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

Dania V. Francis*

October 9, 2013

Abstract

Non-cognitive skills contribute to both academic and labor market success. In the classroom, teachers often rely on their assessments of students' non-cognitive abilities when evaluating overall student performance and when recommending students for advanced or remedial class placement. Given the potential impact on long-term student outcomes, it is important to identify any biases in these perceptions that could disadvantage subgroups of students. In this paper, I estimate demographic subgroup differences in teacher reports of student absenteeism while controlling for administrative records of actual absences. I find evidence that teachers report low-income students as being more absent than their more affluent peers even when they have the same number of recorded absences. Math teachers express a bias in favor of males. I find mixed evidence with regard to student race and ethnicity. Keywords: educational inequality, non-cognitive skills

*Assistant Professor, Department of Economics, UMass Amherst, 1020 Thompson Hall, Amherst, MA 01003, dfrancis@econs.umass.edu, +1(413)545-6354.

1 Introduction

Subjective perceptions that teachers form about students' non-cognitive skills 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 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 perceptions of non-cognitive skills 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

[1] Track placement refers to the practices 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.

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.

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:

$$TeacherReport_{ij} = ActualBehavior_i + TeacherPrior_{ij} + ActualBehavior * TeacherPrior_{ij} + \varepsilon_{ij}$$

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 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 demographic subgroup to that individual student as a less time-consuming means of processing information on a large number of students (Arrow, 1971; Phelps, 1972).

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 et al. (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 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.

[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.

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. 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 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 - 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

[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.

Table 1: Sample Distribution of Race, Ethnicity, Gender and Socioeconomic Status

	ECLS-K (%)	NCERDC (%)	
		Cohort 1	Cohort 2
Gender			
Male	52.2	50.1	49.9
Female	47.8	49.9	50.1
Race/Ethnicity			
American Indian	2.2	1.5	1.5
Asian	3.0	2.1	2.3
Black	17.2	27.7	27.0
Hispanic	18.5	8.6	9.4
White	57.0	57.1	56.4
Multi-Racial	2.0	3.0	3.3
Socioeconomic Status			
Below Poverty Level	20.6	–	–
FRL Eligible	34.4	43.3	43.9
Sample Size	9730	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 10 students for whom no race/ethnicity information was given in the ECLS-K Sample

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.

3.1 Sample Demographics

Descriptors of the demographic compositions of both the ECLS-K and NCERDC samples are presented in Table 1. Separate statistics are presented for the each cohort of NCERDC students. The ECLS-K sample consists of 9,730 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%. 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 17% 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 over 3% to a 1.5%. The North Carolina sample has a larger population of low socioeconomic status (SES) students as proxied by free or reduced lunch (FRL) eligibility.[5] About 34% of students from the ECLS-K sample are eligible for free or reduced price lunch while over 43% of the NCERDC student are eligible.

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) Eligibility Status. *FRL* is an indicator of whether the student is eligible for the federal subsidized lunch program.

[4] 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.

[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.

Table 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.

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 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 teachers 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.

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 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 variables to reflect the possible dichotomous nature of the perceptions question. *More Absent* takes a value of 1 if the

