Model Minorities in the Classroom? Positive 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." Using the context of education, this paper first establishes empirical evidence of this stereotype and then analyzes its consequences. We show that teachers rate Asian students' academic mastery more favorably than observationally similar white students. This positive bias contrasts with teachers' lower likelihood of favoring Black and Hispanic students, even after accounting for academic performance and behavior. Notably, the presence of any Asian student in the classroom exacerbates existing Black-white and Hispanic-white assessment gaps. This suggests that the "model minority" stereotype can have negative spillover effects on other minority groups in spite of its ostensibly positive valence.

JEL: I24, J15

*We are grateful to Peter Arcidiacono, Elizabeth Cascio, Katherine Harris-Lagoudakis, Soumaya Keynes, Jessica Merkle, Emily Owens, Emily Weisburst, Arezou Zaresani, and conference/seminar participants at AEFP, APPAM, SEA, and Syracuse University for helpful comments.

†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/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).

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

This study provides evidence on the prevalence and consequences of positive bias towards Asian students in schools. The view of Asians as model minorities is pervasive in K-12 education given their ability to outperform other racial groups, including white students, on most standardized test measures (Kao 1995, Hsin and Xie 2014). We present evidence for positive bias in teacher assessments of Asian students, before exploring heterogeneity across different subgroups of Asians. Specifically, we examine whether the extent of positive bias varies across Asian ethnic subgroups, as well as between high- versus low-performing Asian students. Finally, this paper analyzes whether the propensity of teachers to favor Asian students has spillover effects, by examining how the presence of any Asian students in the classroom affects teachers' assessments of other under-represented minorities.

To address our research questions, we use administrative data from the North Carolina Education Research Data Center (NCERDC) covering students in grades 3-8 from 2007 to 2013. The NCERDC data have 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 several years, which provides a critical mass of Asian students for meaningful analyses. Second, the data contain information on both blind-scored standardized assessments and subjective, non-blind teacher evaluations of student performance on the same underlying skills.

Our inference is based on the comparison of teachers' reading and math evaluations of students with scores measured along the same scale as standardized (and blindly marked) tests. Standardized test scores provide a benchmark for assessing whether teachers are systemically over-rating or under-rating Asian students relative to other groups, conditional on achievement and individual attributes such as gender and attendance. 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 teachers display significant positive bias towards Asian students, relative to white students with the same underlying standardized test scores and sociodemographic characteristics. Compared to white students, teachers are 3.7 percentage points more likely to give Asian students a higher evaluation (over-rate) than their blind-scored achievement level and 2.6 percentage points less likely to give Asian students a lower subjective evaluation (under-rate) than their standardized test achievement level. These magnitudes correspond to 10% 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 a number of robustness checks to rule out alternative explanations for these racial differences, including hard-to-observe behavioral attributes, the comparability of blind vs. non-blind achievement scales, and racial biases in standardized testing. Additionally, we find heterogeneous effects by student achievement levels and more fine-grained ethnic subgroups. The magnitude of positive bias for Asians relative to white students is more pronounced among high-achieving students. Teachers also display greater positive bias towards Asian students from East and South Asian backgrounds, relative to students from Southeast Asian backgrounds.

Next, our findings suggest positive bias may have negative spillovers –having an Asian student in the classroom decreases the propensity for a teacher to over-rate a Black or Hispanic relative to a white student with comparable 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. Further analyses show that exposure to an average-performing Asian student exacerbates teachers' negative biases towards Black and

Hispanic students in ways not observed when the exposure is to an average-performing Black or Hispanic student, indicating results are not being driven by Asian students increasing the average achievement level of a classroom. These findings support the notion that the presence of Asian students may uniquely 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 Asian Americans being the fastest-growing single racial/ethnic category in the country and explicit calls for more research on this demographic group, scholarship on Asian Americans' educational and labor market trajectories and others' perceptions of Asian Americans is limited in disciplines such as economics and sociology (Altonji & Blank, 1999; Sakamoto, Goyette, & Kim, 2009).[1] We document racial differences in teacher assessments that favor Asians relative to white students, in a manner that sets Asian students apart from other minority groups such as Black and Hispanic students. This lends some empirical credence to the existence of positive stereotypes. Yet the patterns for Asians belie substantial heterogeneity within this group, with diminished positive bias towards low-performing Asian students or Asians from particular ethnic groups (e.g. individuals from Southeast Asia). 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 experiences and achievement (Chiswick, 1983; Lee & Zhou, 2015; Sakamoto et al., 2009).

In addition to documenting the extent of bias towards Asian students, we examine how biases toward different racial and ethnic groups might interact. Potentially detrimental consequences of positive bias are the inclination to believe that the targeted group is essentially or fundamentally different from other groups and also 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 effects on other minority groups. These findings are also consistent with a theoretical conception of stereotypes rooted in representativeness (Kahneman and Tversky 1972,

[1] In economics, for example, relatively little research in economics has focused on Asian Americans and the unique experiences they face. A few studies have looked at Asian American labor market trends in earnings and the factors behind them (Chiswick 1983, Duleep and Sanders 1992, Hilger 2017). 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.

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. 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). Biased teacher expectations can become self-fulfilling prophecies that influence students' academic achievement and attainment (Rosenthal and Jacobson 1968, Jussim and Harber 2016, Hill and Jones 2017, Papageorge, Gershenson, and Kang 2020). While recent papers have looked at discrepancies in teacher expectations across racial and ethnic groups (Botelho, Madeira, and Rangel 2015, Rangel and Shi 2020), there is scarce research investigating bias towards Asians.[2]

In the remainder of the paper, Section 2 presents the NCERDC 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 biases in student 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). One key advantage of the dataset is its scope. The universe of elementary and secondary public school students provides a substantive sample of Asian students for analysis. Another advantage is the ability to link students to their teachers and classrooms in order to examine systemic patterns of teacher evaluations. In this paper, we focus on students in grades 3-8 from 2007-2013.

