Discrimination Goes to School*

Fernando Botelho
Department of Economics, USP

Ricardo Madeira
Department of Economics, USP

Marcos A. Rangel
Department of Economics, USP and
PLAS, Princeton University

January, 2014

Abstract

We investigate the extent to which racial discrimination, in the form of the biased assessment of students, is prevalent within Brazilian schools. Robust evidence is drawn from unique data pertaining to middle-school students and educators. We find that even after holding constant performance in blindly scored official tests of proficiency, teacher-assigned Math grades suffer from bias. Relative to an equally proficient White counterpart, a Black eighth-grader is less likely both to be promoted to high-school (cardinal impact) and to be graded above her classroom-specific average (ordinal impact). These findings suggest that schools may be imposing additional obstacles to the acquisition of productive skills and educational credentials by Blacks. By further detailing the heterogeneity in these differentials, we unveil indications that they result from information asymmetries highlighted in models of dynamic statistical discrimination. Implications for policy are discussed in light of our main findings.

*Funding from CAPES-Brazilian Ministry of Education (*Observatorio da Educacao FEA-USP, Project 3313*) and INEP-Brazilian Ministry of Education (*Nucleo de Estudos da Educacao - Fipe*) are gratefully acknowledged. Rangel also acknowledges funding from CNPq-Brazilian Ministry of Science and Technology (Research Productivity Grant). We benefited from comments received at the NBER Education Meeting, the PAA Meetings, the NEUDC Conference, the Economic Demography Workshop at PAA, and from seminar participants at Duke University, Paris School of Economics, Peabody College-Vanderbilt University, Princeton University, and University of Sao Paulo. Thanks are also due to Kevin Lang for detailed comments on an earlier draft. The paper has been screened to ensure no confidential information is revealed. Errors, omissions, and ideas expressed in the text are the sole responsibility of the authors. These do not represent the views of funding agencies or of the *Secretaria de Educacao do Estado de Sao Paulo* (SEE-SP). Previous versions of this article have circulated under the title: "Discrimination Goes to School? Racial Differences in Performance Assessments by Teachers." For the most recent version please access www.marcos-a-rangel.com.
Please address comments to corresponding author: M. A. Rangel, mrangel@princeton.edu.

"[...] Any instructional regime that is responsive to the needs of individual children and flexible enough not to place ceilings on their progress is likely to benefit all children and, by the very nature of what it means to be responsive, to enhance the opportunities for those who start behind to catch up."
Ronald F. Ferguson (1998) in "*Can Schools Narrow the Test Score Gap?*"

1 Introduction

Evidence of a negative association between individual characteristics used to infer African ancestry and educational attainment abounds.[1] Equally notorious is the resilience of achievement gaps across cohorts of Black and White children (Neal, 2006). These are further emphasized by longitudinal studies showing that Black disadvantages emerge during infancy and remain pretty much intact while children attend school.[2] Because evidence regarding racial differences in *actual* returns to human capital accumulation is scant, a better understanding of obstacles to the acquisition of skills and educational credentials by Blacks seems warranted.

Here, we elect discrimination within racially-integrated schools as a candidate explanation for the patterns described above and subsequently examine its prevalence in Brazil. We recognize that such a phenomenon may manifest itself in many different ways within a classroom. Yet, we focus on one that seems less elusive: a teacher's biased evaluation of students with respect to their scholastic proficiency and aptitude (grading). We employ uniquely detailed data from the state of Sao Paulo covering approximately 350 thousand eighth-graders spread across 11.5 thousand public-school classrooms in 2010. Our inference is based on the implicit comparison between teachers' subject-specific grades and scores from end-of-year standardized (and blindly marked) tests of proficiency covering the same official curriculum delivered in regular classes. We show that portions of the teachers' assessments in Mathematics not explained by proficiency scores are significantly associated with the pupils' race. Our most conservative estimates indicate that there are statistically significant underscoring and under-ranking of Blacks relative to Whites. We find that White students are 4% less likely to be deemed non-competent (below passing grade) than their equally proficient Black

[1] Data portraying such historically-rooted patterns have been drawn from different countries and under a variety of institutional settings. For comparative international studies see Alexander et al. (2001); Herring et al. (2004); Telles (2004); and Telles and Steele (2012).

[2] See Phillips et al. (1998); Hedges and Nowell (1999); Reardon (2008); and Madeira and Rangel (2012). Cautionary notes on these findings can be found in Bond and Lang (2012).

classmates. They are also 5% relatively more likely to be graded above their classroom median. In practice, these effects are equivalent on average to "taxing" Blacks' performance in proficiency tests by 0.04 to 0.08 of one standard deviation. These results are shown to be robust to possible omissions of a students' behavioral attributes and to the incidence of measurement error on scores from standardized tests.

Once the existence of racial gaps in assessments is established, we rely on economic theory to examine why it is the case in our context. We draw from a rich literature on statistical discrimination,[3] but most of our reasoning comes from the recent and insightful dynamic statistical discrimination framework in Lehmann (2011). We map our setting into her study of affirmative action's effect on hiring and promotion rates in the labor market for lawyers by focusing on two main institutional aspects. First, teachers are limited by imperfect technology in the process of scholastic competence's measurement, but once assigned to students of a given level (whose admission is decided by a third party), are still solely responsible for promotion and ranking decisions. Second, due to a number of policies implemented since the late 1990s, a dramatic increase in access to education has been observed. We highlight in particular the adoption of *social promotion* schemes between the fifth and seventh grades. In practice, such policy has disproportionally benefited Blacks (who are over-represented among pupils with lower proficiency) by establishing lenient standards for the admission of students into eighth grade. In other words, social promotion has emulated affirmative action within the Brazilian schools we study. Eighth-grade teachers are well aware of the implications of such policy, and priors regarding Black students' competence may have been downgraded as a result. Therefore, when (well-intentioned) teachers issue report cards for their students, subtle biases are generated (say, when rounding continuous marks into a discrete scale) by the weighted combination of noisy information extracted from exams and stereotyped priors.

We then present evidence on the validity of our theoretical reasoning. Employing a strategy similar in spirit to the one in Altonji and Pierret (2001), we examine whether the duration of interaction between teachers and students produces different assessment patterns. The basic idea is that the longer pupil and teacher interact, the smaller is the role of biased priors that emphasize racial identity. In this regard, our empirical exercises unveil that while gaps in promotion rates

[3] Aigner and Cain (1977); Borjas and Goldberg (1978); Lundberg and Startz (1983); Coate and Loury (1993); Cornell and Welch (1996); Altonji and Pierret (2001); Blume (2006); and Bjerk (2008).

and ranking 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 the instructor before eighth grade. Teachers seem to learn about a student's true "type" over academic years or once they are fully aware of promotion standards previously used.

The implications of these findings are far reaching, and may go beyond level promotion and the relative ranking of students. This is the case because we detect discrimination during the transition between middle and high-schools, at a time when Brazilian parents invariably find themselves in the position of investors relying on the asset-return evaluations of more informed experts. For our purposes, the key element of this reasoning is that teacher communications may steer investment decisions in one way or the other.[4] That is to say; parents (and children themselves) likely update investment (and effort) decisions after extracting information from report cards issued by teachers. Therefore, if children's perceived ability increases the returns or reduces the costs of investments, as in the traditional Beckerian human-capital framework, this mechanism can reinforce racial gaps in the accumulation of human capital. In this case, intra-classroom evaluation biases may very well feed back into attainment, school choice, future scholastic performance and, ultimately, labor market outcomes.[5]

We also believe our results shed new light on the effects of affirmative action on access to education and accumulation of human capital, a theme of prime importance as Brazil adopts racial quotas in access to college and in allocation of publicly-funded scholarships. While the profession has focused on behavioral responses among those favored by such polices (the effort choices of high-schoolers that are granted easier access to college), we advocate that the role of instructors within colleges be considered.[6] As an illustration of our argument, take Arcidiacono et al. (2013). The authors examine a rich data set and argue that affirmative action in admissions to the University of California system led to a mismatch between minority-students' abilities and program requirements in the most selective UC *campi*. This mismatch, in turn, explains the low graduation rates among individuals favored by quotas. We believe that an equally valid argument would be that instructors' priors were affected by the enactment of such policies (stereotyping). Because GPA and course

[4] Lam et al. (2006) examines the effect of performance measurement's precision over high-school dropout behavior in South Africa, for example.

[5] See Mechtenberg (2009) for a formalization of an argument like this.

[6] See Assuncao and Ferman (2013) on the early Brazilian experience with quotas, and Cortes and Zhang (2012) for a discussion in the context of the Top 10% Program in Texas.

performance are intimately connected with drop-out and graduation rates, such policy may have indeed imposed ceilings on the progress of the population it was designed to help by distorting instructors' priors (and therefore subjectively assigned course grades).[7] If this mechanism is at work, Brazil should expect college graduation rates among Blacks to still lag behind those observed among equivalently proficient Whites. This is the first policy conclusion we draw: affirmative action in college admissions may have negative impacts over the population it was designed to help when negatively influencing subjective evaluations of college-level materials' mastery.

Considering the role played by misinformation in the results presented here, and beyond its scientific interest, we also envision three other policy lessons being derived from our analysis. First, curbing teacher rotation can be particularly important for Black students (over and beyond any effect on learning *per se*) because increasing interactions between a group of students and a given teacher diminishes the influence of noise on the evaluation of scholastic proficiency. The more a teacher gets acquainted with a given student, the less relevant the pupil's race becomes for evaluation purposes. Second, direct investment in teacher training with regard to the design of exams and tests may be warranted. Well-designed questions are easier to grade and more likely to differentiate students on the most relevant dimensions of proficiency. Finally, because blindly graded proficiency tests are regularly taken by students under standard "school accountability" systems, and despite the intrinsic noisy nature of such scores, there is no strong reason not to use them to generate individual report cards that could aid teachers in their competence evaluations. Particularly under social promotion schemes like the one we study, this additional information should make teachers better able to evaluate their students without resorting to racially biased priors.

The remainder of this article is organized as follows. Section 2 briefly reviews the literature on teacher perceptions and discrimination. Section 3 discusses the institutional background and describes the data we employ. Section 4 outlines a conceptual framework that guides the empirical analysis we perform. Section 5 presents our empirical strategy and the econometric identification strategy. Results and discussions are presented in Section 6. Section 7 concludes the article.

[7]In many ways this is similar to the original argument in Coate and Loury (1993).

2 Related literature

Despite being the first study to examine racial bias using Brazilian student-level data in this degree of detail, we are well aware that the question of whether teachers treat Black and White children differently is not new. In fact, there is a tradition within the sociology literature of directly examining whether teacher bias is a factor in course-grade assignment in the United States (Bowles and Gintis, 1976; Farkas et al., 1990; Rist, 1973; Rosenthal and Jacobson, 1968; Sexton, 1961). Both large- (Sewell and Hauser, 1980; Williams, 1976) and small- (Leiter and Brown, 1985; Natriello and Dornbusch, 1984) scale empirical studies tend to detect insignificant biases. There is also a considerable number of contributions from the social psychology literature focusing on teacher's perceptions of Black and White children (see Ferguson, 1998, 2003 and references therein), which again only unveils weak relationships between Black stereotypes and measures of discriminatory actions.[8]

Our work complements more recent studies from the education and economics literatures. Shay and Jones (2006) and Dorsey and Colliver (1995), examine the quasi-experimental variation provided by institution-level policy changes regarding anonymity in the grading processes applied to college/graduate students in South Africa and the state of Illinois, respectively. No significant racial differentials were observed. However, these articles do not examine how the blind and non-blind evaluation of the same students are related. Figlio (2005) steps in this direction by examining whether teachers' overall perception of a given student (i.e.: gifted, proficient) is affected by the "Blackness" of her first name, even after controlling for performance in standardized examinations. Using data from a school district in Florida, the author uncovers evidence of lower teacher expectations for those perceived to have African American ancestry.

Another important study detecting discrimination in grading is the one reported by Lavy (2008). The author capitalizes on a natural experiment in Israeli high-schools. He cleverly explores the fact that students take two different examinations that cover the same material during their senior year, and that the grading of each exam happens under different anonymity regimes. Focusing on gender differentials, his findings indicate that male students receive lower marks in the non-blindly graded exams (relative to those blindly scored), and that these differences are larger (in absolute value)

[8]See review of studies in Dovidio et al (1996). Demeis and Turner (1978), unlike most of this literature, find significant discrimination against Blacks in an experimental setting.

than among girls. Blind/non-blind contrasts are also skillfully explored in a randomized control trial designed and implemented by Hanna and Linden (2012). The authors identify small and statistically significant positive differences between blind and non-blind scores for members of lower castes in India (relative to upper castes), which is clear evidence of discrimination.

