“Model Minorities” in the Classroom? Positive Evaluation Bias towards Asian Students and its Consequences*

Ying Shi
Syracuse University†

Maria Zhu
Syracuse University‡

Abstract

The fast-growing demographic group of Asian Americans is often perceived as a “model minority.” This paper establishes empirical evidence of this stereotype in the education context and then analyzes its consequences. We show that teachers rate Asian students’ academic performance more favorably than observationally similar White students. This contrasts with teachers’ lower likelihood of favoring Black and Hispanic students, even after accounting for performance and behavior. Notably, the presence of any Asian student in the classroom exacerbates Black-White and Hispanic-White assessment gaps. This suggests that the “model minority” stereotype can negatively impact other minority groups despite its ostensibly positive connotation.

JEL: I24, J15

Keywords: Teacher evaluation, racial bias, Asian Americans

---

*We are grateful to Peter Arcidiacono, Elizabeth Cascio, Katherine Harris-Lagoudakis, Soumaya Keynes, Jessica Merkle, Emily Owens, Emily Weisburst, Arezou Zaresani, and conference participants at AEFP, APPAM, and SEA for helpful comments. Declaration of interest: none.

†Corresponding author. Shi: Department of Public Administration and International Affairs, 426 Eggers Hall, Syracuse, NY 13244-1020 (Email: yshi78@syr.edu; Phone: 315-443-9442)

‡Zhu: Department of Economics, 110 Eggers Hall, Syracuse, NY 13244-1020 (Email: mzhu33@syr.edu; Phone: 315-443-9043)

1 Introduction

Asian Americans currently represent the single fastest growing racial and ethnic group in the United States (Budiman 2020). They experience a unique profile of racial stereotypes compared to other minority groups in the country. Since the mid-1900s, Asian Americans have been lauded as the nation's "model minority," due to perceived success in assimilation, upward mobility, and educational achievement (Wu 2014). The view of Asians as "model minorities" is pervasive in education given their ability to outperform other racial and ethnic groups on standardized tests and grades (Fejgin, 1995; Hsin & Xie, 2014; Kao, 1995) and record of postsecondary enrollment and attainment in selective institutions (Sakamoto, Goyette, & Kim, 2009).

While this "positive" stereotype is ostensibly beneficial, there is concern that it could carry negative consequences. For example, it may hold individuals in the stereotyped groups to unrealistically high expectations (Ho, Driscoll, and Loosbrock 1998), hinder their performance (Cheryan and Bodenhausen 2016), or constrain stereotyped group members in their pursuit of certain academic and career tracks (Czopp 2010). There also may be negative effects if positive stereotypes for Asians reinforce the notion of fundamental differences across groups or bolster negative stereotypes for other, under-represented minority groups (Kay, Day, Zanna, and Nussbaum 2013).

This study provides evidence on the presence and consequences of positive bias towards Asian students in schools. We first show that teachers rate Asian students higher than same-class White peers with the same performance, before exploring how the magnitude of this assessment differential varies across Asian ethnic subgroups. Finally, this paper analyzes whether the propensity of teachers to favor Asian students has spillover effects, by examining how the presence of an Asian student in the classroom affects teachers' assessments of students from other, under-represented minority groups.

To address our research questions, we use administrative data from the North Carolina Education Research Data Center (NCERDC) covering students in grades 3-8. The NCERDC dataset has two key advantages that make it uniquely well-suited for this study. First, the data include the universe of public-school students in North Carolina over our study period (2007-2013), which provides a significant number of Asian students for meaningful analyses. Second, the data contain two different measures of a student's underlying academic mastery, standardized test scores and teacher assessments, which we use to identify teacher bias.

Both teacher assessments and standardized test scores in math and reading are mapped onto a discrete scale from 1 to 4, which allows us to directly compare these two measures of achievement. Standardized test scores provide a benchmark for assessing whether teachers are systematically over-rating or under-rating Asian students relative to other groups, conditional on student achievement and a rich vector of individual sociodemographic and behavioral attributes. In addition to these controls, our analyses also include classroom-level fixed effects to address any endogeneity in teacher evaluations that could arise at the teacher, year, school, subject, and/or grade level.

Results indicate that teachers display significant positive bias towards Asian students, relative to White students in the same class with the same standardized test scores and sociodemographic characteristics. Compared to White students, teachers are 3.9 percentage points more likely to give Asian students a higher evaluation (over-rate) than the blind-scored achievement level indicated by their standardized test scores and 2.5 percentage points less likely to give Asian students a lower evaluation (under-rate). These magnitudes correspond to 11% and 14% of baseline propensities to over-rate and under-rate students, respectively, indicating that teachers' propensities for favoring Asian students are sizable. We perform several robustness checks to rule out alternative explanations for these racial differences, including accounting for the role of measurement error, hard-to-observe behavioral attributes, the comparability of blind vs. non-blind achievement scales across classes, and racial biases in standardized testing. These effects are sizable and present in math and reading and in both elementary and middle schools, suggesting positive bias towards Asian students is pervasive across subjects and grade levels. Additionally, we find heterogeneous effects by more fine-grained ethnic subgroups. Teachers display greater positive bias towards Asian students from East and South Asian backgrounds, relative to students from Southeast Asian backgrounds.

Next, our findings suggest that there are potential negative spillover effects of exposure to Asian students. Specifically, the presence of an Asian student in the classroom decreases the propensity for a teacher to over-rate a Black or Hispanic student relative to a White student with the same test scores, compared to classrooms without any Asian students. We similarly find a significant increase in the propensity for teachers to under-rate Black students when an Asian student is present in the classroom. We provide evidence that these results are not being driven by Asian students increasing average classroom achievement. These findings support the notion that the presence of Asian students and the positive stereotypes associated with these students may amplify negative biases towards other under-represented minority groups.

This paper makes several contributions to existing research. First, it provides empirical evidence on a fast-growing and understudied demographic group, Asian Americans. Despite the rapid growth of Asian Americans as a share of the population, scholarship on their educational and labor market trajectories is still limited in disciplines such as economics and sociology (Altonji & Blank, 1999; Sakamoto et al., 2009). In economics, studies utilizing different datasets, methods, and timelines show Asian Americans attaining varying degrees of earnings parity with their non-Hispanic White counterparts (Black, Haviland, Sanders, & Taylor, 2008; Chiswick, 1983; Duleep & Sanders, 1992; Mar, 2005; Weinberger, 1998). Increasingly, the evidence points to discrimination as a source of downward pressure on Asian American wages and salaries (Duleep & Sanders, 1992; Hilger, 2017; Mar, 2005).[1] Despite the evidence on labor market discrimination, there is

[1]Duleep and Sanders (1992) find that on average, American-born Asian men in the 1980 Census earn the same as their White counterparts, but the relative wages of these Asian

less documentation of potential differential treatment of Asian Americans during the schooling.[2] We show racial differences in teacher assessments that *favor* Asians relative to White students, in a manner that sets Asian students apart from other, under-represented minority groups. This lends empirical credence to the existence of positive stereotypes.

Notably, the patterns for Asians belie substantial heterogeneity, with diminished positive bias towards Asians from particular ethnic groups (e.g., individuals from Southeast Asian backgrounds) and Asians in urban settings. These findings underscore the need to shift away from a view of Asian Americans as a monolithic group towards one that accommodates a diversity of Asian demographic characteristics and experiences (Chiswick, 1983; Duleep & Sanders, 1992; Lee & Zhou, 2015; Sakamoto et al., 2009; Xie & Goyette, 2004).[3]

In addition to documenting the magnitude of Asian-White teacher rating gaps, we examine how they interact with teacher ratings of other racial and ethnic groups. Potentially detrimental consequences of positive bias include the inclination to believe that the targeted group is fundamentally different from other groups and an increase in the usage of *negative* stereotypes (Kay et al., 2013). Our findings that Black-White and Hispanic-White assessment gaps are exacerbated by exposure to an Asian student in the same classroom illustrate that positive bias towards Asians can have spillover ef-

men fall after conditioning on occupation and industry in a manner that is consistent with some discrimination against these highly-educated employees. Asian American men are also less likely to be in managerial positions, a finding on the so-called "glass ceiling" that is echoed by Mar (2005). Hilger (2017) shows that the upward mobility of Asian Americans is driven primarily by earning gains conditional on education that reflects declining discrimination in the latter half of the twentieth century. Duleep and Sanders (2012) provides evidence that the Civil Rights Act led to a decline in anti-Asian discrimination that contributed to these labor market shifts. Note that given the wide range of data, methods, and models, some studies do not find evidence of discrimination or glass ceilings (see, for example, Sakamoto, Woo, and Yap (2006)).

[2]More recently, Arcidiacono, Kinsler, and Ransom (2020) have focused on discriminatory behaviors that Asian students face relative to White counterparts in the college admissions process.

[3]Proponents of the demographic heterogeneity approach argue for a disaggregation of Asian Americans into more nuanced categories due to differences in access to resources that may shape labor market trajectories (Sakamoto et al., 2009; Xie & Goyette, 2004).

fects on other minority groups. These findings are also consistent with a theoretical conception of stereotypes rooted in representativeness (Kahneman and Tversky 1972, Bordalo, Coffman, Gennaioli, and Shleifer 2016), or the frequency in which a type occurs in a group relative to baseline. If Asian students are perceived as high-achievers under the "model minority" stereotype, their presence may emphasize academic performance and increase the application of negative stereotypes toward other, under-represented minority groups.

Finally, this paper contributes to a growing body of research on the role of teacher expectations as an input into education production. A burgeoning literature shows how teacher expectations can vary by student attributes such as race (Burgess & Greaves, 2013; Lavy, 2008; Ouazad, 2014; Rangel & Shi, 2020) and gender (Lavy, 2008; Lindahl, 2016).[4] While papers increasingly document discrepancies in teacher expectations across select racial and ethnic groups, there is still scarce research investigating bias towards Asians.[5] Teacher expectations matter because they affect student grades and the steering of students towards academic tracks such as gifted and talented programs (Donovan and Cross 2002, Lindahl 2016, Card and Giuliano 2016). Students may also adjust their behaviors and academic trajectories in ways that render teacher expectations as self-fulfilling prophecies (Rosenthal and Jacobson 1968, Ouazad and Page 2013, Jussim and Harber 2016, Lavy and Sand 2018, Lavy and Megalokonomou 2019, Papageorge, Gershenson, and Kang 2020, Hill and Jones 2021). The consequences of teacher expectations endure through postsecondary education in some instances (Papageorge et al. 2020) but are less persistently documented in others (Hill and Jones 2021).[6]

[4]The interaction between teacher and student attributes matters, as congruence in race, gender, or immigration status can manifest in more favorable teacher assessments (Lindahl, 2016; Ouazad, 2014).