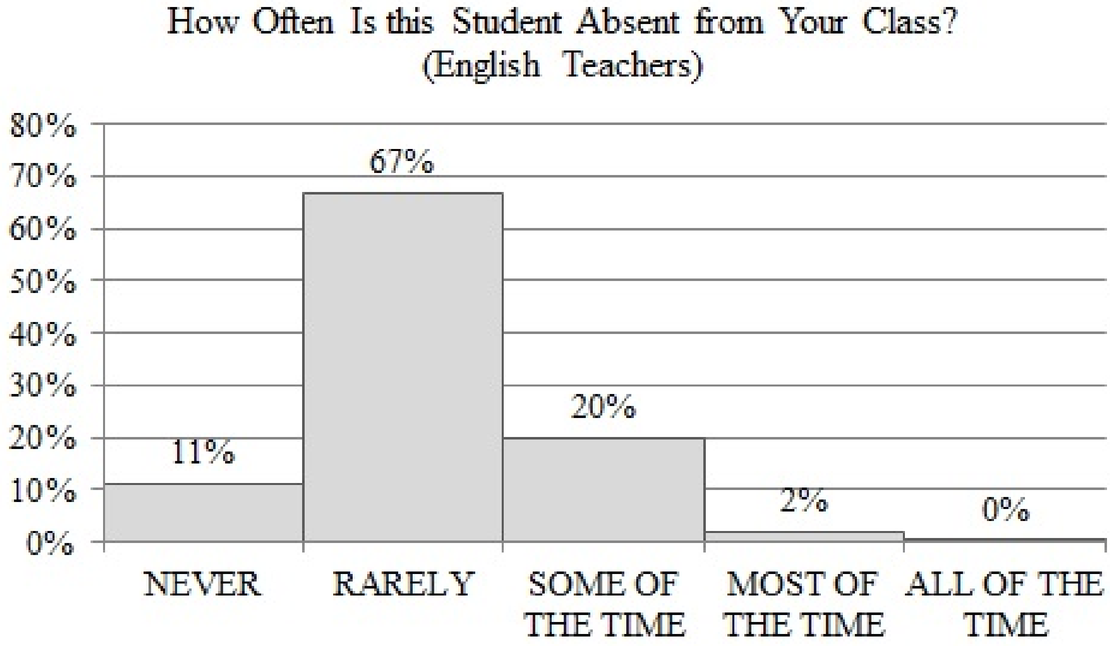

Figure 1: English Teacher Perceptions of Student Absenteeism

student is reported to be absent some of the time, most of 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 2, 3, and 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, FRL eligible students are slightly more likely to be reported as having absences than students who are not FRL eligible.

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. The mean values for absences are relatively stable across

[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.

