected directions, also exist between the average assessment scores of different subgroups of students. Girls score slightly higher on reading than do boys; white students score significantly higher than black and Hispanic students in both subjects and both grade levels.

For the purposes of analysis, we standardize each state-level assessment score to have a mean of zero and standard deviation of one (following the convention of using the individual student level, not the state, standard deviation), which allows for comparison of test scores across subjects, grades, and years. The sample is organized in state-year observations, yielding a maximum of 306 observations for each grade and subject. Not all states administered examinations in all years, and some states did not report assessment scores for all student subgroups. Table C.1 in Appendix C lists which states participated in the NAEP assessments for each year of our sample.

Job loss data are from the Bureau of Labor Statistics' (BLS) Mass Layoff Statistics, which report, for each state when available, the number of workers in a quarter who are affected by mass closings or mass layoffs (defined as 50 or more workers) that last longer than thirty days (BLS does not collect data on layoffs or closings affecting fewer than 50 workers). Data are available from 1995 to 2009. For each year, the BLS reports two measures of workers affected by job loss. The first is the total number of initial claimants (TIC), which reflects the total number of workers who filed unemployment claims after a closing or layoff of 50 or more workers. The second is the total number of separations, which is the number of workers who lost jobs because of a mass closing or mass layoff. A mass closing or mass layoff is defined by BLS as one in which 50 workers from the same firm have filed unemployment insurance claims in a 5-week period.[4] Once BLS classifies that event as a mass closing or mass layoff, it then contacts the firm to gather information about the total number of workers who lost jobs in that event (separations). Separations is our preferred measure since it should capture all workers who experience a mass closing or mass layoff instead of just those workers who then also filed unemployment claims. However, the separations measure is likely to suffer from greater measurement error than TIC because it involves the extra step of contacting companies for further information on events that are identified through initial unemployment claims. As discussed in the Methodology section, we combine these two measures in a two-stage least squares approach in order to reduce measurement error.

Table 4.3: Descriptive Statistics - Job Losses as a Percent of Working Age Population

	Obs.	**Mean**	**Std.Dev.**	**Min**	**Max**
Separations	506	0.71	0.47	0	3.39
Total Initial Claimants	506	0.66	0.47	0	3.66

The variable *Separations* is calculated by dividing the total yearly number of separations in a state by the working age population (ages 25-64) in that state.

The variable *Total Initial Claimants* is calculated by dividing the total yearly number of claimants in a state by the working age population (ages 25-64) in that state.

Source: Bureau of Labor Statistics - http://www.bls.gov/mls/

Table 4.3 presents summary statistics for separations and TIC. For the purposes of our analysis, we express both separations and TIC as a percentage of the working-age population (defined as the number of state residents aged 25 - 64, measured for each state in the 2000 Census) over a one year time period. On average, 0.71% of the working-age population is affected by separations and 0.66% file unemployment claims in a year. The variation in these two measures is roughly the same. Fig-

[4] If a firm has layoffs that occur in multiple sites or divisions within a state, those layoffs are treated as a single, firm-level event if they occur for the same economic reason. If, however, layoffs at different sites occur for different economic reasons, BLS treats those as distinct layoff events, in which case, the layoffs at each site would have to meet the 50 worker threshold to qualify as a mass closing or mass layoff event, and thus to be included in the data.

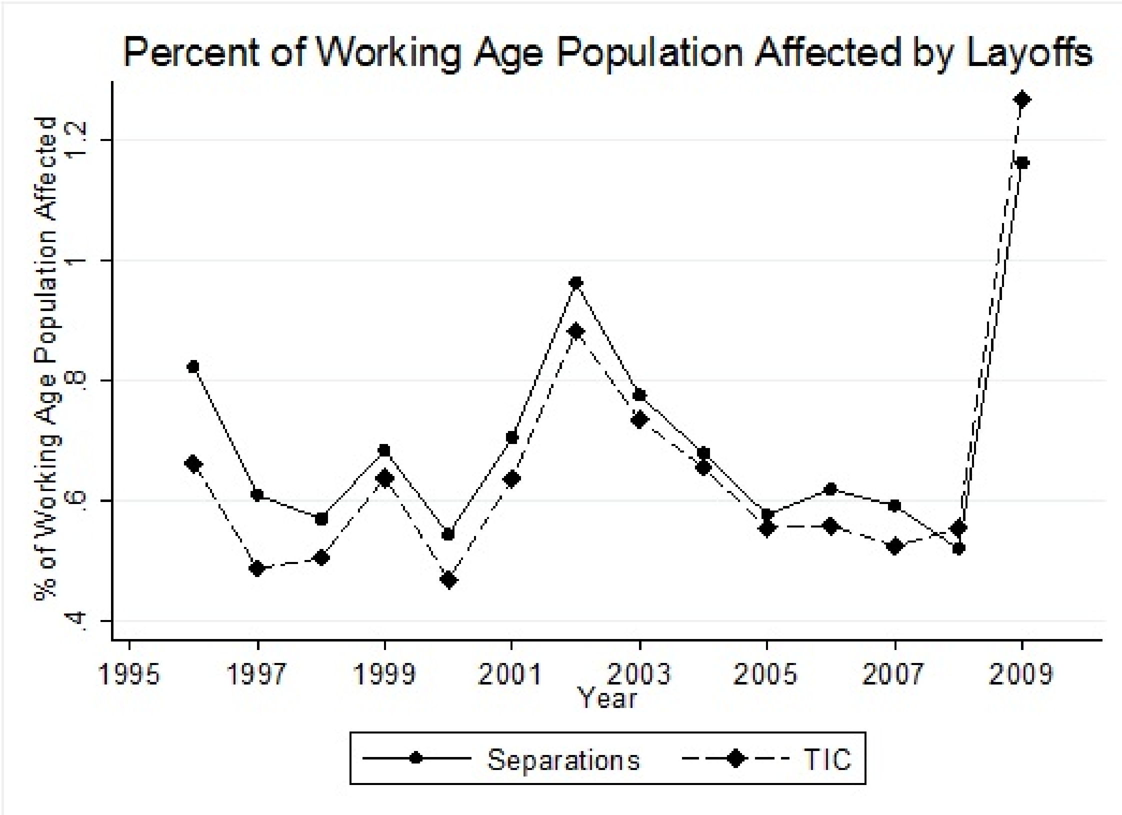

Figure 4.1: National Percent of Working Age Population Affected by Layoffs as Represented by Separations and Total Initial Unemployment Claims. Source: Bureau of Labor Statistics - http://www.bls.gov/mls/

ure 4.1 plots yearly separations and TIC. The measures are highly correlated, with the percentage of workers reported by firms to be affected by separations slightly higher than the percentage of workers who file for unemployment claims in every year except 2008 and 2009.

Figure 4.2 presents the minimum and maximum percent of workers in each state affected by job separations over the 15 years of our panel. There is significant variation in job losses within states over time, as demonstrated by the difference between the minimum and maximum percent affected in each state, as well as between states. The maximum percentage of workers affected by job loss ranges from less than one

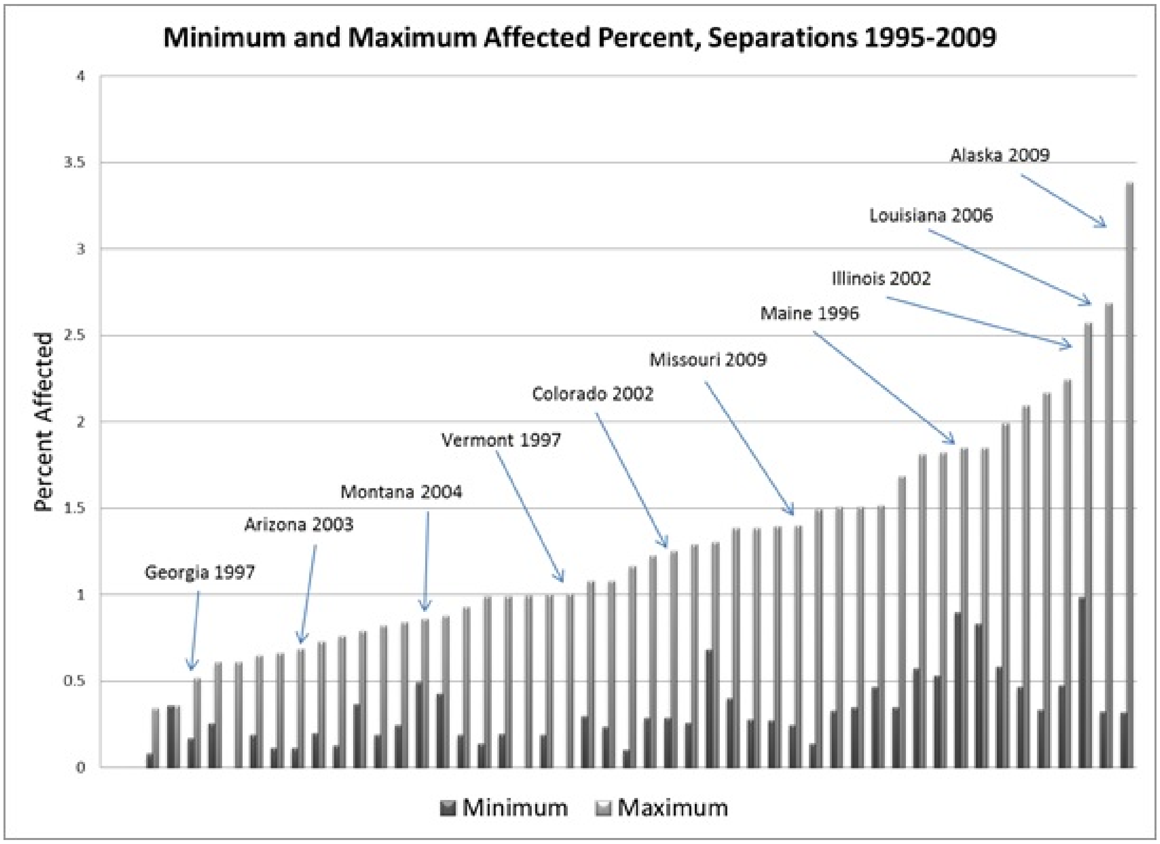

FIGURE 4.2: Minimum and Maximum Percent of Working Age Population Affected by Job Loss, 1995-2009. Source: Bureau of Labor Statistics - http://www.bls.gov/mls/

half of one percent in Maryland in 2009 to nearly 3.5% affected in Alaska in 2009. While the highest observed job loss did occur during the Great Recession (Alaska in 2009), many states experienced their largest losses in the 2002 recession (Colorado, Illinois), or even in years of relatively strong national economic growth, such as 1996 (Maine).