An important feature of the data is the presence of both blind-scored assessments and sub-

[2] An exception is Burgess and Greaves (2013), which juxtaposes teacher assessments across Asian subgroups such as Indian, Chinese, Bangladeshi, and Pakistani.

jective teacher evaluations of student performance along the same scale. Students begin taking standardized achievement tests in math and reading in third grade. These tests are given during the last three weeks of the school year, with questions formulated in a multiple-choice format. Student performance on End-of-Grade (EOG) tests are scored by machine, with raw scores mapped to achievement levels on a 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.[3] We classify standardized test score assessments of math and reading ability as "blind."

Concurrently with the EOG tests, teachers are asked to provide their assessment of each student's achievement level for both math and reading comprehension corresponding to insufficient, inconsistent, consistent, or superior mastery. Specifically, for math and reading subjects, teachers are asked which students "in the [subject] teacher's professional opinion, clearly and consistently exemplifies one of the achievement levels listed." Furthermore, teachers are explicitly instructed to assess students based on mastery, rather than student behavior: "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."

Teachers submit evaluations before knowing standardized test results, and evaluations map to the same 1-4 scale of achievement levels in terms of underlying skill sets being assessed. A key difference between the two assessment methods is that teachers inevitably know which student they are evaluating, which makes their assessments "non-blind." With that knowledge comes

[3] 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

information about and the race and ethnicity of each student, and this analysis focuses on whether this information influences how teachers perceive a student's skill-based achievement level.

2.2 Descriptive Statistics

Table 1 describes the sample of students. Approximately 3% of students are Asian, while the majority of students (54%) are white. Black and Hispanic students make up 27% and 12% of the sample, respectively. 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 30,000 Asian students in our sample, which allows for a statistically rigorous analysis of this group.

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.

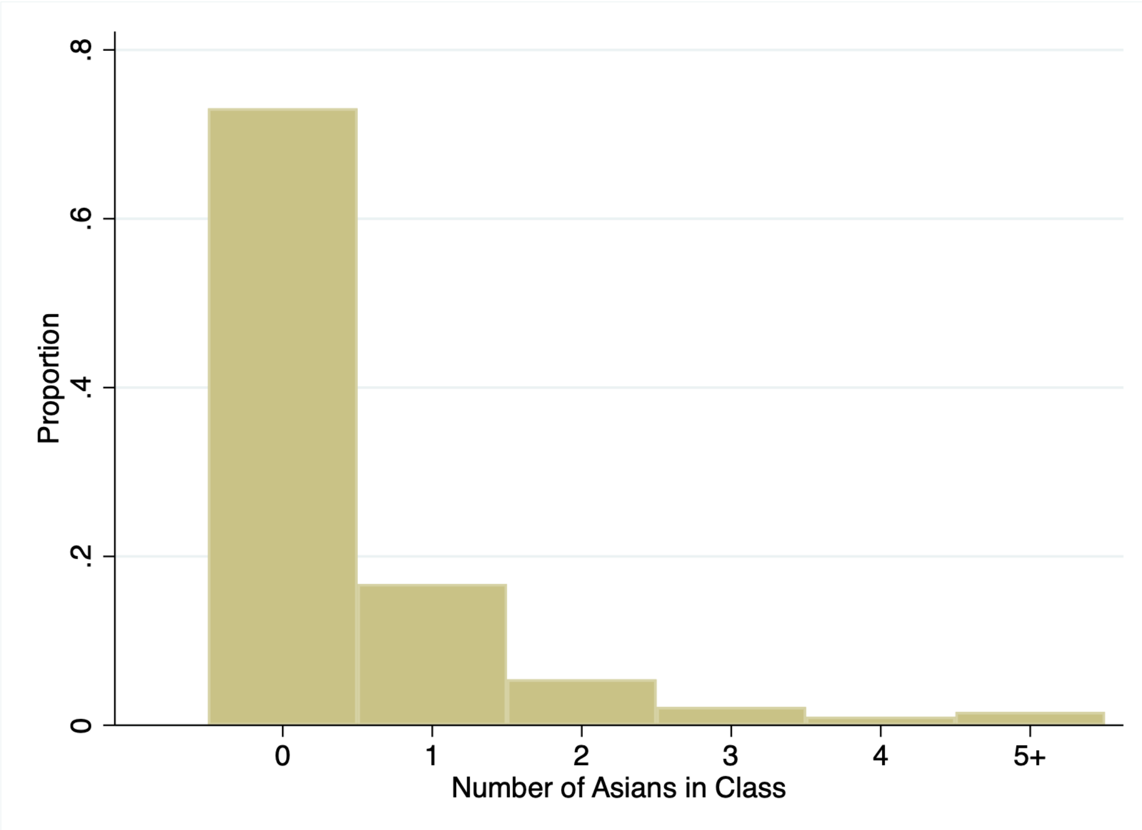
Table 1: Student Characteristics

	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	1,410,448

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 of that value for each year they appear in the data.

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.

Figure 1: Asian Representation across Classes



Observations in these graphs are at the classroom level. The histogram on the left shows the distribution of number of Asians in a classroom, and the histogram on the right shows the distribution of what share of the class Asians make up.

Table 2 details the characteristics of teachers in the sample. Relative to students, teachers are disproportionately white (82% of the sample). Black teachers making up most of remaining teachers at 15% 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 2.76. Overall, 22% of students rank in the top achievement category, level 4. Another 44% of students score at level 3, which represents the plurality of students. Finally, 20% of students score at level 2 and 13% of students score at level 1. Compared to both white and under-represented minority students, Asian

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	50,210

Observations are at the teacher level for teachers teaching grades 3-8 in math or reading classes between 2007-2013. A teacher's experienced calculated as their average number of years of experience over the period they appear in the data.

students have significantly higher average achievement levels and are disproportionately represented in the higher achievement categories. Their average achievement level is 3.10, while the corresponding measures are 3.00, 2.36, and 2.47 for white, Black, and Hispanic students, respectively. The differences in achievement scores between white and Asian students mostly occurs at the top of the distribution. For instance, 40% of Asian students have an achievement level of 4, compared to only 31% of white students.