The discussion presented here plays on three major advantages of our context with respect to other studies in the literature. First, the sheer size of and level of detail in our data base allows us to convey a complete portrait of teacher and student-body characteristics associated with discrimination in actual classroom environments. Teachers grading in 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 incentivizing schemes). Regular teachers, however, feel like entitled gatekeepers, responsible for assigning credentials that will follow a child for life. Second, our study explores both the cardinal and ordinal aspects of discrimination in grading. While acquisition of school credentials is associated with the former (i.e.: passing grade), ordinal features may be particularly important in either school-to-work transitions or high-school admissions that require teacher referrals. In addition, when we consider a smaller reference group, even minor changes (relative to the overall distribution of performance) may have practical importance, as we expect classrooms to be more homogeneous than the population. Finally, and unlike Lavy (2008), we have weak regulation of grading and non-disclosed information regarding standardized test performance to acting parties (teachers or students) before pupils' final assessments are processed. In this way, we explore an environment in which: i) subtle discriminatory behavior is hardly detected by school authorities, and ii) last minute reactions to performance information are not sought by evaluators or by those being evaluated.

3 Data and institutional background

3.1 Student-level data

The Sao Paulo's Secretary of Education has agreed to share with the authors, under cooperation and confidentiality agreements, detailed information on the universe of students and teachers in the state's educational system. Considering only regular primary and secondary schools, official records indicate that enrollment corresponded to approximately 6 million students in 2010. Among eighth-graders, 67% were served by schools directly administered by the state authority, with the remaining share being split between municipal and private institutions. Using confidential individual identifiers we merged information from four distinct sections of the Secretary's data bank: matriculation information, standardized tests of proficiency, transcript records, and teachers' allocation to classrooms.[9] We turn to the description of each one of these.

Matriculation information covers all schools in the state of Sao Paulo, be they private or public. These records are centralized by the Secretary of Education though its role as a regulating agency for private and municipal schools. The centralized matriculation system exists as a way to avoid having parents matriculate their children in more than one school. In the recent past this practice has led to children not being absorbed by the system (as some had taken two or three slots). Matriculation within the public system is also defined in terms of a school's catchment area (districting). Parents apply for a slot and pupils are assigned to the school serving the requested level closest to their residence. The centralization of information offers interesting ways of tracking student mobility within the school system, especially in the case of dropout and migration between or within public and private systems.

Standardized scores were collected in the context of Sao Paulo's Performance Evaluation System - (SARESP- *Sistema de Avaliação de Rendimento do Estado de São Paulo*). The system consists of an annual statewide exam taken by public school students in grades 2 and 4 (elementary school), 6 and 8 (middle school), and 11 (high school). It is directly administered by the state-level authority and has been applied in slightly different formats since 1996. Here, we employ data from its 13th edition (2010), with over 1.5 million test-takers in approximately 5,050 schools. Of this total, 420 thousand were eighth-graders (87.4% attendance rate in this particular level). 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, teaching and pedagogical practices, and perceptions about the school environment, among other issues.

The main purpose of the SARESP exam is to measure the students' proficiency on the subjects assigned to each specific grade/cycle according to a predetermined curriculum, which is imposed on schools by the state authority. The exam in 2010 had three sets of questions covering Math,

[9]The Secretary has never attempted to combine these data. There are different departments in charge of each of these sections, and communication between them is scant. This is the first time these data have been used in an integrated format.

Portuguese language, and Science. For students in grade 8, each exam contained 30 multiple choice questions. Students also produced a written essay. The exams were taken in late November (Spring), close to the end of the academic year during regular-class meeting times in the same classrooms in which students sit for lectures. Students took the exams in two consecutive days. Grading was electronically conducted for the multiple-choice questions with students using a test sheet, which was scanned. After the exam sheets were assigned numerical codes, the essays were graded by a central committee of educators with no direct relation to the students in the system. The State Secretary of Education hired an independent institution that prepared the exam according to predetermined guidelines. To supervise students during the test, teachers from different schools and levels were mobilized, such that students were overseen by an unfamiliar teacher. External observers were also assigned to each school to guarantee the strict obedience of all protocols.

Microdata on these tests' results were made available in the form of proficiency scores in each subject. These scores were computed using Item Response Theory (IRT) methods. Scores were also converted into a (grade-subject-specific) four-step classification system that reflected educators' consensus regarding levels of proficiency (below basic, basic, sufficient, more than sufficient) after the statistical definition of anchor items. Proficiency in the essay portion of the Language exam was reported in identical (yet independent) four-level scales covering each of the four different dimensions of writing ability: theme (ability to keep the text within the proposed theme); vocabulary and pronoun-noun concordance; cohesion and coherence (text organization); and syntax and subject-verb/time concordance. Importantly, individual-level results from SARESP are *never* made publicly available to children, parents, or schools.

We also take advantage of the administrative data set on teacher student assessments. This data set contains detailed information regarding scores and attendance records for all students in schools directly administered by the state's school authority. This is the exact same information delivered to parents every couple of months in the form of report cards. The complete set of report cards available to us includes information on every school subject. In the primary school these correspond to Language (Portuguese), Mathematics, Social Studies, Sciences, Physical Education, and the Arts. In eighth grade, teachers' subject-specialization is complete.

These data resulted from the adoption of an uniform criterion-referenced rule in September 2007. According to such guidelines, all teachers attribute numeric integer grades ranging from 0 to 10, with a passing grade set at 5 points for all disciplines. Attendance in turn is recorded in a 0-100 interval. Interestingly, teachers and school administrators are not given instructions on how to attribute grades as a function of a student's observed proficiency level beyond the guidelines imposed by their uniform school curriculum. The state administration provides pedagogical material and teachers are supposed to evaluate students according to proficiency in its content. Nonetheless, no explicit guidance regarding the design of evaluations (except for questions included at the back of the teacher's booklet) is given, and teachers still have complete autonomy to define evaluation criteria and methods and to allocate students across the 11 grading categories.

The final data set employed in our analysis comes from the records of teacher allocations to classrooms for the years 2007, 2008, 2009, and 2010. These files contain basic demographics (race, age, gender) for all the teachers in the system, and can be linked longitudinally. Combined with the matriculation records, we are able to map all Math teachers with which each student had classes in the three years prior to eighth grade (which, in absence of repetition, corresponds to the entire middle-school cycle). We discuss below how this information can provide important insights into the nature of racial discrimination in grading.

3.2 Racial gaps in Brazil

The discussion of racial differentials in Brazil is somewhat paradoxical. On the one hand, widespread racial mixing in marriage and the desegregation of housing markets have helped spread the view of a Brazilian "haven of racial reconciliation and affinity" (see Richman, 1999). On the other hand, there is overwhelming evidence that such racial tolerance indicators coexist with pertinent differences between Whites and non-Whites (Blacks or Browns) in terms of wages and other measures of economic well-being (see Arias et al., 2004; Campante et al., 2004; and Perry et al., 2006). In fact, the 2005 Human Development Report (United Nations) states that racial difference in economic achievement is one of the main social challenges facing Brazil. The report goes on to suggest that anti-discrimination policies should be central to any poverty reduction program implemented in the country. According to the 2010 Brazilian population census, adult male Whites have 8.4 years of completed education while the corresponding quantity for Blacks is 6.4 years. This lower educational attainment goes hand in hand with log-wage gaps of approximately 0.40 points. These gaps are of equal size when we restrict the sample to the state of Sao Paulo, which is the geographic area of focus for our analysis.

Largely in response to educational attainment differentials and to views such as the one expressed in the United Nations' document, federal and state governments across the country have implemented racial quotas for admission to public universities and in the provision of college-scholarships. Yet, as we discuss below, the most important recent advancements in the closing of racial gaps seem to have indeed come about as a result of *colorblind* social policies.

3.3 Shrinking gaps in attainment

Starting in the mid-1990s and under more favorable macroeconomic conditions, Brazilian policy makers have attacked problems with access to formal education. Both demand- and supply-side initiatives began to be undertaken, including the early steps and expansion of *Bolsa Familia*'s conditional cash-transfer program, and innovations in the allocation of federal budget toward school maintenance and teacher salaries under the *Fundo de Manutencao e Desenvolvimento do Ensino Fundamental (FUNDEF)*. Under this new institutional setting, standard educational policy targets rapidly improved. There was, for example, an unprecedented and significant increase in the rates of enrollment of school-aged children all over the country.

Using data from repeated cross-sections of the Brazilian Household Survey (PNAD), we reproduce in Figure A1 trends in enrollment for children aged 6 or 7 in the state of Sao Paulo between 1989 and 2009 by race (Whites and Blacks). Aggregate enrollment figures went from somewhere around 75% in 1990 to more than 95% (or nearly universal coverage) by 2010. Importantly, from a racial perspective this *democratization* of access to schooling had a major influence on the composition of the student body, increasing the participation of a deprived portion of the population (among which Blacks were overrepresented). In essence, Black-White gaps in enrollment among young children have virtually been eliminated in the state by the end of the period we study.

The absence of racial gaps in initial enrollment does not imply the closing of attainment gaps, however. For that to be the case, we also need retention and drop-out rates to be converging. For the country as a whole, the evidence on this dimension is mixed. In the case of Sao Paulo the patterns seem more favorable because starting in 1998 the state's public school system adopted a social promotion scheme. This policy grouped contiguous primary school grades into two cycles, with retention only occurring at the end of each of them. Cycle 1 encompasses grades 1 to 4 (elementary) and cycle 2 covers grades 5 to 8 (middle school). Under this regulation, a student is promoted to the next level within a cycle if she attends more than 75% of the classes (and has no record of extreme disciplinary problems), irrespective of her mastery of the material covered during the academic year. Insufficient performance can only result in retention at the end of each cycle. In this case, the pupil is supposed to repeat the last grade within that cycle.[10]

We conjecture (but do not directly examine) that social promotion in Sao Paulo is at least in part responsible for a faster convergence in education attainment between Blacks and Whites. Trends are more pronounced than in other parts of the country, and the timing of convergence coincides with policy adoption. Yet what most substantiates this argument is the comparison of year-to-year transition probabilities between middle-schools directly managed by Sao Paulo's school authority, and those run by municipal authorities. The former were all under social promotion during the 2006-2010 period we examine. Meanwhile, among the municipality schools only a small minority were under the same promotion scheme during that time. We find that while racial differences in raw attrition rates are indeed virtually nonexistent in schools that adopt social promotion, they fall in the 4 to 5 p.p. range among those that do not.[11] Here, as in the case of increased access, even if not aimed directly at racial issues, by benefiting students at the bottom of the skill distribution, social promotion had a disproportional effect on primary-school re-enrollment (higher) and retention (lower) rates among Blacks.

We keep these recent trends in racial inclusion in perspective. However, this article focuses on how they are likely to affect the experiences of Black and White children that reach the final grade of middle school, right before racial differentials in enrollment rates and attainment re-emerge among high-schoolers.

3.4 Descriptive statistics

Our working data set was obtained after imposing restrictions based on the availability of both transcripts and test scores data for at least 75% of the students in a given classroom. We also

[10]Several international organizations, including the World Bank, support this policy as an effective way to curb low grade completion and to decrease drop-out rates. The general lines of the argument are that grade retention could adversely affect some of the students' non-cognitive skills (like confidence and self-esteem), increasing anxiety levels and harming their learning capacity. See King et al. (2008).

[11]We use longitudinal matriculation records to compute these, and a description of the data is provided above. It is important to note that they are *unconditional* average transition rates.

restricted our analysis to classrooms with non-homogeneous racial composition and at least ten students. We were left with observations on 352,881 students in 11,475 classrooms. Students that self-declared as Black or White are the main focus of the analysis, but our models are estimated including (and identifying) individuals classified under other races. Table A1 presents descriptive statistics for our working data set. In pretty much every dimension in which we compare Blacks and Whites (and that are later used as control variables in our analysis), the former are in an inferior condition.

Focusing more specifically on scholastic performance, Figures A2 and A3 plot the distribution of test scores and teacher-assigned grades. These represent the main control and the main dependent variables in the econometric exercises that follow, respectively. Even with all of the observed progress in attainment, we can still find sizable differences in achievement between Blacks and Whites in Sao Paulo. For the students in our sample, differentials amount to 0.35 of one standard deviation. A similar pattern is observed in the distribution of teacher-assigned grades, with a disproportionate concentration of Blacks among those earning lower marks. Average differences in grades are approximately 0.6 in a 0-10 scale.