We focus on job loss rather than the state unemployment rate because the unemployment rate can be biased by changes in job-seeking behavior that are confounded with other changes in a community. For example, bad news can discourage workers from looking for work and actually *decrease* the unemployment rate while at the same time increasing community stress and lowering test scores, which would positively bias the estimated relationship between unemployment and test scores. By contrast, firm-level closings and layoffs can more plausibly be viewed as exogenous "shocks" that are driven by the global economy and do not affect test scores other than through their effects on job loss (we also test the exogeneity of these events).

4.4 Methodology

In order to explain the effects of job losses on test scores, we estimate the equation:

$$Score_{st} = \beta JobLoss_{st-1} + \delta_t + \delta_s + \varepsilon \tag{4.1}$$

In this specification, $Score_{st}$ represents the mean scaled test score for students in state s at time t.Separate equations are estimated for each of the four subject-grade combinations, as well as for race and gender subgroup scores. In alternative models, we estimate the equation using scaled percentile scores as dependent variables; these models measure whether the effect of job loss is consistent throughout the test score distribution. $JobLoss_{st-1}$ represents the percent of workers in a state affected by mass layoffs for the year-long period up to and including the quarter the tests were administered (this measure is discussed further below). We also include state fixed effects (δ_s) to account for the possibility that states that have higher job losses on average may also have lower test scores on average, and year fixed effects (δ_t) to account for nation-wide time-varying factors that may affect both job losses and test scores.[5] Observations are weighted by the number of test takers. We report heteroskedasticity-robust standard errors that are clustered at the state level.

[5] The shallow panel of state-year observations is not deep enough to support precise estimates when state-specific time trends are included; results do not vary significantly from those reported here, but are unstable. While the inability to include trends may raise the concern that unobserved changes in states drive both job losses and declining scores, our extensive falsification checks provide no support for this possibility.

To measure job loss, $JobLoss_{st}$, we use a composite of two noisy measures, separations and TIC. Use of either measure on its own is likely to lead to attenuation bias, while a composite based on the correlation between the two can increase the reliability of our estimate of job destruction (Angrist and Pischke, 2009). The noise in our measure of TIC comes from the fact that not all workers who lose jobs file for unemployment. The noise in our measure of separations is due to the fact that, when contacted by the government, employers may not accurately report the number of workers affected by a layoff. Each measure is composed partly of a"true" signal of underlying job destruction, D, and partly of an error term:

$$\begin{aligned} Separations_{st} &= \gamma D_{st} + \varepsilon \\ TIC_{st} &= \sigma D_{st} + u \end{aligned}$$

Where $corr(\varepsilon, u) < 1$

The correlation of the two measures, therefore is:

$$Corr(Separations_{st}, TIC_{st}) = D_{st} + v,$$

Where $v < min(\varepsilon, u)$.

Specifically, we estimate a two-stage least squares specification where, in the first stage, we use TIC to predict separations, and then report the coefficient on $\widehat{Separations}_{st-1}$ in an equation predicting $Score_{st}$. Note that $\widehat{Separations}_{st-1}$ is simply $\gamma\widehat{Corr}(Separations_{st}, TIC_{st})$. Using the estimated correlation of the two measures as our measure of job loss provides a more precise estimate of job destruction than does either measure on its own (Angrist and Pischke, 2009). Using two-stage least squares rather than simply using the correlation as the right-hand side variable in an OLS regression means that our standard errors are automatically adjusted to take into account that $\widehat{Separations}_{st-1}$ is a statistical artifact rather than a direct measurement.

4.5 Results

4.5.1 Main Estimates

Table 4.4 presents the results of estimating the impact of job losses on average test scores. In this table each cell represents the coefficient and standard error on separations derived from estimating equation (4.1) using two-stage least squares for a given subject-grade-subgroup combination. For example, the first cell in the third column is the coefficient derived from estimating the average impact of separations on all students in the sample who took an eighth grade math assessment. The interpretation of this estimate is that job losses that affect one percent of a state's working-age population decrease that state's average eighth-grade math score by 0.076 student-level standard deviations, or by almost three points on average (the student-level standard deviation for the eighth grade math test is 36.4 points).

All 24 point estimates in Table 4.4 are negative, but only estimates for eighth grade math are consistently statistically significant. Results in Table 4.4 suggest two main points. First, math scores are more sensitive to job losses than are reading scores. In both the fourth-grade and eighth-grade samples, the point estimates on math assessments are larger in magnitude than those for reading, and this difference is statistically significant in the eighth grade. Second, eighth grade scores are more sensitive to job losses than are fourth grade scores; point estimates are consistently larger for eighth-grade math than for fourth-grade math. These results are consistent with results from our analysis using county-level job loss and academic performance data from North Carolina (Ananat et al., 2011), which also finds effects of job losses on eighth but not fourth grade test scores. Third, eighth grade math scores decline significantly across all gender and race subgroups. Effects do not vary significantly by race or gender, although point estimates for African-Americans are somewhat larger than for other groups. Effect sizes average -.076 standard deviations and

Table 4.4: Estimation Results - Impact of Job Losses on Student Test Scores

	4th Grade		8th Grade	
	Math	**Reading**	**Math**	**Reading**
All Students	-0.032	-0.013	-0.076***	-0.009
	(0.035)	(0.020)	(0.027)	(0.022)
Student Subgroups				
Gender				
Female	-0.033	-0.003	-0.077***	-0.007
	(0.029)	(0.018)	(0.029)	(0.019)
Male	-0.023	-0.022	-0.072***	-0.022
	(0.038)	(0.024)	(0.027)	(0.025)
Race/Ethnicity				
Black	-0.010	-0.015	-0.109**	-0.042
	(0.042)	(0.035)	(0.049)	(0.041)
Hispanic	-0.014	-0.011	-0.064***	-0.028
	(0.051)	(0.023)	(0.024)	(0.017)
White	-0.049	-0.029	-0.066*	-0.002
	(0.037)	(0.022)	(0.036)	(0.026)

Robust standard errors in parentheses

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

Each cell represents the coefficient and standard error on Separations derived from estimating equation (4.1) for a given subject-grade-subgroup combination. The specification includes both state and year fixed effects.

Source: Bureau of Labor Statistics; National Center for Education Statistics

point estimates range from -.064 (for Hispanics) to -.109 (for African Americans).

The lack of responsiveness of reading scores is consistent with the findings of many school-based interventions (Decker et al., 2004; Abdulkadiroglu et al., 2009; Hoxby and Murarka, 2009; Angrist et al., 2010; Dobbie and Fryer, 2011). It may be that math skills are more highly influenced by factors external to the family, including the school and community contexts, than reading skills, which may be more highly influenced by the family context. It may also be the case that math test scores are more sensitive to recent influences than are reading scores, since it may be easier to isolate and test recently-taught math concepts on an exam than it is to isolate particular reading skills. Further research is needed to understand why math scores may be more responsive than reading scores to changes in the immediate economic circumstances of students' communities.

The lack of responsiveness of fourth grade test scores is also consistent with the developmental literature. Older children who are just entering adolescence are likely more developmentally vulnerable than younger children in the period of middle childhood (Eccles et al., 1993). In addition, families are better able to shield younger children from the effects of job losses; research has shown that as youth age, they become more aware of their families' economic pressures (Mistry et al., 2009). Finally, adolescence is a developmental period marked by the increasing importance of peers (Eccles et al., 1993). Because adolescents are more likely to interact with a peer whose parent has lost a job than are younger children, any effects through peer interactions of community-wide job losses will be stronger for adolescents.

4.5.2 Percentile Test Scores

We have also used two stage least squares to estimate equation (4.1) while replacing average state scores with percentile scores as the dependent variable. The percentile results for eighth graders are presented in Table 4.5 for math test scores and in Table 4.6 for reading scores.[6] As in Table 4.4, each cell presents the coefficient and standard error on separations from a separate regression for each of the various subgroups and for both math and reading. These results follow the same pattern as the results when using average test scores as an outcome. Math scores are typically more responsive to job losses than are reading scores. Point estimates for black students' test scores across the distribution are more negative than are white or Hispanic students'. Notably, effects are quite uniform across the distribution; it does not appear that the effects on average test scores are driven by a few students

[6] The results for fourth-grade students, similar to those for the average test scores outcome measure, do not exhibit statistical significance and are not presented here (available upon request).

"bombing" the test after job losses while other students' scores are stable. Rather, effects on students across the distribution appear to cluster around -.076 standard deviations, ranging from a maximum of -.055 standard deviations for whites at the 90th percentile to a minimum of -.139 standard deviations for blacks at the 50th percentile.