[5]An exception is Burgess and Greaves (2013), which juxtaposes teacher assessments in the English testing system across Asian subgroups such as Indian, Chinese, Bangladeshi, and Pakistani.

[6]Hill and Jones (2021) and Papageorge et al. (2020) use different contexts and identification strategies to examine the impact of differential teacher assessments. The former uses

In the remainder of the paper, Section 2 presents our data and provides an in-depth overview of the blind and non-blind evaluation measures used in the paper. Section 3 discusses the empirical strategy used to identify differences in teacher evaluations across student race. Section 4 presents our results and Section 5 concludes.

2 Data and Descriptive Statistics

2.1 North Carolina Education Data

This study uses statewide administrative records from the North Carolina Education Research Data Center (NCERDC). Student-level data contain sociodemographic information on sex, race and ethnicity, and eligibility for free or reduced lunch. The NCERDC also reports individuals' primary home language, which we use as a proxy to inform more detailed information on students' ethnicities and countries of origin.

Similarly, we observe teacher-level attributes including race, ethnicity, and age. Longitudinal data on when a teacher was first observed in a North Carolina traditional public or charter school allow us to determine teachers' years of experience. Detailed course membership rosters with unique student and teacher IDs enable the linking of student sociodemographic data with teacher records and course attendance. We focus on students in grades 3-8 from 2007-2013, which is the sample for which we observe course membership and assessment information.

An important feature of the data is the presence of both blind-scored assessments and non-blind teacher evaluations of student performance along the same scale. Students take End-of-Grade (EOG) standardized tests in

an instrumental variables strategy with a rich set of fixed effects for elementary and middle school students, while the latter relies on within-student variation in tenth-grade teacher expectations. Hill and Jones (2021) find that teacher evaluations matter for student performance, particularly for earlier grades, although these effects do not persist. Papageorge et al. (2020) document more persistent causal effects through college completion. The mixed evidence on the enduring effects of teacher expectations on student outcomes is consistent with reviews of the literature in social psychology and beyond (Jussim & Harber, 2005).

math and reading from third through eighth grade. These tests are given during the last three weeks of the school year, with questions formulated in a multiple-choice format. Raw student scores on EOG tests are mapped to achievement levels on a discrete scale of 1 to 4 denoting score cutoffs relative to grade-level comparisons. Levels 1 to 4 refer to insufficient mastery, inconsistent mastery, consistent mastery, and superior performance, respectively.[7] We refer to standardized test assessments of math and reading ability as "blind" assessments, since EOG tests are machine-scored, without regard to a student's identity.

Teacher evaluations map to the same four-point scale of achievement levels for each subject. We refer to teacher assessments of students as "non-blind" assessments since teachers inevitably need to know the identity of the student in question in order to evaluate the student. With knowledge of a student's identity comes information about and the race and ethnicity of each student. We examine whether this information influences how teachers perceive a student's skill-based achievement level.

2.2 Teacher Evaluations

Teacher evaluations of student skills in math and reading come from End-of-Grade data files. Concurrently with the state administration of EOG exams, teachers are asked to provide their assessment of each student's skill mastery on the four-point achievement level scale corresponding to insuffi-

[7] A detailed description of each achievement level is as follows:

1. Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.
2. Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.
3. Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.
4. Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work

cient, inconsistent, consistent, or superior mastery. Given the timing, teachers submit evaluations *before* they observe students' end-of-year standardized test results.

There is one stated reason for asking teachers for these evaluations. The state uses the average of these teacher judgments to calibrate cut points in the aforementioned four-point scale. Teacher assessments are only one input, as the state also takes into consideration expert input and standard-setting processes. These assessments are not used for any other purpose, such as teachers' performance evaluations. This implies that teachers lack incentives to misrepresent their assessments of student performance.

In order to interpret racial differences in teacher assessments for students in a given classroom with comparable EOG performance as evidence of bias, we need to establish that teacher ratings aim to measure the same underlying skills as EOG tests. We advance several reasons for a close correspondence in content between these two types of assessments. First, the questionnaire instructions for student evaluations explicitly ask teachers to focus their evaluation on the tested subject. As such, the sequence of signals the teacher receives about a student's science competence should not be an input into their assessment of math mastery, or vice versa. Second, teachers were asked to evaluate students' "absolute" ability. This means that teachers are not judging student performance relative to peers in the same classroom or school, but rather to a common statewide standard that is external to the test. The four-point achievement scales used in teacher and EOG assessments align closely with the North Carolina Standard Course of Study, which defines the curriculum standards for each grade and subject to ensure uniformity across classrooms statewide. Teachers undergo training on standard-based grading to minimize subjectivity, thereby enhancing familiarity with state-defined standard objectives. They furthermore have access to the descriptions of skills associated with each achievement level when they evaluate students.

Accountability pressures also induce teachers to spend greater time preparing their students for standardized exams. To the extent that teachers use practice EOG tests or similar materials, students' aptitude on these assessments likely serve as inputs into both teacher evaluations and the actual EOG test, thereby strengthening their relation to each other. Finally, teachers are explicitly instructed to assess students based on achievement, rather than behavior.[8] This further strengthens the relationship between teacher evaluations and achievement-based EOG scores by minimizing the extent to which teachers consider behavioral or socioemotional factors.

2.3 Descriptive Statistics

Table 1 describes the sample of students. Approximately 3% of students are Asian, while the majority of students (54%) are White. One advantage of the NCERDC data is that even though Asians constitute a relatively small proportion of the overall student body, there are still over 40,000 Asian students in our sample to allow for sufficient statistical power. Black and Hispanic students make up 27% and 12% of the sample, respectively. In our main analysis, we use an indicator for economic disadvantage and the number of days absent in a year as a proxy for behavioral differences that may emerge in the classroom. On average, half of the students in this sample are economically disadvantaged, and students were absent for about 7 days in a given school year.

The relatively small share of Asians in the North Carolina administrative data prompts questions on their distribution, in particular whether they are concentrated in specific classrooms. Figure 1 shows that apart from the 73% of classrooms with no Asian students, the modal case in 17% of classrooms is one Asian student.

Table 2 details the characteristics of teachers in the sample. Relative to students, teachers are disproportionately White (82% of the sample). Most

[8] The prompt given to teachers reads: *"The [subject] teacher should base this response for each student solely on mastery of [subject]. The [subject] teacher may elect to use grades as a starting point in making these assignments. However, grades are often influenced by factors other than pure achievement, such as failure to turn in homework. The [subject] teacher's challenge is to provide information that reflects only the achievement of each student in the subject matter tested."*

Table 1: Student Characteristics

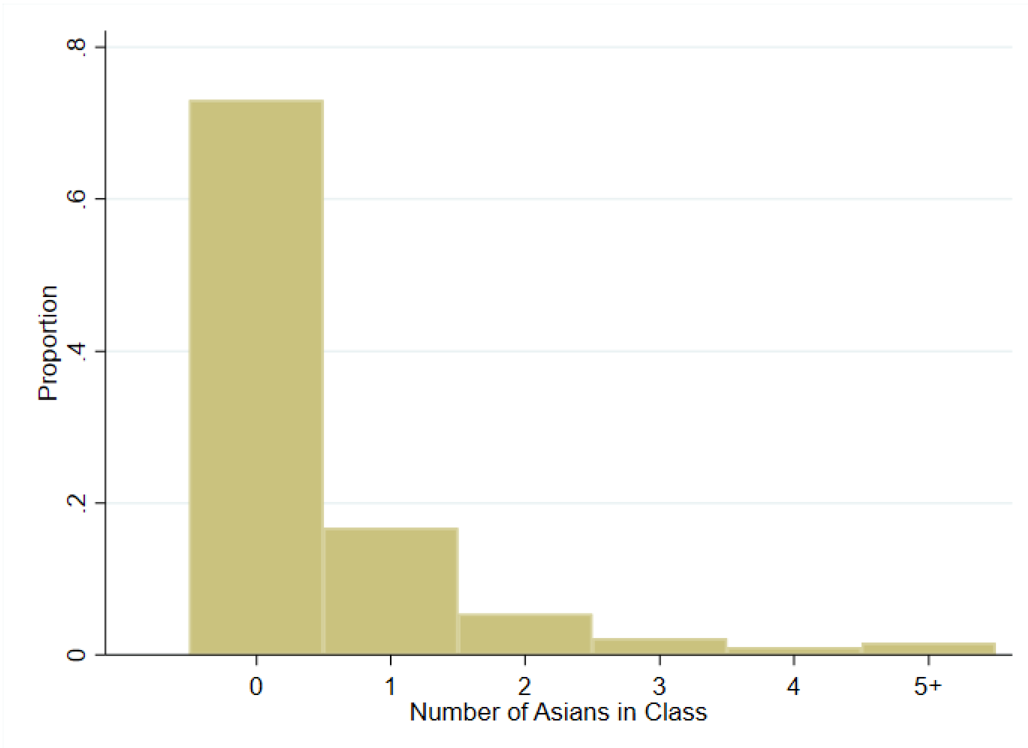
	Mean
White	0.54
Black	0.27
Hispanic	0.12
Asian	0.03
American Indian	0.01
Other race	0.04
Female	0.49
Economically Disadvantaged	0.50
Days Absent	7.04
	(6.32)
N	1,410,653

Observations are at the student level for students in grades 3-8 in math or reading classes between 2007-2013. A student's number of days absent and status as economically disadvantaged are calculated as the average value of that variable for each year they appear in the data.

of the remaining teachers are Black, and Asians comprise only 1% of the teacher sample. Nearly nine out of every ten teachers are female, a proportion in keeping with national statistics of the elementary and middle school teaching workforce that skews heavily towards women. On average, teachers in our sample period have 10.4 years of experience.

To give a sense for how the academic achievement of Asians compares to other students, Table 3 shows the mean and distribution of blind-scored achievement levels by race. The mean blind-scored achievement level in math and reading for students in the sample is 2.80. Overall, 22% of students rank in the top achievement category, level 4. Another 47% of students score at level 3, which represents the plurality of students. Compared to both White and under-represented minority students, Asian students have significantly higher average achievement levels and are disproportionately represented in the higher achievement categories. Their average achievement level is 3.12, while the corresponding measures are 3.03, 2.42,

Figure 1: Asian Representation across Classes



Observations are at the classroom level. The histogram shows the distribution of number of Asians in a classroom.

and 2.52 for White, Black, and Hispanic students, respectively. The difference in achievement scores between White and Asian students is concentrated at the top of the distribution. In our sample, 40% of Asian students have an achievement level of 4, compared to only 31% of White students.