Finally, in Figure A4 we plot the smoothed raw relationship between teacher-assigned grades and test scores 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 and all our empirical estimations are in essence an attempt to verify whether these gaps are indeed there even after we hold constant other productive attributes that make Black and White students different in the eyes of their teachers and address some measurement error challenges. Before examining the data in more detail, we turn to the conceptual framework that guides our interpretation.

4 Conceptual framework

We focus our attention on a stylized description of grading that leads directly into our empirical specifications. The model is by no means general, but rather is used as a rhetorical device to emphasize a particular source of racial differentiation in teachers' assessments. In principle, there are two basic reasons for teachers to systematically mis-evaluate the competence of students with certain characteristics. First, teachers may merely like/dislike people with those traits, imposing rewards/punishments that can take both cardinal and ordinal forms. Second, teachers may attempt to be more sophisticated, evaluating (hard to measure) competence by also using observed characteristics perceived to be correlated with the former. In this case, the characteristics themselves convey information, and can help teachers generate better assessments of a latent "proficiency". These alternative sources of discrimination are well know in the economics literature. The first is a loose representation of taste discrimination (Becker, 1957), whereas the second falls under the realm of statistical discrimination (Arrow, 1971; Phelps, 1972; Aigner and Cain, 1977). In our model we highlight the operation of the latter.

Moreover, the conceptual framework presented here concentrates sole attention on the screening role of eighth-grade instructors, and does not feature discrimination in other dimensions of teacher-student interactions (mentoring, coaching, etc.). The basic intuition is that teachers observe noisy signals of the students' proficiency in Math, and know their behavior in class and racial identities. Due to social promotion in earlier grades, eighth-grade teachers know that a particularly lenient rule for promoting students was used. They also (correctly) assume that social promotion rules disproportionally affected promotion rates among Blacks, and will therefore require from the latter a higher signal of proficiency to deem them minimally competent or to allow them to climb classroom-level rankings.

We, therefore, start by defining an objective function for graders of school work. The model assumes that they operate as statisticians, compelled to maximize the power of the hypothesis test embedded in the evaluation of a student's competence. We impose that teachers weight Type I and Type II errors symmetrically (i.e.: excessive lenience and excessive rigor are equally unwelcome). Evaluation errors can be reduced by exerting more grading effort, something teachers are supposed to dislike, but such services can also be "purchased" in the market at a fixed piece-rate π. Schematically, teacher r chooses a grading/evaluation effort level M_r and at the end of the school year assigns to each student i (in a group of size n_r) a grade g_{ir} that best represents the expected value of her true competence (g^*_{ir}) in order to solve the following optimization problem:

$$\max_{M,g_i} E\left[\sum_{i=1}^{n} u(g_i - g_i^*)\right] - ln\,(M)\,, \tag{1}$$

$$s.t. : I + w = \pi M$$

where we omit teacher-level subscripts for clarity of exposition, and I represents teachers' exogenous income and w stands for salaries paid against the inelastically supplied unit of labor.

We further assume a simple risk-neutral quadratic function for disutility generated by evaluation errors:

$$u(g_i - g_i^*) = -\frac{1}{2}\left(g_i - g_i^*\right)^2 . \tag{2}$$

Importantly, the model allows teachers to broadly define competence. As in Mechtenberg (2009), they acknowledge true proficiency (p_i^*) and other directly observed scholastic attributes ($\mathbf{a_i}$) as elements to be rewarded.[12] That is to say:

$$g_i^* = \alpha_0 + \alpha_1 p_i^* + \mathbf{a_i'}\alpha_{\mathbf{2}} \tag{3}$$

Teachers do not observe true proficiency directly, so we further assume that they collect a sequence of noisy (yet unbiased) signals $s_i^m = p_i^* + u_i^m$. Signals result from formulating and grading tests/exams, and hence we associate them with evaluation effort ($m = 1, 2, ..., M$). The higher the effort, the more signals will be gathered about each student's proficiency. The estimation of proficiency can then be described as a combination of those signals and a prior for mean proficiency solely based on other attributes ($\mathbf{b_i}$) to compute the best linear projection ($E\left[p_i^* | s_i^1, ..., s_i^M, \mathbf{b_i}\right]$):

$$\hat{p}_i^* = \frac{\sigma_{p^*}}{\sigma_{p^*} + \sigma_{\bar{u}}}\bar{s}_i + \frac{\sigma_{\bar{u}}}{\sigma_{p^*} + \sigma_{\bar{u}}}\left[\beta_0 + \mathbf{b_i}'\beta_{\mathbf{2}}\right], \tag{4}$$

where $\bar{s}_i = \frac{\Sigma s_i^m}{M}$, $\sigma_{\bar{u}} = \frac{var(u_i^m)}{M}$ and σ_{p^*} represents the variance of actual proficiency within the student population.[13] The vector $\beta_{\mathbf{2}}$ establishes the relation between characteristics and group's average proficiency (prior's formulation).

Combining all the elements in the model, and defining $\theta = \frac{\sigma_{\bar{u}}}{\sigma_{p^*} + \sigma_{\bar{u}}}$, we reach the following optimal rule for grading:

[12] 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. Our formulation could also allow for racial bias operating directly via teachers' definition of competence (which we would recognize as taste-based discrimination, however). There is an interesting parallel between this variation and racial perception bias regarding others' pain discussed in Trawalter et al. (2012).

[13] At this point we do not take a stand on the elements shared by $\mathbf{a_i}$ and $\mathbf{b_i}$, but elaborate on it in the empirical section below.

$$g_i = (\alpha_0 + \theta\beta_0) + \alpha_1(1-\theta)\bar{s}_i + \mathbf{a_i}'\boldsymbol{\alpha_2} + \alpha_1\theta\mathbf{b_i}'\boldsymbol{\beta_2}. \tag{5}$$

The case of racial discrimination at hand can be illustrated within an example. Let $\mathbf{b_i}$ be a scalar corresponding to an indicator $Black_i$, which identifies racial identity. Assume also that this scalar is not one of the elements in $\mathbf{a_i}$. Therefore:

$$g_i = (\alpha_0 + \theta\beta_0) + \alpha_1(1-\theta)\bar{s}_i + \mathbf{a_i}'\boldsymbol{\alpha_2} + \alpha_1\theta\beta_2 Black_i \tag{6}$$

Notice that in this representation, racial bias is derived from the information about proficiency conveyed by a student's race. A prediction of the model is that improvements in the signal-extraction technology should make race a less relevant element of the grade assignment process. At the same time, the relationship between grades and individual test scores should be strengthened. This would be the case if teachers were to (exogenously) increase grading effort, if new information were distributed to teachers, or if tests were made less noisy. We take versions of this simple model to the data. Further discussions on alternative specifications and identification challenges are presented in the empirical section below.

5 Empirical strategy

The first practical challenge we face in our empirical strategy comes from the way grades are reported. A conceptual issue arises from the heterogeneity in different teachers' application of the grade scale. As in the case of comparing responses using a Likert scale, contrasting grades assigned by different teachers is not clear cut. While a classroom fixed-effect added to the regression accounts for different mean scores across classes, an issue of dispersion remains; that is, even after factoring out the class average, a one point gain in class A can hardly be compared to the same absolute gain in class B if they have different grading standards in the spread of grades. At first we simply put aside this concern and use grades as our dependent variable, but we do so recognizing that (within this scale) measured gaps have both cardinal and ordinal meanings.

To carefully examine these different aspects we focus on two alternative dependent variables. The first is the only really cardinal measure available in our data: an indicator of minimum competence.

This was made common across teachers by the central authority's establishment of a passing grade (set at 5). So, independently of a teacher's choices regarding dispersion of grades within a classroom (or her subjective understanding of one additional point in the scale), it will always be the case that those above or at grade 5 are deemed competent while those below are not. This cardinal notion ought to be common across all classrooms. In the second measure, we only consider the relative position of a student with respect to her classmates, in a metric that makes no attempt to compare students in different classes. In practice, we focus on the empirical variation captured by binary indicators of reference to both the classroom's average and median grades.

A second practical concern is the different natures of the exams applied within the school context by teachers and the standardized tests adopted for external monitoring of learning. In principle, because teachers receive a uniform curriculum from the external examiner, their evaluations should reflect the same skills and cognitive abilities as the external standardized exam. Yet, it is plausible that competence in a given content can be measured by examining performance using different tasks (format). Take the case of Language evaluations, for example. Teachers most likely combine observations regarding reading, writing, and speaking abilities when assessing a student's language competence. Paper-and-pencil standardized tests implemented in our context, however, can only capture reading skills using a multiple choice exam. This is one of the reasons for restricting our analysis to Mathematics: we expect the objectivity inherent in the material to translate itself into skills more easily measured in a test-like format. Of course, it is possible that teacher-designed Math exams also reward reading and writing skills (over and above the Math performance). For precisely that reason we include scores from these two other sections of the standardized examinations as controls in our empirical model.

In essence, and in reference to equation (5) of the conceptual framework proposed above, we explore our information regarding scores in standardized Math exams as a proxy for the average level of proficiency measured by teachers in their own classroom examinations. Meanwhile, other skills also considered relevant by teachers are factored into the *productive attributes* term ($\mathbf{a_i}$). Therefore, we propose the following empirical representation that incorporates teacher/classroom fixed-effects (η_r) and a pupil-level disturbance term (ϵ_{ir}):

$$g_{ir} = \delta_0 + \delta_1 score_{ir} + \mathbf{a_{ir}}'\boldsymbol{\delta_2} + \mathbf{b_{ir}}'\boldsymbol{\delta_3} + \eta_r + \epsilon_{ir}, \tag{7}$$

where $score_{ir}$ is the measure of test performance available in our data that replaces the "theoretical" average level of proficiency captured in teacher-designed examinations ($\bar{s}_{ir}$), and once again $\mathbf{b_{ir}}$ lists elements affecting teachers' priors with regard to proficiency.

To make explicit further challenges to our empirical exercise, the elements in the vector of scholastic attributes ($\mathbf{a_i}$) can also be decomposed into observed and unobserved components:

$$g_{ir} = \delta_0 + \delta_1 score_{ir} + \mathbf{x_{ir}}'\boldsymbol{\delta_{21}} + \mathbf{z_{ir}}'\boldsymbol{\delta_{22}} + \mathbf{b_{ir}}'\boldsymbol{\delta_3} + \eta_r + \epsilon_{ir} \tag{8}$$

where $\mathbf{x_{ir}}$ represents the elements observed both by teachers and the econometrician and $\mathbf{z_{ir}}$ stands for those only observed by the former.

Given that our central objectives reside in inferences regarding δ_1 and $\boldsymbol{\delta_3}$, this simple empirical representation highlights the two main econometric problems we face: i) measurement error in proficiency scores, and ii) unobserved heterogeneity biases.[14]

Measurement error biases result from the fact that despite being associated to the average proficiency measured by teachers, our measure is necessarily noisier and may even have a different underlying mean value. An easy way to understand the discrepancy between the two is to consider that while teachers "observe" results from multiple and heterogeneous tests, the econometrician only observes results from one of them. Those biases directly limit our ability to test the predictions from our conceptual framework. In the exercises below we explore the fact that the individual results of standardized tests taken in previous years by each student are available, and employ a fixed-effects instrumental variables estimation that should bypass the measurement error problem.

Unobserved heterogeneity adds another layer of complications because even in the absence of measurement error in scores, elements of $\mathbf{b_{ir}}$ may very well be related to elements of $\mathbf{z_{ir}}$. In particular, we worry about behavioral indicators that are available to teachers during classroom interactions and are correlated with racial identity. 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: *i*) teacher attendance records, assuming the students that miss more classes are disengaged or poorly behaved even when attending (we used attendance in the first six school-months); *ii*) parent-reported perceptions

[14] For a discussion of the effects of measurement errors and omission biases when using test scores as covariates, see Andrabi et al. (2011).

of student engagement, behavior, and effort in school-related activities; *iii*) student self-reported indicators of class absence and procrastination with homework; and *iv*) Physical Education (PE) grades (in the first semester of classes). PE grades are under the responsibility of a different teacher. Athletic equipment and infrastructures, 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, may be more relevant. Of course, some schools could organize intramural sports competitions, such that athletic traits would carry more weight in the physical education grade, but even if this were so, disciplinary traits should still be a relevant component of evaluations.