Table 4.5: Estimation Results - Impact of Job Losses on Percentile Test Score Outcomes - 8th Grade Math

Percentile	**10th**	**25th**	**50th**	**75th**	**90th**
All Students	-0.078**	-0.081**	-0.077***	-0.071***	-0.064***
	(0.039)	(0.033)	(0.026)	(0.024)	(0.023)
Student Subgroups					
Gender					
Female	-0.087**	-0.087***	-0.076***	-0.069**	-0.071***
	(0.036)	(0.033)	(0.028)	(0.027)	(0.026)
Male	-0.069	-0.075**	-0.077***	-0.077***	-0.055***
	(0.043)	(0.033)	(0.027)	(0.023)	(0.021)
Race/Ethnicity					
Black	-0.101	-0.122**	-0.139***	-0.118***	-0.089*
	(0.066)	(0.051)	(0.043)	(0.043)	(0.046)
Hispanic	-0.058	-0.065**	-0.059***	-0.077***	-0.056*
	(0.041)	(0.027)	(0.022)	(0.024)	(0.029)
White	-0.059	-0.065	-0.068**	-0.062**	-0.055*
	(0.051)	(0.043)	(0.034)	(0.031)	(0.029)

Robust standard errors in parentheses

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

Each cell represents the coefficient and standard error on Separations derived from estimating equation (4.1) for a given subject-grade-subgroup combination. The specification includes both state and year fixed effects.

Source: Bureau of Labor Statistics; National Center for Education Statistics

Table 4.6: Estimation Results - Impact of Job Losses on Percentile Test Score Outcomes - 8th Grade Reading

Percentile	**10th**	**25th**	**50th**	**75th**	**90th**
All Students	-0.048	-0.023	-0.008	0.001	0.009
	(0.034)	(0.029)	(0.021)	(0.015)	(0.015)
Student Subgroups					
Gender					
Female	-0.034	-0.013	0.003	0.009	0.020
	(0.033)	(0.025)	(0.017)	(0.017)	(0.019)
Male	-0.061	-0.028	-0.017	-0.009	-0.004
	(0.042)	(0.035)	(0.026)	(0.020)	(0.017)
Race/Ethnicity					
Black	0.005	-0.030	-0.042	-0.051	-0.055
	(0.054)	(0.039)	(0.038)	(0.044)	(0.042)
Hispanic	-0.076***	-0.033	-0.005	-0.006	0.022
	(0.028)	(0.026)	(0.028)	(0.021)	(0.025)
White	-0.026	-0.008	0.001	0.006	0.018
	(0.051)	(0.035)	(0.023)	(0.019)	(0.016)

Robust standard errors in parentheses

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

Each cell represents the coefficient and standard error on Separations derived from estimating equation (4.1) for a given subject-grade-subgroup combination. The specification includes both state and year fixed effects.

Source: Bureau of Labor Statistics; National Center for Education Statistics

4.6 Robustness Checks

Our main results - that math scores are more sensitive to job losses than reading scores, and that eighth graders are more sensitive to job losses than fourth graders - are robust across subgroups of students and are also robust to percentile outcome measures. In this section we discuss seven other robustness checks (results are summarized in Table 4.7; estimates from all of these checks are available upon request).

First, we estimated the model 51 times, excluding each state and the District of

Columbia individually. While the state fixed effects we include in our model will absorb any persistent relationship between test scores and job loss in a particular state, these specifications test whether severe events in a particular state (such as Hurricane Katrina in Louisiana) that can cause above-average job losses and below-average test scores significantly affect our results. All results are similar when dropping each state, meaning that no single state is driving our results.

Second, we performed a similar exercise excluding each year. While the year fixed effects we include in our model will absorb the effects of any nationwide phenomenon that affected both job loss and test scores in a given year, these specifications further test whether severe events that may have affected both outcomes in only some parts of the country (such as 9/11 on the mid-Atlantic region) significantly affect our results. All results are similar when dropping each year, meaning that no one-time sub-national event is driving our results.

Third, we ran unweighted regressions. While the analysis on which our main results are based weighs each state-year observation by the number of test takers in that state and year (and, where appropriate, subgroup), this analysis treats all state-year observations equally. Whereas our main results can therefore be interpreted as reporting the effects in the typical state in which a student lives (the results most important to a national policymaker), these results can be interpreted as the effects on a state itself (a result important to state policymakers). Analysis conducted using unweighted observations obtains results that are substantially similar to those shown here.

Fourth, we conducted analysis using only subsets of states for which we were not missing data on racial subgroups. Because of geographical variation in the size of the population of black and Hispanic students, some states did not report subgroup scores for either or both black or Hispanic students in some years. In order to test whether the differential estimated responses to job losses experienced by black

Table 4.7: Summary of Robustness Checks

Effect of job losses on eighth grade math scores when:	
Unweighted	-0.076**
	(0.030)
Restricting sample to:	
Balanced panel of states	-0.123***
	(0.031)
State-years with above median unemployment rate	-0.097**
	(0.029)
Controlling for:	
State demographics	-0.075***
(age, race, and education structure)	(0.027)
Test-taker demographics	-0.075***
(race and free lunch status)	(0.026)
State unemployment rate	-0.061***
	(0.024)
State UI claims per capita	-0.063***
	(0.024)
State GDP	-0.073***
	(0.027)
Forclosures	-0.060***
	(0.022)

students compared to white or Hispanic students was driven by larger proportions of black students living in regions that are more sensitive to job loss, we estimated models using only the subsample of state-years for which there were no missing observations for blacks. The racial differences in point estimates of sensitivity to job losses are robust to this specification change. We performed a similar exercise using observations on Hispanic students and obtained similar results.

Fifth, we conducted analysis using only a balanced panel of states for which we are never missing job losses or NAEP scores in years in which the tests are conducted. (Table C.1 in Appendix C lists, for each state, the years in which it reported test scores, job losses, and both.) We did so in order to test whether the estimated responses to job losses are influenced by the inclusion of states that only selectively report job losses or test scores (whose participation decisions in BLS data collection and/or NAEP data collection are perhaps influenced by their economies or by their expected scores). Because we have a shallow panel of only at most six observations for each state-grade-test, this robustness check likely reduces measurement error as well (since state fixed effects are unlikely to be well estimated for states that are observed fewer than six times, meaning that such states will contribute significant noise to our estimates). In fact, while the number of students we observe is reduced by 40% under the restriction that job losses are never missing, the estimated effect of a 1% job loss on eighth-grade math scores increases by 50% to -0.114 standard deviations, and the t-statistic increases as well, to 3.7. Similarly, while the number of students we observe is reduced by 29% under the restriction that NAEP scores are never missing, the estimated effect of a 1% job loss on eighth-grade math scores increases by 26% to -.096 standard deviations, and the t-statistic is stable at 2.8. These results suggest that our estimates are not only robust to but actually strengthened by restricting to a balanced panel.

Sixth, we examined the effects of job loss on test scores using only the subset of state-years in which the state started the year with a high (above the median for the full panel of states) unemployment rate. Given the mechanisms we propose through which we believe job losses affect child academic achievement, we hypothesize that job losses should matter more in times and places when the local economy is already under stress. We believe job losses will have stronger effects on stress, on family and community functioning, and subsequently on test scores, when a high pre-existing unemployment rate makes it more difficult for those who experience job displacement to find a new job. That is in fact what we find: in areas with high current unemployment, the estimated effect of a 1% job loss on eighth-grade math scores increases in magnitude by 28% to -.097 standard deviations, and precision increases as well.

Finally, we have run models including a variety of important time-varying covariates and results do not change. These include characteristics of state populations, including percent of the state population who are minors (under age 18), percent who are elderly (age 65 or older), percent who are white, black, and Hispanic, education structure of the state (percent of adults 25 and older who are high school dropouts, high school graduates, have some college education, and are college graduates), and percent who are poor. We have also estimated our models controlling for characteristics of the student test-taker population, including percent of students who are white, black, and Hispanic and percent of students who are eligible for free or reduced-priced lunch. Finally, we have controlled for underlying economic conditions in the state, including: the average unemployment rate in the year preceding the test; unemployment insurance claims per capita in the year preceding the test; state GDP; and the home foreclosure rate in the state in the year preceding the test.

4.7 Falsification Checks

We conducted falsification checks in which we estimated equation (4.1) using future job losses, i.e. losses in the four quarters following the test, instead of lagged job losses. Significant estimates from these regressions would cast doubt on our identifying assumption that job losses, conditional on state and year fixed effects, can be viewed as exogenous shocks to states. Such results would instead suggest that states that experience above-average job losses in a given year already had declining test scores. However, the results of the falsification checks, which are presented in Table 4.8, are generally small and statistically insignificant. Only two of the twenty coefficients are marginally significant at the 10% level, and of these one is in the unexpected (positive) direction. One of the twenty results presented is significant at the 5% level, and it, too, is in the unexpected (positive) direction. These estimates lend support to the assumption that changes in state test scores do not occur until after job losses occur, and hence the relationship between job losses and test scores can be interpreted causally.