Table 3: Blind-scored Achievement Levels by Race

Blind-scored Achievement	All	White	Asian	Black	Hispanic
Mean	2.76	3.00	3.10	2.36	2.47
Level 4	0.22	0.31	0.40	0.09	0.12
Level 3	0.44	0.47	0.38	0.41	0.43
Level 2	0.20	0.15	0.13	0.29	0.26
Level 1	0.13	0.08	0.09	0.22	0.20
N	16,000,011	8,637,104	389,372	4,184,539	1,892,616

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.

Table 4 compares the propensity for teacher 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, and columns represent the teacher's non-blind achievement rating for the student. Cells denote the proportion of students who were given the corresponding column scores by their teachers, conditional on having a blind achievement score level denoted by the row. Dark shaded cells represent cases in which teachers over-rate students relative to their blind-scored achievement levels, while light shaded cells represent cases in which teachers under-rate students relative to their blind-scored achievement levels.

Table 4: Blind vs. Non-Blind Scores

	Teacher rating							
Blind-scored	**White students**				**Asian students**			
Achievement	Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
Level 1	0.21	0.45	0.32	0.02	0.21	0.41	0.33	0.04
Level 2	0.08	0.33	0.52	0.08	0.06	0.28	0.52	0.14
Level 3	0.02	0.14	0.57	0.26	0.02	0.10	0.50	0.37
Level 4	0.00	0.02	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 shaded areas denote cells for which a teacher over-rated a student relative to their blind scores, and light shaded areas denote cells for which a teacher over-rated a student relative to their blind scores.

Values in the table indicate teachers may be more likely to over-rate Asians and less likely to under-rate Asians relative to white counterparts. 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 of 4 by their teachers, this proportion is 37% for Asian students. Conversely, while 35% of white students who have a blind achievement score of 4 are given a rating lower than 4 by their teachers, this proportion is only 26% for Asian students. Overall, teachers have a 36.4% chance of over-rating if the student is white and a 44.0% chance of over-rating if the student is Asian. Teachers have an 18.3% chance of under-rating if the student is white and a 14.3% chance if the student is Asian. 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 the 99% confidence level.[4]

While Table 4 provides suggestive evidence that teachers may exhibit positive bias towards

[4]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.

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 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 presence of teacher biases in subjective 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 student ability, behavior, and factors governing the assignment of students into classrooms. Our main specification accounts for confounding 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.

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 raw end-of-grade exam score, E_{ic}. This accounts for the possibility that teachers may be systematically more likely to over-assess or under-assess high-performing students relative to lower-performing students, which could bias our estimate of β if the distribution of student achievement scores differs by race. Another potential issue arises from the fact 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 varies by race.[5] In this scenario, differences in teacher assessment relative to achievement

[5]For example, suppose white students who get categorized in achievement level 4 in 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

bins may reflect actual differences in achievement, rather than underlying teacher racial biases. The inclusion of end-of-grade exam score indicators ensures that this phenomenon is not driving our results.

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, since we are looking at differences in assessment outcomes within teachers. 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' propensity to over-rate or under-rate students differs across student racial/ethnic groups, we examine the coefficient of interest β on the vector of student race and ethnicity indicators ($\mathbf{R}_{ic}$), with white students as the reference category. In other words, β captures systemic racial differences in teachers' subjective evaluations within a given class, after adjusting for students' performance on standardized, blind-scored 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 ($\mathbb{1}\{NB < B\}$) their student i in classroom c:

$$O_{ic} = \text{AnyAsian}_c \times \mathbf{R}'_{ic}\gamma + \sum_{j=1}^{J} \mathbf{R}'_{ic}\delta_j + \mathbf{R}'_{ic}\pi + \rho f(E_{ic}) + \mathbf{X}'_{ic}\Gamma + \theta_c + \epsilon_{ic} \tag{2}$$

The above model follows Equation 1 in flexibly controlling for the student's blind-scored raw test performance alongside individual attributes such as the number of days absent, economic dis-

in achievement level 4 have raw test scores well above the cutoff.

advantage, 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-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.[6] We infer a causal interpretation of the parameter of interest (γ) 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 and course, after controlling for time-varying factors such as teacher experience. 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-course fixed effects and controls for teacher experience, classroom attributes such as the white-Black or white-Hispanic achievement gap do not predict whether an Asian student is present (Appendix Table A1).

4 Results

4.1 Teacher Bias toward Asian Students

Table 5 shows the propensity for teachers to over-rate or under-rate students after adjusting for raw standardized test scores and individual characteristics such as days absent. The outcome variable in the first column is an indicator variable for whether a teacher over-rates a student

[6] We furthermore include student race indicators interacted with teacher experience indicator in order to accommodate for the possibility that the composition of Asians in a classroom is correlated with this time-varying teacher attribute.

relative to their blind achievement level, while the outcome variable in the second column is an indicator variable for whether a teacher under-rates a student.[7] The omitted racial group is white students.

Table 5: Racial Differentials in Teacher Assessments

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Asian	0.037***	-0.026***
	(0.002)	(0.001)
Black	-0.028***	0.023***
	(0.001)	(0.001)
Hispanic	-0.024***	0.020***
	(0.001)	(0.001)
American Indian	-0.024***	0.011***
	(0.003)	(0.002)
Other race	-0.008***	0.006***
	(0.001)	(0.001)
N	12,386,507	13,835,115

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. All specifications include class fixed effects, observable student characteristics, and factor variable controls for raw end-of-grade test (interacted with subject) scores. Student characteristics include gender, days absent, and an indicator for economic disadvantage status.

Results indicate teachers are 3.7 percentage points more likely to over-rate Asian students relative to white students with the same standardized test scores and individual characteristics. The magnitude of the effect is sizable, considering the overall propensity to over-rate students is 0.378. Put another way, teachers are more likely to over-rate Asian students relative to white students by nearly 10% of the baseline propensity of being over-rated. We document comparable magnitudes when examining the phenomenon of under-rating. Teachers are 2.6 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. Disaggregating these results by subject shows that the differences in teacher assessments described above are reflected in both math and reading teacher evaluations (Appendix Table B1).