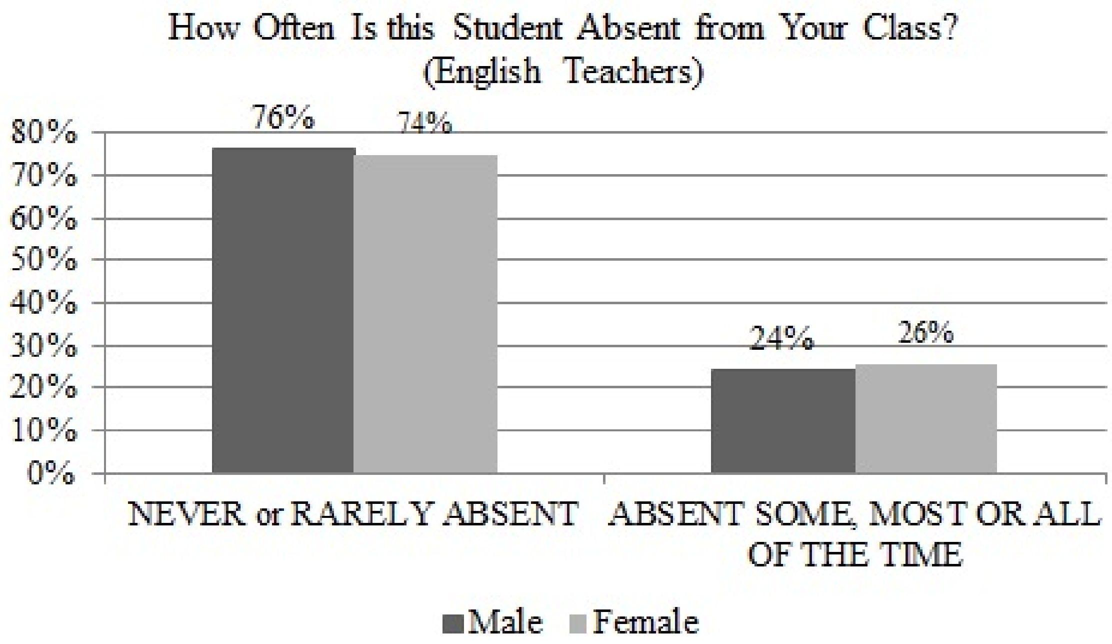

Figure 2: English Teacher Perceptions of Student Absenteeism by Gender

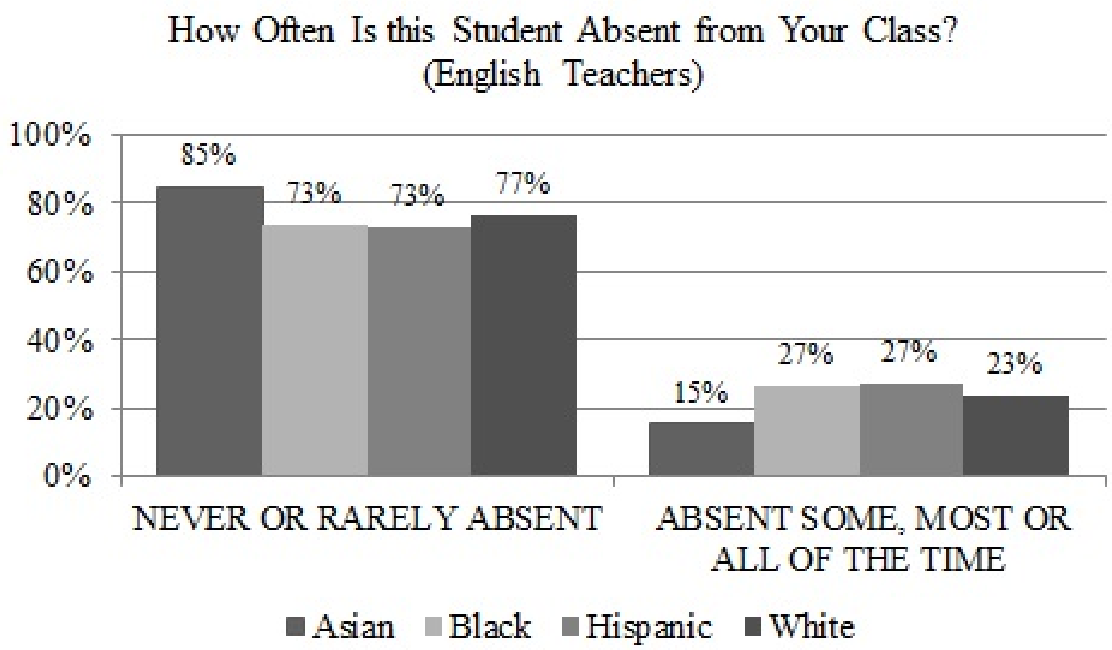

Figure 3: English Teacher Perceptions of Student Absenteeism by Race and Ethnicity

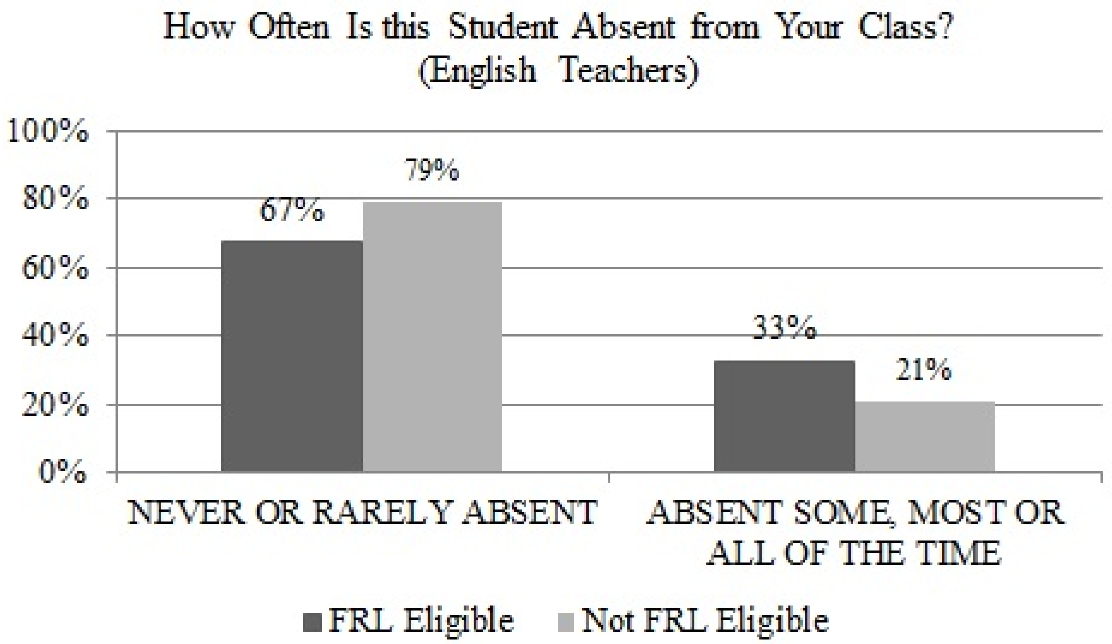

Figure 4: English Teacher Perceptions of Student Absenteeism by FRL Eligibility

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 5 depicts mean values and 95% confidence intervals for recorded absences in the ECLS-K sample by student demographic characteristics. Free or reduced lunch eligible students, on average, have significantly higher absences than students who are not FRL eligible. Male and female students display no significant differences in the average number of absences. While the differences do not reach significance, Asian 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 6 for Cohorts 1 and 2. Again, FRL eligible students have significantly higher absences than non-FRL eligible 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

Table 3: Descriptive Statistics for Recorded Attendance - ECLS-K Sample

	Obs	Mean	Std. Err	Min	Max
ECLS-K Sample					
3rd Grade Absences	8470	6.31	0.15	0	200
5th Grade Absences	8450	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.
Absences are measured across the entire school year.

