American Economic Journal: Applied Economics 2015, 7(4): 37–52
http://dx.doi.org/10.1257/app.20140352

Racial Discrimination in Grading: Evidence from Brazil†

By Fernando Botelho, Ricardo A. Madeira, and Marcos A. Rangel*

We investigate whether racial discrimination in the form of biased assessment of students is prevalent within Brazilian schools. Evidence is drawn from unique administrative data pertaining to eighth-grade students and educators. Holding constant performance in blindly-scored tests of proficiency and behavioral traits we find that blacks have lower teacher-assigned math grades than their white classmates. Heterogeneity in differentials provides evidence both of robustness with respect to omission biases and of compatibility with predictions from models of statistical discrimination. (*JEL* I21, I24, J15, O15)

Indicators of racial tolerance coexist with persistent differences between whites and blacks in terms of wages and schooling in Brazil.[1] While there have been important recent advancements towards the closing of racial gaps in primary-school enrollment brought by social policies that targeted a deprived portion of the population (among which nonwhites were overrepresented), black-white differentials in high school access and graduation remain stark.[2]

In the present article we propose discrimination within schools as a candidate explanation for racial gaps in attainment. We examine its prevalence by focusing on a very specific and relatively understudied aspect: a teacher's biased evaluation of students with respect to their scholastic proficiency and aptitude.[3] The paper employs uniquely detailed and large administrative data from the state of São Paulo covering

* Botelho and Madeira: University of São Paulo, Av. Prof. Luciano Gualberto 908, São Paulo-SP 05508-010 (e-mail: fbotelho@usp.br, and rmadeira@usp.br); Rangel: Duke University, 302 Towerview Drive, Box 90312, Durham, NC 27708 (e-mail: marcos.rangel@duke.edu). Funds for organization of administrative data from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)-Brazilian Ministry of Education (Observatorio da Educacao FEA-USP, Project 3313) and Instituto Nacional de Estudos e Pesquisas Educacionais Anisìo Teixeira (INEP)-Brazilian Ministry of Education (Nucleo de Estudos da Educacao - Fipe) are gratefully acknowledged. We benefited from comments received at the National Bureau of Economic Research (NBER) Education Meeting, the Population Association of America (PAA) Meeting, the Northeastern Universities Development Consortium (NEUDC) Conference, the Economic Demography Workshop at PAA, and from seminar participants at Duke University, Paris School of Economics, Vanderbilt University, Princeton University, Pontifical Catholic University (PUC)-Rio de Janeiro, Escola de Economia de São Paulo-Fundacao Getulio Vargas (EESP-FGV), Stanford University, University of California-Riverside, University of Delaware, University of Illinois at Chicago (UIC), and University of São Paulo (USP). Thanks are due to Helen Ladd and Kevin Lang for providing comments on an earlier draft. The usual caveat applies.

† Go to http://dx.doi.org/10.1257/app.20140352 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] Telles (2004); Arias, Yamada, and Tejerina (2004); Watkins (2005); Perry et al. (2006); and Rangel (2014).

[2] See Madeira and Rangel (2013) for a discussion.

[3] For a review of studies outside economics see Ferguson (1998, 2003) and Dovidio et al. (1996).

approximately 277,000 eighth-graders spread across 10,600 racially-integrated public school classrooms in 2010. Our inference is based on the careful contrasting of teachers' subject-specific grades and scores from an end-of-year standardized and blindly-marked proficiency test covering the same official mathematics curriculum delivered in regular classes.

The econometric analyses show that portions of teachers' assessments in math not explained by proficiency scores are associated with pupils' racial background. This inference remains valid after both controlling for a large number of productive attributes and dealing with the incidence of measurement error on the test scores used as covariates in our estimations. Our most conservative estimates indicate that there are statistically significant underscoring of blacks relative to whites of about 0.02 of one standard deviation (corresponding to a relative grade reduction of 8 percent for the average black). The measured racial gap in promotion rates between equivalently proficient and similarly well-behaved students amounts to a 4.2 percent relative increase in the average probability of retention among blacks. In order to reproduce this latter effect, the average black student's performance in blindly-scored proficiency tests would need to be reduced by 0.13 of one standard deviation. Results are very much in line with the expected subtlety of this particular form of discrimination.

The main challenge to our inference rests on the fact that the correlation between outcomes and easily observed racial indicators can result either from imperfectly informed teachers using race as signals of ability or from the fact that such variables happen to be related to some measure of competence that is used by evaluators but not observed by the econometrician. We provide evidence on the robustness of our results by directly examining propositions from Altonji and Pierret (2001). Under statistical discrimination, the longer the pupil and teacher interact the smaller should be the role of biased priors that emphasize racial identity, and the larger should be the role of hard-to-measure signals of proficiency. This is not expected to occur under a case of omission biases.

Exercises based on these premises unveil that while racial gaps in grades are salient for black and white students attending classes with a teacher for the first time, no significant disparities are found among those that have already had classroom interactions with that instructor before eighth grade. Concomitantly, teachers that interact with students for more than one academic year are more responsive to variations in proficiency captured by the tests scores. Besides being compatible with the predictions of a statistical discrimination model, these findings clearly lend further support to our baseline results' robustness with regards to omission of student-specific characteristics.

The discussion presented here plays on three advantages of our context with respect to recent and very important contributions to the economics literature on the topic (Hanna and Linden 2012, and Lavy 2008).[4] First, while teachers' grading within experimental settings may very well reveal different discriminatory behavior due to the one-shot nature of the event (even when hypothetical biases are curbed by

[4] See also Dorsey and Colliver (1995); Figlio (2005); Shay and Jones (2006); Hinnerich, Höglin, and Johannesson (2011a, 2011b); Burgess and Greaves (2013); and Sprietsma (2013).

incentivizing schemes), we capture actions by regular teachers acting as screeners. Second, there are both weak regulation of grading and nondisclosure of information regarding blindly-scored test performance to acting parties (teachers or students) before pupils' final assessments are processed. Finally, we employ longitudinal information on the relation between teachers and pupils in order to closely examine the robustness of our findings and the potential mechanisms behind grading discrimination.