[7] 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 since it is not possible to under-rate these students.

Consistent with previous literature, teachers are less likely to over-rate and more likely to under-rate Black and Hispanic students relative to both white and Asian students (Burgess & Greaves, 2013; Rangel & Shi, 2020). Notably, the magnitude of teachers' increased propensity to over-rate and decreased propensity to under-rate Asians is larger than the magnitudes of decreased over-rating and increased under-rating for Black and Hispanic students. These results suggest that the degree of positive bias teachers display towards Asian students in subjective evaluations is at least as strong, if not more so, than the negative bias they display towards under-represented minority students.

Robustness

We undertake a number of additional analyses to address concerns that results may be driven by alternative channels. First, 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. As a robustness check, we augment this specification with the lagged number of days absent from the prior year to address the possibility that a student's contemporaneous absences are endogenous with teacher subjective assessments. We also include a control for a vector of 50 different types of disciplinary infractions. Table B2 in the Appendix provides details on the types of infractions students may be reported on, and Table B3 shows the results of our analysis with augmented behavioral controls. Including lagged days absent and the types of disciplinary infractions does not significantly change the results, providing further support that our findings are not being driven by underlying behavioral differences across racial groups.

Another way we test for the role of underlying behavioral differences is to look at how these effects vary by gender. Several studies have established that boys display significantly more early childhood behavior problems on average than girls (Bertrand & Pan, 2013; Diprete & Jennings, 2012; Owens, 2016). If systematic differences in the way teachers assess Asian and white students with the same standardized test scores are driven by unobserved behavioral factors, we would expect these results to be larger for boys relative to girls. Table B5 in the Appendix shows racial assessments by gender. Teachers are more likely to over-rate Asian girls but also more likely to under-rate them, relative to their propensities to over-rate and under-rate Asian boys. These findings are inconsistent with what we would expect to see if racial differences in teacher assessments were driven by underlying behavioral differences, in which case we would expect teachers to be less likely to over-rate Asian girls and more likely to under-rate them.

Aside from potential behavioral differences across racial groups, we explore the comparability of the blind and non-blind achievement scales as a confounding factor. Even though both ostensibly measure math and reading mastery, teachers' standards of mastery may vary depending on the particular school or classroom context.[8] To ensure that we are not mistaking these influences for teacher bias, we construct adjusted distributions of End-of-Grade achievement levels so that they match the distribution of teacher rating levels (on the 1-4 scale) in each class. Using raw EOG 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. Table B4 shows that when the outcome is modified to teacher over- and under-rating relative to these adjusted EOG achievement levels, the estimated coefficients for Asian students remain exactly the same. This strongly suggests that what we interpret to be teacher bias is not confounded by the comparability of blind vs. non-blind achievement scales.

Next, we consider the possibility that our findings are being driven by racial biases in standardized testing, rather than racial biases in teacher evaluations. Theoretically, observed racial patterns in over-rating and under-rating could happen in the absence of any teacher bias if it were the case that tests are negatively biased towards Asian students. In this situation, teacher assessments would reflect true student achievement, and differences in blind and non-blind assessments would be driven by non-blind test biases. To explore this possibility, we look at how teachers' assessments of Asian students relative to white students varies by whether an Asian student reports English as their primary home language. If it were the case that standardized tests display neg-

[8] For instance, teachers with high-performing students may have higher standards for what constitutes a proficient student, independent of state guidelines. Our reliance on classroom fixed effects is designed to absorb these effects common to classrooms.

ative cultural/racial biases towards Asian students, 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. Table B6 in the Appendix shows this analysis, and results 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 English home language Asians. These findings go in the opposite direction of the coefficients we would expect if results were being driven by racial bias in tests, rather than racial bias in teachers, providing support that our findings do in fact reflect teacher bias.

Finally, we include a specification containing lagged standardized test scores, in addition to contemporaneous achievement. Teachers likely form impressions of students over the course of the year, and one concern is that students' achievement levels at the beginning of the year may affect both teacher subjective assessments and teachers' value-added. This could potentially bias estimates if, for example, Asian students are higher-performing on average at the beginning of the school year and teachers invest more effort into bringing up test scores for lower-performing students. The inclusion of lagged test scores controls for initial differences in student achievement that may affect teachers' propensities to over-assess or under-assess students. Table B7 in the Appendix displays this analysis, and results indicate the inclusion of lagged test scores does not reduce the propensity for teachers to over-rate or under-rate Asian students, relative to comparable white students in the class.

4.2 Heterogeneity in Teacher Bias by Asian Subgroup

To gain a deeper understanding of the nature of the teacher bias, we next examine heterogeneity in bias by different Asian subgroups. Specifically, we look at how the degree of bias differs across Asian students by academic achievement, as well as ethnic groups.

Descriptive statistics from Section 2 suggest that positive bias towards Asians may be more pronounced among higher-achieving Asian students. Table 6 takes this hypothesis to the empirical framework described in Section 3 and segments students as "High", "Middle" or "Low"-scoring based on standardized test achievement levels.[9] Note that the omitted racial group is white students and the omitted achievement level is low-performing students.

Table 6: Heterogeneity in Teacher Assessments by Asian Achievement

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Asian×Middle Score	0.045***	-0.015***
	(0.006)	(0.002)
Asian×High Score	0.064***	-0.037***
	(0.005)	(0.003)
Asian	-0.014***	-0.002
	(0.005)	(0.002)
N	12,386,507	13,835,115

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted blind score level: low. Omitted race: white students. Other minority races and score interactions are not displayed in table but are included in regression. All specifications include class fixed effects, observable student characteristics, and factor variable controls for raw end-of-grade test (interacted with subject) scores. Student characteristics include gender, days absent, and an indicator for economic disadvantage status.