Table 4.8: Falsification Results - Impact of Future Job Losses on Student Test Scores

	4th Grade		8th Grade	
	Math	Reading	Math	Reading
All Students	-0.003	0.02	-0.016	0.001
	-(0.026)	-(0.017)	-(0.024)	-(0.013)
	Student Subgroups			
Gender				
Female	-0.013	0.018	-0.007	0.018
	-(0.025)	-(0.014)	-(0.021)	-(0.016)
Male	0.006	0.024	-0.032	-0.007
	-(0.027)	-(0.028)	-(0.024)	-(0.012)
Race/Ethnicity				
Black	0.045*	0.049*	0.01	-0.001
	-(0.019)	-(0.022)	-(0.022)	-(0.030)
Hispanic	-0.117	-0.037	-0.094**	-0.014
	(0.000)	-(0.023)	-(0.028)	-(0.035)
White	-0.023	0.011	-0.014	0.002
	-(0.024)	-(0.013)	-(0.027)	-(0.014)

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

We also conducted falsification checks in which we estimated equation (4.1) using state population counts for 13- and 14-year-olds as the outcome (measured in the American Community Survey), in order to examine whether our results were being driven by migration patterns (see Table 4.9). Significant effects of job losses on migration patterns would affect the interpretation of our results by raising the possibility that some of the effect of job loss on test scores might stem from changes in the composition of students taking the tests. For example, if students with higher test scores were more likely to move out of state in response to job losses, lower test scores may be (at least partially) explained by migration and not changes in student performance. However, outmigration in response to industry downsizing is believed to take an entire generation to complete (Blanchard et al., 1992), and so it is not surprising that we find no relationship between job losses and the total number of 13- and 14-year-olds in a given state the following year. We also find no evidence of changes in test-taker gender or race composition after job losses.

Examining the state population as a whole, again using data from the American Community Survey (ACS), we see no effect of job losses on the share of the population that is elderly or on racial or educational composition of the state. Again, this is consistent with the short time frame of our analysis combined with the fact that outmigration in response to economic changes is a generational process. We do, however, see a marginally significant effect of job losses on the share of the population that is poor, which is consistent with our expectation that job losses increase economic distress. This finding also provides reassurance that our lack of statistically significant effects for other population characteristics reflects an actual lack of migration and not merely noise and imprecision in the ACS data.

Examining the state population as a whole, again using data from the American Community Survey (ACS), we see no effect of job losses on the share of the population that is elderly or on racial or educational composition of the state. Again, this is consistent with the short time frame of our analysis combined with the fact that outmigration in response to economic changes is a generational process. We do, however, see a marginally significant effect of job losses on the share of the population that is poor, which is consistent with our expectation that job losses increase economic distress. This finding also provides reassurance that our lack of statistically significant effects for other population characteristics reflects an actual lack of migration and not merely noise and imprecision in the ACS data.

Table 4.9: Falsification Results - Migration of State and Test-Taker Population

Effects of 1% job loss on:	
State population of 13- and 14-year-olds	-2,938.23
	(3,200.25)
Percent of state population who are:	
Elderly	0.079
	(0.062)
Black	-0.08
	(0.095)
White	0.111
	(0.251)
Hispanic	-0.204
	(0.156)
High school dropouts	0.048
	(0.158)
High school graduates	-0.032
	(0.137)
Has some college	0.028
	(0.099)
College graduates	-0.043
	(0.115)
Poor	0.897*
	(0.492)
Percent of test-takers who are:	
Male	0.256
	(0.318)
White	-0.021
	(1.243)
Black	-0.411
	(0.436)
Hispanic	-0.000
	(0.947)

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

4.8 Interpretation

Policymakers and researchers alike have so far paid little attention to the potential effects of job losses on aggregate test scores, although they do frequently acknowledge the struggles of children facing parental job loss. One likely explanation for this oversight is that observers assume that the aggregate impacts of job loss are simply the difference in outcomes between children whose parents do and do not lose jobs scaled by the share of workers who lose jobs (Kalil and DeLeire, 2002; Hill et al., 2011; Rege et al., 2011). Since these studies find that students who face parental job loss experience outcome declines of .06 to .17 standard deviations relative to students who do not face parental job loss, observers who extrapolate from these studies to predict population-level effects of a 1% job loss would estimate effect sizes of .0006 to .0017 standard deviations. Even taking into consideration that, according to analysis of the ACS, 1.5% of eighth-graders experience job loss within the household when there is a job loss to 1% of the working-age population, estimates would still be .0009 to .00255 standard deviations. Based on such calculations, a policymaker would erroneously conclude that aggregate effects of job losses on test scores are negligible.

Our estimate of .076 standard deviations is more than an order of magnitude larger. We believe that this difference is due to the fact that our methodology captures negative effects of job loss on workers and families who maintain employment but are affected by their friends' and neighbors' job loss and the resulting changes to their communities and classrooms. If we assume that these other children are affected by statewide job loss, albeit less severely than those who experience parental unemployment, it becomes straightforward to reconcile our study with the findings of earlier studies that contrast the two groups of children, as illustrated in Table 4.10.

For example, suppose that the 98.5% of children whose parents do not lose em-

Table 4.10: Calibration: Combinations of Direct and Indirect Effects Consistent with a Population Average Effect of .076 SD from a 1% Job Loss

	A	B	C	D	E	F	G
	Spillover	Direct effect (SD) on 1.5% of population who experience parental job loss	Indirect effect on 98.5% of population who do not experience parental job loss	Difference between those who do and do not experience parental job loss	Papers finding the difference listed in column D	Estimated population effect (SD) when extrapolating to aggregate effect from column D (i.e. when assuming spillover = 0)	Share of true effect missed when extrapolating to aggregate effect from column D (i.e. when assuming spillover = 0)
(1)	0.00	5.067	0.000	5.067		0.076	0.0%
(2)	0.10	0.670	0.067	0.603		0.009	88.1%
(3)	0.20	0.358	0.072	0.287		0.004	94.3%
(4)	0.30	0.245	0.073	0.171	Hill et al.	0.003	96.6%
(5)	0.33	0.222	0.074	0.148	Kalil and DeLeire	0.002	97.1%
(6)	0.40	0.186	0.074	0.111		0.002	97.8%
(7)	0.50	0.150	0.075	0.075		0.001	98.5%
(8)	0.56	0.135	0.075	0.060	Rege et al.	0.001	98.8%
(9)	0.80	0.095	0.076	0.019		0.000	99.6%
(10)	1.00	0.076	0.076	0.000		0.000	100.0%

"Spillover" refers to effects on children whose parents do not experience job loss as a percentage of the true direct effect on children whose parents experience job loss.

ployment, but who are indirectly affected either at home or at school, experience test score declines that are one-third the magnitude of the decline experienced by the 1.5% of students who experience household job loss, a scenario displayed in row 5 of Table 4.10. In that case, a .222 standard deviation decrease in math scores among students who parents lose jobs would imply a .074 standard deviation decrease among other students, and the combination of these effects would produce a .076 standard deviation decrease in the state. The combination would also produce a .15 standard deviation decrease in the test scores of children experiencing parental job loss relative to other children, exactly the estimate that Kalil and DeLeire (2002) report for this difference. Note that extrapolating from the Kalil and DeLeire estimate by assuming that children whose parents do not lose employment experience zero declines, however, would miss 97% of the total impact in this scenario, since impacts one-third the size occur to a group nearly 100 times larger.

Comparing our estimates to those in two other studies that use the same approach as Kalil and DeLeire (2002), by Hill et al. (2011)[7] and Rege et al. (2011), suggests effects on children who do not experience parental job loss that are 30-56% the size of the effects on those who do experience parental job loss (see rows 4 and 8 of Table 4.10). These magnitudes, while striking, are consistent with the effects of downturns on adults who maintain employment relative to those who lose employment; for example, Dooley et al. (1988) find that a one-standard-deviation increase in the local unemployment rate increases an individual's psychological distress by one-fourth as much as does personal job loss.

It is not necessary to find large spillovers plausible in order to be struck by the share of aggregate effects missed when focusing only on children who experience parental job loss, however. As can be seen in row 2 of Table 4.10, even if the indirect impact is only one-tenth the size of the direct impact of job loss, analysis focusing solely on children who experience parental job loss would miss 88% of the aggregate impact of job destruction. Readers who are skeptical of the mere existence of any effects on children who do not experience household job loss should note that zero effect of job losses on the test scores of children who are not directly impacted is

[7] We use the OLS estimates from Hill et al., as their OLS empirical strategy is most comparable to the strategy that we and the other papers we discuss employ.

implausible, as that would require a 5-standard-deviation decline in the test scores of children who experience parental job loss (row 1 of Table 4.10).

4.9 Mechanisms

Given the evidence for sizeable aggregate declines in test scores, of a magnitude plausible only if driven both by directly and indirectly affected children, it is worth exploring through what mechanisms job losses may drive declines, a task to which we turn next. We consider four possibilities: changes in income; changes in school resources; changes in behavior; and changes in stress.

Family income

Family income appears to have causal effects on children's academic achievement. Dahl and Lochner (2012) find that an increase in family income of 20% increases test scores by .06 standard deviations. This magnitude, however, suggests that not all of the decline in test scores after job losses can be caused by income declines. A job loss to 1% of a state's working-age population leads, according to an analysis of the ACS, to a decline in the mean income of households containing a 13- or 14-year-old of only about 2% (effects greater than 4% can be ruled out with 95% confidence). Thus, while income losses are almost certainly a partial cause of the decline in test scores, additional mechanisms must be at work.

School Resources

Another possible reason for declining test scores after job losses is that, in the face of downturns, school budgets are reduced. We examine this possibility in Table 4.11, which presents data from the National Center for Education Statistics's National Public Education Financial Survey. A job loss to 1% of a state's working-age population has small negative but insignificant effects on state and local educational

Table 4.11: Estimated Results - Impact of Job Losses on Per Pupil School Finance

	State Revenues	Local Revenues	Federal Revenues	Instructional Expenditure	Support Expenditure	Total Ed Expenditure
	-147.798	25.181	75.315**	-124.612	-24.218	-190.217
	(165.129)	(92.002)	(30.035)	(123.329)	(59.629)	(213.867)
Mean	6,740***	5,404***	883***	7,272***	4,136***	13,110***
	(278)	(187)	(63)	(215)	(115)	(363)

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

revenues per pupil. It has a significant, but positive, effect on transfers from the federal government, consistent with the fact that federal education support to the states is partially compensatory. However, the effect on federal payments is quite small ($75 per pupil); the fact that it is significant while the effect on state revenues is not, despite being larger at -$147, likely reflects the fact that federal payments are measured with less error.