Ultimately, our main empirical model consists of the following generalized formulation (for the ease of exposition we assume $\mathbf{b_{ir}} = \mathbf{x_{ir}}$):

$$g_{ir} = \delta_0 + \delta_1 score_{ir} + \mathbf{x_{ir}}' \left[\boldsymbol{\delta_{21}} + \boldsymbol{\delta_3} \right] + \tilde{\eta}_r + \tilde{\epsilon}_{ir}, \tag{9}$$

where race, gender, age, auxiliary test scores (those from the language sections of the exam), parental socio-demographics, and our proxies for behavior are considered elements of the vector $\mathbf{x_{ir}}$ and the remaining elements of $\mathbf{z_{ir}}$ not observed by the econometrician are either absorbed by the classroom fixed-effects or by the disturbance term (considered orthogonal to covariates in the fully controlled model).

We also extend this analysis to explore the heterogeneity of the parameters in (9), according to teacher and student-body characteristics. In particular we pay attention to the amount of knowledge a given teacher has about her pupils. Tenure in a given school and duration of classroom-like interactions for a given student-teacher pair are main candidates here. In this way we examine some of the main predictions from our statistical discrimination conceptual framework. In particular, and in the spirit of Altonji and Pierret (2001), we test whether differentials in teacher-assigned grades diminish (or even disappear) as a teacher's experience with some children increases. If coefficients are shown to be sensitive to this interaction, it is an indication that statistical discrimination may be at play in the study environment.

6 Results

6.1 General results (OLS)

Our initial estimations are derived from the specification in (9). Table 1 presents the results, illustrating the effect of the addition of controls over racial differentials in our four alternative dependent variables.[15] Panel A focuses on the Black-White gaps in final grades (0-10 scale). Group averages are presented in column 1. Considering all of the students in our sample, Whites are graded at 6.1 on average while grades among Blacks average 5.5. 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 the racial gaps in our context. In column 4, individual demographic characteristics (gender and a second order polynomial on age) are included, and estimated racial gaps are reduced. This reflects the relations between age, gender and race within a given school level, with Blacks being slightly older on average and males being significantly more likely to self-declare as Blacks. Estimated gaps are further reduced as we include Math and language examination scores in columns 5 and 6, respectively, but are surely not explained away. In column 7 family background measures are included. At this point, conditional gaps are approximately 20 to 25% of the unconditional ones.

Table 2 turns to a careful examination of the potential effects of unobserved heterogeneity. In these specifications we add our proxies for a child's behavioral attributes, over and above those indirectly captured by family socio-economic background. We present results that employ a type of proxy at time (self-reported, parent-reported, school-reported) and also when all of them are included together (in column 7). An inspection of the direct effects of these behavioral aspects indicates significant results that go in the expected direction. Holding performance in tests and socio-demographics constant, Math grades improve (and significantly do so) when the child attends a higher proportion of classes, when she gets higher grades in physical education, when parents report her as dedicated to and motivated with school work and, ultimately, when she herself declares that she does not procrastinate on finishing her homework.[16] Notice, however, that in all cases there are

[15] 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.

[16] These coefficient estimates are not shown in Table 1 to preserve space, and are available upon request.

clear signs that behavioral aspects do not explain away the gaps in grading.

In Table 3 we resort to information on the past year's grades as a final control variable, with the hopes of capturing both Math abilities and behavioral aspects relevant to the teacher that were not previously captured. Despite the reduction in size, estimated gaps are still statistically significant. Indeed, they are significant even when we employ the more stringent Schwarz criterion for significance.[17] Ultimately we find that Blacks' average Math grades are 0.06 points below those of equally proficient and well-behaved Whites. This amounts to 11% of the unconditional gaps or a 1% decrease in average Black grades.[18]

The pattern described above is also seen in the other columns of Table 3, where we examine the racial gaps in minimum-competence probability and relative positioning within the classroom (above average and median grades, respectively). In all cases we find significant differences between Blacks and Whites that look the same in a myriad of dimensions that capture both proficiency and other productive attributes relevant for Math competence. According to these estimates, the chance of a Black student being deemed incompetent in Math is 0.77 percentage points higher than that for a White student. This represents a decrease of 5% in the chance of promotion (for the average Black student). The effects we capture over ordinal measures are equally sizable. Blacks are consistently and significantly less likely to be ranked at the top of their Math classes. The probability of beating the average or the median grade in their classroom is approximately 1.8 percentage points lower. In other words, the reduction in the probability of being "honored" for their Math grades amounts to 5 to 6.6% for the average Black eighth-grader in our data.

We also examine an alternative explanation for the existence of gaps in teacher-assigned grades for individuals that look identical both in the proficiency and in the behavioral dimensions. If proficiency tests measure end-of-year accumulated knowledge while teacher grades measure average performance in examinations during the year, and if Blacks and Whites have different learning curves, we can observe patterns as the ones above. In particular, this would be the case if Blacks' proficiency levels were catching up over the academic year with White's. We falsify this explanation

[17] Considering our very large sample, the Schwarz criterion, which sets critical values of significance as a function of sample sizes is indeed more appropriate to judge the significance of results.

[18] One may also argue that some of our control variables are the result of discrimination in their own right, inducing our models to underestimate the size of Black-White gaps. We see merit in such argumentation, but prefer to be as conservative as possible in our empirical exercises, restricting the analysis that follows to the use of a fully controlled model.

looking at our data in two ways. First we examine grading patterns at different points of the school year and find similar effect-sizes as the ones above when focusing on mid-term evaluations (they should be larger if the argument is correct). Second we also find that Black and White proficiency scores are not converging over time when focusing on a longitudinal sample of students followed between 6th and 8th grades.[19]

To gather a sense of the size of these effects and (possibly) the mechanisms behind grading discrimination, we explore two simple simulations. We start by converting the blindly graded proficiency scores into a classroom-specific 0-10 scale. Conversion is undertaken by: i) computing the difference between the score of student i and the minimum score in her classroom; ii) dividing this quantity by the difference between the maximum and minimum score in that classroom, and; iii) distributing this quantity in the range given by the teacher-assigned grades in that classroom. We then add a simulated discriminatory rounding routine while converting from a continuous into a discrete grade scale. This is done by assuming that every time a Black student's score lies in the $[h+0.45, h+0.54)$, where h is an integer in the 0-9 scale, the resulting grade is necessarily h (Simulation 1).[20]. In Simulation 2 we round down the grades of Blacks in the $[h+0.45, h+0.60)$ instead. White students in the exact same situation have $h+1$ as their assigned grades, following unbiased rounding rules. The econometric model above is then re-estimated using these simulated grades as dependent variables. Table 4 reports the comparison between estimated Black-White gaps under the actual and the simulated situations. Interestingly, we find that in the four dimensions of grading we examine, actual differences between Blacks and Whites are similar to (and sometimes even larger than) the ones in the simulated exercise. This is an interesting finding, as it gives us a notion that the results we find are in the ballpark of what happens when such a subtle discriminatory action is imposed. This exercise provides some confidence that the results we uncover are not ignorable, being consistent with the homogeneous employment of biased rounding by large proportion of teachers in the sample, for example.

[19] We cannot rule out, however, that the gaps we uncover result from differential levels of motivation depending on the student's race and on the nature of proficiency exams. If Blacks take standardized tests more seriously than in-class examinations relative to their White counterparts we would expect to find results like the ones above. This is indeed a possibility, but one for which we do not have a direct empirical implication to be tested using our data.

[20] Another interesting variation for this simulation would be the allowance of rounding rules that are different across the grade range, that is; for different values of h.

6.2 *Dealing with measurement error (2SLS)*

In Table 5, we tackle a different issue, examining the robustness of our findings to the presence of measurement errors on the 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 lagged values of 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 very correlated with current ones. Once 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 effects are still significant in all cases, leaving most of our qualitative conclusions unaffected. Quantitatively, Black-White differentials in teacher-assigned grades are equivalent to a reduction of 0.04 to 0.08 standard deviations in proficiency scores.[21] Still, according to these exercises, all else constant, Black students are 4% relatively less likely to be promoted than their White classmates. Whites are also 3% (5%) relatively more likely to be graded above their classroom median (average) than equally proficient and apt Blacks. Once again, these small effects are very much in line with the subtleties we expect to permeate racial discrimination in grading.

We have established the prevalence of racial gaps in assessments, but have not yet uncovered any particular supportive mechanism. To do so, we explore the existence of heterogeneity in the size of racial differentials and its relation to some teacher or student-body characteristics. In doing so we try to keep close track of the predictions from the conceptual framework presented above.

Before moving into the comparison across different data strata, we present in Table A2 a summary of the main effects under three different samples. The first reproduces what we have seen above (the full sample), and the second and third represent samples for which teacher unique-identifiers and teacher responses to questionnaires are available, respectively. By looking into the results under these subsamples, we see that no clear sample selection bias emerges.

[21] Interestingly, these effects fall in the same range as those estimated experimentally by Hanna and Linden (2012) for the case of castes in India.

6.3 Heterogeneity in racial gaps by teachers' evaluation practices

The first stratification we examine reiterates our understanding that the gaps in evaluation we measure are not generated by unobserved heterogeneity biases. To gather further insight 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. These questions were posed in an independent manner, so that there they are not excludable categories. In Table 6, we explore these responses to stratify teachers in three (not necessarily distinct) groups. Those that believe objective methods are very important, those to whom objective methods are not important, and those to whom subjective/observational methods are very important. Strikingly, we find no evidence that these groups discriminate against Blacks with different intensities. In fact, if anything, larger effects are found among those that believe in the objective evaluation of students. In our opinion, this is the first indication that imperfect information plays a central role in our findings: racial bias seems to occur more frequently among those that are trying to extract the most out of their noisy measures of proficiency.

6.4 Heterogeneity in racial gaps by students' characteristics

In Table 7 we concentrate attention in strata defined by student's characteristics. We investigate heterogeneity in racial differentials according to gender, age (relative to median age in student population) and proficiency (measured by past results in Language standardized exam). In most cases we uncover that racial differentials are larger and more likely to be significant among boys, older children and those with lower proficiency. Incidentally, together with the case of Blacks, these groups are the most likely to have been favored by social promotion, reaching eighth grade without proper mastery of materials. In our line of reasoning these results conform with the idea that especially among these subgroups, racial identification is used by teachers as a marker of lower average proficiency when formulating their priors regarding students' competence.

6.5 Heterogeneity in racial gaps by teachers' characteristics

Table 8 is solely based on teacher demographics (obtained from official assignment records). We re-estimate our model using fixed-effects instrumental variables techniques for different strata according to race (Panel A) and gender (Panel B). We see that no clear pattern emerges from these. In the case of race, which is examined here to investigate in-group biases, Black teachers are less likely to exhibit gaps in promotion rates and grades, but are as likely as White teachers to discriminate in ordinal dimensions. Together these findings are incompatible with the idea of taste discrimination (at least in its simplest format) against members of the out-group. In any case, it is hard to empirically design tests that can actually rule out the role of racist tastes in Black-White differentials, but this is also true for the whole literature on the topic.

In Panel B we see that differentials for male and female teachers are not obviously distinct either, except for the fact that female teachers drive the depressed promotion rates among Blacks. Nonetheless, this finding coexists with similar differences in terms of final grades. One way to reconcile these two elements is that grade reductions are happening at different points of the grade distribution for male and female teachers. This would be the case if male teachers were biasing grades among high scorers while female teachers were biasing grades among low performing students, for example. In terms of ordinal differences, grades assigned by male teachers seem to be slightly more racially biased.

6.6 Heterogeneity in racial gaps by teachers' knowledge of school's population

To more directly examine the role of imperfect information we stratify the sample considering the exposure of different teachers to the school population. This is done by examining strata defined by the knowledge of the community surrounding the school (teachers that live in the school's neighborhood and those who do not) and by tenure in the school (less than two years or more than three years). These two exercises yield different insights into our question. Living in the school neighborhood does not seem to provide teachers with a particularly more knowledgeable view of their students (Panel A). Nonetheless, when we investigate the case of tenure variations, the results are quite striking. Panel B estimates indicate that teachers less acquainted with the school's population seem to be the ones most likely to differentiate students according to race when evaluating competence. This is seen not only in terms of the Black negative coefficient but also on the smaller effects of proficiency tests over measures of teacher-assigned competence. These are exactly the predictions for the effect of increased noise within the conceptual framework presented above.

Table 10 makes the same point exploring information on pupil-teacher matches, by exploring the longitudinal information on students' and teachers' assignment to classrooms. We actually map the individual-level acquaintance level between every student and their current teacher. It is clear from these estimates that the longer teacher and student interact, the smaller the role of biases in emphasizing racial identity. In other words, this empirical exercise reveals that while gaps in grades, promotion rates and ranking 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 the instructor before eighth grade. This is our main indication that imperfect information lies at the heart of the discrimination results we estimate.