Results indicate that positive bias among teachers is most pronounced among Asian students in the high-scoring category. Relative to high-scoring white students, teachers are more likely to over-rate high-scoring Asian students by 6.4 percentage points. Conversely, they are 3.7 percentage points less likely to under-rate high-achieving Asians relative to their similarly-performing white peers. The extent to which teachers over-rate and under-rate Asian students in the middle-scoring category is attenuated, but still sizable and significantly different from zero, at 4.5 and -1.5 percentage points, respectively. In contrast, there is less evidence that teachers differentially favor low-scoring Asian students. They are 1.4 percentage points *less* likely to over-rate low-achieving Asians relative to low-achieving whites and do not differ in their propensity to under-rate these two groups. This pattern for heterogeneous effects by achievement levels is consistent with the idea that high-achieving Asians confirm teachers' ex-ante expectations of Asian students as model

[9] In the over-rating sample comprising students whose blind achievement levels are 1, 2, or 3, we classify students as "Low", "Middle", and "High", respectively. In the under-rating sample covering students scoring at achievement levels 2, 3, and 4, we classify them as "Low", "Middle", and "High", respectively.

minorities, which leads their subjective evaluations to be more biased along stereotypical dimensions.

Next, we look at heterogeneity in teacher bias across different Asian ethnic groups. Existing studies examining the educational and labor market trajectories of Asians typically classify them into a single category, even when research demonstrates substantial differences in schooling and earnings across Asian ethnic groups (Chiswick, 1983). Grouping all Asians into a monolithic category potentially disguises these differential experiences and trajectories. 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.[10]

Table 7 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 7 also provides descriptive statistics for Asian students by home language subgroup. Consistent with previously documented 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.

Next we analyze teacher assessments across Asian subgroups using home language as a proxy for ethnicity. Table 8 shows substantial heterogeneity in bias across subgroups. Compared to white students, teachers are 5.7 percentage points more likely to over-rate South Asian students, 4.3 percentage more likely to over-rate East Asian students, and 1.8 percentage points more likely to over-rate Southeast Asian students. A Wald test of coefficients indicates the coefficients between both South Asians and Southeast Asians and East Asians and Southeast Asians are significantly different at the 1% level, suggesting systemic differences in teacher assessments towards Southeast Asian students relative to peers speaking a home language commonly associated with countries

[10]Table C2 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.

Table 7: Asian Subgroups by Home Language Status

	N	Percent	% FRL	Math scores	Reading scores
East Asian	2256	6.09	0.19	1.15	0.41
South Asian	2353	6.35	0.22	0.86	0.48
Southeast Asian	5522	14.90	0.69	0.31	-0.32
Other Asian	2192	5.91	0.68	-0.32	-0.63
Asian (English)	20937	56.49	0.30	0.71	0.46
Asian: Missing Language	3800	10.25	0.45	0.54	0.21
Total/average	37060	100.00	0.38	0.57	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. All specifications include class fixed effects, observable student characteristics, and factor variable controls for raw end-of-grade test (interacted with subject) scores.

including China, Japan, South Korea, India, and Pakistan. In terms of under-rating, teachers are 3.0 percentage points less likely to under-rate South Asians, 2.1 percentage points less likely to under-rate East Asians, and 1.5 percentage points less likely to under-rate Southeast Asians, compared to white students. A Wald coefficient test rejects the hypothesis that teachers have the same propensity to under-rate South Asians and Southeast Asians at the 1% level.

A key advantage to using home language information to proxy for Asian ethnic subgroup is that we are able to obtain detailed ethnic information at the individual level. However, a drawback of this approach is that a large portion of the sample reports English as their primary home language. For robustness, we also analyze subgroup heterogeneity using a second approach that relies 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 Appendix C.

4.3 Spillover Effects on Under-Represented Minorities

Despite the positive valence of categorizing Asian students as a "model minority", such stereotyped views may have adverse intrapersonal and interpersonal consequences, for example by reinforcing the possibility of fundamental differences across groups and increasing the usage of

Table 8: Differentials in Teacher Assessments by Home Language

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
East Asian	0.043***	-0.021***
	(0.008)	(0.004)
South Asian	0.057***	-0.030***
	(0.006)	(0.004)
Southeast Asian	0.018***	-0.015***
	(0.003)	(0.003)
Other Asian	-0.035***	0.021***
	(0.008)	(0.008)
Asian: English	0.053***	-0.033***
	(0.003)	(0.002)
Asian: Missing Language	0.038***	-0.030***
	(0.005)	(0.003)
N	12,386,507	13,835,115

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. Note: other minority races are included in regression, although they are not displayed in table. All specifications include class fixed effects, observable student characteristics, and factor variable controls for raw end-of-grade test (interacted with subject) scores.

negative stereotypes (Kay et al., 2013). Table 9 begins to investigate how exposure to Asian students affects teachers' assessments of students from *other*, under-represented, 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-course level, including time-invariant preferences in assessments toward students of different racial and ethnic groups. The within-teacher design addresses concerns that Asian students sort into certain teachers' classrooms non-randomly on the basis of characteristics such as teacher credentials, training, or race.

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, such that restricting to one Asian student maintains a fairly representative sample. Table 9 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, compared to when no Asian students are present. Teachers are less likely to over-rate Black and Hispanic students by 0.5 and 0.6 percentage points, respectively. To place these magnitudes in context, this increases the baseline racial disparities in over-rating by one-fifth or more for Black and Hispanic students (see Table 5). When we turn to under-rating, the presence of an Asian student in the same classroom increases this propensity among teachers 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 corresponding change in under-rating for Hispanic students.

Table 9: Effect of Exposure to One Asian Student

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Black×Any Asian	-0.005***	0.004**
	(0.002)	(0.002)
Hispanic×Any Asian	-0.006**	0.003
	(0.002)	(0.002)
American Indian×Any Asian	0.005	0.003
	(0.010)	(0.008)
Other×Any Asian	-0.004	0.002
	(0.004)	(0.003)
Class FE	Y	Y
Race×teacher-school-course FE	Y	Y
N	11,065,708	11,971,495

*** p<0.01, ** p<0.05, * p<0.1. Sample includes grades 3-8 from 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. All specifications include controls for raw test scores, gender, economic disadvantage, absences, and interactions between race and years of teacher experience. Standard errors are clustered at the teacher level.