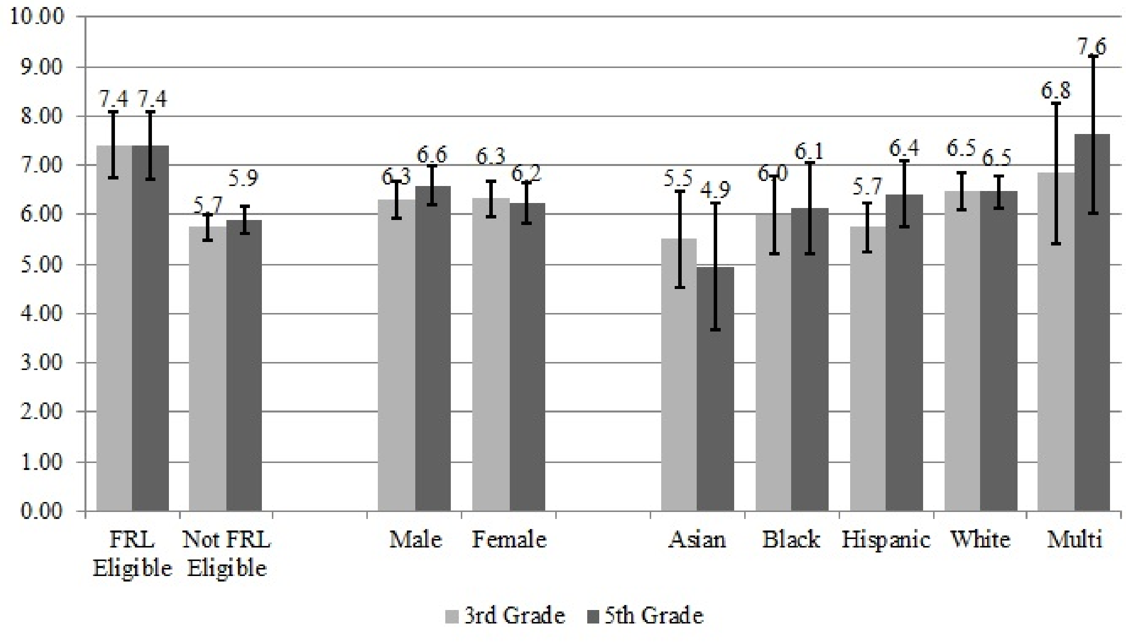

Figure 5: Mean Values and Confidence Intervals for Recorded School-Year Absences by Demographic Characteristics - ECLS-K Sample

on average than all other students, while multi-racial students have slightly higher absences.