7 Conclusions

In this article, we empirically detect racial discrimination within racially-integrated Brazilian eighth grade public-school classrooms. Math teachers' assessments of students with respect to scholastic proficiency and aptitude (grading) are found to be biased. White students are 4% less likely to be deemed non-competent (below passing grade) than their equally proficient and equivalently well-behaved Black classmates. The former are also 5% relatively more likely to be graded above their classroom median. Such effects are equivalent to "taxing" Blacks' performance in proficiency tests by 0.04 to 0.08 of one standard deviation. These results are shown robust to possible omissions of a students' behavioral attributes and to the incidence of measurement errors on scores from standardized tests. It turns out that well intentioned teachers issue report cards for their students with subtle biases (incurred when rounding continuous marks into a discrete scale, for example) and end up adding obstacles to the acquisition of skills and educational credentials by Blacks.

We find that these biases likely result from imperfect information and statistical discrimination or, in other words, from the weighted combination of noisy proficiency signals extracted from exams and stereotyped priors. In the case explored here, stereotyping seems to have resulted from lenient standards for admission of students into eighth grade (which have disproportionally benefited Blacks). Improvements in the signal-extraction "technology" available to teachers make race a less relevant element of the grade assignment process and, at the same time, strengthen the relationship between grades and individual proficiency scores. This shown to be the case in our data.

Our findings lead to important policy lessons. First, curbing teacher rotation (which is very high in our context) can be particularly important for Black students because beyond their likely influence over learning, increased interactions between a group of students and a given teacher diminishes the influence of noise on the evaluation of scholastic proficiency. The more a teacher gets acquainted with a given student, the less relevant the latter's race becomes for evaluation purposes. Second, direct investment in the training of teachers with regards to the design of exams and tests may be warranted when attempting to curb discriminatory outcomes. Third, educational governing bodies should promote the clear communication of standardized test results at the individual level to teachers as a way of widening their information set about students' abilities. Finally, our results point to important nuances on the overall impact of affirmative action policies in admission to college (or social promotion schemes for that matter) in environments where the progress of those targeted by the policy depends on continued subjective evaluations of performance.

In scientific terms, the results presented here also indicate 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 understand the inner workings of discrimination within schools. We leave this for our future research on the topic.

References

[1] Aigner, D. and G. Cain (1977); "Statistical theories of discrimination in the labor market". *Industrial and Labor Relations Review*, 30, 175-187.

[2] Alexander, N.; A Guimarães; C. Hamilton; L. Huntley and W. James (2001); "Beyond Racism: Race and Inequality in Brazil, South Africa, and the United States," L. Rienner Publisher, Boulder, CO.

[3] Altonji, J. and C. Pierret (2001); "Employer Learning and Statistical Discrimination". *The Quartely Journal of Economics*, Vol.116. No. 1 (Feb.2001), pp. 313-350.

[4] Andrabi, T.; J. Das; A. Khwaja and T. Zajonc (2011); "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics", *American Economic Journal: Applied Economics*, 3(3): 29-54.

[5] Arcidiacono, P; E. Aucejo; P. Coate and V. J. Hotz (2013); "Affirmative Action and University Fit: Evidence from Proposition 209" unpublished manuscript, Duke University.

[6] Arias, O.; G Yamada and L.Tejerina (2004); "Education, Family Background and Racial Earnings Inequality in Brazil", manuscript, Inter-American Development Bank, Washington, DC, USA.

[7] Arrow (1971); "Some Mathematical Models of Race in the Labor Market" in Pascal, A. (ed.) Racial Discrimination in Economic Life (Lexington, MA: Lexington Books) 187-204.

[8] Assuncao, J. and B. Ferman (2013); "Does Affirmative Action Enhance or Undercut Investment Incentives? Evidence from Quotas in Brazilian Public Universities", unpublished manuscript, Sao Paulo School of Economics, Fundacao Getulio Vargas.

[9] Becker, G (1957); The Economics of Discrimination, The University of Chicago Press.

[10] Bjerk, D. (2008); "Glass Ceilings or Sticky Floors? Statistical Discrimination in a Dynamic Model of Promotion and Hiring", *The Economic Journal*, 118 (July).

[11] Blume, L. (2006); "The Dynamics of Statistical Discrimination", *The Economic Journal*, Vol 116, November, pp. F480-F498.

[12] Bond, T. and K. Lang (2012); "The Evolution of the Black-White Test Schore Gap in Grades K-3: The Fragility of Results". NBER Working Paper Series, Working Paper 17960.

[13] Borjas, G. and A. Goldberg (1978); "Biased Screening and Discrimination in the Labor Market". *The American Economic Review*, Vol. 68, No. 5 (Dec.,1978), pp. 918-922.

[14] Bowles, S. and H. Gintis (1976); Schooling in Capitalist America. New York: Basic Books.

[15] Campante, F.; A. Crespo and P. Leite (2004); "Wage Inequality across Races in Brazilian Urban Labor Markets: Regional Aspects", Revista Brasileira de Economia, Vol. 58 (2), pp. 185-210.

[16] Coate and G. Loury (1993); "Will Affirmative-Action Policies Eliminate Negative Stereotypes?", *The American Economic Review*, American Economic Association, vol. 83(5), pages 1220-40, December.

[17] Cornell, B. and I. Welch (1996); "Culture, Information, and Screening Discrimination". *Journal of Political Economy*, 1996, vol. 104, no. 3

[18] Cortes, K. and L. Zhang (2012); "The Incentive Effects of the Top 10% Plan" unpublished manuscript, The Bush School of Government and Public Service.

[19] Demeis, D. and R. Turner (1978); "Effects of Students' Race, Physical Attractiveness, and Dialect on Teachers' Evaluations". Contemporary Educational Psychology 3 (1):77-86.

[20] Dorsey, J. K. and J. Colliver (1995); "Effect of Anonymous Test Grading on Pasing Rates as Related to Gender and Race". Academic Medicine 70 (4): 321-23.

[21] Dovidio, J. F.; J. Brigham; B. Johnson and S. Gaertner (1996); "Stereotyping, Prejudice, and Discrimination: Another Look" in N. Macrae; C. Stangor and M. Hewstone (eds.) Stereotypes and Stereotyping, New York, Guilford.

[22] Farkas, G.; R. Grobe; D. Sheehan and Y. Shuan (1990); "Cultural Resources and School Success: Gender, Ethnicity, and Poverty Groups within an Urban School District" *The American Sociological Review*, 1990, Vol.55 (February: 127-142).

[23] Ferguson, R. (1998); "Can Schools Narrow the Black-White Score Gap?" in C. Jencks and M. Phillips (eds.) The Black-white Test Score Gap, The Brookings Institution.

[24] Ferguson, R. (2003); "Teachers' Perceptions and Expectations and the Black-White Test Score Gap". Urban Education 38 (4): 460-507.

[25] Figlio, D. (2005); "Names, Expectations and the Black-White Test Score Gap", NBER Working Paper No. 11195.

[26] Gelbach, J. (2009); "When Do Covariates Matter? And Which Ones, and How Much?"; Working Paper 09-07, University of Arizona 2009.

[27] Hanna, R. and L. Linden (2012); "Discrimination in Grading". *American Economic Journal: Economic Policy*, 4(4): 146-168.

[28] Hedges, L. and Nowell, A. (1999); "Changes in the Black-White Gap in Achievement Test Scores".*Sociology of Education*, 72:2, 111-135.

[29] Herring. C; V. Keith and K. Horton (2004); "Skin Deep: How Race and Complexion Matter in the Color-Blind Era" Institute for Research on Race and Public Policy.

[30] King, E.; P. Orazem and E. Paterno (2008); "Promotion With and Without Learning", World Bank Policy Research Working Paper, 10.1596/1813-9450-4722.

[31] Lam, D.; C. Ardington and M. Leibbrandt (2006); "Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa", University of Cape Town, Working Paper Number 56.

[32] Lavy, V. (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.

[33] Lehmann, Jee-Yeon K. (2011); "Job Assignment and Promotion Under Statistical Discrimination: Evidence from the Early Carrers of Lawyers". Unpublished manuscript, University of Houston.

[34] Leiter, J. and J. Brown (1985); "Determinants of Elementary School Grading". Sociology of Education 58: 166-80.

[35] Lundberg, S. and R. Startz (1983); "Private Discrimination and Social Intervention in Competitive Labor Markets". *The American Economic Review*, LXXIII, 340-347

[36] Madeira, R. and M. A. Rangel (2012); "Racial Achievement Gaps in Another America: Discussing Schooling Outcomes and Affirmative Action in Brazil" in J. Clarke (ed) Closing the Achievement Gap from an International Perspective, Springer Verlag.

[37] Metchemberg, L. (2009); "Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages", *The Review of Economic Studies*, 76, 1431âĂŞ1459.

[38] Natriello, G. and S. Dornbusch (1984); "Bringing Behavior Back In: The Effects of Student Characteristics and Behavior on the Classroom Behavior of Teachers." American Educational Research Jounal 20: 29-43.

[39] Neal, D. (2006); "Why has the Black-White Skill Convergence Stopped?" Handbook of the Economics of Education, Volume 1.

[40] Perry, G.; O. Arias; J. H. López; W. Maloney and L. Servén (2006); "Poverty Reduction and Growth: Virtuous and Vicious Circles", The World Bank.

[41] Phelps, E. (1972); "The Statistical Theory of Racism and Sexism", *The American Economic Review*, American Economic Association, vol. 62(4), pages 659-61, September.

[42] Phillips M.; J. Crouse and J. Ralph (1998); "Does the Black-White Test Score Gap Widen After Children Enter School?" in C. Jencks and M. Phillips (eds.) The Black-white Test Score Gap, The Brookings Institution.

[43] Reardon, S. (2008); "Differential Growth in the Black-White Achievement Gap During Elementary School Among Initially High-and-Low-Scoring Students". Institute for Research on Education Policy and Practice, Working Paper: 2008-07.

[44] Rist, R. (1973); The Urban School: A Factory for Failure. Cambridge, MA: MIT Press.

[45] Rosenthal, L. and R. Jacobson (1968); Pygmalion in the Classroom. New York: Holt.

[46] Sexton, P. (1961); Education and Income. New York: Viking.

[47] Sewell, W. and R. Hauser (1984); "The Wisconsin Longitudinal Study of Social and Psychological Factors in Aspiration and Achievements". Pp. 59-100 in Research in Sociology of Education and Socialization, vol.1, edited by Alan C. Kerchkhoff. Greenwich, CT: JAI Press.

[48] 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.

[49] Telles, E. (2004); "Race in Another America: The Significance of Skin Color in Brazil" Princeton University Press.

[50] Telles, E. and L. Steele (2012); "Pigmentocracy in the Americas: How is Educational Attainment Related to Skin Color?", Americas Barometer Insights, 2012, Number 73.

[51] Trawalter, S.; K. Hoffman and A. Waytz (2012); "Racial Bias in Perceptions of Others' Pain". PLOS ONE, November, Volume 7, Issue 11.

[52] United Nations (2005); "Brazil: National Human Development Report", New York and Brasilia, Brazil.

[53] Williams, T. (1976); "Teacher Prophecies and the Inherittance of Inequality". *Sociology of Education*, 49: 223-36.

Table 1: Racial Differentials in Mathematics Teacher End-of-Year Assessments

OLS Fixed-Effects Estimations

	Raw means	Raw diff.	Conditional Differences				
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Panel A: End-of-Year Grades (0-10 Scale)							
White	6.056**						
	(0.008)						
Black	5.499**	-0.556**	-0.490**	-0.375**	-0.222**	-0.126**	-0.101**
	(0.011)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Math test (z-score)					0.583**	0.379**	0.370**
					(0.003)	(0.003)	(0.003)
Panel B: Minimum Competence Indicator (grade>=5), in %							
White	90.573**						
	(0.123)						
Black	85.209**	-5.364**	-5.027**	-3.463**	-2.270**	-1.178**	-1.082**
	(0.231)	(0.213)	(0.202)	(0.198)	(0.195)	(0.190)	(0.207)
Standardized test					5.542**	3.393**	3.270**
score in Math					(0.077)	(0.072)	(0.072)
Panel C: Above classroom average grade (%)							
White	49.192**						
	(0.160)						
Black	36.759**	-12.432**	-13.451**	-10.274**	-6.582**	-3.778**	-3.004**
	(0.292)	(0.302)	(0.294)	(0.288)	(0.272)	(0.262)	(0.284)
Standardized test					14.741**	9.978**	9.756**
score in Math					(0.093)	(0.100)	(0.100)
Panel D: Above classroom median grade (%)							
White	40.305**						
	(0.141)						
Black	26.986**	-13.319**	-13.340**	-10.424**	-6.655**	-3.928**	-3.101**
	(0.250)	(0.278)	(0.282)	(0.277)	(0.261)	(0.252)	(0.273)
Standardized test					14.672**	10.033**	9.832**
score in Math					(0.091)	(0.098)	(0.097)
Controls							
Classroom FE			Y	Y	Y	Y	Y
Child demographics				Y	Y	Y	Y
Test scores in Math					Y	Y	Y
Other standardized scores (includes essay)						Y	Y
Parental demographics/SES							Y

Notes: Standard-errors are clustered at the classroom level. Sample consists of 352,881 students in 11,475 classrooms. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. Age is measured in months and included as second-order polynomial; standardized Math test scores are z-scores (at the population level for Panels A and B and at classroom level for Panels C and D) and are included as second order polynomial; other standardized test scores include reading and essay portions of Language tests (the multiple-choice test results in language are also interacted with Math scores polinomyal); parental demographics include maternal education, maternal age, maternal skin color, home ownership, number of cars owned, number of restrooms in house, and geographic region of maternal birth.

