RACIAL DISCRIMINATION GOES TO SCHOOL: EVIDENCE FROM BRAZIL*†

Fernando Botelho

Ricardo Madeira

Marcos A. Rangel

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 Mathematics 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 median (ordinal impact). These findings suggest that schools may be imposing additional obstacles to the acquisition of educational credentials by Blacks. By further detailing heterogeneity in these differentials, we unveil indications that they result from incomplete information issues highlighted in models of statistical discrimination and made particularly salient by the adoption of social promotion schemes in our context.

JEL: I21; J15; and I24.
Keywords: racial discrimination, schooling, grades, standardized tests, social promotion.

*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). The *Secretaria de Educacao do Estado de Sao Paulo - SEE/SP* officially provided access to the data. We benefited from comments received at the NBER Education Meeting, the PAA Meeting, 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, PUC-Rio, Sao Paulo School of Economics, Stanford University, University of California at Riverside, University of Delaware, University of Illinois at Chicago, and University of Sao Paulo. The paper has been screened to ensure no confidential information is revealed. Part of the analysis presented here was performed while Rangel was a visiting scholar at Princeton University. He is thankful for the hospitality of the Research Program in Development Studies and of the Program in Latin American Studies. 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 *SEE/SP*. Previous versions of this article have circulated under the title: "Discrimination Goes to School? Racial Differences in Performance Assessments by Teachers."

†Corresponding author: Marcos A. Rangel, Duke University, 194 Rubenstein Hall, 302 Towerview Drive, Durham, NC 27708-0239, USA; (919) 613-7340; marcos.rangel@duke.edu

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 *expected* 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 (i.e.: grading). We employ uniquely detailed data from the state of Sao Paulo covering approximately 277 thousand eighth-graders spread across 10.6 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.

The analyses show that portions of teachers' assessments in Mathematics not explained by proficiency scores are associated with pupils' racial background. Our most conservative estimates indicate that there are statistically significant underscoring and under-ranking of Blacks relative to Whites. The measured racial gap in promotion rates between equivalently proficient and well-behaved students corresponds to a 4.1% increase in the retention probability for the average Black. Focusing exclusively on the ordinality aspect we also uncover a gap that translates into a 4.5% reduction on the probability of Blacks being graded above the classroom median. In practice, these effects are equivalent to "taxing" an average student's performance in proficiency tests by 0.03 to 0.04 of one standard deviation. These results are shown robust to possible omissions of behavioral

[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 (2013). Cautionary notes on these findings can be found in Bond and Lang (2012).

attributes and to the likely incidence of measurement error on scores from standardized tests used in our estimations.

Once the existence of racial gaps in assessments is established, we rely on economic theory to examine why that is the case in our context. We draw from a rich literature on statistical and screening discrimination.[3] We map our setting into these studies by focusing on two main institutional aspects. First, teachers are limited by imperfect screening technology in the process of scholastic competence's measurement and, once assigned to students of a given level (whose admission is decided by a third party), are solely responsible for promotion and ranking decisions. Second, due to a number of policies implemented since the late 1990's, a dramatic increase in access to public education has been observed. We highlight in particular the adoption of *social promotion* schemes between the fifth and seventh grades. In practice, by establishing lenient standards for the admission of students into eighth grade, such policy has disproportionally benefited Blacks (who are over-represented among pupils with lower proficiency). In other words, social promotion has emulated affirmative action within the Brazilian school system we study. Eighth-grade teachers were well aware of the implications of such policy, and priors regarding students' competence may have been downgraded as a result. Therefore, we hypothesize that when teachers issue report cards for their students, subtle biases may be generated by the weighted combination of noisy information extracted from their own screening exams and stereotyped priors.

We then present evidence on the validity of such 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 and the larger is the role of hard-to-measure signals of proficiency.[4] In this regard, our empirical exercises unveil that while gaps in 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 that instructor before eighth grade. Teachers seem to learn about a student's true "type" over academic years and once they are

[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); Bjerk (2008); and Lehmann (2011).

[4]The same notion of interactions and learning is also central in Lundberg and Startz (2007).

fully aware of grading standards previously used (i.e.: their own).[5]

We believe our results can shed new light on the effects of affirmative action over 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 targeted by such polices (the effort choices of high-schoolers that are granted easier access to college), we indirectly 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, could explain 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 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 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 competent Whites in the more competitive fields. This is an indirect policy implication we draw: affirmative action in college admissions (or social promotion schemes for that matter) may have negative impacts over the population it was designed to help when negatively influencing subsequent subjective evaluations of competence.