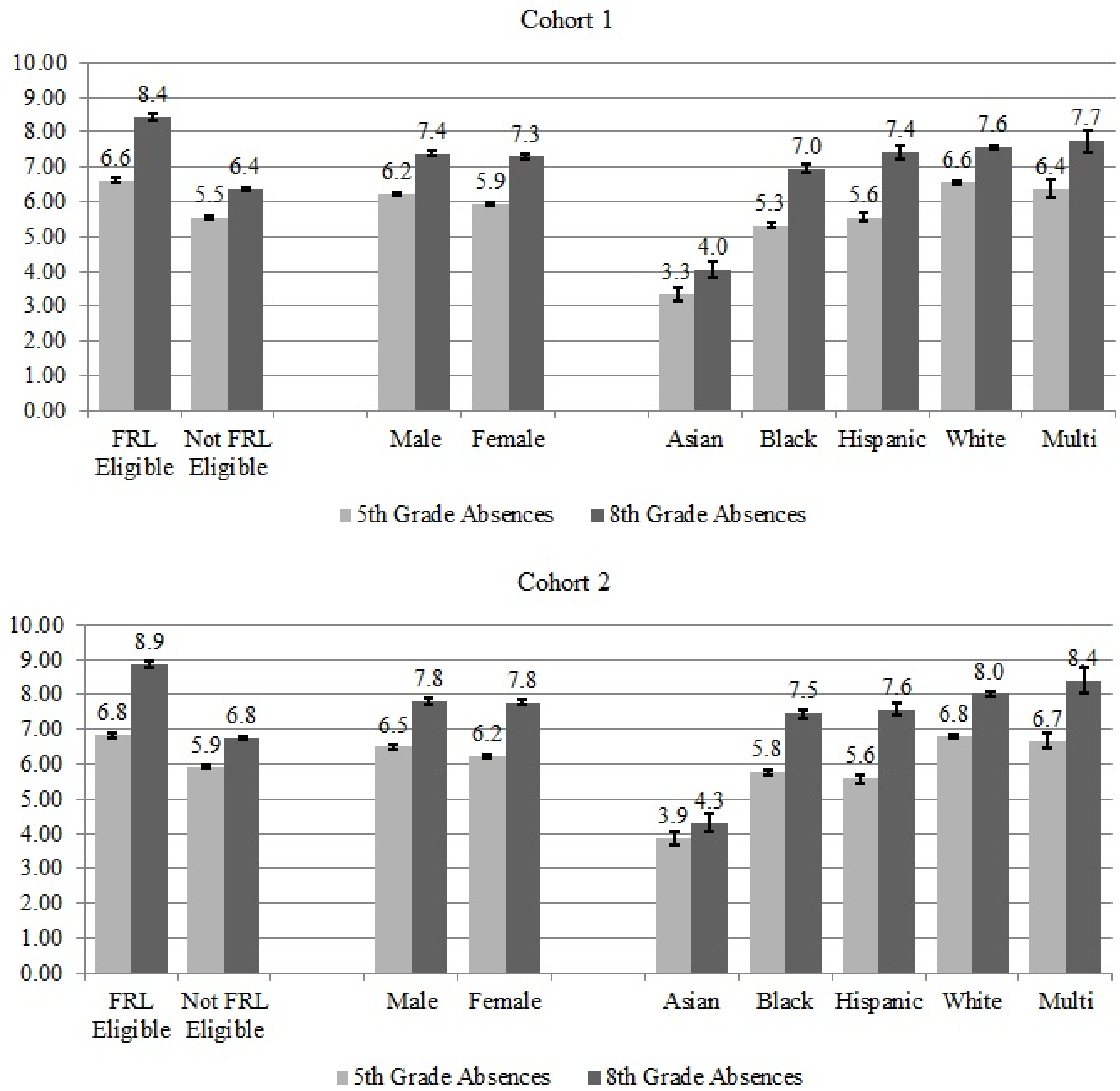

Figure 6: Mean Values and Confidence Intervals for Recorded School-Year Absences by Demographic Characteristics - NCERDC Sample.

The descriptive statistics seem to indicate that FRL eligible students 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 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:

$$pAbsence^{*}_{ij} = \gamma_0 + \gamma_1 AbsenceRecord_i + \Gamma_2 X_i + \varepsilon_i \quad (1)$$

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 *X* is a vector of the student specific demographic variables - *American Indian*, *Asian*, *Black*, *Hispanic*, *Multi-Racial*, *Male*, and *FRL* - described previously 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 (X_i^{NC}) as follows:

$$absences8_i^{NC} = \beta_0 + \beta_1 absences5_i^{NC} + \Phi X_i^{NC} + \varepsilon_i \quad (2)$$

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 (X_i^{EC}):[7]

$$absences5_i^{EC} = \beta_0 + \beta_1 absences3_i^{EC} + \Phi X_i^{EC} + \varepsilon_i \quad (3)$$

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

$$\widehat{absences}8_i = \widehat{\beta_0^{NC}} + \widehat{\beta_1^{NC}} absences5_i^{EC} + \widehat{\Phi^{NC}} X^{EC} \quad (4)$$

Similarly, for the alternative strategy:

$$\widehat{absences}8_i = \widehat{\beta_0^{EC}} + \widehat{\beta_1^{EC}} absences5_i^{EC} + \widehat{\Phi^{EC}} X^{EC} \quad (5)$$

[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.

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

$$pAbsence^*_{ij} \quad = \quad \gamma_0 + \gamma_1 \widehat{absences}8_i + \Gamma_2 X_i + \varepsilon_i \tag{6}$$

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, 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.

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 student demographic characteristics 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 student demographic characteristics. 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 student demographic characteristics, 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.

4.1 First Stage

The first stage results are presented in Table 4. The first column presents estimates derived from estimating equation (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 (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. 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. (Stock et al., 2002) for determining the strength of an instrument.

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

Table 5: Marginal Effects from Probit Estimation of MoreAbsent on Simulated Absences and Student Demographics - English Teachers

	Raw	ECLS-K	NCERDC
Simulated Absences	–	0.024***	0.019***
		(0.002)	(0.002)
Black	-0.049**	0.008	0.018
	(0.024)	(0.028)	(0.034)
Hispanic	0.003	0.019	0.022
	(0.023)	(0.023)	(0.021)
Asian	-0.112***	-0.032	-0.035
	(0.028)	(0.034)	(0.038)
American Indian	0.092	0.085	0.055
	(0.066)	(0.073)	(0.061)
Multi-Racial	0.042	0.064	0.125**
	(0.057)	(0.052)	(0.058)
Male	-0.018	-0.023	0.011
	(0.015)	(0.016)	(0.018)
FRL	0.136***	0.090***	0.066**
	(0.027)	(0.026)	(0.026)
Observations	8,130	8,130	8,130

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

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 that the relationship between prior and future absences is marginally decreasing, very slightly.