Table 2: Racial Differentials in Mathematics Teacher End-of-Year Assessments - controls for behavior

OLS Fixed-Effects Estimations

	Raw means	Raw diff.	Conditional Differences				
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Panel A: End-of-Year Grades (0-10 Scale)							
White	6.056**						
	(0.008)						
Black	5.499**	-0.556**	-0.101**	-0.084**	-0.080**	-0.106**	-0.084**
	(0.011)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Math test (z-score)			0.370**	0.354**	0.349**	0.310**	0.296**
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Panel B: Minimum Competence Indicator (grade>=5), in %							
White	90.573**						
	(0.123)						
Black	85.209**	-5.364**	-1.082**	-0.971**	-0.873**	-1.152**	-1.018**
	(0.231)	(0.213)	(0.207)	(0.206)	(0.206)	(0.199)	(0.198)
Standardized test			3.270**	3.125**	3.070**	2.456**	2.357**
score in Math			(0.072)	(0.071)	(0.071)	(0.068)	(0.068)
Panel C: Above classroom average grade (%)							
White	49.192**						
	(0.160)						
Black	36.759**	-12.432**	-3.004**	-2.542**	-2.422**	-3.127**	-2.495**
	(0.292)	(0.302)	(0.284)	(0.281)	(0.280)	(0.271)	(0.268)
Standardized test			9.756**	9.344**	9.232**	8.360**	7.971**
score in Math			(0.100)	(0.098)	(0.098)	(0.095)	(0.094)
Panel D: Above classroom median grade (%)							
White	40.305**						
	(0.141)						
Black	26.986**	-13.319**	-3.101**	-2.623**	-2.531**	-3.216**	-2.561**
	(0.250)	(0.278)	(0.273)	(0.270)	(0.268)	(0.262)	(0.259)
Standardized test			9.832**	9.423**	9.311**	8.535**	8.136**
score in Math			(0.097)	(0.095)	(0.095)	(0.093)	(0.091)
Controls							
Classroom FE			Y	Y	Y	Y	Y
Child demographics			Y	Y	Y	Y	Y
Test scores in Math			Y	Y	Y	Y	Y
Other standardized scores (includes essay)			Y	Y	Y	Y	Y
Parental demographics/SES			Y	Y	Y	Y	Y
Behavior (Self reported)				Y			Y
Behavior (Parental reports)					Y		Y
Behavior (Based on school reports)						Y	Y

Notes: Standard-errors are clustered at the classroom level. Sample consists of 352,881 students in 11,475 classrooms. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. Age is measured in months and included as second-order polynomial; standardized Math test scores are z-scores (at the population level for Panels A and B and at classroom level for Panels C and D) and are included as second order polynomial; other standardized test scores include reading and essay portions of Language tests (the multiple-choice test results in language are also interacted with Math scores polinomyal); parental demographics include maternal education, maternal age, maternal skin color, home ownership, number of cars owned, number of restrooms in house, and geographic region of birth; self-reported behavior are "not skipping classess" and "not procrastinating with homework" indicators; parental reports of behavior are based on perceptions "good behavior in school", "effort with school work" and "interest in school activities"; school reports on behavior include absence rates in Math and Language and physical education grades during the first half of the academic year.

Table 3: Racial Differentials in Grading in Math Teacher Assessments - additional controls for past grades

Fixed-Effects Estimations

	Numerical grade (0-10)	+ Past grades	Minimum competence	+ Past grades	Above class average	+ Past grades	Above class median	+ Past grades
Black average	5.50		85.21		36.76		26.99	
Black vs White diff.	-0.084**	-0.056**	-1.018**	-0.767**	-2.495**	-1.738**	-2.561**	-1.799**
	(0.008)	(0.007)	(0.198)	(0.196)	(0.268)	(0.259)	(0.259)	(0.250)
Math test (z-score)	0.296**	0.218**	2.357**	1.685**	7.971**	6.089**	8.136**	6.224**
	(0.003)	(0.003)	(0.068)	(0.067)	(0.094)	(0.091)	(0.091)	(0.089)

Notes: Standard-errors are clustered at the classroom level. Sample consists of 352,881 students in 11,475 classrooms. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. Past grades in Math and Language are 2009 grades measured in relation to corresponding classroom mean. See notes in Table 2 for full list of controls.

Table 4: Comparative Between Actual Impacts and Impacts Based on Simulated Biased Rounding of Grades

	Numerical grade (0-10)			Minimum competence			Above class average			Above class median		
	Actual	Simulated 1	Simulated 2	Actual	Simulated 1	Simulated 2	Actual	Simulated 1	Simulated 2	Actual	Simulated 1	Simulated 2
Black average	5.50	5.46	5.42	85.21	71.05	70.32	36.76	40.77	39.87	26.99	29.08	28.21
Black vs White diff.	-0.056**	-0.058**	-0.095**	-0.767**	-0.849**	-1.582**	-1.738**	-1.401**	-2.272**	-1.799**	-1.188**	-2.043**
	(0.007)	(0.004)	(0.004)	(0.196)	(0.193)	(0.193)	(0.259)	(0.188)	(0.193)	(0.250)	(0.183)	(0.190)

Notes: Standard-errors are clustered at the classroom level. Sample consists of 352,881 students in 11,475 classrooms. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. See notes in Table 1 and Table 2.

Table 5: Racial Differentials in Grading in Math Teacher Assessments - instrumental variables

OLS and Instrumental-Variables Fixed-Effects Estimations

	Numerical grade (0-10)		Minimum competence		Above class average		Above class median	
	FE	IV-FE	FE	IV-FE	FE	IV-FE	FE	IV-FE
Black average	5.50		85.21		36.76		26.99	
Black vs White diff.	-0.056**	-0.035**	-0.767**	-0.577**	-1.738**	-1.138**	-1.799**	-1.218**
	(0.007)	(0.008)	(0.196)	(0.204)	(0.259)	(0.282)	(0.250)	(0.271)
Math test (z-score)	0.218**	0.853**	1.685**	7.078**	6.089**	25.550**	6.224**	25.220**
	(0.003)	(0.026)	(0.067)	(0.653)	(0.091)	(0.964)	(0.089)	(0.911)
First-stage statistics								
F-statistic 1st stage								
Math score		9,599		9,599		10,189		10,189
Math score squared		2,731		2,731		1,714		1,714
Math score * Language score		4,882		4,882		4,326		4,326
Math score sq * Language score		3,873		3,873		4,588		4,588
Language score		18,928		18,928		18,928		18,928
Kleibergen-Paap rk LM statistic		863.34		863.34		463.44		463.44
Kleibergen-Paap rk Wald statistic		939.46		939.46		486.14		486.14
Kleibergen-Paap Wald rk F statistic		187.84		187.84		97.20		97.20

Notes: Standard-errors are clustered at the classroom level. Sample consists of 352,881 students in 11,475 classrooms. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. See notes in Tables 2 and 3.

Table 6: Stratifications by teacher's opinion regarding best practices for assessing students

Instrumental-Variables Fixed-Effects Estimations

	Numerical grade (0-10)			Minimum competence			Above class average			Above class median		
	Objective	Non-Objective	Subjective	Objective	Non-Objective	Subjective	Objective	Non-Objective	Subjective	Objective	Non-Objective	Subjective
Black average	5.521**	5.490**	5.512**	85.474**	85.240**	85.514**	36.952**	36.219**	36.321**	27.244**	26.918**	26.968**
Black vs White diff.	-0.046**	-0.033**	-0.039**	-0.580	-0.623*	-0.558*	-1.804**	-0.898*	-1.260**	-1.660**	-1.113**	-1.439**
	(0.014)	(0.011)	(0.010)	(0.352)	(0.284)	(0.257)	(0.514)	(0.386)	(0.362)	(0.494)	(0.367)	(0.346)
Math test (z-score)	0.796**	0.892**	0.827**	5.623**	6.832**	5.778**	25.151**	26.775**	25.437**	25.355**	25.786**	24.923**
	(0.045)	(0.037)	(0.032)	(1.177)	(0.908)	(0.824)	(1.729)	(1.280)	(1.161)	(1.631)	(1.204)	(1.089)
Observations	108,854	186,663	216,164	108,854	186,663	216,164	108,854	186,663	216,164	108,854	186,663	216,164
Classrooms	3,517	6,091	7,033	3,517	6,091	7,033	3,517	6,091	7,033	3,517	6,091	7,033

Notes: Standard-errors are clustered at the classroom level. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. All first stage statistics indicate strong instruments and are available upon request. See notes in Table 2.

Table 7: Stratifications by student's demographic characteristics

Instrumental-Variables Fixed-Effects Estimations

	Numerical grade (0-10)		Minimum competence		Above class average		Above class median	
Panel A: Student's gender								
	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl
Black vs White diff.	-0.048**	-0.019	-0.682*	-0.285	-1.107**	-1.077*	-1.487**	-0.782
	(0.011)	(0.013)	(-0.288)	(0.279)	(0.360)	(0.475)	(0.339)	(0.468)
Math test (z-score)	0.766**	0.936**	7.974**	6.049**	24.242**	27.888**	22.636**	28.589**
	(0.042)	(0.034)	(1.191)	(0.719)	(1.609)	(1.276)	(1.472)	(1.248)
N	*179,214*	*146,916*	*179,214*	*146,916*	*179,214*	*146,916*	*179,214*	*146,916*
Classrooms	*11,475*	*11,472*	*11,475*	*11,472*	*11,475*	*11,472*	*11,475*	*11,472*
Panel B: Student's age								
	Above median age	Below median age	Above median age	Below median age	Above median age	Below median age	Above median age	Below median age
Black vs White diff.	-0.047**	-0.022	-0.904**	-0.203	-1.293**	-1.014*	-1.221**	-1.281**
	(0.011)	(0.012)	(0.302)	(0.277)	(0.384)	(0.441)	(0.364)	(0.423)
Math test (z-score)	0.851**	0.858**	7.974**	6.616**	25.009**	27.294**	24.684**	26.402**
	(0.036)	(0.037)	(0.950)	(0.879)	(1.296)	(1.489)	(1.210)	(1.424)
N	*183,917*	*168,906*	*183,917*	*168,906*	*183,917*	*168,906*	*183,917*	*168,906*
Classrooms	*11,465*	*11,354*	*11,465*	*11,354*	*11,465*	*11,354*	*11,465*	*11,354*
Panel C: Student's past proficiency								
	Below median past score	Above median past score	Below median past score	Above median past score	Below median past score	Above median past score	Below median past score	Above median past score
Black vs White diff.	-0.040**	-0.025*	-0.570	-0.537*	-1.137**	-0.810	-1.125**	-0.875
	(0.012)	(0.013)	(0.306)	(0.249)	(0.405)	(0.460)	(0.371)	(0.459)
Math test (z-score)	0.382*	0.768**	-2.363	10.157**	16.973**	24.983**	19.923**	22.969**
	(0.184)	(0.041)	(4.750)	(0.942)	(5.610)	(1.734)	(5.211)	(1.659)
N	*176,389*	*176,451*	*176,389*	*176,451*	*176,389*	*176,451*	*176,389*	*176,451*
Classrooms	*11,450*	*11,449*	*11,450*	*11,449*	*11,450*	*11,449*	*11,450*	*11,449*

Notes: Standard-errors are clustered at the classroom level. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. See notes in Table 2.