Considering the role played by misinformation in the results presented here, and beyond its scientific interest, we also draw three other lessons for education policy 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

[5]Tests of learning in the context statistical discrimination can also be seen in Autor and Scarborough (2008); Lange (2007); List (2004); and Farber and Gibbons (1998).

[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.

[7]In many ways this is similar to the original stereotyping-affirmative action nexus proposed by Coate and Loury (1993) in the context of labor markets.

purposes.[8] 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, the generation of individual report cards 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. Above all, public schools and respective education authorities could do a better job on their use of performance information in order to maximize efficiency.[9] The reduction of grading discrimination of the sort we uncover would just be an added bonus.

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.[10] 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.[11]

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

[8]This would come in addition to the benefical effects of grade/subject-specific teacher experience measured by Ost (2014).

[9]In fact, in the last few years a number of private high-schools started adopting "big data" analysis in order to maximize performance of their students in the ENEM (Brazilian SAT) exams. Improved ENEM scores are known for having important impacts over a school's reputation and revenues.

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

[11]See Mechtenberg (2009) for a formalization of an argument like this. See also Lundberg and Startz (1983), who are explicit in modeling human capital investments' response to the presence of discrimination.

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.

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.[12]

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 that 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

[12] 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.

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) 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 are entitled gatekeepers, being (and feeling) 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), in our context there are both weak regulation of grading and non-disclosure of information regarding standardized test performance to acting parties (teachers or students) before pupils' final assessments are processed. In this way, the present paper explores 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 primary, middle and high-school students in 2010. Among eighth-graders, 67% were served by schools directly administered by the state authority, with the remaining share being evenly split between municipal and private institutions. Using confidential individual identifiers we merged information from four distinct sections of the Secretary's data bank: matriculation information, teachers' allocation to classrooms, transcript records and standardized tests of proficiency.[13] 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 (which should correspond to intra-state migration in the case of public schools), and calssroom assignments over the years. Our working data set covers the 2007-2012 period.

Records of teacher allocations to classrooms for the years 2007, 2008, 2009, 2010 and 2011 complement the analysis. 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 whom each student had classes in the three years prior

[13]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.

to eighth grade (which, except in case of retention, 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.

We also take advantage of the administrative data set on teachers' assessments of individual students between 2007 and 2011. 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 eighth grade, in which teachers are fully specialized by subject, these correspond to Language (Portuguese), Mathematics, History, Geography, Sciences, Physical Education, and the Arts.

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 percentage points (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 technology and methods and to allocate students across the 11 grading categories.

The final data set employed in our analysis provides results from standardized scores fielded in the context of Sao Paulo's Performance Evaluation System - (SARESP- *Sistema de Avaliacao de Rendimento do Estado de Sao 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). Here we employ data from the 10th to the 13th editions (2007 to 2010), with over 1.5 million test-takers in approximately 5,050 schools covered in the latter year. 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 questions covering Math and Portuguese language. For students in eighth grade, each exam contained 30 multiple choice questions. The exams were taken in late November (Spring), right before the end of the academic year, during regular-class meeting times, and 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. 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. Importantly, individual-level results from SARESP are *never* made publicly available to children, parents, or schools.[14]

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

[14]For years prior to 2010 we were also granted access to IRT-scores from proficiency tests in Science and evaluations of an essay-based portion of the Language exam. The latter covered 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.

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, in the provision of college-scholarships, and in public-sector hiring. Yet, as we discuss below, the most important recent advancements towards the (potential) closing of racial gaps seem to have indeed come about as a result of *colorblind* social policies.

3.2.1 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), Madeira and Rangel (2013) show trends in enrollment for children aged 6 or 7 in the state of Sao Paulo between 1989 and 2009 by race (Whites and Blacks/Browns). 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 increased 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 non-Whites 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 a closing of attainment gaps, however. For that to be the case, retention and drop-out rates should converge. For the country as a whole, the evidence on this dimension is mixed. In the case of Sao Paulo the patterns seem more favorable. We conjecture (but do not directly examine) that the adoption of a *social promotion* scheme in Sao Paulo is at least in part responsible for a faster convergence in education attainment between Blacks and Whites. Starting in 1998 such 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.[15]

[Figure I here]

In fact, trends are more pronounced in Sao Paulo than in other parts of the country, and the timing of convergence coincides with the policy's adoption. Yet what most substantiates this argument is the comparison of year-to-year attrition 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. Figure I reproduces a simple computation of Black/White relative survival probabilities in both school systems. Assuming parity at 5th grade (one Black student per White student) we see that within the system adopting social promotion Blacks' relative attrition is lower than within municipal systems in every single year examined.[16] Figure II reinforces this understanding by showing that the reduction in attrition associated with social promotion occurs even conditional on Math proficiency levels. 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.