4.2 Absences

Marginal effects derived from probit estimations relating teacher reports of student absences to student demographic characteristics are presented in Tables 5 - 7 for English, math and science teachers respectively. The marginal effects are calculated at the mean value of the continuous variable *AbsentHat*, 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 11 to 14% less likely to be reported as being absent than the reference group - white students. The results for FRL eligible students are also consistent across all three subjects. Students who are FRL eligible are about 14 to 18% more likely to be reported as absent than non-FRL eligible students. Math and science teachers tend to perceive male students more positively than female students. Among math teachers, being male, relative to female, decreases the likelihood of being reported as absent by about 6% while the corresponding likelihood for science teachers is about 4%. Science teachers also tend to report that the attendance of Hispanic students is more favorable than that of white students on average, while English teachers tend to report that the attendance of black 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-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 more 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 eligible students remain highly significant even after controlling for the instrumented measures of recorded absences, though they decrease slightly

Table 6: Marginal Effects from Probit Estimation of MoreAbsent on Simulated Absences and Student Demographics - Math Teachers

	Raw	**ECLS-K**	**NCERDC**
Simulated Absences	–	0.029***	0.014***
		(0.003)	(0.003)
Black	-0.035	0.058	-0.039
	(0.044)	(0.054)	(0.046)
Hispanic	0.033	0.054	0.058
	(0.029)	(0.035)	(0.037)
Asian	-0.113***	-0.006	-0.026
	(0.038)	(0.048)	(0.050)
American Indian	-0.026	-0.012	0.030
	(0.073)	(0.100)	(0.064)
Multi-Racial	-0.026	-0.069	0.027
	(0.064)	(0.057)	(0.073)
Male	-0.056***	-0.046**	-0.032
	(0.019)	(0.019)	(0.020)
FRL	0.138***	0.097***	0.110***
	(0.030)	(0.031)	(0.029)
Observations	8,130	8,130	8,130

Two-way demographic interaction terms are also included in all models

Standard errors in parentheses

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

in magnitude. FRL eligible students are anywhere from 7 to 14% more likely than students who are not FRL eligible to be reported as being absent by their teachers. This indicates that, while FRL eligible students have more absences on average than students who are not FRL eligible, the difference is not great enough to explain away the less favorable teacher reports of FRL eligible students' absenteeism. In other words, there is evidence of bias against students who are FRL eligible. This bias appears to be strongest among science teachers.

Table 7: Marginal Effects from Probit Estimation of MoreAbsent on Simulated Absences and Student Demographics - Science Teachers

	Raw	**ECLS-K**	**NCERDC**
Simulated Absences	–	0.020***	0.020***
		(0.003)	(0.004)
Black	-0.056	0.009	0.002
	(0.047)	(0.058)	(0.049)
Hispanic	-0.064**	-0.052*	-0.043
	(0.030)	(0.031)	(0.032)
Asian	-0.141***	-0.070*	-0.089**
	(0.033)	(0.036)	(0.036)
American Indian	-0.087	0.001	-0.026
	(0.085)	(0.081)	(0.099)
Multi-Racial	0.035	0.151*	0.095
	(0.119)	(0.091)	(0.112)
Male	-0.039*	-0.016	-0.012
	(0.023)	(0.023)	(0.023)
FRL	0.177***	0.133***	0.135***
	(0.036)	(0.035)	(0.039)
Observations	8,130	8,130	8,130

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

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 5% less likely to be reported as having absences than female students by math teachers in the ECLS-K sample. The difference in the NCERDC sample is about 3%, but it is not statistically significant.