This article is organized as follows. Section I briefly describes the data we employ. Section II presents our empirical strategy. Results are presented in Section III. Section IV concludes.

I. Data

In the analyses that follow we employ administrative data from the state of São Paulo. The state's Secretary of Education collects detailed information on the universe of students and teachers in its educational system. Considering only regular students, official records indicate that enrollment corresponded to six million primary, middle, and high-school pupils in 2010. Among those in the last year of middle school (eighth grade), 67 percent were served by schools directly administered by the state authority, with the remaining share being evenly split between municipal and private institutions. Using confidential individual identifiers we merged information from four distinct sections of the Secretary's data bank: transcript records, standardized tests of proficiency, matriculation information, and teachers' allocation to classrooms

In order to examine grading discrimination we take advantage of the administrative dataset on teachers' assessments of individual students. This dataset contains detailed information regarding grades and attendance records for all eighth grade students in schools directly administered by the state's school authority. According to official guidelines, all teachers assign numeric integer grades ranging from 0 to 10, with a passing grade set at 5 points for all disciplines. Attendance is recorded in percentage points (0–100 interval). Teachers and school administrators are not given instructions on how to attribute grades as a function of a student's observed competence level beyond the guidelines imposed by their uniform school curriculum.

We merge these data with results from standardized scores in the context of São Paulo's Performance Evaluation System—(SARESP—*Sistema de Avaliacao de Rendimento do Estado de São Paulo*). Here we employ data from the tenth to the 13th editions (2007 to 2010), with 420,000 eighth grade test-takers (87.4 percent attendance rate) in the latter year. The exam is based on multiple-choice questions covering math and reading. Microdata on these tests' results were made available in the form of proficiency scores in each subject.[5] As an integral part of the testing procedures, parents, students, and teachers also answer a survey that covers socioeconomic status, demographics (including race), study habits, and evaluation practices, among other issues.

[5] These scores were computed using Item Response Theory (IRT) methods. Individual-level test results, past or current, are *never* made available to children, parents, teachers, or school principals.

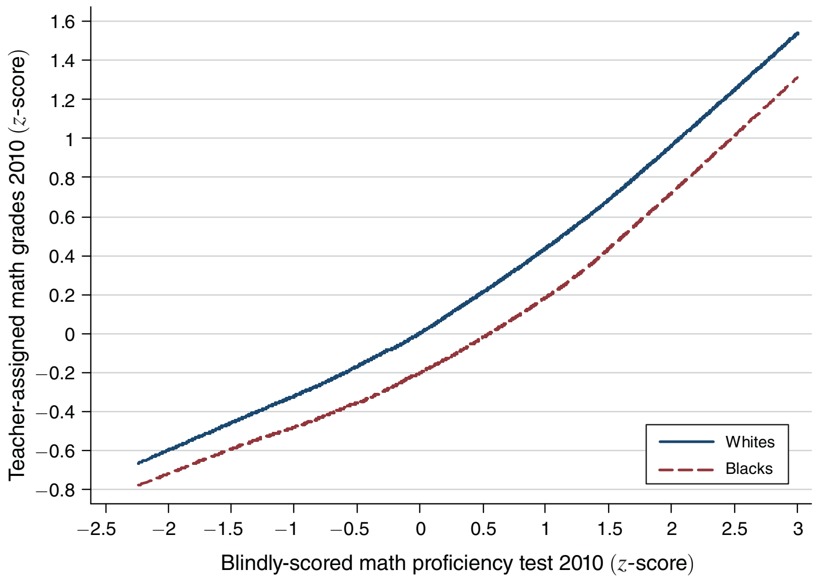

Figure 1. Smoothed Unconditional Relation between Teacher-Assigned Grades and Blindly-Scored Tests, Mathematics, Eighth Graders, 2010

We complement information of these datasets with matriculation records that allow tracking of students within the school system, and across classrooms over the years 2007 to 2012; and records of teacher allocations to classrooms for the years 2007 to 2011, containing basic demographics (race, age, gender). Combined these two datasets allow mapping of all teachers with whom each student had classes in the three years prior to eighth grade.

Our working dataset is obtained after imposing restrictions based on the availability of both transcripts and (concurrent and past) test scores data for at least 75 percent of the students in a given eighth-grade classroom at the end of 2010. We also restricted our analysis to classrooms with nonhomogeneous racial composition (at least one black and one white student) and 15 or more students. We were left with observations on 277,444 students in 10,614 classrooms across 3,511 schools. Students that self-declared as black or white are the main focus of the analysis.[6]

As expected, in pretty much every dimension in which we contrast blacks and whites within our sample (and that are later used as control variables in our analysis), the former are unfavorably compared to the latter (Table A1). In Figure 1 we plot the smoothed unconditional relationship between teacher-assigned grades and performance in blindly-scored tests in our data. This figure summarizes the main exercise of this article. For every level of test performance, blacks receive lower grades from their teachers. The econometric strategy described below is an attempt

[6] As in most exercises in the social sciences that consider race, we implicitly assume that those that discriminate (teachers/employers) and those that are discriminated (pupils/workers) agree on the racial classification captured in the records. We identify as black all students that have been declared as such in any survey or enrollment documentation between 2005 and 2012.

to verify whether these gaps are significant after we hold constant other productive attributes that make black and white students different in the eyes of their teachers and address measurement error issues.