[15] 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 hampering their learning process. See King et al. (2008).

[16] 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.

[Figure II here]

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 dramatically increase at the high-school level.

3.2.2 Descriptive statistics

Our working data set was obtained after imposing restrictions based on the availability of both transcripts and (concurrent and past) test scores data for at least 75% of the students in a given 8th-grade classroom at the end of 2010. We also restricted our analysis to classrooms with non-homogeneous racial composition and at least fifteen students. We were left with observations on 277,444 students in 10,614 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 (most of which are mixed-race individuals). We identify as Black students that have been declared as such in any survey or enrollment documentation between 2005 and 2012. Table A.I, in the Appendix, presents descriptive statistics for our working data set. It is easy to see that, as expected, 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 unfavorably compared to the latter.

Focusing more specifically on scholastic performance, Figure III plots the cumulative distribution function of test scores (left) and teacher-assigned grades (right). 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 (without conditioning on classroom fixed-effects). A similar pattern is observed in the distribution of teacher-assigned grades, with a disproportionate concentration of Blacks among those obtaining lower marks. Average differences in grades are approximately 5.6 in a 0-100 scale.

[Figure III here]

Finally, in Figure IV we plot the (Lowess) 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 both hold constant other productive attributes that make Black and White students different in the eyes of their teachers and address measurement error challenges. However, before examining the data in more detail, we turn to a simple conceptual framework that guides our estimations and orients the interpretation of results.

[Figure IV here]

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 "competence". 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.

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 have access to noisy signals of the students' proficiency in Math, and observe both their behavior in class and their racial identities. 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 could be reduced by exerting more screening effort, something we implicitly assume teachers either dislike (utility costs) or are reluctant to "purchase" in the market (monetary/opportunity costs), or have limited access to (school authorities may cap the number of tests that can be applied to students in a given year).[17]

Schematically, teacher r inelastically employs a grading/evaluation effort level T_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} taking into consideration i's unobservable true competence (g^*_{ir}) in order to solve on expectation the following optimization problem:

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

where we omit teacher-level subscripts for clarity of exposition.

We impose symmetry and tractability by adopting a simple quadratic function for the disutility generated by evaluation errors:

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

Importantly, we allow 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.[18] That is to say:

$$g^*_i = \alpha_1 p^*_i + \mathbf{a_i'}\boldsymbol{\alpha_2} \tag{3}$$

Teachers do not observe true proficiency directly, so we further assume that they collect a sequence of noisy (yet unbiased) signals $s^t_i = p^*_i + u^t_i$. Signals result from formulating and grading tests/exams, and hence we associate them with evaluation effort ($t = 1, 2, ..., T$). The higher the

[17] One could also conceive a technological constraint here: teaching and testing are complementary activities.

[18] 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).

effort, the more signals will be gathered about each student's proficiency. Teachers' estimator of proficiency can then be described as a combination of those signals and a prior for mean proficiency:

$$\hat{p}_i^* = \frac{\sigma_{p^*}}{\sigma_{p^*} + \sigma_{\bar{u}}} \bar{s}_i + \frac{\sigma_{\bar{u}}}{\sigma_{p^*} + \sigma_{\bar{u}}} \beta_1, \tag{4}$$

where $\bar{s}_i = \frac{\Sigma s_i^t}{T}$, $\sigma_{\bar{u}} = \frac{var(u_i^t)}{T}$ and σ_{p^*} represents the variance of actual proficiency within the student population, while β_1 indicates the average student's proficiency (prior).

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:

$$g_i = \theta \alpha_1 \beta_1 + (1 - \theta) \alpha_1 \bar{s}_i + \mathbf{a_i}' \alpha_\mathbf{2}. \tag{5}$$

From this formulation there are two ways in which statistical racial differentiation can be depicted. The first, rational stereotyping, is based on the idea that attributes including race ($\mathbf{b_i}$) can be informative in the computation of proficiency's best linear projection $E\left[p_i^* | s_i^1, ..., s_i^T, \mathbf{b_i}, \mathbf{a_i}\right]$.[19] In other words, the formualtion of priors regarding group's average proficiency encompasses the use of other characteristics.[20] The case of racial discrimination at hand can be illustrated within our context. Due to social promotion in earlier grades, eighth-grade teachers know that a particularly lenient rule for promoting students was used. They likely assume that such scheme disproportionally affected promotion rates among Blacks. In the absence of any other information teachers will therefore have lower expectations regarding the latter's proficiency levels. If we let $\mathbf{b_i}$ be a scalar corresponding to an indicator $Black_i$ not included in $\mathbf{a_i}$, we can amend the optimal grading equation to:

$$g_i = \theta \alpha_1 \beta_1 + (1 - \theta) \alpha_1 \bar{s}_i + \mathbf{a_i}' \alpha_\mathbf{2} + \theta \alpha_1 \beta_2 Black_i. \tag{6}$$

The second (and not mutually exclusive) possibility is that racial biases materialize as screening discrimination. This is the case when the reliability of proficiency signals collected by teachers is

[19] 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.