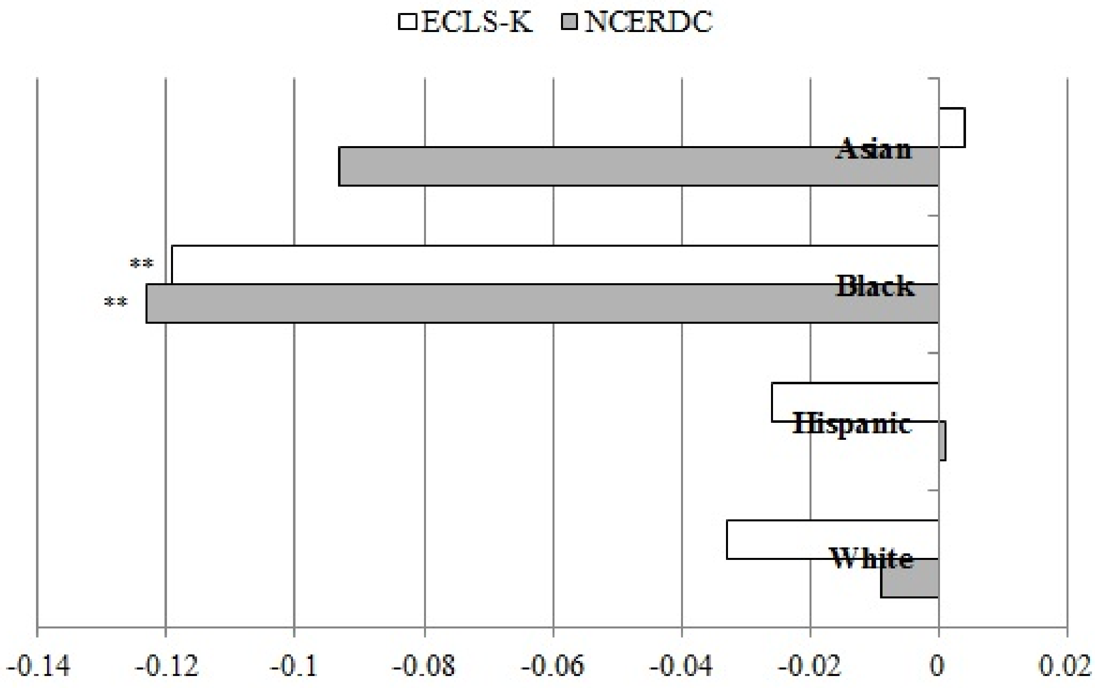

Figure 7: Marginal Effects of Probit Estimation of Math Teacher Reports of Male Absence Relative to Female Absence, by Race (**p<0.05)

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 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 A in the Appendix. The largest, and most statistically significant results are for black males relative to black females. Black males are roughly 12% less likely to be viewed as absent by math teachers than black females in both the ECLS-K and NCERDC samples. No other results in the figure are statistically different from zero.

Marginal effects from probit estimation of teacher reports of FRL eligible students relative to non-eligible students are presented in Figure 8 by race. The data underlying the figure are presented in Table B in the appendix. 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

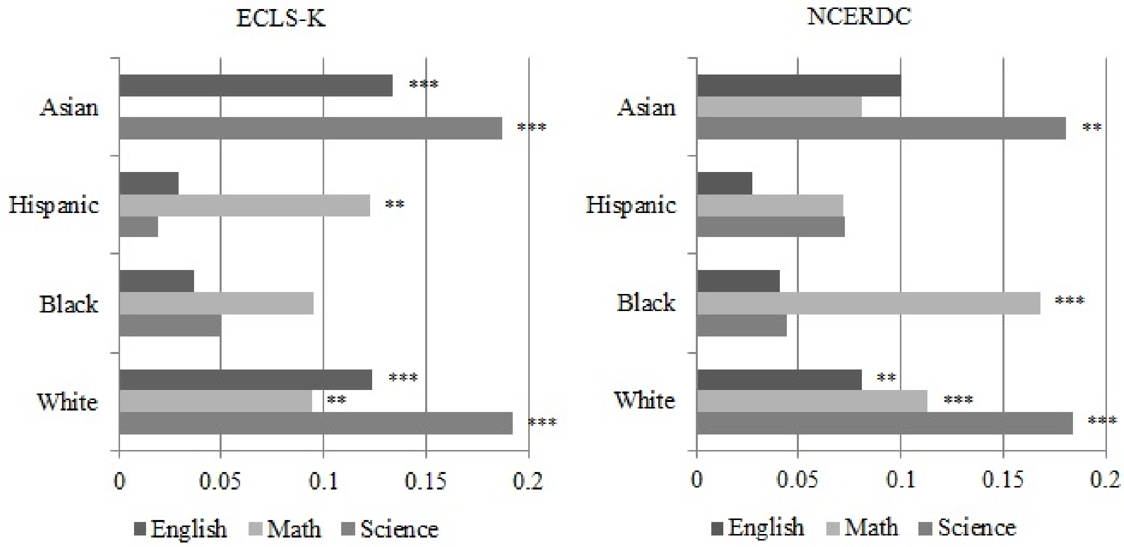

Figure 8: Marginal Effects of Probit Estimation of Teacher Reports of FRL Eligible Student Absence Relative to Non-Eligible Student Absence, by Race (*** p<0.01, ** p<0.05)

simulated absences. They are anywhere from 8% to 19% 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 point estimates ranging from 13% to 19%. There are no clear patterns when looking at low-income black and Hispanic students. Based on the ECLS-K estimates, math teachers view low-income Hispanic students as being significantly more absent than more affluent Hispanic students with an increased likelihood of about 12%. The comparable point estimate based on the NCERDC estimate is 7%, but is not significantly different from zero. Conversely, low-income black students are about 17% more likely than more affluent black students to be reported as being absent by math teachers, based on the NCERDC estimates, but only 9% more likely based on the ECLS-K estimates, which are not significantly different from zero.