II. Empirical Strategy

The main dependent variable in the exercises that follow is an end-of-year teacher's assigned grade in mathematics. The reason for restricting our analysis to mathematics is conceptual: we expect the material to translate itself into skills more easily measured in a test-like format. Therefore, our main covariate is expected to do a better job in capturing the proficiencies kept in check by teachers with their own screening mechanisms. In contrast, our empirical strategy is not well suited for studying language grades. This is the case because our test-based measure of proficiency is solely based on text comprehension while instructors are required to also evaluate grammar and vocabulary, among other topics.

We have converted the grades into z-scores to facilitate reading of results. In order to aid the interpretation of the practical impacts of our main results we also present estimates using two alternative binary dependent variables. The first is a cardinal measure, an indicator of minimum competence (grades above or equal to 5), while the second is an ordinal measure which we generate using an indicator function for grades above the classroom's median.

The empirical specification is based on the implicit characterization of teachers as screeners of scholastic competence. We expect teachers to acknowledge proficiency in math as well as other directly observed scholastic attributes.[7] They are seen as having access to noisy signals of a student's proficiency in math, while observing both her behavior in class and her racial identity. Implicitly assuming costs for the reduction of noise in proficiency measurements, we conceive that rational stereotyping can emerge via race being used as a signal in the computation of teachers' estimate of proficiency. In other words, the formulation of priors regarding certain group's average proficiency can find its way into the final assessment of competence.[8]

Therefore, we propose the following empirical representation that incorporates teacher/classroom fixed effects (η_r) and a pupil-level disturbance term (ϵ_{ir}):

$$g_{ir} = \alpha_1 f(scores_{ir}) + \mathbf{x}'_{ir}\,\alpha_{21} + \mathbf{z}'_{ir}\,\alpha_{22} + \mathbf{b}'_{ir}\,\alpha_3 + \eta_r + \epsilon_{ir}, \tag{1}$$

where $f(scores_{ir})$ is a function test performance available in our data that stands for the level of proficiency captured in teacher-designed examinations, and $\mathbf{b}_{ir}$ lists elements affecting teachers' priors with regard to proficiency. In order to make explicit challenges to our empirical exercise, the elements in the vector of productive scholastic attributes were also decomposed into observed and unobserved components,

[7] Mechtenberg (2009) refers to the latter as *attitudes*, which we envision as a broad concept that includes habits, styles, behavior, and any other personality trait deemed *productive* by teachers and possibly valued by parents and future employers.

[8] Ben-Zeev et al. (2014) provides interesting laboratory-based evidence of racialized recall biases. In particular, black man are remembered as lighter when subjects are offered a counter-stereotypic stimulus (regarding educational attainment).

with $\mathbf{x}_{ir}$ representing elements observed both by teachers and the econometrician and $\mathbf{z}_{ir}$ standing for those only observed by the former. Given that our central objective is to consistently estimate α_3, this simple empirical representation highlights the two main threats to the internal validity of our econometric estimations: (a) measurement error in proficiency scores, and; (b) unobserved heterogeneity.

Measurement error biases result from the fact that despite being associated to the proficiency measured by teachers, our measure is necessarily noisier. An easy way to understand the discrepancy between the two is to consider that while teachers draw observations from multiple and heterogeneous tests, the econometrician only observes results from one of them. Measurement error in test scores will affect our estimates of α_3 due to the multiple-regression's mechanics. We therefore explore the fact that the individual results of blindly-scored standardized tests taken in previous years by each student are available in our data and employ a fixed effects instrumental variables estimation to deal with the measurement error problem at hand.[9]

Unobserved heterogeneity adds another layer of complications because elements of $\mathbf{b}_{ir}$ may very well be related to elements of $\mathbf{z}_{ir}$. In particular, we worry about competencies not captured by proficiency tests, such as behavioral indicators that are available to teachers during classroom interactions and are correlated with racial identity.[10] We take this very seriously and, in the exercises below, consider a number of proxies for behavior in an attempt to check the sensitivity of our results. We have explored information correlated with behavior from different sources such as: past grades (lagged dependent variable) and an indicator of retention in the previous year; teacher attendance records, assuming those missing more classes are disengaged or rowdy when attending (we use attendance to language classes to avoid feedback effects); parent-reported perceptions of student engagement, behavior, and effort in school-related activities; student self-reported indicators of class absence and procrastination with homework; and physical education (PE) grades.[11]

Ultimately, our main empirical model consists of regressing grades on race, gender, age, parental sociodemographics, and our proxies for behavior. $f(scores_{ir})$ is introduced as a fourth-order polynomial on math scores, a linear term for reading scores, and interactions between those.[12]

[9] Since we have also obtained access to past proficiency tests covering natural sciences' material, we are able to perform overidentification tests.

[10] It would also be problematic if blacks were to take standardized tests more seriously than in-class examinations relative to their white counterparts. This stereotype-threat-like argument is indeed valid, but one for which we do not have a direct empirical implication to be tested using our data. Such caveat also plagues the whole literature on racial and gender gaps in test scores/grades. See Cornwell, Mustard, and Van Parys (2013) for unobserved heterogeneity challenges in the context of gender differentials in grading.

[11] PE grades are under the responsibility of a different teacher. Athletic equipment and infrastructure, such as fields and tracks, are not available in most schools, and students usually perform simple calisthenics and routines during classes. In eighth grade, for instance, one can hardly argue that grades are assigned as a function of athletic skills. Instead, other traits often valued by teachers, such as obedience, respect for the other students, and the capacity to respond to simple commands, are likely more relevant.

[12] The use of either splines or indicator variables after discretizing the scales does not alter the inferences we perform. Whenever F-tests indicated that the fourth-order elements were not significant, we opted for presenting results based on a more parsimonious third-order polynomial.