Table 4 compares the propensity for teachers to under-rate or over-rate Asian students, compared to their propensity to do so for White students. Rows denote a student's blind-scored achievement level based on standardized test performance, and columns represent the teacher's non-blind achievement rating for the student. Cells denote the proportion of students at each teacher-rated level, conditional on a given blind-scored achievement level. Dark (light) shaded areas denote cells for which a teacher over-rates (under-rates) a student relative to their blind-scored achievement levels.

Table 2: Teacher Characteristics

	Mean
White teacher	0.82
Black teacher	0.15
Hispanic teacher	0.01
Asian teacher	0.01
Other teacher race	0.01
Female teacher	0.88
Teacher experience (years)	10.39
	(9.67)
N	50,215

Observations are at the teacher level for teachers teaching grades 3-8 in math or reading classes between 2007-2013. Teacher experience is calculated as the average number of years of experience over the period the teacher appears in the data.

Table 3: Blind-scored Achievement Levels by Race

Blind-scored Achievement	All	White	Asian	Black	Hispanic
Mean	2.80	3.03	3.12	2.42	2.52
Level 4	0.22	0.31	0.40	0.09	0.12
Level 3	0.47	0.48	0.39	0.44	0.45
Level 2	0.20	0.14	0.13	0.29	0.26
Level 1	0.11	0.06	0.08	0.19	0.17
N	16,004,741	8,639,535	389,432	4,185,749	1,893,326

Observations represent blind-graded, standardized test scores in math and reading for students from 2007-2013. Two-sample t-test results indicate the mean blind-scored achievement of Asians is significantly larger from that of each of the other racial groups at a 99% confidence level.

Values in Table 4 indicate teachers may be more likely to over-rate Asians and less likely to under-rate Asians relative to White peers. These patterns are especially stark for high-achieving students as measured by blind-scored achievement levels. For example, while 26% of White students who have a blind achievement level score of 3 are rated at an achievement level

Table 4: Blind vs. Non-Blind Scores

Blind-scored Achievement	Teacher rating: White students Level 1	Level 2	Level 3	Level 4	Asian students Level 1	Level 2	Level 3	Level 4
Level 1	0.22	0.45	0.31	0.02	0.22	0.42	0.32	0.04
Level 2	0.08	0.34	0.50	0.07	0.07	0.29	0.51	0.13
Level 3	0.02	0.15	0.58	0.26	0.02	0.12	0.50	0.36
Level 4	0.00	0.03	0.32	0.65	0.00	0.02	0.24	0.74

Table aggregates math and reading evaluations. Cells represent the share of students who got a blind-score in the row value that were evaluated by their teachers at the column value. Dark (light) shaded areas denote cells for which a teacher over-rate (under-rate) a student relative to their blind-scored achievement levels

of 4 by their teachers, this proportion is 36% for Asian students. Conversely, while 35% of White students who have a blind achievement score of 4 are given a lower rating by their teachers, this proportion is only 26% for Asian students. Overall, teachers are 8 percentage points more likely to over-rate an Asian relative to a White student, relative to a baseline probability of over-rating among White students of 34%. Teachers are 4 percentage points less likely to under-rate Asian students, relative to the rate of under-rating among White students of 19%. Two-sample t-tests reveal that the probability of a teacher to over-rate or under-rate an Asian student differs significantly from their propensity to do so for a White student at a 99% confidence level.[9]

While Table 4 provides suggestive evidence that teachers may exhibit positive bias towards Asian students relative to White students, these numbers should not be interpreted as causal because they do not control for any underlying differences between White and Asian students themselves or differences in factors affecting their assignment to particular schools, teachers, and classes that may affect assessment scores. The next section discusses in detail potential endogeneity concerns of causal interpretations of these correlations and presents the empirical strategy used to identify the

[9]We exclude students with a blind score of 4 in the measurement of over-rating and students who score of 1 in the measurement of under-rating since these students mechanically cannot be over-rated or under-rated.

presence of teacher biases in student evaluation.

3 Empirical Strategy

Cross-tabulations of subjective teacher assessments and blind-scored standardized test outcomes are unlikely to reflect teacher bias without adjusting for precise student ability, behavior, and conditions governing the assignment of students into classrooms. Our main specification accounts for these factors by estimating the following linear probability model:

$$O_{ic} = \mathbf{R}'_{ic}\beta + \alpha f(E_{ic}) + \mathbf{X}'_{ic}\Omega + \eta_c + \epsilon_{ic} \tag{1}$$

where O_{ic} represents the outcome of interest for student i in class c. We look at two different outcomes: whether the teacher's non-blind assessment level is *higher* or *lower* than the student's blind-scored assessment level based on standardized test performance. Given blind ($B \in \{1, 2, 3, 4\}$) and non-blind ($NB \in \{1, 2, 3, 4\}$) student assessments, O_{ic} denotes $\mathbb{1}\{NB > B\}$ and $\mathbb{1}\{NB < B\}$, respectively. Students who score a 4 are not included in the over-rating sample since it is mechanically infeasible to over-rate these students. Analogously, students who score a 1 are excluded from the under-rating sample.

This regression framework addresses multiple potential confounding factors in order to isolate racial differences in assessment attributed to teacher bias (as captured by the coefficient on student race indicators $\mathbf{R}'_{ic}$). First, Equation 1 flexibly controls for a student's end-of-grade exam score, E_{ic}, using subject-, year-, and grade-specific raw test score fixed effects. These controls account for possible differences in the distribution of student achievement scores by race. The inclusion of end-of-grade exam score indicators also addresses the concern that assessment categories are fairly coarse, since students are placed in one of four test-score bins, so it may be that student distributions within bins vary by race.[10] In this scenario, differences in

[10]For example, suppose White students who get categorized in achievement level 4 in

teacher assessments relative to achievement bins may reflect actual differences in achievement, rather than underlying teacher racial biases.

The vector $\mathbf{X}'_{ic}$ controls for a set of observable characteristics, including student gender, number of days absent during the year, and whether the student is economically disadvantaged. These variables address the possibility that different student racial groups consist of different compositions along these characteristics, which may subsequently affect teacher assessments. In particular, if there are unobserved behavioral components that affect assessment, this may be captured by number of days a student is absent during the year.

Finally, the addition of a class fixed effect, η_c, means identification comes from *within-classroom* variation in teacher assessments. The fixed effect accounts for the possibility that Asian students are disproportionately concentrated in classrooms with more- or less-lenient teachers relative to White counterparts. It also accounts for any classroom-specific shocks that may affect learning, as well as changes across testing standards over time.

To determine how teachers' propensities to over-rate or under-rate differ across student racial and ethnic groups, we examine the coefficient of interest β on the vector of student race and ethnicity indicators ($\mathbf{R}_{ic}$), using White students as the reference category. β captures racial differences in teachers' subjective evaluations within a given class, after adjusting for students' performance on blind-scored standardized tests and behavioral proxies. We interpret this differential as teacher racial bias in assessments.

Next, we augment our empirical specification to test for spillover effects of exposure to any Asian students in the classroom. As before, the outcome variable O_{ic} denotes whether the teacher is over-rating ($\mathbb{1}\{NB > B\}$) or under-rating ($\mathbb{1}\{NB < B\}$) student i in classroom c:

$$O_{ic} = \mathbf{R}'_{ic}\pi + (\mathbf{R}'_{ic}\text{AnyAsian}_c)\Phi + \mathbf{R}'_{ic}\delta_j + \rho f(E_{ic}) + \mathbf{X}'_{ic}\Gamma + \theta_c + \epsilon_{ic} \quad (2)$$

blind test scores tend to have raw test scores that are right at the cutoff between bins or achievement level 3 and 4, while Asian students categorized in achievement level 4 have raw test scores well above the cutoff.

The above model follows Equation 1 in flexibly controlling for the student's blind-scored test performance using subject-, year-, and grade-specific score fixed effects, alongside individual attributes such as the number of days absent, economic disadvantage, and gender. The use of θ_c absorbs classroom-level shocks such as shared disruptions to learning and teacher preferences for grading that are common to all students.

This specification departs from the base model in the inclusion of an interaction term between student race and whether there is at least one Asian student in the classroom (AnyAsian_c). Since it is highly plausible that classroom racial composition relates to school and teacher characteristics due to the sorting of students into classrooms, we also include a full set of student race indicators interacted with teacher-school-grade-course fixed effects (δ_j). These absorb fixed differences in the likelihood of having at least one Asian student across teachers in a given school and course type (e.g. fifth grade math). The residual variation in AnyAsian_c is then within teacher and course. We infer a causal interpretation of the parameters of interest (Φ) as the effect of exposure to any Asian student on racial differences in teacher assessments, with a focus on Black-White and Hispanic-White gaps.

Our empirical strategy assumes idiosyncratic variation in exposure to at least one Asian student for a teacher in a given school, grade, and course. We advance that this is a plausible assumption given natural population variation in the presence of students of a particular racial or ethnic group. We also restrict the analytic sample to only classrooms with zero or one Asian student so that results are not identified off of classrooms with larger concentrations of Asian students. To further assess the validity of our assumption, we examine the relationship between having one Asian student and class characteristics using classroom-level data. Conditional on teacher-school-grade-course fixed effects, classroom attributes such as the Black-White or Hispanic-White achievement gap do not predict whether an Asian student is present (Appendix Table A1).

4 Results

4.1 Racial Differences in Teacher Assessments

Table 5 shows racial differences in teacher evaluations after adjusting for raw standardized test scores, individual characteristics, and class fixed effects. The outcome variable in the first column is an indicator for whether a teacher over-rates a student relative to their blind-scored achievement level, while the second outcome variable is an indicator for whether a teacher under-rates a student. Students who score a 4 are not included in the over-rating sample since it is mechanically infeasible to over-rate these students. Analogously, students who score a 1 are not included in the under-rating sample. The omitted racial group is White students.

Results indicate teachers are 3.9 percentage points more likely to over-rate Asian students relative to White students in the same class with the same standardized test scores and individual characteristics. The magnitude is sizable, considering the effect is equivalent to nearly 11 percent of the baseline propensity of being over-rated. We document comparable magnitudes when examining the phenomenon of under-rating. Teachers are 2.5 percentage points less likely to under-rate Asian students relative to White student counterparts who are observationally similar. This translates to a magnitude of 14% of the baseline propensity of being under-rated.[11]

In contrast to the favored ratings of Asian students, analogous racial differentials in teacher assessments go in the opposite direction for Black and Hispanic students, a finding that is consistent with previous literature on subjective teacher evaluations of under-represented minority students (Burgess & Greaves, 2013; Rangel & Shi, 2020). Notably, the magnitudes of teachers' increased propensity to over-rate and decreased propensity to

[11] Another way of interpreting the extent to which teachers favor Asian students is to run the same model with the four-point teacher evaluation scale as the dependent variable rather than over- or under-rating indicators. We find that teachers confer Asian students a level of achievement that is 0.06 higher than same-scoring White peers in the same classroom. This is a sizable difference given that Asian students' mastery as measured via EOG achievement levels is only 0.09 higher than White peers.