4.3 Teacher Subgroups

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 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.

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

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 9, 10, and 11 - for English teachers, math teachers and science teachers respectively. In the figures, each row represents a separate estimation of equation (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 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 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, FRL eligible students are consistently reported as being more absent than students who are not FRL eligible 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 eligible students reach statistical significance are some of the categories with the largest populations of teachers

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

- female teachers, white teachers, fully certified teachers, teachers with higher than a bachelor's degree, and teachers whose parents have a high school degree.

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

Among math teachers (Figure 10), male students are reported to have less absences than female students across the board, again significantly so among female 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 eligible students are still reported as having more absences across the board, with perhaps the exception of black teachers.

With respect to science teachers (Figure 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, FRL eligible students are viewed as having more absence than non FRL eligible 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.

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 male 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.

Finally, 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.

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 FRL 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 might be suggestive evidence that math teachers report that low-income black and Hispanic students are more absent than their more affluent counterparts, even after controlling for simulated absence, 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 are rated more favorably with respect to their behavior. The results in this paper suggest that it may not be bias that produces those results, but rather that students may simply behave more favorably when pared 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.

Acknowledgements

I would like to thank Elizabeth O. Ananat, Charles T. Clotfelter, William A. Darity, Jr., Anna Gassman-Pines, and Jacob L. Vigdor for helpful comments on earlier drafts of this paper.

References

(2010). *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) edition.

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.

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(5):521–549.

Condron, D. J. (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems*, 54(1):139–160.

Darity Jr., W. and Jolla, A. (2009). Desegregated schools with segregated education. In Hartman, C. and Squires, G., editors, *The Integration Debate: Futures for American Cities*, pages 99–117. Routledge, New York.

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(4):290–307.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2, Papers and Proceedings of the One Hundred Seventeenth Annual Meeting of the American Economic Association, Philadelphia, PA, January 7-9, 2005):158–165.

Downey, D. B. and Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, 77(4):16.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education*, 54(3):151–162.

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(3):547–561.

Feng, L. (2009). Opportunity wages, classroom characteristics, and teacher mobility. *Southern Economic Journal*, 75(4):1165–1190.

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.

Goldwater, O. D. and Nutt, R. L. (1999). Teachers' and students' work-culture variables associated with positive school outcome. *Adolescence*, 34(136):653–653.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). Why public schools lose teachers. *The Journal of Human Resources*, XXXIX(2):326–354.

Harris, A. (2006). I (don't) hate school: Revisiting oppositional culture theory of blacks' resistance to schooling. *Social Forces*, 85(2):797–833.

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(4):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):1–14.

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(2):131–155.

Lleras, C. and Rangel, C. (2009). Ability grouping practices in elementary school and african american/hispanic achievement. *American Journal of Education*, 115(February 2009):279 – 304.

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

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*.

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

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(4):659–661.

Pope, D. and Sydnor, J. (2010). A new perspective on stereotypical gender differences in test scores. *Journal of Economic Perspectives*, 24:95–108.

Scafidi, B., Sjoquist, D. L., and Stinebrickner, T. R. (2007). Race, poverty and teacher mobility. *Economics of Education Review*, 26:145–159.

Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology*, 69(5):797–811.

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(4):518–529.

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(8):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(3):201–204.

Van Ewijk, R. (2011). Same work, lower grade? student ethnicity and teachers subjective assessments. *Economics of Education Review*, 30(5):1045–1058.

Appendix

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

	ECLS-K	NCERDC
White Male	-0.033	-0.009
	(0.025)	(0.025)
Black Male	-0.119**	-0.123**
	(0.059)	(0.058)
Hispanic Male	-0.026	0.001
	(0.058)	(0.057)
Asian Male	0.004	-0.093
	(0.072)	(0.078)

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

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

	ECLS-K			NCERDC		
	English	Math	Science	English	Math	Science
White	0.124***	0.094**	0.192***	0.081**	0.113***	0.184***
	(0.042)	(0.041)	(0.050)	(0.039)	(0.042)	(0.054)
Black	0.037	0.095	0.050	0.041	0.168***	0.044
	(0.039)	(0.066)	(0.085)	(0.047)	(0.052)	(0.068)
Hispanic	0.029	0.123**	0.019	0.027	0.072	0.073
	(0.040)	(0.060)	(0.048)	(0.028)	(0.066)	(0.052)
Asian	0.134***	-0.005	0.187***	0.100	0.081	0.181**

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