In order to provide further evidence of the robustness of our findings we directly explore propositions presented by Altonji and Pierret (2001).[13] Under statistical discrimination the longer pupil and teacher interact, the smaller should be the role of biased priors that emphasize racial identity and the larger should be the role of hard-to-measure signals of proficiency. There are no reasons to believe that this should occur if results are entirely due to omission biases. In practice we implement this extension by including interactions of variables that measure teacher-student acquaintance level with race indicators and (functions of) test scores.

III. Results

A. *Basic Model*

Table 1 presents results illustrating the impact of additional controls over racial differentials and over the marginal effect of proficiency scores (measured at the average-black performance level) that we estimate.[14] Group averages are presented in column 1. Considering all of the students in our sample, whites have grades averaging 0.10 while for blacks the average is -0.24 of one standard deviation. This difference is relatively unaffected by the inclusion of classroom fixed effects (column 3), indicating that racial segregation in assignment to classrooms or schools is unlikely to be behind black-white gaps. In column 4, individual demographic characteristics (gender and a second order polynomial on age) and the polynomial for contemporaneous standardized math scores are included. Measured racial gaps are, not surprisingly, significantly reduced. Indeed, a large share of the competence differences seen by teachers is captured by performance in blindly-scored standardized tests of proficiency.

In column 5 we include past year's math grade as additional control variables (lagged dependent variable), with the intention of capturing child-specific time-invariant abilities and competence aspects relevant to all teachers (past and present). Controlling for this produces a dramatic reduction both in the measured racial differentials and on the marginal impact of test scores. Family background, including maternal education, region of birth, and age; home-ownership; number of bathrooms in dwelling; and number of cars owned, are all added in column 6. Proxies for a child's behavioral attributes (self-reported, parent-reported, school-reported), over and above those indirectly captured by past grade and family socioeconomic background, are included in column 7. Despite their significance in explaining grades, the inclusion of both family background and behavioral aspects has minimal impact on our estimates of racial gaps, suggesting that at this point very little is left out of the model.[15]

[13] Tests of learning in the context of statistical discrimination can be seen in Lundberg and Startz (2007), Autor and Scarborough (2008), Lange (2007), List (2004), and Farber and Gibbons (1996).

[14] The sequential inclusion of controls should not be taken as representative of the influence they exert over the gaps we want to measure. See Gelbach (2009) for a methodological discussion.

[15] F-tests are shown for covariates included in the models. An inspection of the direct effects of these behavioral traits indicates significant results that go in the expected direction. Holding performance in tests and sociodemographics constant, math grades improve (and significantly do so) when a child attends a higher proportion of classes, when she gets higher grades in physical education, when parents report her as dedicated to and motivated

Table 1—Unconditional and Conditional Racial Differentials in Math Grades (z-scores)—OLS Estimations

		Raw black-white gaps	—	Conditional black-white gaps			
	Averages (1)	gaps (2)	FE (3)	FE (4)	FE (5)	FE (6)	FE (7)
White	0.10						
Black	−0.24	−0.338***	−0.296***	−0.094***	−0.045***	−0.043***	−0.040***
		(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)
Proficiency in math (z-score)				0.208***	0.113***	0.111***	0.099***
				(0.004)	(0.003)	(0.003)	(0.003)
Controls (*f-stats* [*p-value*])							
Classroom fixed effects		No	Yes	Yes	Yes	Yes	Yes
Child demographics		No	No	1,541.26	517.00	478.63	439.99
				[0.00]	[0.00]	[0.00]	[0.00]
Performance in standardized tests		No	No	4,966.81	1,871.98	1,837.96	1,677.03
				[0.00]	[0.00]	[0.00]	[0.00]
Past grade in math (2009)		No	No	No	34,196.52	32,888.87	20,002.51
					[0.00]	[0.00]	[0.00]
Family background		No	No	No	No	148.04	20.58
						[0.00]	[0.00]
Behavioral traits		No	No	No	No	No	1,318.06
							[0.00]
Order of polynomial on scores				4th	4th	4th	4th

Notes: Standard errors in parentheses are clustered at the classroom level. Sample consists of 277,444 students in 10,614 classrooms. Marginal effect of proficiency scores evaluated at the mean proficiency level (for the black population) are presented. Controls consist of classroom fixed effects, child's gender and age polynomial (second order), a fourth-order (third-order) polynomial function of concurrent math z-scores interacted with reading z-scores. Family background includes maternal education, age, region of birth (in or out of state), home ownership, ownership of automobiles, and number of bathrooms in the household. Behavioral traits include reports of parents regarding child's interest for school work, effort regarding studies and overall behavior. They also include physical education grades and language classes attendance rates for the first half of the school year, and an indicator for retention in the past year. Finally, self-reported measures of behavior are included with indicators of procrastination with homework, class-skipping, and interest in extracurricular math activities.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

In Table 2 (columns 4 and 5) we tackle the problem of measurement error on the proficiency score variables. As discussed above, because these are used as covariates in our analysis, biases on the estimation of *all* parameters are expected. We therefore employ polynomials of lagged blindly-scored test scores (resulting from tests taken in the most recent school year prior to the current one) as instrumental variables. Reflecting the cumulative nature of proficiency exams, past scores are strongly correlated with current ones. Moreover, overidentification tests suggest we have no obvious reason to distrust the validity of the sets of instruments employed and, therefore, also hint to the absence of unobserved heterogeneity issues.[16] Once

with school work and, ultimately, when she herself declares not to procrastinate on finishing homework. These estimates are available upon request.

[16] Of course, Hansen's test remains only a necessary and not a sufficient condition for exogeneity. See first-stage summary statistics in the online Appendix Table WA1.