Table 5: Racial Differentials in Teacher Assessments

	Over-rate ($NB > B$)	Under-rate ($NB < B$)
Asian	0.039***	-0.025***
	(0.002)	(0.001)
Black	-0.027***	0.024***
	(0.001)	(0.001)
Hispanic	-0.022***	0.021***
	(0.001)	(0.001)
American Indian	-0.022***	0.011***
	(0.003)	(0.002)
Other race	-0.006***	0.006***
	(0.001)	(0.001)
N	12,383,439	14,147,869

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

under-rate Asians are at least as large, if not more so, than the extent of decreased over-rating and increased under-rating for Black and Hispanic students.

Next, we disaggregate our results by math and reading classes to determine if racial differentials are more pronounced in a particular subject. Appendix Table A2 shows that teachers are also more likely to favor Asian students relative to observationally comparable White peers in both math and reading. Coefficients for reading are larger in magnitude. This may reflect the relatively more subjective nature of reading or English language arts instruction, which leaves more room for interpretation relative to the problem-based nature of mathematics. Separate analyses by grade level find similar patterns of teacher assessments in both elementary and middle schools (Appendix Table A3). Overall, these findings indicate that teachers' positive bias towards Asian students is pervasive across grades and subjects.[12]

Robustness Checks

We undertake a number of analyses to address concerns that our results are consistent with alternative explanations. Specifically, we examine the roles of measurement error, differences in assessment standards across classes, unobserved behavioral characteristics, and racial biases in standardized testing that may potentially influence our results.

First, we address the issue that test scores may measure underlying ability with error and that findings on the Asian-White assessment gap may be partially attributable to this measurement error. This could be the case under the assumptions of racial differences in underlying skill distributions and uncorrelated errors (Hanushek & Rivkin, 2009). In this situation, White students who are observed as high-achieving will be more likely than observationally similar Asian students to be actually low-achieving. If teacher ratings reflect students' true achievement, teachers may be less likely to classify White students as high-achieving than Asian students, even in the absence of bias.

We examine the robustness of our results to measurement error concerns using an instrumental variables approach in Table B1. The first column replicates our main findings on racial differentials, while the second column shows that Asian-White gaps in teacher assessments are robust to including standardized test scores as a linear control interacted with the subject and

[12]One aspect of these differential racial assessments is teacher optimism, in which teachers inflate student assessments relative to objective measures of classroom performance. While teachers have been shown to be optimistic across student groups (Papageorge et al., 2020), we demonstrate here that teachers' optimism is more pronounced towards Asian relative to White students.

grade level. Column 3 instruments for test scores using contemporaneous scores from the other subject, such as using same-year math scores to instrument for reading achievement. Under this specification, teachers are even more likely to favor Asians in over-rating, while teachers are only slightly less likely to under-rate Asians compared to the OLS specification that does not correct for measurement error. One drawback of this first instrument is that it potentially suffers from an overly restrictive assumption of uncorrelated errors across contemporaneous subjects. For example, student illness and learning disruptions common to both subjects in a given year can contribute to correlated errors. Given these concerns, we next instrument using lagged achievement scores for the same subject. This specification enables accounting for measurement error under the assumption of uncorrelated errors over time and is not subject to concerns about contemporaneous shocks raised above. Results in column 4 of Table B1 show coefficients that are very similar to the first instrument. Finally, we instrument for contemporaneous test scores using lagged other subject test scores in column 5, which is perhaps even more likely to satisfy the assumption of uncorrelated errors. Once more, the likelihood that teachers over-rate Asian students relative to White students does not attenuate when taking measurement error into account. The Asian-White gap in under-rating is also robust across instrumental variables specifications. Taken together, the evidence suggests that measurement error in standardized testing does not explain our main findings.

Second, we address the concern that comparisons of blind and non-blind scores may be capturing differences in assessment standards across classes. Teachers' standards of mastery may vary depending on the particular school or classroom context, and this could generate racial gaps in teacher assessments in the presence of non-random sorting of students by race across schools and/or classrooms within schools. For instance, teachers with high-performing students may have higher standards for what constitutes a proficient student, independent of state guidelines. If this were the case, students in high-performing classes will be less likely to be over-rated than students in lower-performing classrooms with the same underlying ability, as measured by raw End-of-Grade test scores. While the inclusion of classroom fixed effects in our analyses control for differences across classroom in the outcome variables, which capture a measure of the difference between blind and non-blind scores, they do not address classroom-level differences in baseline scores. To ensure that we are not mistaking these influences for teacher bias, we construct adjusted distributions of blind test-based achievement levels within each class to match the distribution of teacher rating levels (on the same four-point scale). Specifically, using raw EOG test scores, we place the same number of the class's students into each blind-scored achievement level as observed in the corresponding teacher rating scale and re-run our analysis on teacher over-rating and under-rating by race. Results in Table B2 shows that when we modify the outcomes of teacher over- and under-rating relative to these adjusted EOG achievement levels, the estimated coefficients for Asian students are very similar to the unadjusted coefficients. This strongly suggests that what we interpret to be teacher bias is not confounded by the comparability of blind vs. non-blind achievement scales.

Third, we consider the possibility that systematic differences in teacher assessments of Asian and White students with the same standardized test score arise due to differences in unobserved behavioral characteristics, rather than teacher bias. As mentioned in Section 2, teachers are explicitly instructed to assess students solely on their mastery of the subject matter tested. Nevertheless, it is possible that students' behavioral traits inadvertently influence teachers' assessment of mastery. The main results in Table 5 include an indicator variable for the number of days a student is absent during the year as a proxy for behavior. We augment our main specification with additional behavioral controls in Table B4 in the Appendix. First, we include controls for the lagged number of days absent to address the possibility that a student's contemporaneous absences are endogenous to teachers' subjective assessments. Next, we estimate our main specification on the sample of students without any prior disciplinary infractions on their record.[13] Reassuringly, including lagged days absent and restricting the sample to students without prior disciplinary infractions do not significantly change results, providing further support that our findings are not being driven by underlying behavioral differences across racial groups.[14]

Finally, we explore the possibility that our findings are driven by racial biases in standardized testing, rather than in teacher biases in evaluations. Theoretically, observed racial patterns in over-rating and under-rating are consistent with a scenario of standardized tests displaying negative cultural/racial bias towards Asian students in the absence of any teacher bias. If this were the case, we expect these results to be exacerbated for Asian students who do not speak English as their primary home language (relative to those who do speak English as a primary home language) for a couple of reasons. First, research indicates bilingual children may face especially large structural disadvantages with regards to standardized tests (Valdés & Figueroa, 1994). Additionally, home language can be seen as a proxy for assimilation, with the assumption that Asian students who speak English at home are less likely to suffer from cultural or Asian-specific racial biases that may be embedded in standardized tests. Our robustness check examines whether it is the case that gaps are larger for Asian students who do not speak English as their primary home language. Results in Appendix Table B5 indicate that teachers are actually *more* likely to over-rate Asian students who report English as their primary home language and *less* likely to under-rate them. These findings go in the opposite direction of the coefficients we would expect if results were being driven by racial bias in tests, providing further support that our findings reflect teacher bias. One potential concern is that this interpretation of results does not take into consideration Asian students whose families come from countries where English is widely spoken and who might have unique cultural backgrounds despite

[13] Table B3 shows the full list of disciplinary infraction types.

[14] We have also run specifications using the full sample of students and including detailed controls on the types and numbers of incidents in each category a student received during the school year. Results using this specification find that our main findings are robust to behavioral controls as well.

speaking English at home. As a further check, we also run our analysis on a subset of counties in which the Asian population is least likely to be from Asian countries where English is widely spoken and find that our results are robust to this.[15]

4.2 Heterogeneity in Teacher Assessments

Grouping Asian students into a single category potentially disguises their diverse experiences and trajectories. Existing studies examining the educational and labor market trajectories of Asians often rely on monolithic categories, even when research demonstrates substantial differences in schooling and earnings across Asian ethnic groups (Chiswick, 1983). In response, we take advantage of existing, albeit limited, data to investigate the extent to which teacher bias may vary across Asian ethnic groups. The NCERDC does not contain direct information on a student's background beyond general racial and ethnic markers (White, Asian, Black, Hispanic, etc.), so we proxy for ethnic subgroups using two complementary methods. In the preferred specification, we rely on NCERDC data reporting a student's primary home language and use that information to classify Asian students into three regional subgroups: East Asian, Southeast Asian, and South Asian.[16]

Table 6 shows the breakdown of Asian students in the sample by home language. Slightly over half of Asians in the sample report English as their primary language. Table 6 also provides descriptive statistics for Asian students by home language subgroup. Consistent with previously docu-

[15]We use detailed race information in ACS data from 2007-2013 to calculate what share of Asians in each county come from an Asian country that reports English as an official language, which includes India, Pakistan, Singapore, and the Philippines. Next, we rerun our specification of heterogeneity in teacher bias towards Asian students by English home language status using students from the subset of counties in which proportion of the Asian population that are from Asian countries where English is the official language is below the median.

[16]Table C1 in the Appendix details the languages corresponding to each category. Most languages under the East Asian group are spoken in China, Japan, and South Korea. The majority of individuals in the South Asian group speak languages prevalent in India, Pakistan, and Bangladesh. The Southeast Asian group includes languages commonly spoken in Vietnam, Cambodia, Laos, Indonesia, Thailand, Malaysia, Philippines, and Burma.

mented patterns (Sakamoto et al., 2009), East Asian and South Asian students report a higher socioeconomic status than Southeast Asian students. They also have higher average math and reading scores.

Table 6: Asian Subgroups by Home Language Status

	N	Percent	% FRL	Math scores	Reading scores
East Asian	4,153	10.70	0.22	1.10	0.46
South Asian	2,468	6.36	0.22	0.89	0.49
Southeast Asian	5,682	14.64	0.69	0.03	-0.33
Other Asian	2,299	5.93	0.67	-0.28	-0.59
Asian (English)	2,0726	53.42	0.30	0.72	0.46
Asian: Missing Language	3,471	8.95	0.45	0.53	0.21
Total/average	38,799	100.00	0.38	0.59	0.26

Observations denote unique students in grades 3-8 between 2007-2013 who identify as Asian. Classification by subgroup based on home language. For students who appear in the data for multiple years, we use the average economically disadvantaged status and average math/reading z-scores across years.