Next, Table 10 analyzes how changes in teacher evaluations of under-represented minorities in response to the presence of Asian students vary depending on the academic achievement of the Asian student. In particular, we inquire whether the role of Asians in exacerbating teachers' negative biases toward under-represented minorities is driven by high-achieving (stereotype-conforming) Asian students. We use lagged test scores as a measure for achievement to address potential endogeneity concerns with teacher expectations and Asian student performance.[11] Coef-

[11] With lagged test scores, the sample becomes students in grades 4-8, rather than 3-8, since we do not observe lagged scores for students in grade 3.

ficients on the interactions between race variables and Any Asian 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 *average* Asian student (defined as scoring at the statewide mean). 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-s.d. increase in the Asian student's achievement.

Table 10: Effect of Exposure to One Asian Student, by Achievement

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Black×Any Asian	-0.005**	0.004**
	(0.002)	(0.002)
Hispanic×Any Asian	-0.005**	0.002
	(0.003)	(0.002)
Black×Asian Lagged Z-score	-0.008***	0.001
	(0.002)	(0.002)
Hispanic×Asian Lagged Z-score	-0.007***	0.004**
	(0.002)	(0.002)
Class FE	Y	Y
Race×teacher-school-course FE	Y	Y
N	9,083,421	9,852,670

*** p<0.01, ** p<0.05, * p<0.1. Sample includes grades 4-8 from 2007-2013, and is limited to classrooms that have either zero or one Asian student. Any Asian is a binary variable, while Lagged Z-score is the Asian student's standardized lagged z-score. Models also include interactions between Asian students, 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 raw test scores, gender, economic disadvantage, absences, and interactions between race and years of teacher experience. Standard errors are clustered at the teacher level.

Table 10 shows that exposure to an average-achieving Asian student *decreases* the propensity for teachers to over-rate both Black and Hispanic students by 0.5 percentage points relative to white students in the class. As such, even the presence of an Asian student who performs at the state average is exacerbating 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 0.7 percentage points, relative to white students. These effects are more muted overall when looking at teacher under-rating. Exposure to an average Asian student decreases teachers' propensity to under-rate Black students by about 0.4 percentage points, with no evidence of 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 patterns of teacher assessments. Tables D1 and D2 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 achievement level 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 average. To ensure that this is not driven by the lower average performance of Black students, 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 0.5 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 is unique. Exposure to mediocre-performing Asian students 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 fastest growing 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 are more likely to over-rate Asian students and less likely to over-rate their Black and Hispanic peers relative to observationally similar white students in the same classrooms. Teacher assessment patterns setting 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 et al., 2015; Card & Giuliano, 2016; Hill & Jones, 2017; Lindahl, 2016; Papageorge et al., 2020).

We investigate 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 potentially 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 valence. 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.[12] Despite theory and evidence from mostly lab settings that positively stereotyped group members may change their academic expectations and orienta-

[12] 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)).

tion 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 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*.

Bertrand, M., & Pan, J. (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics*, *5*(1), 32–64.

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.

Diprete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*, 1–15.

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.

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.

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

Hill, A., & Jones, D. (2017). Self-Fulfilling Prophecies in the Classroom. *Working Paper*.

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. (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.

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.

Owens, J. (2016). Early Childhood Behavior Problems and the Gender Gap in Educational Attainment in the United States. *Sociology of Education*, *89*(3), 236–258.

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.

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

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

APPENDIX

A Empirical Strategy and Exposure to Any Asian Student

Table A1: Variation in Exposure to Any Asian Student

	Presence of an Asian Student	
	(1)	(2)
White-Black Math Achievement Gap	0.010***	0.004
	(0.003)	(0.003)
White-Black Reading Achievement Gap	0.007**	0.003
	(0.003)	(0.003)
White-Hispanic Math Achievement Gap	0.016***	0.003
	(0.003)	(0.003)
White-Hispanic Reading Achievement Gap	-0.004	-0.000
	(0.003)	(0.002)
Teacher-school-course FE	N	Y
N	322,648	313,676

*** p<0.01, ** p<0.05, * p<0.1. Class-level sample includes grades 4-8, 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/ethnic groups. All specifications include indicators for teacher experience. Standard errors are clustered at the teacher level.

B Racial Differentials in Teacher Assessments: Robustness Checks

We begin by showing racial differentials in teacher assessments by subject. Table B1 separately examines the propensity of over-rating and under-rating for math and reading. As in the pooled sample, teachers are more likely to favor Asian students relative to observationally comparable white peers in both subjects. Coefficients for reading are sometimes larger, perhaps due to the relatively subjective content.

Table B1: Racial Differentials in Teacher Assessments, by Subject

	Math		Reading	
	Over-rate ($B > NB$)	Under-rate ($B < NB$)	Over-rate ($B > NB$)	Under-rate ($B < NB$)
Asian	0.027***	-0.021***	0.042***	-0.030***
	(0.002)	(0.002)	(0.003)	(0.002)
Black	-0.019***	0.014***	-0.036***	0.030***
	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.021***	0.020***	-0.027***	0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.021***	0.006**	-0.026***	0.016***
	(0.003)	(0.003)	(0.005)	(0.003)
Other race	-0.003**	0.002	-0.011***	0.009***
	(0.001)	(0.001)	(0.002)	(0.001)
N	5,353,129	6,343,566	7,033,369	7,491,540

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students.

Table B2 displays the list of disciplinary infractions a student may have been reported for, as well as the frequency with which they occur in the data. 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. School districts are required by the state to report all incidents that result in an out-of-school suspension, referral to an alternative school or program, or an expulsion. However, many schools also report more minor incidents as well. We include all incidents reported in our analyses using disciplinary infraction controls. Differences in type of incidents reported across schools are not a concern since identification of our results comes from within-classroom variation.