TABLE 2—UNCONDITIONAL AND CONDITIONAL RACIAL DIFFERENTIALS IN MATH GRADES (z-SCORES)—OLS AND IV ESTIMATIONS

						Alternative dependent variable	
	Averages (1)	Raw gaps wrt whites (2)	OLS-FE (3)	IV-FE (4)	IV-FE (5)	IV-FE 1{Proficient} (6)	IV-FE 1{Above median} (7)
White	0.10						
Black	−0.24	−0.338***	−0.040***	−0.020***	−0.020***	−0.568***	−1.137***
		(0.007)	(0.004)	(0.005)	(0.005)	(0.205)	(0.274)
Proficiency in math			0.099***	0.526***	0.498***	4.209***	28.776***
(z-score)			(0.003)	(0.165)	(0.029)	(1.102)	(1.576)
Over-ID test				0.694	0.672	0.396	1.032
(j-statistic [p-value])				[0.4048]	[0.4124]	[0.5289]	[0.3096]
Controls (f-*stats* [p-*value*])							
Classroom fixed effects		No	Yes	Yes	Yes	Yes	Yes
Child demographics		No	439.99	1,672.07	1,680.62	529.13	884.13
			[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Performance in standardized tests		No	1,677.03	11,539.81	11,554.66	2,041.47	7,791.16
			[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Past grade in math (2009)		No	20,002.51	8,137.89	8,708.05	1,020.63	6,525.9
			[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Family background		No	20.58	225.52	226.11	122.9	106.2
			[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Behavioral traits		No	1,318.06	18,474.02	18,595.35	5,307.87	12,148.66
			[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Order of polynomial on scores			4th	4th	3rd	3rd	3rd

Notes: Standard errors in parentheses are clustered at the classroom level. Sample consists of 277,444 students in 10,614 classrooms. Marginal effect of proficiency scores evaluated at the mean proficiency level (for the black population) are presented.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

measurement error is accounted for, we encounter smaller racial differentials and at the same time larger slope parameters in the relation between math grades and math test scores. The racial gaps are still statistically significant after this correction even if we employ the more stringent Schwarz criterion.[17]

We find that blacks' average math grades are 0.02 of one standard deviation below those of equally proficient and well-behaved whites. This amounts to 6 percent of the unconditional gaps or to a relative grade reduction of 8 percent for the average black. By taking the ratio of estimated coefficients we see that the black-white differentials in teacher-assigned grades are also equivalent to a marginal reduction of 0.04 of one standard deviation in proficiency scores. These effects are very much in line with the subtleties we expect to permeate racial discrimination in grading.

Columns 6 and 7 reproduce these exercises with a focus on meaningful binary variables that summarize cardinal and ordinal aspects of the gaps in grades, respectively.

[17] One may argue that some of our control variables are the result of grading discrimination in their own right, inducing models to underestimate the size of black-white gaps. We see merit in such argumentation, since biased grading within the school year can indeed induce students to "misbehave," but prefer to be as conservative as possible in our empirical exercises. We restrict the analysis that follows to the use of a fully controlled model.

Table 3—Conditional Racial Differentials in Math Grades (z-scores) by Subgroups Based on Teachers' Characteristics—IV Estimations

	Black-white gap	Proficiency in math (z-score)	Over-ID test j-statistic [p-value]	Sample of classrooms	Sample of students
Panel A. Checking for sample selection issues					
Full sample	−0.020***	0.498***	0.672	10,614	277,444
	(0.005)	(0.029)	[0.4124]		
Teachers responding to questionnaire	−0.019***	0.484***	1.321	8,925	233,750
	(0.005)	(0.031)	[0.2504]		
Panel B. Stratification by evaluation methods					
Objective teacher	−0.023***	0.461***	0.118	3,305	86,485
	(0.009)	(0.058)	[0.7312]		
Subjective teacher	−0.021***	0.446***	0.369	6,548	171,727
	(0.006)	(0.036)	[0.5436]		
Panel C. Stratification by teachers' race					
White teacher	−0.020***	0.499***	2.189	7,153	187,717
	(0.006)	(0.033)	[0.1390]		
Black/mixed teacher	−0.016	0.412***	0.132	1,772	46,031
	(0.012)	(0.080)	[0.7167]		
Panel D. Stratification by teachers' experience					
Older teacher	−0.021***	0.446***	0.976	5,775	151,377
	(0.006)	(0.037)	[0.3231]		
Younger teacher	−0.014	0.557***	0.636	3,150	82,373
	(0.009)	(0.055)	[0.4253]		
Panel E. Stratification by teachers' tenure in school					
Long-tenure teacher	−0.011*	0.521***	0.095	6,078	160,261
	(0.006)	(0.034)	[0.7575]		
Short-tenure teacher	−0.036***	0.367***	1.925	2,894	74,673
	(0.010)	(0.067)	[0.1653]		

Notes: Standard errors in parentheses are clustered at the classroom level. Objective and subjective teacher classifications come from survey responses in which teachers indicate their opinion about the importance of different evaluation methods. Teachers' race comes from survey's self-reported race. Older and younger are defined using top and bottom third of the age distribution of teachers in the system. Long-tenure are teachers with at least three years of school experience while short-tenure are those in the school for no more than two years. See notes in Table 1.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

According to these estimates, the measured racial gap in promotion rates between equivalently proficient and well-behaved students corresponds to a 4.2 percent increase in the eighth-grade retention probability for the average black (or 0.6 percentage points in absolute terms). Focusing exclusively on the ordinality aspect we also estimate a gap that translates into a 1.1 percentage point (or 4.3 percent) reduction on the probability of blacks being graded above the classroom median.

B. *Robustness*

We explore expected heterogeneity in the size of racial differentials and its relation to some teacher characteristics to further examine the robustness of our findings to the omission of behavioral characteristics. In Table 3, before moving into the comparison across different data strata, in panel A we present a summary of the main effects under the full sample and under the subsample for which we have additional teacher characteristics (from survey questionnaires). The contrast between these indicate that we should not necessarily expect selection biases when dealing with the smaller sample.