Next we analyze teacher assessments across Asian subgroups using home language as a proxy for ethnicity. Table 7 shows substantial heterogeneity in the extent of teacher assessment gaps across subgroups. Compared to their assessments of White students, teachers are 5.7 percentage points more likely to over-rate South Asian students, 4.4 percentage more likely to over-rate East Asian students, and 2.2 percentage points more likely to over-rate Southeast Asian students. A Wald test indicates the coefficients between South Asians and Southeast Asians, and East Asians and Southeast Asians are significantly different at the 1% level, suggesting systematic differences in teacher assessments towards Southeast Asian students relative to peers speaking a home language commonly associated with countries including China, Japan, South Korea, India, and Pakistan. In terms of under-rating, estimates suggest that teachers are somewhat less likely to under-rate South Asian relative to both East Asian Southeast Asian students, although coefficient estimates are not statistically different between any of the three groups.

A key advantage to using home language information to proxy for Asian ethnic subgroup is that we are able to infer detailed ethnic information at

Table 7: Differentials in Teacher Assessments by Home Language

	Over-rate $(NB > B)$	Under-rate $(NB < B)$
East Asian	0.044***	-0.013***
	(0.008)	(0.004)
South Asian	0.057***	-0.024***
	(0.006)	(0.004)
Southeast Asian	0.022***	-0.017***
	(0.003)	(0.002)
Other Asian	-0.021***	0.023***
	(0.008)	(0.007)
Asian: English	0.053***	-0.032***
	(0.003)	(0.001)
Asian: Missing Language	0.039***	-0.032***
	(0.005)	(0.003)
N	12,383,439	14,147,869

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. Other minority races are included in regression, although they are not displayed in the table. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

the individual level. However, a drawback of this approach is that a large portion of the sample reports English as their primary home language, and we are unable to infer detailed ethnic information for these students. We therefore analyze subgroup heterogeneity using a second approach based on Census ethnicity data. Specifically, we proxy for Asian subgroup concentration using the relative shares of East Asian, South Asian, and Southeast Asians in the county in which a school is located. This approach finds similar evidence of heterogeneity in teacher bias across Asian subgroups, with teachers being more positively biased towards South Asians and East Asians, relative to Southeast Asians. More details and results of this analysis can be found in Table C2 in Appendix C.

Next, we assess whether results differ by school location. Specifically, we examine whether the degree of positive bias towards Asians varies for teachers in an urban versus more rural setting. Table C3 in the Appendix augments our main specification with an interaction term for whether the school is based in a city (relative to a rural, town, or suburban location). Our findings reveal that teachers in cities are less positive towards Asian students: they are 1.7 percentage points less likely to over-rate Asian students and 0.9 percentage points more likely to under-rate Asian students than counterparts teaching in non-city settings. Upon closer examination, Table C4 shows that Asian students in cities have relatively lower socioeconomic and academic outcomes than White peers compared to Asian students outside of cities. This suggests that positive stereotyping towards Asian students may be lower in urban areas because Asians in these areas tend to conform less to the "model minority" stereotype, perhaps because of different compositions by ethnic subgroups.[17]

4.3 Spillover Effects on Under-Represented Minorities

Despite the positive connotation of categorizing Asian students as a "model minority," such stereotypes may have adverse intrapersonal and interpersonal consequences, for example by reinforcing the notion of fundamental differences across groups and increasing the usage of negative stereotypes

[17] In addition to examining heterogeneity by the detailed ethnic categories of Asian students and by school attributes, we conducted further analyses to examine the role of teacher characteristics. For example, teachers of a given racial and ethnic group or experience level may be more prone to classroom racial biases. We examine whether the extent of racial differentials is associated with teacher race, age, and experience and do not find any evidence that these attributes have significant bearing on teacher assessments towards Asian students. Note that due to the very small number of Asian teachers in our sample, we did not have enough statistical power to check for the role of racial congruence on our results. Such race match effects have been demonstrated in select contexts for Asian American students (see, for example: Lusher, Campbell, and Carrell (2018).

(Kay et al., 2013). Table 8 investigates how exposure to Asian students affects teachers' assessments of students from *other* minority groups, relative to White peers with similar academic and behavioral records. Identification is based on variation in exposure to a single Asian student for a teacher in a given school who instructs a particular course (e.g., 5th grade math). Our models thus control for teacher attributes that are fixed at the teacher-school-grade-course level, including time-invariant preferences in assessments toward students of different racial and ethnic groups. The within-teacher-course design addresses concerns involving non-random sorting of Asian students into classrooms on the basis of characteristics such as teacher race and course rigor.

To gauge the effect of exposure to any Asian student, we restrict the analysis to classrooms with zero or one Asian student only. Figure 1 documents that the modal case in the context of any exposure is a single Asian student, with classrooms having up to one Asian student making up 90% of the sample. Table 8 shows that the presence of any Asian student in the classroom significantly decreases a teacher's propensity to over-rate Black and Hispanic students relative to White students, relative to when no Asian students are present in the same teacher's classroom. Teachers are less likely to over-rate Black and Hispanic students by 0.4 and 0.5 percentage points, respectively. To place these magnitudes in context, this widens the Black-White and Black-Hispanic racial disparities in over-rating by approximately one-sixth to one-quarter (see Table 5). The presence of an Asian student in the same classroom increases the teacher propensity to under-rate by 0.4 percentage points among Black students. The relative change is on par with the magnitudes observed for over-rating. We do not find a significant corresponding change in under-rating for Hispanic students.

One potential concern is that rather than negative spillover effects, results from Table 8 may be driven by Asian students raising the achievement level in a class, which could subsequently affect teachers' assessment standards . Table 9 analyzes how changes in teacher evaluations of under-represented minority students vary depending on the academic achieve-

Table 8: Effect of Exposure to One Asian Student

	Over-rate ($NB > B$)	Under-rate ($NB < B$)
Black×Any Asian	-0.004**	0.004***
	(0.002)	(0.002)
Hispanic×Any Asian	-0.005**	0.002
	(0.002)	(0.002)
American Indian×Any Asian	0.001	0.005
	(0.008)	(0.008)
Other×Any Asian	-0.003	0.000
	(0.004)	(0.003)
Class FE	Y	Y
Race×teacher-school-grade-course FE	Y	Y
N	11,095,439	12,304,738

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students across all racial groups in grades 3-8 between 2007-2013 and is limited to classrooms that have either zero or one Asian student. Any Asian is a binary variable indicating that the classroom had one Asian student. The omitted category is White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

ment of the Asian student in class. In particular, we inquire whether exposure to an average-performing Asian student can exacerbate teachers' negative biases toward under-represented minorities, or if the negative spillover phenomenon found in Table 8 is driven by exposure to high-achieving Asian students. We use lagged test scores, normalized within the population of all students, as a measure for achievement to address potential endogeneity concerns with teacher expectations and Asian student performance.[18] Coefficients on the interactions between race variables and Any Asian are

[18] Sample sizes are smaller because we do not observe lagged scores for students in grade 3 and must restrict analyses to students in grades 4-8.

interpreted as differences in teacher propensities to over-rate or under-rate students in this racial group relative to White students who are also exposed to the same *average* Asian student (defined as scoring at the statewide mean for a given grade and year). The coefficients on the interactions between race variables and Asian lagged achievement scores are interpreted as the difference in teacher propensities to over-rate or under-rate students in this racial group relative to White students who are also exposed to the same Asian student, for each one-standard-deviation increase in the Asian student's achievement.

Exposure to an average-achieving Asian student decreases the propensity for teachers to over-rate both Black and Hispanic students by 0.4 percentage points relative to White students in the same class (Table 9). As such, even the presence of a mediocre-performing Asian student exacerbates existing inequalities in teacher assessments for these minority groups. This effect is reinforced when exposure is to higher-performing Asian students, with a one standard deviation increase in the Asian student's achievement decreasing the propensity for teachers to over-rate Black and Hispanic students by 0.8 and 0.6 percentage points, relative to White students. Effects are more muted overall when looking at teacher under-rating. Exposure to an average Asian student increases teachers' propensity to under-rate Black students by 0.4 percentage points. In contrast, there is no evidence of exposure to Asian students having a significant change in the propensity for teachers to under-rate Hispanic students.

To ensure that the consequences of exposure to Asian students is distinct from that of other minority groups, we examine whether the presence of a single Black or Hispanic student leads to similar spillover effects on teacher assessments. Tables D1 and D2 in the Appendix restrict the sample to classes with zero or a single Black student to show how exposure affects teachers' ratings of Hispanic students, as well as how these effects vary by the academic performance of the Black student. In contrast to the results on Asian students, the presence of a Black student induces no measurable changes in teachers' assessment behavior towards Hispanic students on av-

Table 9: Effect of Exposure to One Asian Student, by Achievement

	Over-rate $(NB > B)$	Under-rate $(NB < B)$
Black×Any Asian	-0.004**	0.004**
	(0.002)	(0.002)
Hispanic×Any Asian	-0.004*	0.003
	(0.002)	(0.002)
Black×Asian Lagged Z-score	-0.008***	0.002
	(0.002)	(0.002)
Hispanic×Asian Lagged Z-score	-0.006**	0.003
	(0.002)	(0.002)
Class FE	Y	Y
Race×teacher-school-grade-course FE	Y	Y
N	9,179,023	10,189,401

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students across all racial groups in grades 4-8 between 2007-2013 and is limited to classrooms that have either zero or one Asian student. Any Asian is a binary variable indicating that the classroom had one Asian student. The omitted category is White students. Any Asian is a binary variable, while Lagged Z-score is the Asian student's standardized lagged z-score normalized within the population of all students in a give grade and year. Models include interactions between American Indian students, and students of other racial or ethnic groups with the Any Asian and Asian Lagged Z-score variables. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

erage. To distinguish exposure effects of an average Black student from the influence of academic performance, we estimate both level and slope effects of exposure based on lagged achievement. Table D2 shows that an average-performing Black student has no effect on the propensity of teachers to over-rate or under-rate Hispanic students, a result that is distinct from the racial disparity-exacerbating effects of exposure to an average-performing Asian student. Similarly, there are no significant changes in spillover effects along the achievement gradient for the Black student. We supplement these findings by examining exposure to a single Hispanic student in Tables D3 and D4. There is no evidence that teachers change their assessment behavior towards Black students as a result of having an average-performing Hispanic student. Table D4 shows that a one standard deviation increase in achievement level of the Hispanic student in the classroom decreases the propensity for teachers to over-rate Black students by 0.6 percentage points, but there is no significant effect on the under-rating of Black students. Taken together, these results suggest that the spillover consequences of Asian students are unique. Exposure to an average-performing Asian student is sufficient for worsening existing disparities in Black-White teacher assessments, while the same does not hold for exposure to other, under-represented, minority groups.