Likewise, the effects of job losses on instructional expenditures, support expenditures, and total educational expenditures per pupil are negative but insignificant and small. The point estimate for the change in per pupil education expenditures, -$190, represents less than 1.5% of per pupil spending. There is little evidence that changes in school spending of this magnitude have measureable effects on student achievement (Hanushek, 2003), and it is highly unlikely that they can account for the large and significant effects on test scores that we find.

Behavior

In his classic book *When Work Disappears*, William JuliusWilson (1996) argues that local job losses lead from idleness among adults to increased antisocial role-modeling behavior, such as drug abuse and violence, and from there to youth antisocial behav-

Table 4.12: Effects of Job Loss on Youth Behaviors

Ever used alcohol	-0.0236***
	(0.0081)
Ever used marijuana	-0.0149
	(0.0094)
Ever had sex	0.0024
	(0.0118)
Number of recent sex partners	-0.0514**
	(0.0245)
Used contraception last time had sex	0.1249***
	(0.0124)
Carried a weapon to school in the last year	-0.0049
	(0.0035)
Physical fight in the last year	0.0037
	(0.0074)
Felt unsafe at school in the last year	0.0059
	(0.0046)
Suicidal thoughts in the last year	0.0046
	(0.0055)
Suicidal plans in the last year	0.0083*
	(0.0048)

Robust standard errors in parentheses

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

Each cell represents the coefficient and standard error on Separations. The specification includes state and year fixed effects.

ior, disengagement from and failure in school. We examine the possibility of such a "downward spiral of behavior" using data on teen behaviors taken from the Youth Risk Behavior Survey (YRBS). The YRBS is fielded biannually in February through April of odd years by the Centers for Disease Control, and is the definitive source of widely-reported statistics such as the share of teens who are sexually active or who have experimented with illegal drugs.

Table 4.12 presents results from regressions estimating the effect of a job loss to 1% of the working-age population on youth behaviors. Students are significantly less likely to report having consumed alcohol and insignificantly less likely to have used marijuana after job losses. They report significantly fewer recent (over the last 3 months) sex partners and are significantly more likely to have used contraception the last time they had sex. There is no change in the probability of having carried a weapon to school in the past year, or in having engaged in a physical fight. These results are not consistent with a "downward spiral of behavior"; rather, youth appear to have better behaviors during downturns. We find no evidence that antisocial behavior trends can account for the decline in test scores.

Distress

The last row of Table 4.12 shows the effects of job losses on the probability that a youth has made plans to commit suicide over the past year, a behavior also tracked in the YRBS. There is a significant rise in this activity of about 8%, from a base of 1 in 10 teens. This suggests emotional distress among youth during downturns on par with increases in adult distress. Moreover, it suggests that, as among adults (Fenwick and Tausig, 1994), distress is widespread rather than concentrated among those who experience household job loss; in order for directly impacted youth to drive these findings entirely, half of them would have to have made plans for suicide in the past year, an implausible magnitude. Because mental health problems can inhibit learning (Fergusson and Woodward, 2002) this mechanism may account for part of the decline in test scores we observe.

4.10 Conclusion

This paper finds that students experience sizeable declines in eighth-grade math test scores in the wake of economic downturns. We argue that students who do not experience parental job loss, as well as those who do, are hurt by downturns. The failure in previous research to capture effects on the former group of students means that inferences drawn from that research have understated aggregate effects of downturns by an order of magnitude. When correctly measured, aggregate effects are comparable to effects of policy interventions, such as Tennessee STAR (Word et al., 1990), that have generated enormous policy interest. States with large job losses (we observe maximum losses of 3.4%) are predicted to experience average test score declines of over 25% of a standard deviation, or nearly 10 points. The magnitude of these effects suggests that costs to students from downturns are a relevant consideration, along with other costs of recessions, for policymakers considering economic stimulus and other policies to mitigate effects of the business cycle.

In addition, in this era of greatly increased focus on school accountability for student performance, education policy makers and leaders should be cognizant of the external factors that can negatively influence student achievement. Given the accountability standards enacted in NCLB legislation, even small changes in average test scores could have large implications, if they change schools' proficiency levels. In the 2009-10 school year, 38% of schools failed to make Adequate Yearly Progress (AYP) as mandated under NCLB (CEP, 2011). Under conservative assumptions, we estimate that a state that experienced a downturn leading to job losses to 1% of its workers (a magnitude that we observe in most states during our panel) would have had only 35% of its schools fail to achieve AYP in the absence of a downturn, an 8% decline.[8]

Statewide job losses, which result from factors external to schools such as pressure from globalization and macroeconomic conditions, can significantly influence student achievement and are well beyond the control of teachers and school administrators. The significant effect these losses can have on schools' abilities to meet accountability goals suggests that policymakers may want to consider recent economic change when defining whether a school is meeting accountability targets.

[8] This calculation assumes: that school-level test score standard deviations are closer to student than to state standard deviations (30 points, compared to 36 for students and 9 for states, a conservative assumption); and that school averages are normally distributed.

Appendix A

Data Appendix for Chapter 2

Table A.1: Sample Averages, 2005-2006 School Year for Elementary Schools

	All	Treatment	Comparison
N	5567	1648	3919
Total Enrollment	554	618	528
Outcomes			
% Fully Credentialed	96.8	95.2	97.4
% Not Fully Credentialed	4.4	6.4	3.5
Average Experience (yrs)	12.8	11.6	13.3
% with MA	32.8	29.2	34.3
% with BA	66.4	69.8	64.9
Pupil/Teacher Ratio	19.3	19.3	19.4
API Score	765	664	808
English Language Arts Z-Score	0.04	-0.97	0.46
Math Z-Score	0.07	-0.85	0.46
Algebra Z-Score	–	–	–
Student Demographics			
% ELL	28.5	48.1	20.3
% FRL	53.8	82.7	41.7
% American Indian	1.2	1.4	1.1
% Asian	10.9	5.3	13.2
% Black	7.3	9.6	6.3
% Hispanic	45.5	70.6	34.9
% White	32.6	11.9	41.4
% Multi Ethnic	2.5	1.2	3.1
Teacher Demographics			
% Male	14.7	17.5	13.5
% Female	85.3	82.5	86.4
% American Indian	0.6	0.6	0.5
% Asian	6.2	6.0	6.3
% Black	3.6	6.2	2.5
% Hispanic	15.1	27.2	10.0
% White	73.3	58.5	79.5
% Multi Ethnic	1.2	1.5	1.1

Sample Averages for the treatment and comparison groups are significantly different from each other for all variables except Pupil/Teacher Ratio, Percent of Teachers who are American Indian, and Percent of Teachers who are Asian.

Source: California DataQuest Management System (http://dq.cde.ca.gov/dataquest/)

Table A.2: Sample Averages, 2005-2006 School Year for Middle Schools

	All	Treatment	Comparison
N	1298	396	902
Total Enrollment	901	1021	848
Outcomes			
% Fully Credentialed	92.1	87.0	94.3
% Not Fully Credentialed	10.2	15.9	7.6
Average Experience (yrs)	12.1	10.6	12.7
% with MA	35.2	32.4	36.4
% with BA	63.3	65.8	62.2
Pupil/Teacher Ratio	22.6	22.5	22.7
API Score	728	622	775
English Language Arts Z-Score	-0.06	-1.10	0.39
Math Z-Score	-0.02	-1.00	0.40
Algebra Z-Score	0.36	-0.45	0.71
Student Demographics			
% ELL	20.2	34.6	13.9
% FRL	49.2	76.3	37.3
% American Indian	1.0	0.7	1.1
% Asian	11.1	6.5	13.1
% Black	8.7	13.9	6.4
% Hispanic	44.4	68.1	34.0
% White	33.1	10.1	43.3
% Multi Ethnic	1.7	0.8	2.0
Teacher Demographics			
% Male	36.6	41.0	34.7
% Female	63.3	58.8	65.3
% American Indian	0.7	0.8	0.6
% Asian	5.4	6.3	5.1
% Black	5.7	11.6	3.1
% Hispanic	11.8	19.8	8.3
% White	74.8	59.1	81.6
% Multi Ethnic	1.6	2.5	1.3

Sample Averages for the treatment and comparison groups are significantly different from each other for all variables except Pupil/Teacher Ratio and Percent of Teachers who are American Indian.

Source: California DataQuest Management System (http://dq.cde.ca.gov/dataquest/)

Table A.3: Sample Averages, 2005-2006 School Year for High Schools

	All	Treatment	Comparison
N	1283	445	838
Total Enrollment	1428	1272	1511
Outcomes			
% Fully Credentialed	90.1	85.4	92.6
% Not Fully Credentialed	11.8	16.9	9.1
Average Experience (yrs)	12.6	12.0	13.0
% with MA	38.5	36.4	39.6
% with BA	58.7	60.5	57.7
Pupil/Teacher Ratio	22.4	22.6	22.3
API Score	687	583	743
English Language Arts Z-Score	-0.10	-1.09	0.43
Math Z-Score	0.02	-0.78	0.28
Algebra Z-Score	-0.22	-0.97	0.12
Student Demographics			
% ELL	14.4	22.9	9.9
% FRL	37.6	54.1	28.9
% American Indian	1.7	1.9	1.6
% Asian	10.1	6.7	11.9
% Black	8.4	13.3	5.9
% Hispanic	37.4	53.0	29.2
% White	40.0	23.0	49.0
% Multi Ethnic	2.3	2.1	2.4
Teacher Demographics			
% Male	47.3	47.2	47.4
% Female	52.6	52.7	52.5
% American Indian	0.9	1.0	0.9
% Asian	4.9	5.4	4.7
% Black	4.9	9.1	2.7
% Hispanic	12.0	16.9	9.4
% White	75.5	65.2	81.0
% Multi Ethnic	1.7	2.3	1.4

Sample Averages for the treatment and comparison groups are significantly different from each other for all variables except Pupil/Teacher Ratio, Percent of Students who are American Indian or Multi-Ethnic, Percent of Teachers who are Male/Female, and Percent of Teachers who are American Indian.