Table B3 shows results for analyses using additional proxies for underlying behavioral characteristics of students. Columns (1) and (3) replicate the main specification, which includes categori-

Table B2: 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
Posession 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. Note: we restrict this list to the 50 most frequently occurring infraction types in the data.

cal variables controlling for the number of days a student was absent, using the set of observations for which we observe lagged days absent from the prior year. Columns (2) and (4) augment this specification with categorical variables controlling for lagged days absent from the prior year. Columns (3) and (6) further augment this specification with a set of controls for 50 different types of disciplinary infractions. Coefficients barely move with the inclusion of additional behavioral proxies.

Table B3: Robustness check: Augmented behavioral controls

	(1) Over-rate $(B > NB)$	(2) Over-rate $(B > NB)$	(3) Over-rate $(B > NB)$	(4) Under-rate $(B < NB)$	(5) Under-rate $(B < NB)$	(6) Under-rate $(B < NB)$
Asian	0.039***	0.036***	0.035***	-0.027***	-0.025***	-0.025***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Black	-0.029***	-0.031***	-0.028***	0.023***	0.024***	0.021***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.024***	-0.026***	-0.026***	0.020***	0.022***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.024***	-0.024***	-0.023***	0.012***	0.012***	0.011***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Other	-0.008***	-0.008***	-0.007***	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
Disciplinary Infraction Controls	N	N	Y	N	N	Y
N	11,832,973	11,832,936	11,832,933	13,240,924	13,240,911	13,240,896

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. Each disciplinary infraction is measured as a categorical variable that accounts for the number of infractions of that type that the student accrued over the year.

Table B4 considers the possibility that teachers may have higher or lower standards for mastery than the test score-based achievement levels adopted statewide. To ensure better comparability of the non-blind and blind achievement scales, we construct an adjusted EOG achievement level distribution where the number of students belonging to each of the four levels in a given class matches the number of students given the corresponding level by their teacher. The first two columns replicates findings in Table 5, while Columns (3) and (4) reconstruct the over-rating and under-rating outcomes using the adjusted achievement level distribution. Notably, the coefficients on Asian students are the same as before, suggesting that racial differences are not explained by the comparability of achievement scales.

Table B4: Robustness Check: Adjust No. of Students in Each EOG Level to Match Teacher Ratings

	(1) Over-rate $(B > NB)$	(2) Under-rate $(B < NB)$	(3) Over-rate $(B > NB)$	(4) Under-rate $(B < NB)$
Asian	0.037***	-0.026***	0.037***	-0.026***
	(0.002)	(0.001)	(0.002)	(0.001)
Black	-0.028***	0.023***	-0.026***	0.021***
	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.024***	0.020***	-0.028***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.024***	0.011***	-0.017***	0.015***
	(0.003)	(0.002)	(0.003)	(0.002)
Other race	-0.008***	0.006***	-0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Adjusted Ach. Level Distribution	N	N	Y	Y
N	12,386,507	13,835,115	11,858,966	14,972,325

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. 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. All models furthermore include class fixed effects, observable student characteristics, and factor variable controls for raw end-of-grade test (interacted with subject) scores. Student characteristics include gender, days absent, and an indicator for economic disadvantage.

Table B5 analyzes how the propensity for teachers to over-rate or under-rate Asian students varies by gender. Results indicate teachers are 0.8 percentage points more likely to over-rate Asian girls relative to Asian boys. They are 0.9 percentage points more likely to under-rate Asian girls relative to Asian boys. Other minority groups and their interactions with gender are included in the regression in Table B5 as well, although they are not displayed in the table.

Table B6 analyzes how the propensity for teachers to over-rate or under-rate Asian students varies by whether or not the student's primary home language is English. Results indicate teachers are 1.7 percentage points more likely to over-rate Asian students who have English as a primary home language compared to Asians who report a language other than English as their primary home language. Conversely, teachers are 1.0 percentage points less likely to under-rate Asian students who have English as a primary home language compared to Asians who report a language other than English as their primary home language. Other minority groups and their interactions with English home language are included in the regression in Table B6 as well, although they are

Table B5: Robustness check: Gender Heterogeneity in Racial Differentials in Teacher assessments

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Asian	0.033***	-0.031***
	(0.003)	(0.002)
Asian×Female	0.008**	0.009***
	(0.004)	(0.002)
Female	0.035***	-0.032***
	(0.001)	(0.000)
N	12,386,506	13,835,103

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. Other minority groups and their interactions with gender are included in the regression, although they are not displayed in the table.

not displayed in the table.

Table B6: Robustness check: Restrict to Students who Report English Home Language

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Asian	0.032***	-0.021***
	(0.003)	(0.002)
Asian×English	0.017***	-0.010***
	(0.004)	(0.003)
English	0.014***	-0.005***
	(0.001)	(0.001)
N	12,386,507	13,835,115

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. Other minority groups and their interactions with English home language are included in the regression, although they are not displayed in the table.

Table B7 looks at how the inclusion of lagged end-of-grade standardized test controls (in addition to contemporaneous test controls) affects estimates of the propensity for teachers to over-rate or under-rate Asians, relative to white students. Columns (1) and (2) replicate the main baseline specification in Table 5 on the sample of students for whom we observe lagged test scores.[13] Columns (3) and (4) augment the baseline specification with lagged student test score controls. Both lagged and contemporaneous test scores are controlled for flexibly using indicator variables.

[13]Notable, standardized end-of-grade testing does not begin for students until third grade, so we do not observe lagged scores for any third graders in the sample.

Results indicate teachers' increased propensity to over-rate Asians and decrease propensity to under-rate Asians is not driven by initial differences in achievement at the beginning of the year. For the subset of students for whom we observe lagged scores, teachers are 4.5 percentage points more likely to over-rate an Asian student without lagged score controls and 6.0 percentage points more likely to over-rate an Asian student with lagged score controls. Teachers are 2.8 percentage points less likely to under-rate an Asian student without lagged score controls and 3.4 percentage points less likely to under-rate an Asian student with lagged score controls.