5 Conclusion

Limited research exists on Asian Americans, despite their increasing prominence in K-12 education and status as the single fastest growing racial demographic group in the United States. This study provides evidence for the treatment of Asian Americans as "model minorities" in elementary and secondary schools. We show that teachers, when tasked with assessing student mastery in a subject, rate Asian students more favorably relative to White students in the same class with the same standardized test scores. The assessment advantages conferred upon Asian students are persistent across grade levels and subjects and are robust to accounting for factors such as measurement error and behavioral differences. Crucially, teacher assessment patterns that set Asians apart from other groups of under-represented minorities can have lasting consequences given the influence of teacher expectations on students' own behaviors and longer-term academic trajectories (Botelho, Madeira, & Rangel, 2015; Card & Giuliano, 2016; Hill & Jones, 2021; Lindahl, 2016; Papageorge et al., 2020).

We investigate potential consequences of this so-called positive bias by examining the extent to which teacher assessments of Asian students might interact with their judgment of students belonging to other minority groups. Our finding that exposure to an Asian student widens both Black-White and Hispanic-White assessment gaps indicates the negative consequences of positive bias towards Asian students. The presence of Asian students amplifies differences in teacher judgment of minority groups vis-a-vis White students, thereby magnifying existing racial differences. These findings recall small-scale studies demonstrating that positive stereotypes reinforce beliefs in the biological underpinnings of group differences and the application of negative stereotypes (Kay et al., 2013) and suggest the potential for negative spillover effects of biases with an ostensibly positive connotation. To the extent that stereotypes are based on representative generalizations that are exaggerated to provide the greatest differentiation in a given context (Bordalo et al., 2016; Kahneman & Tversky, 1972), stereotypical judgment for Black and Hispanic students may be most salient when faced with a high-performing Asian student.

Taken together, our results underscore the existence and potential pitfalls of positive biases. Future work can explore the long-term consequences of positive biases for Asian students themselves, building on previous research that establish substantial intrapersonal and interpersonal costs of receiving positive stereotypes.[19] Despite theory and evidence from mostly lab settings that positively stereotyped group members may change their academic expectations and orientation towards particular academic or career tracks (Czopp, 2010; Ho et al., 1998), little research links these short-term changes in expectations and behaviors to long-run academic outcomes. A related topic that merits additional research is the extent of differential responses among individuals who conform in varying degrees to positive stereotypes of the larger group; namely, shifting away from a monolithic conception of Asian students to distinguish between the academic responses

[19]Previous studies have shown that the targets of such biases are more likely to experience psychological distress and depersonalization and are less likely to seek help from others (e.g. Gupta, Szymanski, and Leong (2011)).

of Asian subgroups.

References

Altonji, J. G., & Blank, R. M. (1999). *Race and gender in the labor market* (Handbook of Labor Economics). Elsevier.

Arcidiacono, P., Kinsler, J., & Ransom, T. (2020). Asian American Discrimination in Harvard Admission. *Working Paper*.

Black, D. A., Haviland, A. M., Sanders, S. G., & Taylor, L. J. (2008, July). Gender Wage Disparities among the Highly Educated. *Journal of Human Resources*, *43*(3), 630–659.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016, November). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753–1794.

Botelho, F., Madeira, R. A., & Rangel, M. A. (2015). Racial Discrimination in Grading: Evidence from Brazil. *American Economic Journal: Applied Economics*, *7*(4), 37–52.

Budiman, A. (2020). *Asian Americans are the fastest-growing racial or ethnic group in the U.S. electorate* [Pew Research Center].

Burgess, S., & Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics*, *31*(3), 535–576.

Card, D., & Giuliano, L. (2016). Can Tracking Raise the Test Scores of High-Ability Minority Students? *American Economic Review*, *106*(10), 2783–2816.

Cheryan, S., & Bodenhausen, G. V. (2016). When Positive Stereotypes Threaten Intellectual Performance: The Psychological Hazards of "Model Minority" Status:. *Psychological Science*.

Chiswick, B. R. (1983). An Analysis of the Earnings and Employment of Asian-American Men. *Journal of Labor Economics*, *1*(2), 197–214.

Czopp, A. M. (2010). Studying is lame when he got game: racial stereotypes and the discouragement of Black student-athletes from schoolwork. *Social Psychology of Education*, *13*(4), 485–498.

Donovan, S., & Cross, C. (2002). *Minority Students in Special and Gifted Education*. National Academy Press.

Duleep, H. O., & Sanders, S. (1992). Discrimination at the Top: American-Born Asian and White Men. *Industrial Relations*, *31*(3), 416–432.

Duleep, H. O., & Sanders, S. G. (2012, June). *The Economic Status of Asian Americans Before and After the Civil Rights Act* (SSRN Scholarly Paper No. ID 2089668). Rochester, NY: Social Science Research Network.

Fejgin, N. (1995). Factors Contributing to the Academic Excellence of American Jewish and Asian Students. *Sociology of Education*, *68*(1), 18–30.

Gupta, A., Szymanski, D. M., & Leong, F. T. L. (2011). The "model minority myth": Internalized racialism of positive stereotypes as correlates of psychological distress, and attitudes toward help-seeking. *Asian American Journal of Psychology*, *2*(2), 101–114.

Hanushek, E. A., & Rivkin, S. G. (2009, March). Harming the best: How schools affect the black-white achievement gap. *Journal of Policy Analysis and Management*, *28*(3), 366–393.

Hilger, N. (2017). Upward Mobility and Discrimination: the Case of Asian Americans. *NBER Working Paper 22748*.

Hill, A. J., & Jones, D. B. (2021, June). Self-Fulfilling Prophecies in the Classroom. *Journal of Human Capital*, 000–000.

Ho, C. P., Driscoll, D. M., & Loosbrock, D. L. (1998). Great Expectations: The Negative Consequences of Falling Short1. *Journal of Applied Social Psychology*, *28*(19), 1743–1759.

Hsin, A., & Xie, Y. (2014). Explaining Asian Americans' academic advantage over whites. *Proceedings of the Natironal Acadmy of Sciences of the United Staets of America*, *111*(23), 8416–8421.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *9*(2), 131–155.

Jussim, L., & Harber, K. D. (2016). Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies:. *Personality and Social Psychology Review*.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.

Kao, G. (1995). Asian Americans as Model Minorities? A Look at Their Academic Performance. *American Journal of Education*, *103*(2), 121–159.

Kay, A. C., Day, M. V., Zanna, M. P., & Nussbaum, A. D. (2013). The insidious (and ironic) effects of positive stereotypes. *Journal of Experimental Social Psychology*, *49*(2), 287–291.

Lavy, V. (2008, October). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, *92*(10–11), 2083–2105.

Lavy, V., & Megalokonomou, R. (2019, June). *Persistency in Teachers' Grading Bias and Effects on Longer-Term Outcomes: University Admissions Exams and Choice of Field of Study* (Tech. Rep. No. w26021). Cambridge, MA: National Bureau of Economic Research.

Lavy, V., & Sand, E. (2018, November). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, *167*, 263–279.

Lee, J., & Zhou, M. (2015). *The Asian American achievement paradox*. New York: Russell Sage Foundation.

Lindahl, E. (2016). Are teacher assessments biased? – evidence from Sweden. *Education Economics*, *24*(2), 224–238.

Lusher, L., Campbell, D., & Carrell, S. (2018, March). TAs like me: Racial interactions between graduate teaching assistants and undergraduates. *Journal of Public Economics*, *159*, 203–224.

Mar, D. (2005). Asian Americans in the Labor Market: Public Policy Issues. *AAPI Nexus Journal*, *3*(2).

Ouazad, A. (2014, May). Assessed by a Teacher Like Me: Race and Teacher Assessments. *Education Finance and Policy*, *9*(3), 334–372.

Ouazad, A., & Page, L. (2013, September). Students' perceptions of teacher biases: Experimental economics in schools. *Journal of Public Economics*, *105*, 116–130.

Papageorge, N., Gershenson, S., & Kang, K. M. (2020). Teacher Expectations Matter. *Review of Economics and Statistics*, *102*(2), 234–251.

Rangel, M., & Shi, Y. (2020). First Impressions: The Case of Teacher Racial Bias. *Working Paper*.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the Classroom. *The Urban Review*, *3*(1), 16–20.

Sakamoto, A., Goyette, K. A., & Kim, C. H. (2009). Socioeconomic Attainments of Asian Americans. *Annual Review of Sociology*, *35*, 255–276.

Sakamoto, A., Woo, H., & Yap, K.-L. (2006). Are Native-born Asian Americans Less Likely To Be Managers? Further Evidence on the Glass-ceiling Hypothesis – AAPI Nexus Journal. *AAPI Nexus*, *4*(1), 13–37.

Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Westport, CT, US: Ablex Publishing.

Weinberger, C. J. (1998, January). Race and Gender Wage Gaps in the Market for Recent College Graduates. *Industrial Relations: A Journal of Economy and Society*, *37*(1), 67–84.

Wu, E. (2014). *The Color of Success: Asian Americans and the Origins of the Model Minority*. Princeton University Press.

Xie, Y., & Goyette, K. A. (2004). *A Demographic Portrait of Asian Americans*. Russell Sage Foundation.

APPENDIX

A Additional Tables

Table A1: Variation in Exposure to Any Asian Student

	Presence of an Asian Student	
	(1)	(2)
White-Black Math Achievement Gap	0.014***	0.005
	(0.003)	(0.003)
White-Black Reading Achievement Gap	0.004	0.003
	(0.003)	(0.003)
White-Hispanic Math Achievement Gap	0.016***	0.003
	(0.003)	(0.003)
White-Hispanic Reading Achievement Gap	-0.003	-0.001
	(0.003)	(0.003)
Teacher-school-grade-course FE	N	Y
N	318,815	309,193

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Class-level sample includes grades 4-8 from 2007-2013, and is limited to classrooms that have either zero or one Asian student. Achievement gaps are computed as the difference in the average lagged math and reading z-scores across racial or ethnic groups.