Source: California DataQuest Management System (http://dq.cde.ca.gov/dataquest/)

Table A.4: LAUSD Sample Averages, 2005-2006 School Year

	All	**Treatment**	**Comparison**
N	450	207	243
Total Enrollment	758	939	603
		Outcomes	
API Score	732	660	792
% Fully Credentialed	96.3	95.7	96.8
% Not Fully Credentialed	4.0	4.7	3.4
Average Experience (yrs)	11.8	11.2	12.4
% with MA	30.7	30.2	31.1
% with BA	68.2	68.8	67.8
Pupil/Teacher Ratio	19.0	19.1	18.8
		Student Demographics	
% ELL	45.0	59.0	33.1
% FRL	77.4	91.7	65.1
% American Indian	0.4	0.2	0.5
% Asian	7.0	1.9	11.4
% Black	11.6	12.2	11.0
% Hispanic	69.4	84.4	56.7
% White	11.6	1.3	20.3
% Minority	80.1	96.6	67.7
		Teacher Demographics	
% Male	18.5	21.9	15.5
% Female	81.5	78.0	84.5
% American Indian	0.6	0.4	0.8
% Asian	12.2	9.4	14.6
% Black	11.9	15.1	9.1
% Hispanic	30.9	41.2	22.2
% White	44.4	34.0	53.2
% Minority	42.8	56.3	31.4

Asian includes Filipinos and Pacific Islanders

Minority is a combination of Black and Hispanic

Sample Averages for the treatment and comparison groups are significantly different from each other for all variables except percent of students who are black, and percent of teachers who have bachelors or masters degrees.

Source: California DataQuest Management System (http://dq.cde.ca.gov/dataquest/)

Appendix B

Data Appendix for Chapter 3

Table B.1: Marginal Effects of Probit Estimation of Math Teacher Reports of Male Absence Relative to Female Absence, by Race

	ECLS-K	NCERDC
White Male	-0.016	-0.024
	(0.026)	(0.025)
Black Male	-0.137**	-0.153***
	(0.062)	(0.059)
Hispanic Male	-0.012	0.006
	(0.060)	(0.056)
Asian Male	0.007	-0.030
	(0.083)	(0.072)

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table B.2: Marginal Effects of Probit Estimation of Science Teacher Reports of FRL Student Absence Relative to Non-FRL Student Absence, by Race

	ECLS-K			NCERDC		
	English	Math	Science	English	Math	Science
White	0.105***	0.106**	0.215***	0.104**	0.069*	0.175***
	(0.040)	(0.042)	(0.052)	(0.041)	(0.040)	(0.050)
Black	-0.015	0.136**	-0.016	0.032	0.088	0.041
	(0.043)	(0.058)	(0.088)	(0.039)	(0.067)	(0.087)
Hispanic	0.044	0.151**	-0.032	0.026	0.103*	0.027
	(0.034)	(0.069)	(0.053)	(0.040)	(0.061)	(0.047)
Asian	0.138**	0.000	0.228***	0.123**	-0.035	0.178***
	(0.059)	(0.090)	(0.084)	(0.049)	(0.078)	(0.065)

Standard errors in parentheses

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

Appendix C

Data Appendix for Chapter 4

Table C.1: States Reporting both Test Score and Job Loss Data

Year	1996	1998	2000	2002	2003	2005	2007	2009
Test Subject	Math	Reading	Math	Reading	Both	Both	Both	Both
Alabama	T, J	T	T,J	T,J	T	T	T,J	T,J
Alaska	T, J			J	T,J	T	T,J	T,J
Arizona	T	T	T,J	T,J	T,J	T,J	T,J	T,J
Arkansas	T, J	T	T,J	T,J	T,J	T	T,J	T,J
California	T, J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Colorado	T, J	T		J	T,J	T	T,J	T,J
Connecticut	T, J	T	T,J	T,J	T,J	T,J	T,J	T,J
Delaware	T	T,J		T,J	T	T	T	T,J
D.C.	T	T	T	T	T	T	T	T
Florida	T, J	T,J	J	T,J	T,J	T,J	T,J	T,J
Georgia	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Hawaii	T	T,J	T,J	T,J	T	T	T,J	T,J
Idaho	J	J	T	T,J	T,J	T,J	T,J	T,J
Illinois	J	J	T,J	J	T,J	T,J	T,J	T,J
Indiana	T,J	J	T,J	T,J	T,J	T,J	T,J	T,J
Iowa	T	J	J	J	T,J	T,J	T	T,J
Kansas	J	T	T,J	T,J	T,J	T,J	T,J	T,J
Kentucky	T	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Louisiana	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Maine	T,J	T,J	T	T,J	T	T	T	T,J
Maryland	T,J	T,J	T,J	T,J	T,J	T	T,J	T,J
Massachusetts	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Michigan	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Minnesota	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Mississippi	T,J	T,J	T,J	T,J	T,J	T,J	T	T,J
Missouri	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Montana	T,J	T	T	T	T,J	T	T	T,J
Nebraska	T		T	T	T,J	T	T	T
Nevada	J	T,J	T,J	T,J	T	T	T	T,J
New Hampshire				J	T	T	T	T
New Jersey	J	J	J	J	T,J	T,J	T,J	T,J
New Mexico	T	T,J	T	T,J	T	T	T,J	T,J
New York	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
North Carolina	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
North Dakota	T	J	T	T	T,J	T,J	T	T
Ohio	J	J	T,J	T,J	T,J	T,J	T,J	T,J
Oklahoma		T,J	T	T,J	T,J	T	T,J	T,J

Continued on next page

Table C.1 – continued from previous page

Year	1996	1998	2000	2002	2003	2005	2007	2009
Test Subject	Math	Reading	Math	Reading	Both	Both	Both	Both
Oregon	T	T	T,J	T,J	T,J	T,J	T,J	T,J
Pennsylvania	J	J	J	T,J	T,J	T,J	T,J	T,J
Rhode Island	T	T	T	T,J	T	T	T	T
South Carolina	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
South Dakota			J		T	T	T	T
Tennessee	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Texas	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Utah	T	T	T,J	T,J	T	T	T,J	T
Vermont	T,J	J	T	T	T	T	T	T
Virginia	T,J	T,J	T,J	T,J	T,J	T,J	T,J	T,J
Washington	T,J	T,J	J	T,J	T,J	T,J	T,J	T,J
West Virginia	T,J	T	T	T	T,J	T	T,J	T
Wisconsin	T,J	T,J	J	J	T,J	T,J	T,J	T,J
Wyoming	T	T	T	T	T	T	T	T
# of States Reporting Both	28	25	28	36	38	30	37	41

T - NAEP test score data are available for a given state-year

J - BLS mass closing or mass layoff data are available for a given state-year

Source: Bureau of Labor Statistics - http://www.bls.gov/mls/

National Center for Education Statistics - http://nces.ed.gov/nationsreportcard/

Bibliography

(2010a), *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) Kindergarten Through Fifth Grade Parent and Teacher Social Rating Scale (SRS) Items and Restricted-Use Data Files*, U.S. Department of Education. National Center for Education Statistics, Washington, DC, (nces 2010071) edn.

(2010b), *NAEP: Measuring Student Progress Since 1964*, U.S. Department of Education. National Center for Education Statistics, Washington, DC.

(2011), "How Many Schools Have Not Made Adequate Yearly Progress?" Tech. rep., Center on Education Policy, Washington, D. C.

Aaronson, D., Barrow, L., and Sanders, W. (2007), "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 25, 95–135.

Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T. J., and Pathak, P. (2009), "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," Tech. rep., National Bureau of Economic Research.

Ananat, E. O., Gassman-Pines, A., and Gibson-Davis, C. (2011), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, chap. The Effects of Local Employment Losses on Children's Educational Achievement, pp. 299–313, New York, Russell Sage Foundation.

Angrist, J. and Lavy, V. (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 114, 533–575.

Angrist, J. and Pischke, S. (2009), *Mostly Harmless Econometrics: An Empiricists Companion*, Princeton, NJ: Princeton University Press.

Angrist, J., Dynarski, S., Kane, T., Pathak, P., and Walters, C. (2010), "Who benefits from KIPP?" Tech. rep., National Bureau of Economic Research.

Arrow, K. (1971), "The Theory of Discrimination," Princeton University.

Bishop, K., Dudley, P. M., Mihaly, K., and Murphy, A. (2005), "The Lasting Impact of Teachers: An Analysis of Teacher Race and Student Achievement," Duke University.

Black, D. A., McKinnish, T. G., and Sanders, S. G. (2003), "Does the Availability of High-Wage Jobs for Low-Skilled Men Affect Welfare Expenditures? Evidence from Shocks to the Steel and Coal Industries," *Journal of Public Economics*, 87, 1921–1942.

Black, D. A., McKinnish, T. G., and Sanders, S. G. (2005a), "The Economic Impact Of The Coal Boom And Bust," *The Economic Journal*, 115, 449–476.

Black, D. A., McKinnish, T. G., and Sanders, S. G. (2005b), "Tight Labor Markets and the Demand for Education: Evidence from the Coal Boom and Bust," *Indus. & Lab. Rel. Rev.*, 59, 3.