In the first set of stratifications (panel B) we examine if the gaps in evaluation we measure are not generated by unobserved heterogeneity biases (associated with student behavior) captured in the subjectivity of a teacher's evaluation method. We explore a section of the questionnaire answered by teachers in the context of SARESP, in which opinions regarding the importance of objective instruments of evaluation (tests and exams) and also the importance of using more observational methods (classroom behavior, students' motivation, oral examinations, etc.) were gathered. We find no evidence that these groups discriminate against blacks with different intensities.

Panels C to E are solely based on teacher demographics and work experience. We reestimate our model for different strata according to teacher's race, age, and school tenure. We cannot rule that point estimates for racial gaps are the same across groups. The exception is tenure: among teachers working in a given school for more years racial gaps in grading are smaller while the elasticity to standardized tests is larger. These latter findings motivate a more stringent test of the role of information about students over the racial gaps we estimate.

The robustness of our basic formulation to omission biases is further examined using a theory-based test of statistical discrimination put forth by Altonji and Pierret (2001). We explore data on pupil-teacher matches by utilizing the longitudinal information on students' and teachers' assignment to classrooms. We map the individual-level acquaintance level between every student and their current teacher. In this case longer interactions should increase signal to noise ratios, increasing the marginal effect of (posterior) proficiency measures at the same time it reduces the one related to characteristics used to construct priors (race).

It is clear from estimates in Table 4 (panel A) that longer-term teacher-student interactions produce smaller (and insignificant) grading gaps associated with racial identity.[18] In other words, this empirical exercise reveals that while black-white gaps in grades and rankings are salient for students attending classes with a teacher for the first time, no significant disparities are found among those that have already had classroom interactions with that instructor before eighth grade. It is also the case that acquaintance of teacher and students increases the weight given to proficiency scores on the determination of grades. Both these differences (in intercept and in slope) are shown as statistically significant in column 3. In practice, black students that have not interacted with their current teacher before eighth grade have their grades diminished by what is equivalent to a reduction of 0.05 of one standard deviation in the proficiency tests' performance. The gap is zero for those known to the teacher. This is our main indication that omission biases cannot be the main driving force behind our results. It also suggests that imperfect information lies at the heart of the discrimination results we estimate.[19]

[18] Notice that, to the extent these are individual-level variables, these models explore within-classroom/within-acquaintance variation in a way that emulates a difference-in-differences formulation.

[19] Evidence in favor of this learning argument can also be seen in different strata, as reproduced in the online Appendix Table WA2).

TABLE 4—CONDITIONAL RACIAL DIFFERENTIALS IN MATH GRADES (z-SCORES) AND LEARNING VIA TEACHER-STUDENT INTERACTIONS—IV ESTIMATIONS

	Teacher interacting with students over multiple academic years (1)	Teacher interacting with students for the first academic year (2)	Difference (3)=(2)−(1)
Panel A. Learning by grading using duration of math teacher-student interactions			
Black-white gap in math grades	−0.005	−0.024***	−0.019*
	(0.010)	(0.006)	(0.012)
Proficiency in math (z-score)	0.603***	0.461***	−0.142**
	(0.053)	(0.035)	(0.064)
Panel B. Falsification of learning by grading using duration of language teacher-student interactions			
Black-white gap in math grades	−0.020*	−0.021***	−0.001
	(0.011)	(0.006)	(0.012)
Proficiency in math (z-score)	0.494***	0.500***	0.006
	(0.049)	(0.035)	(0.060)
Panel C. Falsification of learning by grading using duration of future math teacher-student interactions			
Black-white gap in math grades	−0.024	−0.019***	0.005
	(0.015)	(0.005)	(0.016)
Proficiency in math (z-score)	0.546***	0.484***	−0.062
	(0.143)	(0.029)	(0.147)

Notes: Standard errors in parentheses are clustered at the classroom level. Math teachers are classified as having previous interactions with a given student if assigned to the student's past classrooms between 2007 and 2009 (panel A). The same is used in defining past interactions with language teachers (panel B). Future interaction is defined by students having the same math teacher in 2011 as they had in 2010 (panel C). See additional notes in Table 1.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

The robustness of our learning argument can be further put to the test by examining an alternative explanation for such findings. In particular, since classroom fixed effects are expected to deal with teacher-level omitted characteristics, we focus on investigating if the assignment of teachers to students captured in our proposed measure of knowledge is not simply revealing that black-white differences in students' (omitted) behavioral characteristics are smaller among those selected into longer-term interactions. Results in panel B of Table 4 strongly reject both this threat by presenting evidence that neither *math*-grade racial gap nor its relation to proficiency are a function interaction time between a student and her *language* teacher.[20] We reach the exact same conclusion when we employ the identity of future math teachers to measure acquaintance levels. Teachers that *will* spend more time with a given student do not discriminate more or less today than those that will not (Table 4, panel C). Alternatively, students that will spend more time interacting with their current math teachers in the future do not have their racial identity playing a role on evaluations that is different than for those that will not.[21]

[20] It is important to emphasize that in our context the group of students is common for math and language teachers, so if there were a particular rule for allocating students to the same teacher year after year this should be true for both subjects.

[21] Even though it is implicit in these exercises, for completeness, we present descriptive statistics of racial gaps in (observed) characteristics of students that do and do not spend more than one year interacting with a given

We conclude our analysis by verifying that our finding regarding learning and consequent reduction in racial gaps are not a result of the omission (at the level of the interaction effect) of other characteristics in our econometric specification. In particular, we examine if by including interactions between teacher-student relation indicators and other control variables we are able to eliminate the differences observed in the race coefficient. We see no reason to believe this is the case (online Appendix, Table WA5).

Taken together, our auxiliary findings clearly substantiate the robustness of our basic results to unobserved heterogeneity biases.