Table A2: Racial Differentials in Teacher Assessments, by Subject

	Math		Reading	
	Over-rate $(NB > B)$	Under-rate $(NB < B)$	Over-rate $(NB > B)$	Under-rate $(NB < B)$
Asian	0.025***	-0.018***	0.047***	-0.030***
	(0.002)	(0.002)	(0.002)	(0.002)
Black	-0.015***	0.014***	-0.036***	0.032***
	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.018***	0.021***	-0.025***	0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.015***	0.007***	-0.027***	0.014***
	(0.003)	(0.003)	(0.004)	(0.003)
Other	-0.001	0.002	-0.009***	0.009***
	(0.001)	(0.001)	(0.002)	(0.001)
N	5,355,635	6,449,731	7,027,795	7,698,129

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

Table A3: Racial Differentials in Teacher Assessments, by Grade Level

	Elementary		Middle	
	Over-rate $(NB > B)$	Under-rate $(NB < B)$	Over-rate $(NB > B)$	Under-rate $(NB < B)$
Asian	0.035***	-0.022***	0.043***	-0.026***
	(0.003)	(0.002)	(0.002)	(0.001)
Black	-0.026***	0.024***	-0.028***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.016***	0.019***	-0.028***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.022***	0.006*	-0.022***	0.017***
	(0.005)	(0.003)	(0.003)	(0.002)
Other	-0.002	0.005***	-0.010***	0.006***
	(0.002)	(0.002)	(0.001)	(0.001)
N	5,841,809	6,670,457	6,541,370	7,477,136

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students enrolled in an elementary school (grades 3-5) or middle school (grades 6-8) in 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

B Robustness Checks

Table B1: Role of Measurement Error

			Instrumental variables		
			Other	Lagged	Lagged Other
	OLS	OLS	Subject	Subject	Subject
	(1)	(2)	(3)	(4)	(5)
Over-rate: $(NB > B)$					
Asian	0.039***	0.041***	0.059***	0.055***	0.061***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	12,383,439	12,383,460	12,309,504	9,780,015	9,755,459
Under-rate: $(NB < B)$					
Asian	-0.025***	-0.027***	-0.022***	-0.022***	-0.021***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
N	14,147,869	14,147,869	14,096,865	11,333,174	11,308,970
Individual characteristics	Y	Y	Y	Y	Y
Test score FE × subject × grade × year	Y				
Linear test score × subject × grade		Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. Column 2 includes the interactions of subject, grade, and test scores entered linearly. Column 3 instruments for test scores using the other subject test score, measured contemporaneously (i.e., instrument math scores using reading scores). Column 4 instruments for test scores using the lagged same-subject score, while Column 5 instruments for test scores using the lagged other-subject score. The top panel on over-rating excludes students whose EOG scores are at achievement level 4. The bottom panel on under-rating excludes students whose EOG scores are at achievement level 1. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

Table B2: Comparability of Blind vs. Non-Blind Achievement Scales

	(1) Over-rate $(NB > B)$	(2) Under-rate $(NB < B)$	(3) Over-rate $(NB > B)$	(4) Under-rate $(NB < B)$
Asian	0.039***	-0.025***	0.037***	-0.025***
	(0.002)	(0.001)	(0.002)	(0.001)
Black	-0.027***	0.024***	-0.024***	0.021***
	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.022***	0.021***	-0.025***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.022***	0.011***	-0.015***	0.014***
	(0.003)	(0.002)	(0.003)	(0.002)
Other	-0.006***	0.006***	-0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Adjusted Ach. Level Distribution	N	N	Y	Y
N	12,383,439	14,147,869	11,833,212	14,964,346

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. Columns (3) and (4) use raw EOG test scores to put students into adjusted achievement levels such that the number of students per class in each level is the same as the number of students at each of the four teacher rating levels. Outcomes Columns (3) and (4) are indicator variables for whether the teacher rating level is higher or lower than the *adjusted* blind-scored achievement levels based on EOG performance. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

Table B3: Disciplinary Infractions List

Infraction	Frequency
Disruptive behavior	1,693,620
Bus misbehavior	806,673
Insubordination	643,970
Aggressive behavior	642,685
Fighting	582,034
Inappropriate language/disrespect	537,929
Disrespect of faculty/staff	435,807
Other school defined offense	253,873
Other	169,684
Bullying	132,511
Theft	119,418
Excessive tardiness	101,421
Disorderly conduct	80,255
Dress code violation	78,637
Skipping class	71,356
Late to class	62,470
Cell phone use	62,076
Communicating threats	61,960
Skipping school	60,386
Inappropriate items on school property	54,307
Assault on student	50,019
Property damage	48,119
Harassment–verbal	47,428
Harassment–sexual	39,740
Possession of a weapon (excluding firearms/explosives)	36,941
Honor code violation	31,200
Truancy	25,818
Being in an unauthorized area	22,959
Leaving school without permission	20,634
Excessive display of affection	18,708
Falsification of information	18,333
Leaving class without permission	18,169
Unlawfully setting a fire	17,469
Assault on student w/o weapon and not resulting in injury	17,290
Misuse of school technology	17,095
Gang activity	12,167
Possession of tobacco	10,437
Possession of controlled substance–marijuana	9,872
Affray	8,561
Cutting class	7,844
Immunization	7,800
Repeat Offender	7,115
Assault–other	6,356
Assault on school personnel not resulting in injury	6,057
Possession of counterfeit items	5,729
Use of tobacco	5,408
Mutual sexual contact between two students	3,562
Alcohol possession	3,082
Hazing	2,805
Possession of controlled substance–other	2,717

Table displays list of disciplinary infractions that students can be reported for, as well as the frequency with which each infraction appears in the sample. A given student may have been reported for multiple types of infractions over the course of the year, and it is also possible for a student to be reported for the same infraction multiple times over the course of the year. Note: we restrict this list to the 50 most frequently occurring infraction types in the data.

Table B4: Augmented Behavioral Controls

	Full Sample		No Infrac.	Full Sample		No Infrac.
	(1)	(2)	(3)	(4)	(5)	(6)
	Over-rate	Over-rate	Over-rate	Under-rate	Under-rate	Under-rate
	$(NB > B)$	$(NB > B)$	$(NB > B)$	$(NB < B)$	$(NB < B)$	$(NB < B)$
Asian	0.039***	0.038***	0.038***	-0.025***	-0.024***	-0.023***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Black	-0.027***	-0.029***	-0.028***	0.024***	0.026***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.022***	-0.024***	-0.024***	0.021***	0.022***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.022***	-0.022***	-0.025***	0.011***	0.012***	0.012***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Other	-0.006***	-0.006***	-0.005***	0.006***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Days Absent	Y	Y	Y	Y	Y	Y
Lagged Days Absent	N	Y	Y	N	Y	Y
N	12,383,439	11,830,085	9,594,778	14,147,869	13,539,473	11,534,820

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

Table B5: Restrict to Students who Report English Home Language

	Full Sample		Restricted Sample	
	Over-rate $(NB > B)$	Under-rate $(NB < B)$	Over-rate $(NB > B)$	Under-rate $(NB < B)$
	(1)	(2)	(3)	(4)
Asian	0.031***	-0.019***	0.026***	-0.007
	(0.003)	(0.002)	(0.005)	(0.005)
Asian×English	0.020***	-0.012***	0.021***	-0.020***
	(0.004)	(0.002)	(0.007)	(0.005)
English	0.007***	-0.005***	0.011***	-0.008***
	(0.001)	(0.001)	(0.002)	(0.002)
N	12,383,439	14,147,869	3,579,117	3,995,412

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. Other minority groups and their interactions with English home language are included in the regression, although they are not displayed in the table. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$. In the estimations for the restricted sample in columns (3) and (4), we only include the students in the subset of counties in which proportion of the Asian popuation that are from Asian countries where English is the official language is below the median, calculated using 2007-2013 ACS data.

C Heterogeneity in Teacher Assessments

Table C1 shows how NCERDC self-reported primary home languages were categorized into the ethnic subgroups of East Asian, South Asian, and Southeast Asian. In addition to the home languages in the table, some Asian students in this sample also reported English as a primary home language or a non-English language that was not identifiable as a language associated with an Asian ethnic subgroup (e.g. Italian or Swahili). Table 6 in the paper shows the breakdown of Asian students in the sample by reported language categories.

Table C1: NCERDC Home Language Code Classification

Subroup	**Language Codes**
East Asian	Chinese (Mandarin), Chinese (Cantonese), Chinese (Zhongwen), Chinese (Shanghai/Wu), Chinese (Taiwan), Chinese, Japanese, Korean
South Asian	Gujarati, Hindi, Punjabi/Panjabi, Tamil, Telugu, Urdu, Bengali, Bihari, Hindi/Indian/Urdu, Kannada, Kashmiri, Pushto/Eastern Pashto, Saurashtra/Sowrashtra, Sindhi, Marathi, Oriya, Hindko
Southeast Asian	Vietnamese, Burmese, Cambodian/Khmer, Cebuano, Indonesian, Hmong/Hmong-Mien/Hmogie/Chaug, Koho, Rade, Tagalog/Filipino, Lahu, Lao/Laotian, Tai/Eastern Tai, Malay/Bahasa Malaysia, Malayalam, Thai/Ta/Thaiklang, Jarai, Mnong, Chin

Classification of Asian students into subgroups based on NCERDC self-reported home language.

As an alternative approach, we use county-level Asian subgroup population to proxy for students' ethnicities. Data comes from the American Community Survey (ACS) from 2007-2013. For each county, we measure the average aggregate Asian population over that time frame, as well as the Asian population broken down by subgroup (East Asian, South Asian, and Southeast Asian). We use the proportion of Asians of a given subgroup in the county as a proxy for how likely an Asian student is from a given subgroup. One limitation of this approach is that the data are rather coarse –unlike in our preferred approach, we do not observe ethnicity data at the individual level. Furthermore, the ACS only has individual county-level data for the 25 largest counties in North Carolina, out of 50 total. The remaining smaller counties are aggregated into one category. The benefit of this approach though, is that we are able to circumvent the issue that many Asians in our sample are English-speaking, which created identification issues in the home language approach.

Table C2 shows results using county-level Asian ethnic shares as a subgroup proxy. As in the home language approach, results indicate that conditional on the share of Asians in a county, an increase in the share of East and/or South Asians relative to Southeast Asians increases the propensity that teachers will over-rate an Asian student, relative to a White student with the same standardized test score. A 10 percentage point increase in the share of Asians in a county that are East Asian, relative to Southeast Asian, increases the propensity that a teacher will over-rate an Asian student by 0.6 percentage points. A 10 percentage point increase in South Asian share increases the propensity that a teacher over-rates a Southeast Asian student by 0.5 percentage points. A Wald test of coefficients shows that the effect of proportion East Asian and proportion South Asian are not statistically different from one another.