Table B7: Robustness check: Include Both Lagged and Contemporaneous Achievement Scores

	(1) Over-rate $(B > NB)$	(2) Under-rate $(B < NB)$	(3) Over-rate $(B > NB)$	(4) Under-rate $(B < NB)$
Asian	0.045***	-0.028***	0.060***	-0.034***
	(0.002)	(0.001)	(0.002)	(0.001)
Black	-0.031***	0.023***	-0.014***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.025***	0.019***	-0.009***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
American Indian	-0.025***	0.016***	-0.016***	0.010***
	(0.004)	(0.002)	(0.003)	(0.002)
Other race	-0.009***	0.006***	-0.003*	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Controls for current scores	Y	Y	Y	Y
Controls for lagged scores	N	N	Y	Y
N	9,775,926	11,095,737	9,775,922	11,095,735

*** p<0.01, ** p<0.05, * p<0.1. SE clustered at teacher level. Omitted category: white students. There are slightly fewer observations in specifications that control for lagged scores because a few singletons get dropped.

C Additional Subgroup Analyses

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 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 C1 shows analysis 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 increase the propensity that a teacher will over-rate an Asian student by 0.6 percentage points, and 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. 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.

Table C1: Racial Differentials in Teacher Assessments by ACS Asian Subgroup

	Over-rate ($B > NB$)	Under-rate ($B < NB$)
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 teacher level. Omitted category: white students. Note: 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 category is proportion of Southeast Asians.

Table C2: 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.

Table C3: American Community Survey Ancestry Code Classification

Subroup	**Ancestry Codes**
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

Classification of counties into subgroups shares based on ACS self-reported ancestry.

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. This parallels analyses on the consequences of having an Asian student present in the classroom. In Table D1, the binary variable of interest takes on a value of one if there is a Black student in the class. We furthermore include classroom fixed effects and race interacted with teacher-school-course fixed effects. Thus, identification comes from variation in how a given teacher (for a particular school and course combination) assesses students of different races in the same classroom based on whether a Black student is present. 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 9.

Table D1: Effect of Exposure to One Black Student

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Hispanic×Any Black	0.003	-0.004
	(0.004)	(0.003)
Class FE	Y	Y
Race×teacher-school-course FE	Y	Y
N	2,900,330	3,925,208

*** p<0.01, ** p<0.05, * p<0.1. Sample includes grades 3-8 from 2007-2013, and is limited to classrooms that have either zero or one Black student. All specifications furthermore include interactions between race and teacher experience, and controls for raw test scores, gender, economic disadvantage, and absences. Models also include interactions between students of other racial or ethnic groups with the Any Black variable. Standard errors are clustered at the teacher level.

Next, Table D2 examines whether the effect of Black students on teacher biases toward Hispanic students varies based on the achievement level 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, the null effects do not appear to change based on the student's academic achievement.

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

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Hispanic×Any Black	0.003	-0.001
	(0.005)	(0.004)
Hispanic×Black Lagged Z-score	-0.001	0.002
	(0.004)	(0.003)
Class FE	Y	Y
Race×teacher-school-course FE	Y	Y
N	2,252,614	3,115,827

*** p<0.01, ** p<0.05, * p<0.1. Sample includes grades 4-8 from 2007-2013, and is limited to classrooms that have either zero or one Black student. Any Black is a binary variable, while Lagged Z-score is the Black student's standardized lagged score. Models also include interactions between students of other racial or ethnic groups with the Any Black and Black Lagged Z-score variables. The coefficients on the interactions between race variables and Any Black are interpreted as the difference in teacher propensities to overrate or underrate students in this racial group relative to white students who are also exposed to the same average Black student. The coefficients on the interactions between race variables and Black Lagged Z-score are interpreted as the difference in teacher propensities to overrate or underrate students in this racial group relative to white students who are also exposed to the same Black student, for each one-s.d. increase in the Black student's achievement. All specifications include controls for raw test scores, gender, economic disadvantage, absences, and interactions between race and years of teacher experience. Standard errors are clustered at the teacher level.

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 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 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. An increase in one standard deviation of the Hispanic student in the class decreases the propensity for a teacher to over-rate Black students in

Table D3: Effect of Exposure to One Hispanic Student

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Black×Any Hispanic	0.002	-0.000
	(0.002)	(0.002)
Class FE	Y	Y
Race×teacher-school-course FE	Y	Y
N	5,083,235	6,240,139

*** p<0.01, ** p<0.05, * p<0.1. Sample includes grades 3-8 from 2007-2013, and is limited to classrooms that have either zero or one Hispanic student. All specifications furthermore include interactions between race and teacher experience, and controls for raw test scores, gender, economic disadvantage, and absences. Models also include interactions between students of other racial or ethnic groups with the Any Hispanic variable. Standard errors are clustered at the teacher level.

the class by 0.5 percentage points. There are no significant effects of having a Hispanic student in the class on teachers' propensities to under-rate Black students.

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

	Over-rate $(B > NB)$	Under-rate $(B < NB)$
Black×Any Hispanic	0.002	0.001
	(0.003)	(0.002)
Black×Hispanic Lagged Z-score	-0.005**	0.001
	(0.002)	(0.002)
Class FE	Y	Y
Race×teacher-school-course FE	Y	Y
N	4,002,942	4,991,746

*** p<0.01, ** p<0.05, * p<0.1. Sample includes grades 4-8 from 2007-2013, and is limited to classrooms that have either zero or one Hispanic student. Any Hispanic is a binary variable, while Lagged Z-score is the Hispanic student's standardized lagged score. Models also include interactions between students of other racial or ethnic groups with the Any Hispanic and Hispanic Lagged Z-score variables. The coefficients on the interactions between race variables and Any Hispanic are interpreted as the difference in teacher propensities to overrate or underrate students in this racial group relative to white students who are also exposed to the same average Hispanic student. The coefficients on the interactions between race variables and Hispanic Lagged Z-score are interpreted as the difference in teacher propensities to overrate or underrate students in this racial group relative to white students who are also exposed to the same Hispanic student, for each one-s.d. increase in the Hispanic student's achievement. All specifications include controls for raw test scores, gender, economic disadvantage, absences, and interactions between race and years of teacher experience. Standard errors are clustered at the teacher level.