Blanchard, O. J., Katz, L. F., Hall, R. E., and Eichengreen, B. (1992), "Regional Evolutions," *Brookings Papers on Economic Activity*, 1992, 1–75.

Blanchflower, D. G. and Oswald, A. J. (1995), *The Wage Curve*, MIT Press.

Boozer, M. and Rouse, C. (2001), "Intraschool Variation in Class Size: Patterns and Implications," *Journal of Urban Economics*, 50, 163–189.

Borman, G. D. and Dowling, N. M. (2008), "Teacher Attrition and Retention: A Meta-Analytic and Narrative Review of the Research," *Review of Educational Research*, 78, 367–409.

Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005), "Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students," *The American Economic Review*, 95, 166–171.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., and Wyckoff, J. (2010), "The Effect of School Neighborhoods on Teacher Career Decisions," *Working Paper*.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., and Wyckoff, J. (2011), "The Role of Teacher Quality in Retention and Hiring: Using Applications to Transfer to Uncover Preferences of Teachers and Schools," *Journal of Policy Analysis and Management*, 30, 88–110.

Boykin, A. W., Tyler, K. M., and Miller, O. (2005), "In Search of Cultural Themes and Their Expressions in the Dynamics of Classroom Life," *Urban Education*, 40, 521–549.

Card, D., Dobkin, C., and Maestas, N. (2009), "Does Medicare Save Lives?" *The Quarterly Journal of Economics*, 124, 597.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2011), "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," *NBER Working Paper*, 17699.

Clark, A., Knabe, A., and Rätzel, S. (2010), "Boon or Bane? Others' Unemployment, Well-Being and Job Insecurity," *Labour Economics*, 17, 52–61.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. (2005), "Who Teaches Whom? Race and the Distribution of Novice Teachers," *Economics of Education Review*, 24, 377–392.

Clotfelter, C. T., Ladd, H., Vigdor, J., and Wheeler, J. (2006-2007a), "High-Poverty Schools and the Distribution of Teachers and Principals," *North Carolina Law Review*, 85, 1345–1380.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007b), "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects," *Economics of Education Review*, 27, 673–682.

Clotfelter, C. T., Glennie, E., Ladd, H., and Vigdor, J. (2008), "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina," *Journal of Public Economics*, 92, 1352–1370.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2010), "Teacher Mobility, School Segregation, and Pay-Based Policies to Level the Playing Field," *Education Finance and Policy*, Forthcoming.

Coelli, M. B. (2011), "Parental Job Loss and the Education Enrollment of Youth," *Labour Economics*, 18, 25–35.

Condron, D. J. (2007), "Stratification and Educational Sorting: Explaining Ascriptive Inequalities in Early Childhood Reading Group Placement," *Social Problems*, 54, 139–160.

Conger, R. and Elder, G. H. (1994), *Families in Troubled Times: Adapting to Change in Rural America*, Aldine de Gruyter.

Dahl, G. B. and Lochner, L. (2012), "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," *The American Economic Review*, 102, 1927–1956.

Darity, W. and Frank, D. (2003), "The Economics of Reparations," *The American Economic Review*, 93, 326–329.

Darity Jr., W. and Jolla, A. (2009), "Desegregated Schools With Segregated Education," in *The Integration Debate: Futures for American Cities*, eds. C. Hartman and G. Squires, pp. 99–117, Routledge, New York.

Darity Jr, W. A., Lahiri, B., and Frank, D. V. (2010), "Reparations for African-Americans as a Transfer Problem: A Cautionary Tale," *Review of Development Economics*, 14, 248–261.

Dauber, S. L. and Epstein, J. L. (1993), "Parents' Attitudes and Practices of Involvement in Inner-City Elementary and Middle Schools," in *Families and Schools in a Pluralistic Society*, ed. N. F. Chavkin, chap. 2, pp. 53–72, SUNY Press, Albany, NY.

Dauber, S. L., Alexander, K. L., and Entwisle, D. R. (1996), "Tracking and Transitions through the Middle Grades: Channeling Educational Trajectories," *Sociology of Education*, 69, 290–307.

Decker, P. T., Mayer, D. P., and Glazerman, S. (2004), "The Effects of Teach For America on Students: Findings from a National Evaluation," Tech. rep.

Dee, T. S. (2005), "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review*, 95, 158–165.

Dobbie, W. (2011), "Teacher Characteristics and Student Achievement: Evidence from Teach for America," *Working Paper*.

Dobbie, W. and Fryer, R. G. (2011), "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 3, 158–187.

Dooley, D. and Catalano, R. (1984), "The Epidemiology of Economic Stress," *American Journal of Community Psychology*, 12, 387–409.

Dooley, D., Catalano, R., and Rook, K. S. (1988), "Personal and Aggregate Unemployment and Psychological Symptoms," *Journal of Social Issues*, 44, 107–123.

Downey, D. B. and Pribesh, S. (2004), "When Race Matters: Teachers' Evaluations of Students' Classroom Behavior," *Sociology of Education*, 77, 16.

Downey, G. and Coyne, J. C. (1990), "Children of Depressed Parents: an Integrative Review." *Psychological bulletin*, 108, 50.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., and Mac Iver, D. (1993), "Development During Adolescence: The Impact of Stage Environment Fit on Young Adolescents' Experiences in Schools and in Families." *American Psychologist*, 48, 90.

Eder, D. (1981), "Ability Grouping as a Self-Fulfilling Prophecy: A Micro-Analysis of Teacher-Student Interaction," *Sociology of Education*, 54, 151–162.

Edwards, R. (2008), "American Time Use over the Business Cycle," Tech. rep., mimeo, Queens College and the Graduate Center, City University of New York.

Ehrenberg, R. G., Goldhaber, D. D., and Brewer, D. J. (1995), "Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988," *Industrial and Labor Relations Review*, 48, 547–561.

Elder Jr, G. H., Eccles, J. S., Ardelt, M., and Lord, S. (1995), "Inner-City Parents under Economic Pressure: Perspectives on the Strategies of Parenting," *Journal of Marriage and the Family*, pp. 771–784.

Feng, L. (2009), "Opportunity Wages, Classroom Characteristics, and Teacher Mobility," *Southern Economic Journal*, 75, 1165–1190.

Fenwick, R. and Tausig, M. (1994), "The Macroeconomic Context of Job Stress," *Journal of Health and Social Behavior*, pp. 266–282.

Fergusson, D. M. and Woodward, L. J. (2002), "Mental Health, Educational, and Social Role Outcomes of Adolescents with Depression," *Archives of General Psychiatry*, 59, 225.

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

Forster, M., Sundell, K., Morris, R. J., Karlberg, M., and Melin, L. (2010), "A Randomized Controlled Trial of a Standardized Behavior Management Intervention for Students With Externalizing Behavior," *Journal of Emotional and Behavioral Disorders*.

Francis, D. V. (2012), "Sugar and Spice and Everything Nice? Teacher Perceptions of Black Girls in the Classroom," *The Review of Black Political Economy*, 39, 311–320.

Goldhaber, D. (2008), "Teachers Matter, But Effective Teacher Quality Policies are Elusive," in *Handbook of Research in Education Finance and Policy*, eds. H. F. Ladd and E. B. Fiske, chap. 9, pp. 146–165, Routledge, New York.

Goldwater, O. D. and Nutt, R. L. (1999), "Teachers' and students' work-culture variables associated with positive school outcome," *Adolescence*, 34, 653–653.

Guarino, C. M., Santibanez, L., and Daley, G. A. (2006), "Teacher Recruitment and Retention: A Review of the Recent Empirical Literature," *Review of Educational Research*, 76, 173–208.

Hamre, B. K. and Pianta, R. C. (2003), "Early Teacher–Child Relationships and the Trajectory of Children's School Outcomes through Eighth Grade," *Child Development*, 72, 625–638.

Hanushek, E., Kain, J., O'Brien, D., and Rivkin, S. (2005), "The market for teacher quality," .

Hanushek, E. A. (1997), "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, 19, 141–164.

Hanushek, E. A. (2003), "The Failure of Input-based Schooling Policies," *The Economic Journal*, 113, F64–F98.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004), "Why Public Schools Lose Teachers," *The Journal of Human Resources*, XXXIX, 326–354.

Harris, A. (2006), "I (Don't) Hate School: Revisiting Oppositional Culture Theory of Blacks' Resistance to Schooling," *Social Forces*, 85, 797–833.

Harris, D. N. and Sass, T. R. (2011), "Teacher Training, Teacher Quality and Student Achievement," *Journal of Public Economics*, 95, 798–812.

Hill, H. D., Morris, P. A., Castells, N., and Walker, J. T. (2011), "Getting a Job is Only Half the Battle: Maternal Job Loss and Child Classroom Behavior in Low-Income Families," *Journal of Policy Analysis and Management*, 30, 310–333.

Hoxby, C. (2000), "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *The Quarterly Journal of Economics*, 115, 1239–1285.

Hoxby, C. M. and Murarka, S. (2009), "Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement," Tech. rep., National Bureau of Economic Research.

Hughes, J. N., Gleason, K. A., and Zhang, D. (2005), "Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students," *Journal of School Psychology*, 43, 303–320.

Hughes, J. N., Luo, W., Kwok, O.-M., and Loyd, L. K. (2008), "Teacher-Student Support, Effortful Engagement, and Achievement: A 3-Year Longitudinal Study," *Journal of Educational Psychology*, 100, 1–14.

Imbens, G. W. and Lemieux, T. (2008), "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615–635.

Jackson, C. K. (2010), "Match quality, worker productivity, and worker mobility: Direct evidence from teachers," Tech. rep., National Bureau of Economic Research.

Jacob, B. A. (2007), "The Challenges of Staffing Urban Schools with Effective Teachers," *The Future of Children*, 17, 129–153.