Conversely, a 10 percentage point increase in the share of Asians in a county that are South Asian, relative to Southeast Asian, decreases the propensity that a teacher will under-rate an Asian student by 0.6 percentage points. We find no statistically significant effect of an increase in East Asian share on the propensity that a teacher under-rates a Southeast Asian student. A Wald test of coefficients shows that the effect of proportion East Asian and proportion South Asian are not statistically different from one another at the 5% level but are different at the 10% level.

Next, Table C3 examines whether racial gaps in teacher assessment differ for teachers in an urban versus more rural setting. To do so, we augment our main specification with an interaction term for whether the school is based in a city (relative to a rural, town, or suburban location). Results indicate teachers in cities are less positive towards Asian students: they are 1.7 percentage points less likely to over-rate Asian students and 0.9 percentage points more likely to under-rate Asian students than counterparts teaching in non-city settings. Teachers in cities are also less positive towards Black

Table C2: Racial Differentials in Teacher Assessments by ACS Asian Subgroup

	Over-rate $(NB > B)$	Under-rate $(NB < B)$
Asian	0.014	-0.011**
	(0.009)	(0.005)
Asian×Proportion Asian	-0.027**	0.031***
	(0.013)	(0.007)
Asian×Proportion East|Asian	0.006**	-0.002
	(0.002)	(0.001)
Asian×Proportion South|Asian	0.005***	-0.006***
	(0.002)	(0.001)
Class FE	Y	Y
Race×teacher FE	Y	Y
N	12,386,507	13,835,115

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. Other minority races and interactions with Asian share and Asian subgroup shares are included in regression, although they are not displayed in table. Coefficients represent the effect of a 10 percentage point increase in proportion of interest. The omitted Asian subgroup share is proportion of Southeast Asians. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Asian subgroups are classified using reported ancestry data from the ACS (East Asian: Chinese, Cantonese, Japanese, Okinawan, Korean, Taiwanese. South Asian: Bengali, Nepali, Asian Indian, Punjabi, Pakistani, Sri Lankan. Southeast Asian: Burmese, Cambodian, Filipino, Indonesian, Laotian, Hmong, Malaysian, Thai, Vietnamese). Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher overrating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$. Classification of counties into subgroups shares based on ACS self-reported ancestry data from 2007-2013.

and Hispanic students.

To help understand what might be driving results in Table C3, Table C4 provides descriptive information for students in city vs. non-city settings by race. We see that Asian students in cities have relatively lower socioe-

Table C3: Racial Differentials in Teacher Assessments, City vs. Non-city

	Over-rate $(NB > B)$ (1)	Under-rate $(NB < B)$ (2)
Asian×City	-0.017***	0.009***
	(0.004)	(0.003)
Black×City	-0.012***	0.009***
	(0.002)	(0.001)
Hispanic×City	-0.014***	0.008***
	(0.002)	(0.002)
American Indian×City	-0.008	0.011*
	(0.007)	(0.006)
Other×City	-0.010***	0.005**
	(0.003)	(0.002)
Asian	0.045***	-0.029***
	(0.003)	(0.002)
Black	-0.023***	0.020***
	(0.001)	(0.001)
Hispanic	-0.020***	0.018***
	(0.001)	(0.001)
American Indian	-0.022***	0.009***
	(0.003)	(0.002)
Other race	-0.006***	0.005***
	(0.001)	(0.001)
N	12,261,316	14,016,886

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013. Omitted category: White students. Omitted teacher race: White teachers. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

conomic and academic outcomes than White peers compared to Asian students in rural, town, or suburban locations. Asians are 16 percentage points more likely to be economically disadvantaged than White peers in cities, while they are only 7 percentage points more likely to be economically disadvantaged in non-city locations. The average lagged math score of Asian students in cities is .12 standard deviations higher than that of White peers, while the corresponding measurement is .26 standard deviations higher in non-city locations. The average lagged reading score of Asian students in cities is .17 standard deviations lower than that of White peers, while the corresponding measurement is .02 standard deviations lower in non-city locations. Overall, information in Table C4 suggests that positive stereotyping towards Asian students may be lower in urban areas because Asians in these areas tend to conform less to the "model minority" stereotype.

Table C4: Descriptive Student Statistics by Race, City vs. Non-city

	White		Asian		Black		Hispanic	
	City	Non-City	City	Non-City	City	Non-City	City	Non-City
Econ disadvantaged	0.21	0.34	0.37	0.41	0.73	0.75	0.83	0.84
	(0.41)	(0.48)	(0.48)	(0.49)	(0.44)	(0.43)	(0.38)	(0.37)
Lagged math score	0.56	0.26	0.68	0.52	-0.46	-0.47	-0.26	-0.21
	(0.93)	(0.91)	(1.11)	(0.99)	(0.90)	(0.88)	(0.93)	(0.89)
Lagged reading score	0.56	0.27	0.39	0.25	-0.41	-0.43	-0.41	-0.35
	(0.90)	(0.91)	(1.09)	(0.98)	(0.92)	(0.90)	(0.97)	(0.93)

Observations are at the student-year level for students in grades 3-8 in math or reading classes between 2007-2013. Lagged test scores are measured as z-scores.

D Spillover Effects: Exposure to Black or Hispanic Students

Tables D1 and D2 examine the spillover effects on Hispanic students of exposure to a single Black student. In Table D1, the variable Any Black takes on a value of one if there is a single Black student in the class. We furthermore include classroom fixed effects and race interacted with teacher-school-grade-course fixed effects. Identification comes from variation in how a teacher in a particular school, grade, and course assesses students of different races based on whether they had a Black student in their class. Table D1 also includes interactions between having a Black student in the classroom and indicators for all remaining racial and ethnic groups, even though those variables are not displayed. Results indicate the presence of a Black student in a classroom has no measurable impact on the propensity of teachers to over-rate or under-rate Hispanic students, compared to classrooms with zero Black students. This stands in contrast with the results for Asian student exposure in Table 8.

Next, Table D2 examines whether the effect of Black students on teacher biases toward Hispanic students varies based on the academic performance of the Black student. The presence of an average-performing Black student in the class has no effect on the propensity for a teacher to over-rate or under-rate Hispanic students, as shown in the first row of coefficients. Moreover, whether the Black student is high- or low-achieving has no bearing on teachers' assessments of same-class Hispanic students.

We conduct parallel analyses on the spillover effects on Black students of exposure to a single Hispanic student in Tables D3 and D4. We find that exposure to a single Hispanic student in the classroom has no significant effect on teachers' likelihood of over-rating or under-rating Black students, compared to classrooms without any Hispanic students. This echoes the lack of spillover effects from having a Black student in class and is distinct from the impact of Asian student exposure.

Table D4 examines whether the effect of Hispanic students on teacher

Table D1: Effect of Exposure to One Black Student

	Over-rate $(NB > B)$ (1)	Under-rate $(NB < B)$ (2)
Hispanic×Any Black	0.004	0.002
	(0.004)	(0.003)
Class FE	Y	Y
Race×teacher-school-grade-course FE	Y	Y
N	2,892,721	3,982,062

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013 and is limited to classrooms that have either zero or one Black student. Any Black is a binary variable indicating that the classroom had one Black student. Omitted category: White students. Models include interactions between Asian students, American Indian students, and students of other racial or ethnic groups with the Any Black variable. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

biases toward Black students depends on the baseline performance of the Hispanic student. While the presence of an average-scoring Hispanic student does not affect teachers' propensities to over-rate Black students, the coefficient on the interaction of Black and lagged Hispanic test scores indicates that there are heterogeneous effects by achievement level. A one standard deviation increase in the Hispanic student's performance decreases the propensity for a teacher to over-rate Black students in the class by 0.6 percentage points. There are no significant effects of having a Hispanic student in the class on teachers' propensities to under-rate Black students.

Table D2: Effect of Exposure to One Black Student, by Achievement

	Over-rate $(NB > B)$ (1)	Under-rate $(NB < B)$ (2)
Hispanic×Any Black	0.004	-0.001
	(0.005)	(0.004)
Hispanic×Black Lagged Z-score	0.001	0.000
	(0.004)	(0.003)
Class FE	Y	Y
Race×teacher-school-grade-course FE	Y	Y
N	2,312,475	3,234,926

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 4-8 between 2007-2013 and is limited to classrooms that have either zero or one Black student. Any Black is a binary variable indicating that the classroom had one Black student. Omitted category: White students. Any Black is a binary variable, while Lagged Z-score is the Black student's standardized lagged z-score normalized within the population of all students in a given grade and year. Models include interactions between Asian students, American Indian students, and students of other racial or ethnic groups with the Any Black and Black Lagged Z-score variables. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

Table D3: Effect of Exposure to One Hispanic Student

	Over-rate $(NB > B)$ (1)	Under-rate $(NB < B)$ (2)
Black×Any Hispanic	0.001	-0.001
	(0.002)	(0.002)
Class FE	Y	Y
Race×teacher-school-grade-course FE	Y	Y
N	5,063,580	6,355,783

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 3-8 between 2007-2013 and is limited to classrooms that have either zero or one Hispanic student. Any Hispanic is a binary variable indicating that the classroom had one Hispanic student. Omitted category: White students. Models include interactions between Asian students, American Indian students, and students of other racial or ethnic groups with the Any Black variable. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$.

Table D4: Effect of Exposure to One Hispanic Student, by Achievement

	Over-rate $(NB > B)$ (1)	Under-rate $(NB < B)$ (2)
Black×Any Hispanic	0.001	-0.000
	(0.002)	(0.002)
Black×Hispanic Lagged Z-score	-0.006***	0.002
	(0.002)	(0.002)
Class FE	Y	Y
Race×teacher-school-grade-course FE	Y	Y
N	4,102,091	5,201,211

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at the teacher level. Sample comprises students in grades 4-8 between 2007-2013 and is limited to classrooms that have either zero or one Hispanic student. Any Hispanic is a binary variable indicating that the classroom had one Hispanic student. Omitted category: White students. Any Hispanic is a binary variable, while Lagged Z-score is the Hispanic student's standardized lagged z-score normalized within the population of all students in a give grade and year. Models include interactions between Asian Students, American Indian students, and students of other racial or ethnic groups with the Any Hispanic and Hispanic Lagged Z-score variables. All specifications include controls for observable student characteristics, class fixed effects, and raw end-of-grade test score fixed effects interacted with subject, grade, and year. Student characteristics include gender, days absent, and an indicator for economic disadvantage. The sample of students in the assessment of teacher over-rating includes those with $B \in \{1, 2, 3\}$. The sample of students in the assessment of teacher under-rating includes those with $B \in \{2, 3, 4\}$..