Table 8: Stratifications by teacher's demographic characteristics

Instrumental-Variables Fixed-Effects Estimations

	Numerical grade (0-10)		Minimum competence		Above class average		Above class median	
Panel A: Teacher's race								
	Black/Brown teachers	White teachers	Black/Brown teachers	White teachers	Black/Brown teachers	White teachers	Black/Brown teachers	White teachers
Black average	5.479**	5.506**	85.654**	85.135**	38.339**	36.304**	27.293**	26.863**
Black vs White diff.	-0.035	-0.036**	0.092	-0.820**	-1.552*	-1.139**	-1.635**	-1.211**
	(0.019)	(0.009)	(0.444)	(0.234)	(0.655)	(0.319)	(0.603)	(0.310)
Math test (z-score)	0.834**	0.856**	4.794**	7.398**	26.964**	25.107**	23.654**	25.559**
	(0.058)	(0.029)	(1.462)	(0.746)	(2.763)	(1.026)	(2.477)	(0.985)
N	*66,248*	*275,008*	*66,248*	*275,008*	*66,248*	*275,008*	*66,248*	*275,008*
Classrooms	*2,134*	*8,967*	*2,134*	*8,967*	*2,134*	*8,967*	*2,134*	*8,967*
Panel B: Teacher's gender								
	Male teachers	Female teachers	Male teachers	Female teachers	Male teachers	Female teachers	Male teachers	Female teachers
Black average	5.488**	5.504**	85.786**	84.934**	37.155**	36.571**	26.383**	27.273**
Black vs White diff.	-0.045**	-0.031**	-0.457	-0.639*	-1.497**	-0.992**	-1.243**	-1.228**
	(0.014)	(0.010)	(0.353)	(0.251)	(0.491)	(0.348)	(0.472)	(0.335)
Math test (z-score)	0.754**	0.903**	4.709**	8.227**	22.249**	27.150**	21.649**	26.936**
	(0.041)	(0.033)	(1.017)	(0.842)	(1.634)	(1.202)	(1.549)	(1.133)
N	*109,706*	*243,175*	*109,706*	*243,175*	*109,706*	*243,175*	*109,706*	*243,175*
Classrooms	*3,562*	*7,913*	*3,562*	*7,913*	*3,562*	*7,913*	*3,562*	*7,913*

Notes: Standard-errors are clustered at the classroom level. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. See notes in Table 2.

Table 9: Stratifications by teacher's knowledge of school population

Instrumental-Variables Fixed-Effects Estimations

	Numerical grade (0-10)		Minimum competence		Above class average		Above class median	
	knows	does not	knows	does not	knows	does not	knows	does not
Panel A: Teacher's knowledge of community around school (live in neighborhood vs. does not)								
Black average	5.486**	5.555**	85.170**	85.856**	35.702**	39.232**	26.780**	27.907**
Black vs White diff.	-0.039**	-0.034	-0.502*	-0.869	-1.352**	-0.931	-1.406**	-1.130
	(0.010)	(0.019)	(0.255)	(0.450)	(0.350)	(0.671)	(0.337)	(0.632)
Math test (z-score)	0.870**	0.819**	6.797**	5.194**	26.741**	23.785**	26.247**	22.760**
	(0.032)	(0.061)	(0.824)	(1.481)	(1.109)	(2.699)	(1.054)	(2.495)
Observations	*230676*	*63904*	*230676*	*63904*	*230676*	*63904*	*230676*	*63904*
Classrooms	*7472*	*2106*	*7472*	*2106*	*7472*	*2106*	*7472*	*2106*
Panel B: Teacher's tenure in school (more than 3 years vs less than 2)								
Black average	5.474**	5.553**	84.731**	86.406**	35.896**	37.736**	26.884**	27.392**
Black vs White diff.	-0.026*	-0.061**	-0.469	-0.805*	-0.765*	-2.177**	-0.783*	-2.353**
	(0.011)	(0.015)	(0.277)	(0.366)	(0.379)	(0.535)	(0.360)	(0.515)
Math test (z-score)	0.891**	0.777**	7.138**	4.843**	27.605**	22.558**	26.506**	23.642**
	(0.035)	(0.050)	(0.892)	(1.219)	(1.179)	(2.033)	(1.101)	(1.949)
Observations	*200945*	*95284*	*200945*	*95284*	*200945*	*95284*	*200945*	*95284*
Classrooms	*6508*	*3128*	*6508*	*3128*	*6508*	*3128*	*6508*	*3128*

Standard-errors are clustered at the classroom level. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. See notes in Tables 2 and 3.

Table 10: Exploring individual level relationship with teacher

Instrumental-Variables Fixed-Effects Estimations

	Numerical grade (0-10)		Minimum competence		Above class average		Above class median	
	Knows teacher	Does not know teacher	Knows teacher	Does not know teacher	Knows teacher	Does not know teacher	Knows teacher	Does not know teacher
Black average	5.59	5.47	86.64	84.82	36.29	36.85	27.25	26.91
Black vs White diff.	-0.011	-0.041**	0.026	-0.724**	-0.272	-1.432**	-0.518	-1.413**
	(0.016)	(0.009)	(0.411)	(0.227)	(0.579)	(0.317)	(0.549)	(0.302)
Math test (z-score)	0.883**	0.858**	8.514**	10.031**	25.637**	24.904**	24.115**	23.459**
	(0.027)	(0.024)	(0.694)	(0.591)	(0.931)	(0.810)	(0.870)	(0.767)
Observations	350,755		350,755		350,755		350,755	
Classrooms	11,402		11,402		11,402		11,402	

Standard-errors are clustered at the classroom level. Graphic representation of significance values is made more stringent, with ** p<0.01, and * p<0.05. See notes in Tables 2 and 3.

Table A1: Descriptive statistics

	White (42.2%)		Black (9.8%)		Other races (48.0%)	
Standardized test performances						
Math z-score	0.136	(0.0045)	-0.213	(0.0054)	-0.095	(0.0035)
Language z-score	0.168	(0.0044)	-0.279	(0.0057)	-0.098	(0.0037)
Essay section 1 (max score)	0.380	(0.0023)	0.259	(0.0028)	0.301	(0.0020)
Essay section 2 (max score)	0.391	(0.0024)	0.262	(0.0029)	0.305	(0.0020)
Essay section 3 (max score)	0.252	(0.0019)	0.143	(0.0022)	0.183	(0.0015)
Essay section 4 (max score)	0.197	(0.0017)	0.103	(0.0019)	0.136	(0.0013)
Family Background						
Home ownership	0.544	(0.0021)	0.514	(0.0032)	0.441	(0.0023)
Automobiles (nr. 0-4)	0.590	(0.0030)	0.417	(0.0040)	0.385	(0.0024)
Mom completed college	0.047	(0.0007)	0.032	(0.0010)	0.022	(0.0004)
Mom some college	0.031	(0.0005)	0.024	(0.0008)	0.017	(0.0003)
Mom completed high school	0.222	(0.0014)	0.169	(0.0022)	0.149	(0.0012)
Child demographics						
Age	14.914	(0.0026)	15.096	(0.0049)	15.064	(0.0030)
Male	0.490	(0.0013)	0.624	(0.0027)	0.500	(0.0013)
Behavioral indicators						
Does not procastinate on homework	0.248	(0.0014)	0.185	(0.0022)	0.175	(0.0012)
Does not skip classes	0.691	(0.0022)	0.643	(0.0031)	0.568	(0.0025)
Grade in PE (2nd bi-monthly evaluation)	6.916	(0.0120)	6.567	(0.0152)	6.604	(0.0121)
Grade in PE (1st bi-monthly evaluation)	6.908	(0.0119)	6.564	(0.0155)	6.610	(0.0123)
Attendance rate Math (2nd bi-monthly evaluation)	88.774	(0.0506)	87.486	(0.0788)	87.303	(0.0558)
Attendance rate Math (1st bi-monthly evaluation)	91.040	(0.0539)	89.936	(0.0749)	89.960	(0.0569)
Parents report of level of pupil interest in studies (0-10)	5.939	(0.0179)	5.399	(0.0244)	4.884	(0.0215)
Parents report of low effort in studies	0.158	(0.0011)	0.190	(0.0022)	0.143	(0.0011)
Parents report of high effort in studies	0.160	(0.0012)	0.135	(0.0020)	0.129	(0.0011)
Parents report of bad behavior in school	0.057	(0.0006)	0.088	(0.0015)	0.057	(0.0006)
Parents report of good behavior in school	0.424	(0.0018)	0.318	(0.0027)	0.319	(0.0017)

Notes: Standard-errors in parentheses are clustered at the classroom level.

Table A2: Impact of sample restrictions over estimated gaps

Fixed-Effects Estimations

	Numerical grade (0-10)			Minimum competence			Above class average			Above class median		
	Full sample	With teacher ID's	With teacher questionnaire	Full sample	With teacher ID's	With teacher questionnaire	Full sample	With teacher ID's	With teacher questionnaire	Full sample	With teacher ID's	With teacher questionnaire
Black average	5.50	5.50	5.50	85.21	85.20	85.33	36.76	36.74	36.51	26.99	26.98	27.04
Black vs White diff.	-0.056**	-0.056**	-0.058**	-0.767**	-0.775**	-0.742**	-1.738**	-1.782**	-1.834**	-1.799**	-1.808**	-1.892**
	(0.007)	(0.007)	(0.008)	(0.196)	(0.197)	(0.214)	(0.259)	(0.260)	(0.284)	(0.250)	(0.251)	(0.273)
Math test (z-score)	0.218**	0.218**	0.221**	1.685**	1.687**	1.732**	6.089**	6.097**	6.285**	6.224**	6.237**	6.382**
	(0.003)	(0.003)	(0.003)	(0.067)	(0.067)	(0.073)	(0.091)	(0.091)	(0.100)	(0.089)	(0.089)	(0.097)
Observations	*352,881*	*350,755*	*294,580*	*352,881*	*350,755*	*294,580*	*352,881*	*350,755*	*294,580*	*352,881*	*350,755*	*294,580*
Classrooms	*11,475*	*11,402*	*9,578*	*11,475*	*11,402*	*9,578*	*11,475*	*11,402*	*9,578*	*11,475*	*11,402*	*9,578*

Notes: Fixed-effects estimations. See notes in Table 2.

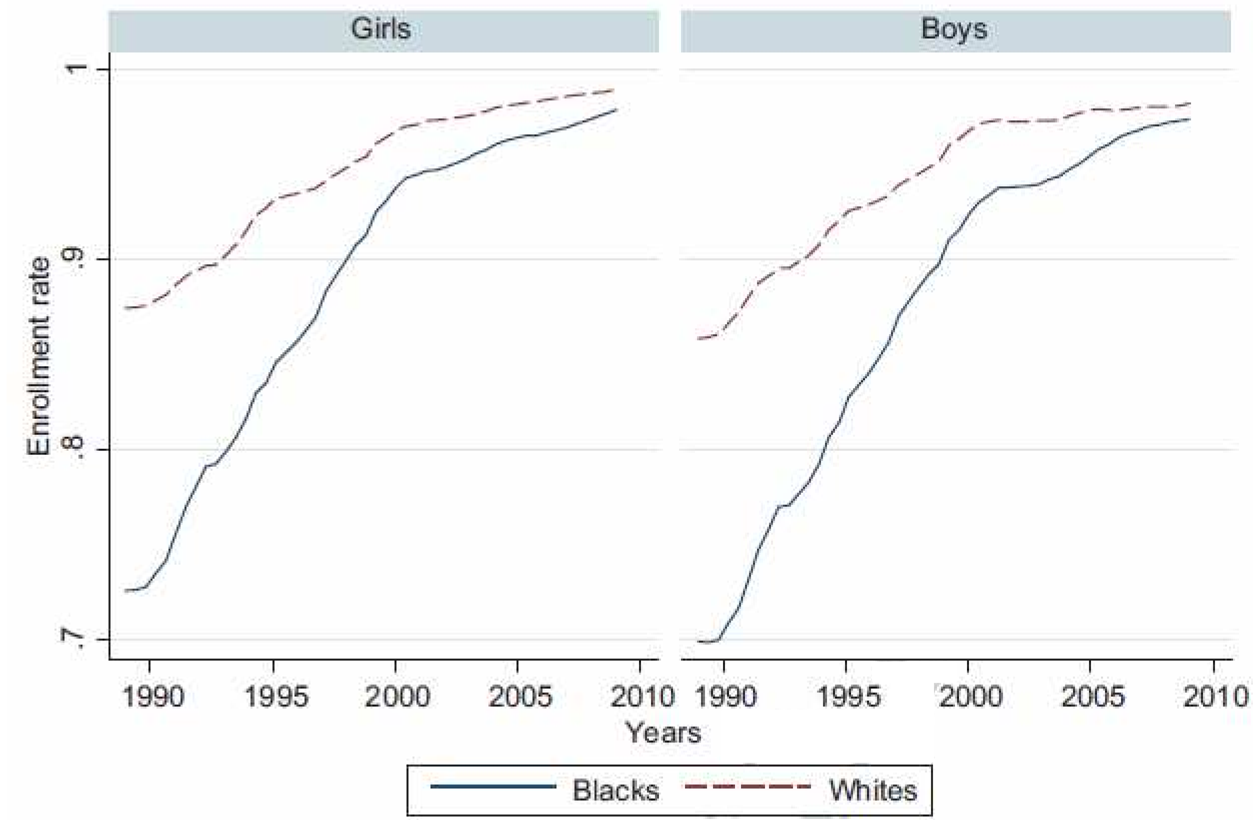

Figure A1: Trends in enrollment rates for Blacks and Whites (1989-2009)