Jacobson, L. S., LaLonde, R. J., and Sullivan, D. G. (1993), "Earnings Losses of Displaced Workers," *The American Economic Review*, pp. 685–709.

Jones, L. P. (1988), "The Effect of Unemployment on Children and Adolescents," *Children and Youth Services Review*, 10, 199–215.

Jussim, L. and Harber, K. D. (2005), "Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies," *Personality and Social Psychology Review*, 9, 131–155.

Kalil, A. and DeLeire, T. (2002), "Parental Job Loss and Early Adolescent Adjustment in Black and White Families," Tech. rep., Working Paper 282 Northwestern University/University of Chicago: Joint Center for Policy Research.

Kalil, A. and Wightman, P. (2010), "Parental Job Loss and Family Conflict," *National Center for Family and Marriage Research, Bowling Green State University, Working paper No. WP-10*, 7.

Kalil, A. and Ziol-Guest, K. (2011), "The Great Recession and Married Parents Use of Time," in *Society for Research in Child Development Biennial Meeting. Montreal, Quebec, Canada*.

Krueger, A. B. (1999), "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 114, 497–532.

Lai, F., Sadoulet, E., and de Janvry, A. (2011), "The Contributions of School Quality and Teacher Qualifications to Student Performance Evidence from a Natural Experiment in Beijing Middle Schools," *Journal of Human Resources*, 46, 123–153.

Lankford, H., Loeb, S., and Wyckoff, J. (2002), "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis," *Educational Evaluation and Policy Analysis*, 24, 37–62.

Levine, P. B. (2011), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, chap. How Does Parental Unemployment Affect Childrens Educational Performance?, New York, Russell Sage Foundation.

Lleras, C. and Rangel, C. (2009), "Ability Grouping Practices in Elementary School and African American/Hispanic Achievement," *American Journal of Education*, 115, 279 – 304.

Louie, V. and Holdaway, J. (2009), "Catholic schools and immigrant students: A new generation," *The Teachers College Record*, 111, 783–816.

Ludwig, J. and Miller, D. L. (2007), "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *The Quarterly Journal of Economics*, 122, 159–208.

Luechinger, S., Meier, S., and Stutzer, A. (2010), "Why Does Unemployment Hurt the Employed? Evidence from the Life Satisfaction Gap Between the Public and the Private Sector," *Journal of Human Resources*, 45, 998–1045.

McKee-Ryan, F., Song, Z., Wanberg, C. R., and Kinicki, A. J. (2005), "Psychological and Physical Well-Being During Unemployment: a Meta-Analytic Study." *Journal of Applied Psychology*, 90, 53.

McLoyd, V. C. (1998), "Socioeconomic Disadvantage and Child Development." *American Psychologist*, 53, 185–204.

McLoyd, V. C., Jayaratne, T. E., Ceballo, R., and Borquez, J. (1994), "Unemployment and Work Interruption among African American Single Mothers: Effects on Parenting and Adolescent Socioemotional Functioning," *Child Development*, 65, 562–589.

Miller, C. (2009), "Teacher Sorting and Own-Race Teacher Effects in Elementary School," Honors, Stanford University, Stanford, CA.

Mistry, R. S., Benner, A. D., Tan, C. S., and Kim, S. Y. (2009), "Family Economic Stress and Academic Well-Being among Chinese-American Youth: The Influence of Adolescents Perceptions of Economic Strain." *Journal of Family Psychology; Journal of Family Psychology*, 23, 279–290.

Morris, P. A. and Gennetian, L. A. (2004), "Identifying the Effects of Income on Children's Development using Experimental Data," *Journal of Marriage and Family*, 65, 716–729.

Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., and Keltikangas-Jarvinen, L. (2011), "Gender Differences in Teachers' Perceptions of Students' Temperament, Educational Competence, and Teachability," *British Journal of Educational Psychology*.

Murnane, R. J. and Steele, J. L. (2007), "What Is the Problem? The Challenge of Providing Effective Teachers for All Children," *The Future of Children*, 17, 15–43.

Nichols, A. (2011), "rd 2.0: Revised Stata Module for Regression Discontinuity Estimation," .

Nye, B., Konstantopoulos, S., and Hedges, L. V. (2004), "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis*, 26, 237–257.

Oakes, J. (2005), *Keeping Track: How Schools Structure Inequality*, Yale University Press, New Haven, 2 edn.

Ochsen, C. (2008), “Subjective Well-Being and the Duration of Aggregate Unemployment in Europe,” Tech. rep., Thünen-Series of Applied Economic Theory, Working Paper #97.

Ochsen, C. and H., W. (2006), “The Social Costs of Unemployment: Accounting for Employment Duration,” Tech. rep., Thünen-Series of Applied Economic Theory, Working Paper #60.

Oreopoulos, P., Page, M. E., and Stevens, A. H. (2008), “The Intergenerational Effect of Worker Displacement,” Tech. rep., National Bureau of Economic Research.

Oreopoulos, P., Von Wachter, T., and Heisz, A. (2012), “The Short-and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates,” Tech. rep., National Bureau of Economic Research.

Ouazad, A. (2008), “Assessed by a Teacher Like Me: Race, Gender, and Subjective Evaluations,” *INSEAD Working Paper No. 2008/57/EPS*.

Phelps, E. S. (1972), “The Statistical Theory of Racism and Sexism,” *The American Economic Review*, 62, 659–661.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., and Morrison, F. J. (2008), “Classroom Effects on Childrens Achievement Trajectories in Elementary School,” *American Educational Research Journal*, 45, 365–397.

Pope, D. G. and Sydnor, J. R. (2010), “Geographic Variation in the Gender Differences in Test Scores,” *The Journal of Economic Perspectives*, 24, 95–108.

Rege, M., Telle, K., and Votruba, M. (2011), “Parental Job Loss and Children’s School Performance,” *The Review of Economic Studies*, 78, 1462–1489.

Rockoff, J. E. (2004), “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data,” *The American Economic Review*, 94, 247–252.

Ruggles, S., Sobek, M., ALEXANDER, T., Genadek, K., Goeken, R., and Schroeder, M. (2004), “Integrated Public Use Microdata Series (IPUMS): Version 3.0 [Machine-readable database],” *University of Minnesota, Minneapolis, available at http://usa. ipums. org/usa*.

Scafidi, B., Sjoquist, D. L., and Stinebrickner, T. R. (2007), “Race, Poverty and Teacher Mobility,” *Economics of Education Review*, 26, 145–159.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin Company, Boston.

Staiger, D. O. and Rockoff, J. E. (2010), "Searching for Effective Teachers with Imperfect Information," *The Journal of Economic Perspectives*, 24, 97–117.

Steele, C. M. and Aronson, J. (1995), "Stereotype threat and the intellectual test performance of African Americans," *Journal of personality and social psychology*, 69, 797–811.

Stevens, A. H. (1997), "Persistent Effects of Job Displacement: The Importance of Multiple Job Losses," *Journal of Labor Economics*, pp. 165–188.

Stevens, A. H. and Schaller, J. (2011), "Short-Run Effects of Parental Job Loss on Children's Academic Achievement," *Economics of Education Review*, 30, 289–299.

Stock, J., Wright, J., and Yogo, M. (2002), "A survey of weak instruments and weak identification in generalized method of moments," *Journal of business and economic statistics*, 20, 518–529.

Sullivan, D. and Von Wachter, T. (2009), "Job Displacement and Mortality: An Analysis Using Administrative Data," *The Quarterly Journal of Economics*, 124, 1265–1306.

Tyler, K. M., Boykin, A. W., and Walton, T. R. (2006), "Cultural considerations in teachers' perceptions of student classroom behavior and achievement," *Teaching and Teacher Education*, 22, 998–1005.

Tyler, K. M., Dillihunt, M. L., Boykin, A. W., Coleman, S. T., Scott, D. M., Tyler, C. M. B., and Hurley, E. A. (2008), "Examining Cultural Socialization Within African American and European American Households," *Cultural Diversity and Ethnic Minority Psychology*, 14, 201–204.

Van Ewijk, R. (2011), "Same Work, Lower Grade? Student Ethnicity and Teachers Subjective Assessments," *Economics of Education Review*, 30, 1045–1058.

Wiley, C. (2000), "A Synthesis of Research on the Causes, Effects, and Reduction of Strategies of Teacher Stress," *Journal of Instructional Psychology*, 27, 80–87.

Wilson, W. J. (1996), *When Work Disappears: The World of the Urban Poor*, New York: Alfred Knopf.

Word, E., Johnston, J., Bain, H., Fulton, B., Zaharias, J., Achilles, C., Lintz, M., Folger, J., and Breda, C. (1990), "The State of Tennessees Student/Teacher Achievement Ratio (Star) Project Final Summary Report," *Tennessee State Department of Education*.

Biography

Dania Veronica Francis (née Frank) was born and raised in southern California. In 1997, she graduated from the California Academy of Math and Science in Carson, California prior to attending Smith College in Northampton, Massachusetts. She earned a bachelor of arts in Economics from Smith College in 2001, graduating *cum laude*. In 2003, Dania earned a masters degree in Economics from Harvard University. Her Doctorate was awarded from Duke University's Sanford School of Public Policy in 2013.

Dania was awarded a Dean's Graduate Fellowship at Duke University. She also received funding through the William T. Grant Foundations Scholar Mentoring Program. During her time in graduate school she published articles in *The American Economic Review*, *The Review of Development Economics*, and *The Review of Black Political Economy* (Darity and Frank, 2003; Darity Jr et al., 2010; Francis, 2012). In the fall of 2013 she will be joining the faculty of the University of Massachusetts at Amherst as an Assistant Professor in the Department of Economics, with an adjunct appointment in Afro-American Studies.