IV. Conclusions

In this article, we empirically detect racial discrimination within racially integrated eighth grade public school classrooms in Brazil. Math teachers' assessments of students with respect to scholastic competence are found to be biased. White students are less likely to be deemed noncompetent (below passing grade) than their equally proficient and equivalently well-behaved black classmates. The racial gap we estimate is equivalent to approximately one-third of the raw (within-classroom) difference in grades associated with having a mother with a college degree or more versus a mother with a high school degree only. These results are shown robust to possible omissions of a students' behavioral attributes and to the incidence of measurement error on scores from standardized tests.

We also find that these racial biases most likely result from imperfect information and statistical discrimination or, in other words, from the weighted combination of noisy proficiency signals extracted from teacher-designed exams and stereotyped priors. In the context explored here, rational stereotyping may have resulted from lenient standards for admission of students into eighth grade (which have disproportionally benefited blacks) embedded in a social promotion scheme adopted by São Paulo's schools. It turns out that well-intentioned teachers issue report cards for their students with subtle biases (possibly incurred when rounding continuous marks into a discrete scale) and, in this way, may end up adding obstacles to the acquisition educational credentials by blacks.

The results presented here also suggest that well-designed randomized control trials focusing on the amount, type, and timing of information about individual students available to teachers can go a long distance in helping us better understand the inner workings of grading discrimination within schools. We leave these for our future research on such topics.

teacher in Table WA3 of the online Appendix. We see no indication that racial gaps are dramatically different across these strata. We also show in Table WA4 that controlling for an interacted term on the estimated propensity to have repeated teacher-student interactions (as a function of teacher and students' observed characteristics) does not alter these results.

Table A1—Descriptive Statistics

	White mean (SE) (1)	Black mean (SE) (2)	Black-white gap diff. (SE) (3)	Black-white gap (FE) diff. (SE) (4)
Grades and tests				
Math test 2010 (z-score)	0.10	−0.24	−0.338	−0.296
	(0.005)	(0.007)	(0.007)	(0.006)
Passed math in 2010 (0–100)	91.58	86.33	−5.246	−4.922
	(0.123)	(0.239)	(0.225)	(0.214)
Above class math median in 2010 (0–100)	39.64	26.15	−13.486	−13.510
	(0.149)	(0.270)	(0.304)	(0.310)
Math test 2009 (z-score)	0.11	−0.24	−0.346	−0.301
	(0.005)	(0.007)	(0.007)	(0.006)
Blind test in math 2010 (z-score)	0.36	0.00	−0.354	−0.249
	(0.005)	(0.006)	(0.007)	(0.006)
Blind test in reading 2010 (z-score)	0.15	−0.29	−0.445	−0.343
	(0.005)	(0.006)	(0.007)	(0.006)
Previously taken blind test in math (0–500)	217.76	202.00	−15.756	−11.521
	(0.180)	(0.242)	(0.259)	(0.250)
Previously taken blind test in reading (0–500)	215.87	197.43	−18.435	−14.018
	(0.192)	(0.262)	(0.275)	(0.265)
2008 blind test in sciences (0–500)	238.24	216.27	−21.965	−16.501
	(0.233)	(0.321)	(0.339)	(0.325)
Demographics				
Boy	0.48	0.61	0.135	0.134
	(0.001)	(0.003)	(0.003)	(0.003)
Age in months (centered at pop. mean)	−0.04	0.10	0.138	0.128
	(0.002)	(0.005)	(0.005)	(0.005)
Family background				
Mom no schooling	0.02	0.03	0.015	0.013
	(0.001)	(0.001)	(0.001)	(0.001)
Mom completed high school	0.22	0.17	−0.051	−0.037
	(0.002)	(0.002)	(0.003)	(0.003)
Mom college dropout	0.03	0.02	−0.006	−0.003
	(0.001)	(0.001)	(0.001)	(0.001)
Mom college graduate	0.04	0.03	−0.014	−0.008
	(0.001)	(0.001)	(0.001)	(0.001)
Home ownership	0.54	0.51	−0.035	−0.030
	(0.002)	(0.004)	(0.003)	(0.003)
Autos in household	0.58	0.41	−0.177	−0.119
	(0.003)	(0.004)	(0.005)	(0.004)
Bathrooms in dwelling	1.07	0.90	−0.171	−0.111
	(0.004)	(0.006)	(0.006)	(0.005)

(*Continued*)

Table A1—Descriptive Statistics (*Continued*)

	White mean (SE) (1)	Black mean (SE) (2)	Black-white gap diff. (SE) (3)	Black-white gap (FE) diff. (SE) (4)
Behavioral traits (proxies)				
Retained in 8th grade in 2009, 0–1	0.01 (0.001)	0.01 (0.001)	0.008 (0.001)	0.006 (0.001)
Well behaved (0–1, parents' report)	0.43 (0.002)	0.32 (0.003)	−0.110 (0.003)	−0.092 (0.003)
Poor behavior (0–1, parents' report)	0.05 (0.001)	0.08 (0.002)	0.028 (0.002)	0.027 (0.002)
High effort (0–1, parents' report)	0.16 (0.001)	0.13 (0.002)	−0.026 (0.002)	−0.023 (0.002)
Low effort (0–1, parents' report)	0.15 (0.001)	0.18 (0.002)	0.028 (0.003)	0.024 (0.003)
Level of interest (0–10, parents' report)	5.89 (0.020)	5.32 (0.027)	−0.576 (0.025)	−0.391 (0.022)
PE grade z-score (1st bimonthly eval.)	0.07 (0.007)	−0.11 (0.008)	−0.176 (0.007)	−0.086 (0.006)
PE grade z-score (2nd bimonthly eval.)	0.07 (0.006)	−0.11 (0.008)	−0.179 (0.007)	−0.089 (0.006)
Attendance (0–100, 1st bimonthly eval.)	91.70 (0.052)	90.64 (0.077)	−1.066 (0.069)	−0.528 (0.058)
Attendance (0–100, 2nd bimonthly eval.)	89.54 (0.049)	88.32 (0.080)	−1.220 (0.076)	−0.714 (0.069)
Skip classes (0–1, self-report)	0.69 (0.002)	0.64 (0.003)	−0.053 (0.003)	−0.031 (0.003)
Not late with homework (0–1, self-report)	0.25 (0.002)	0.18 (0.002)	−0.067 (0.003)	−0.054 (0.003)
Participation in math Olympics (0–1, self-report)	0.45 (0.003)	0.37 (0.003)	−0.075 (0.003)	−0.045 (0.003)

Notes: Standard errors in parentheses are clustered at the classroom level. Sample consists of 123,362 white students and 28,163 black students.