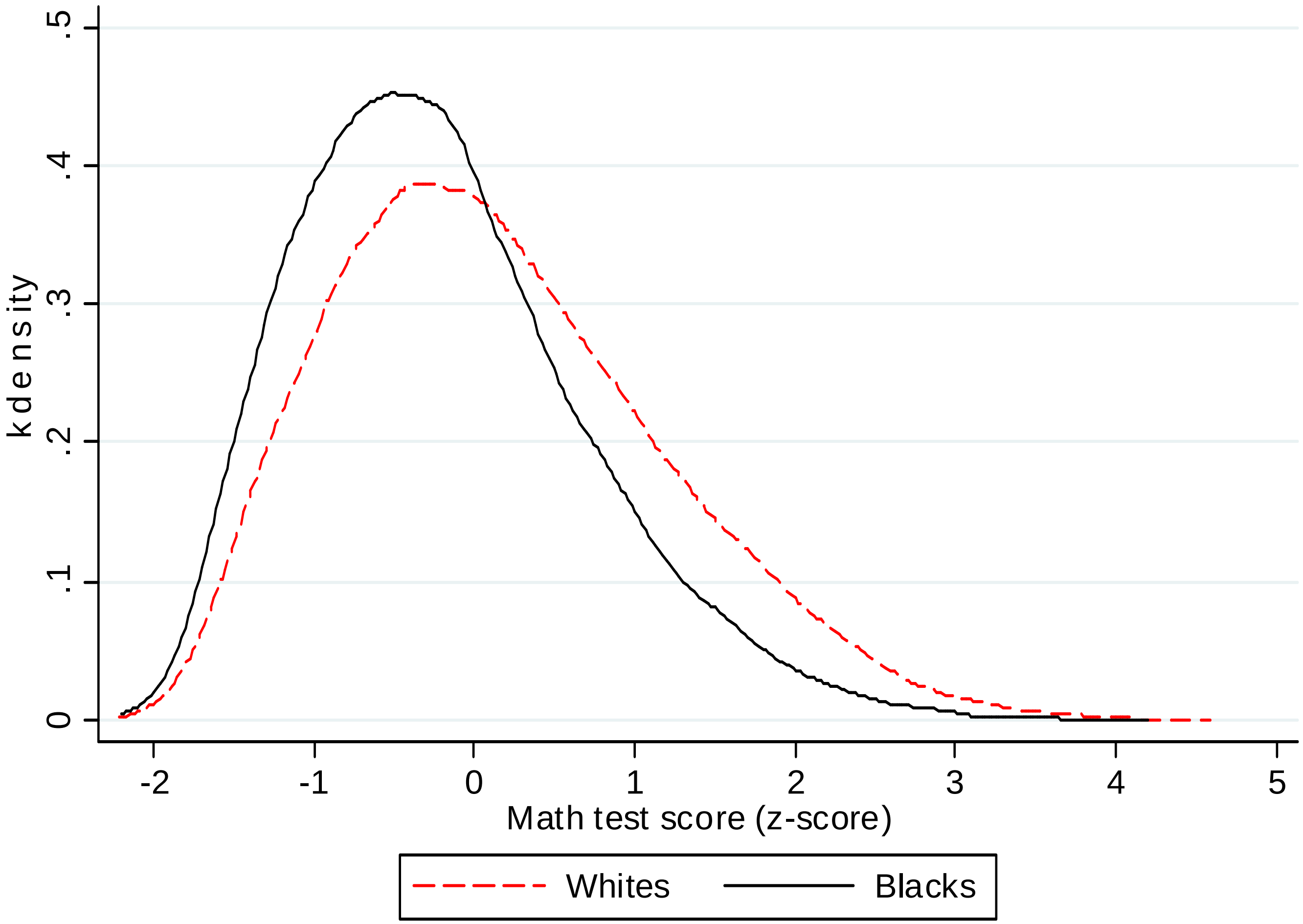

Figure A2: Kernel Density Estimates, Math Test Scores – 8th graders 2010

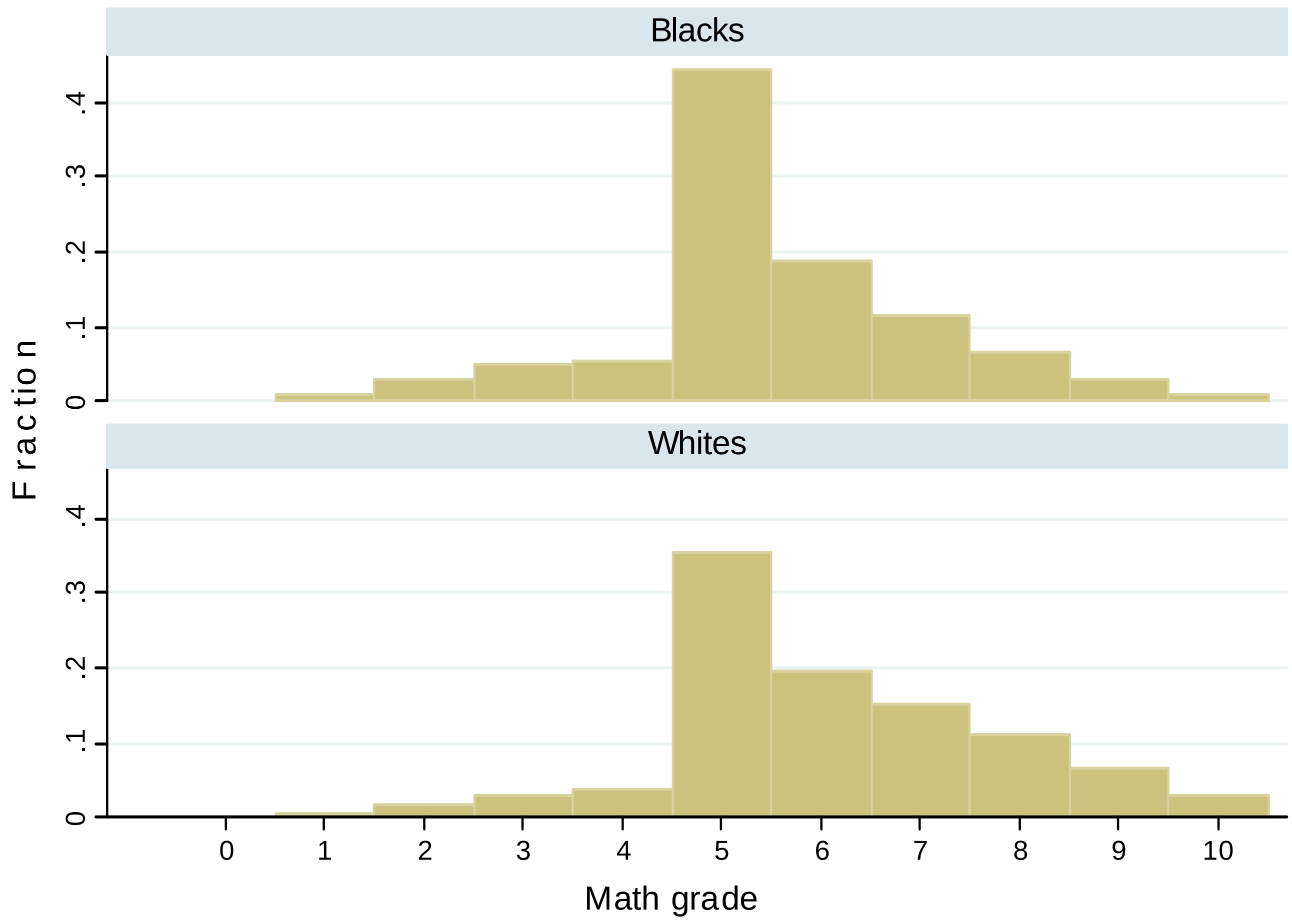

Figure A3: Histogram, Teacher-Assigned Math Grades – 8th graders 2010

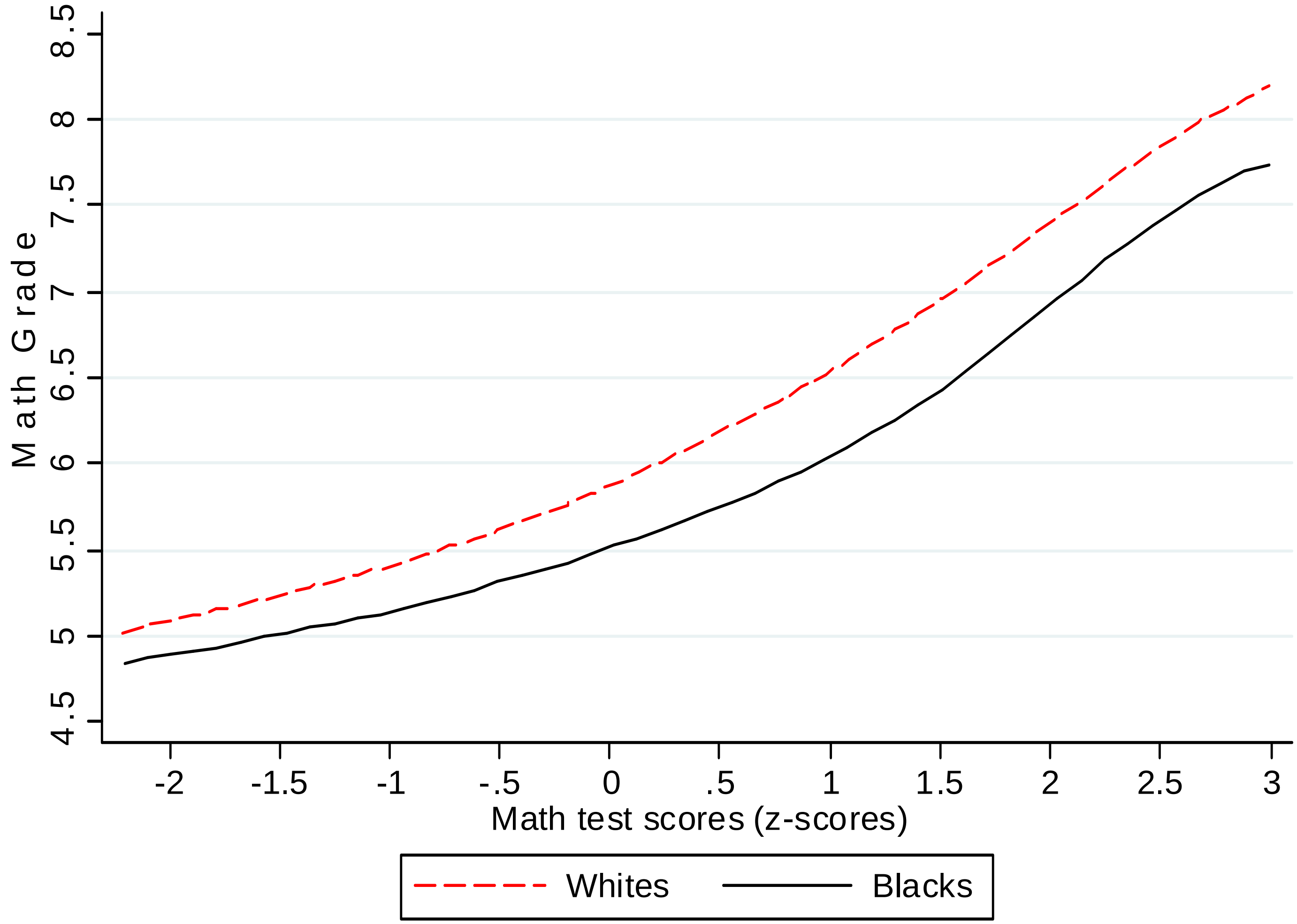

Figure A4: Raw relation between teacher-assigned grades and test scores (local polynomial smooth)