REFERENCES

Altonji, Joseph, and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116 (1): 313–50.

Arias, Omar, Gustavo Yamada, and Luis Tejerina. 2004. "Education, Family Background and Racial Earnings Inequality in Brazil." *International Journal of Manpower* 25 (3–4): 355–74.

Autor, David H., and David Scarborough. 2008. "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments." *Quarterly Journal of Economics* 123 (1): 219–77.

Ben-Zeev, Avi, Tara C. Dennehy, Robin I. Goodrich, Branden S. Kolarik, and Mark W. Geisler. 2014. "When an 'Educated' Black Man Becomes Lighter in the Mind's Eye: Evidence for a Skin Tone Memory Bias." *SAGE Open* 4 (1): 1–9.

Botelho, Fernando, Ricardo A. Madeira, and Marcos A. Rangel. 2015. "Racial Discrimination in Grading: Evidence from Brazil: Dataset." *American Economic Journal: Applied Economics.* http://dx.doi.org/10.1257/app.20140352.

Burgess, Simon, and Ellen Greaves. 2013. "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities." *Journal of Labor Economics* 31 (3): 535–76.

Cornwell, Christopher, David B. Mustard, and Jessica Van Parys. 2013. "Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School." *Journal of Human Resources* 48 (1): 236–64.

Dorsey, J. Kevin, and Jerry A. Colliver. 1995. "Effect of Anonymous Test Grading on Passing Rates as Related to Gender and Race." *Academic Medicine* 70 (4): 321–23.

Dovidio, John F., John C. Brigham, Blair T. Johnson, and Samuel L. Gaertner. 1996. "Stereotyping, Prejudice, and Discrimination: Another Look." In *Stereotypes and Stereotyping,* edited by C. Neil Macrae, Charles Stangor, and Miles Hewstone, 276–322. New York: Guilford Press.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics* 111 (4): 1007–47.

Ferguson, Ronald. 1998. "Can Schools Narrow the Black-White Score Gap?" In *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips, 318–74. Washington, DC: Brookings Institution.

Ferguson, Ronald. 2003. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." *Urban Education* 38 (4): 460–507.

Figlio, David. 2005. "Names, Expectations and the Black-White Test Score Gap." National Bureau of Economic Research (NBER) Working Paper 11195.

Gelbach, Jonah. 2009. "When Do Covariates Matter? And Which Ones, and How Much?" University of Arizona Working Paper 09-07.

Hanna, Rema N., and Leigh L. Linden. 2012. "Discrimination in Grading." *American Economic Journal: Economic Policy* 4 (4): 146–68.

Hinnerich, Bjorn, Erik Höglin, and Magnus Johannesson. 2011a. "Are Boys Discriminated in Swedish High Schools?" *Economics of Education Review* 30 (4): 682–90.

Hinnerich, Bjorn, Erik Höglin, and Magnus Johannesson. 2011b. "Ethnic Discrimination in High School Grading: Evidence from a Field Experiment." Economic Research Institute, Stockholm School of Economics (SSE/EFI) Working Paper Series in Economics and Finance 733.

Lange, Fabian. 2007. "The Speed of Employer Learning." *Journal of Labor Economics* 25 (1): 1–35.

Lavy, Victor. 2008. "Do Gender Steriotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92 (10–11): 2083–2105.

List, John A. 2004. "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field." *Quarterly Journal of Economics* 119 (1): 49–89.

Lundberg, Shelly, and Richard Startz. 2007. "Information and racial exclusion." *Journal of Population Economics* 20 (3): 621–42.

Madeira, Ricardo, and Marcos A. Rangel. 2013. "Racial Achievement Gaps in Another America: Discussing Schooling Outcomes and Affirmative Action in Brazil." In *Closing the Achievement Gap from an International Perspective,* edited by Julia Clarke, 127–62. New York: Springer.

Metchemberg, Lydia. 2009. "Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages." *Review of Economic Studies* 76 (4): 1431–59.

Perry, Guillermo, Omar Arias, J. Humberto López, William F. Maloney, and Luis Servén. 2006. *Poverty Reduction and Growth: Virtuous and Vicious Circles.* Washington, DC: World Bank.

Rangel, Marcos A. 2014. "Is Parental Love Colorblind? Human Capital Accumulation within Mixed Families." *Review of Black Political Economy* 42 (1–2): 57–86.

Shay, S., and B. Jones. 2006. "Anonymous Examination Marking at University of Cape Town: The Quest for an 'Agonising-Free Zone.'" *South Africa Journal of Higher Education* 20 (4): 528–46.

Sprietsma, Maresa. 2013. "Discrimination in grading: Experimental evidence from primary school teachers." *Empirical Economics* 45 (1): 523–38.

Telles, Edward E. 2004. *Race in Another America: The Significance of Skin Color in Brazil.* Princeton: Princeton University Press.

Watkins, Kevin. 2005. *Brazil: National Human Development Report.* New York: United Nations Development Programme (UNDP).