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

a function of race. Lang (1986) raised this as possible result of communication difficulties between Whites (teachers) and Blacks (students), while Lundberg and Startz (2007) suggest that they are the outcome of differential rates of social interaction. In our model screening discrimination would be embeded on race-specific signal-to-noise ratios: θ_1 and $\theta_1 + \theta_2 Black_i$. Under these circumstances, the model would deliver the following decision rule:

$$g_i = \theta_1 \alpha_1 \beta_1 + (1 - \theta_1 - \theta_2 Black_i)\alpha_1 \bar{s}_i + \mathbf{a_i}'\alpha_\mathbf{2} + \theta_2 \alpha_1 \beta_1 Black_i, \tag{7}$$

where the practical distinction with respect to Equation (6) would solely come from the inclusion of race-specific effects of average proficiency signals (slopes).

Notice that in any of these representations, racial bias is derived from the imprecision in the information about proficiency contained in the signals. It follows 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

5.1 Practical issues

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. Nonetheless, we also focus solely on ordinal aspects by present results based on the converting of grades assigned by teachers into classroom-specific percentile rankings.

In order to faciliate the interpretation of the practical impacts of our main results we also present two alternative binary 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 an indicator for grades above the classroom's median.

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 (past essays and concurrent Language tests) as controls in our empirical model.

5.2 *Econometric issues*

In essence, and in reference to Equations (5) and (7) of the conceptual framework proposed above, we explore our information regarding scores in standardized Math and Language 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}):

$$g_{ir} = \delta_1 f(scores_{ir}) + \mathbf{a_{ir}}'\boldsymbol{\delta_2} + \mathbf{b_{ir}}'\boldsymbol{\delta_3} + \eta_r + \epsilon_{ir}, \tag{8}$$

where $f(scores_{ir})$ is a polynomial function of the measures 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:

$$g_{ir} = \delta_1 f(scores_{ir}) + \mathbf{x_{ir}}'\boldsymbol{\delta_{21}} + \mathbf{z_{ir}}'\boldsymbol{\delta_{22}} + \mathbf{b_{ir}}'\boldsymbol{\delta_3} + \eta_r + \epsilon_{ir} \tag{9}$$

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.[21]

Measurement error biases result from the fact that despite being associated to the average proficiency measured by teachers, our measure is necessarily noisier. An easy way to understand the discrepancy between the two is to consider that while teachers "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 the aforementioned conceptual framework. In the exercises below we explore the fact that the individual results of standardized tests in Math

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

and Language taken in previous years by each student are available in our data, and employ a fixed-effects instrumental variables estimation that should bypass the measurement error problem. Since we also have access to past proficiency tests covering Natural Sciences' material, we are in addition able to perform overidentification tests.

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 to Language classes in the first half of the academic year); *ii*) parent-reported perceptions 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 half of the academic year). 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 $\mathbf{b_{ir}} = \mathbf{x_{ir}}$):

$$g_{ir} = \delta_1 f(scores_{ir}) + \mathbf{x_{ir}}' \left[\boldsymbol{\delta_{21}} + \boldsymbol{\delta_3}\right] + \tilde{\eta}_r + \tilde{\epsilon}_{ir}, \qquad (10)$$

where race, gender, age, essay scores, parental socio-demographics, and our proxies for behavior are considered elements of the vector $\mathbf{x_{ir}}$ while 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. $f(scores_{ir})$ is estimated as fourth-order polinomyals of Math scores, a linear term for Language scores, and interactions between those. Whenever F-tests indicated that the fourth-order was not significant we opted for presenting results based on a more parsimonious third-order polynomial.[22]

5.3 Learning

We also extend this analysis to explore the heterogeneity of the parameters in (10), according to teacher and student-body characteristics. In particular we pay attention to the amount of knowledge a given teacher has about each of her pupils. Social interaction in the school neighborhood, tenure in a given school and duration of classroom-like interactions for a given student-teacher pair are our main candidates here. In this way we examine the central prediction from our statistical discrimination conceptual framework: learning of students' true types should preclude the use of race as an indicator of scholastic competence.

In practice, and in the spirit of Altonji and Pierret (2001), we test whether racial differentials in teacher-assigned grades diminish (or even disappear) as a teacher's information regarding a student improves. By the same token we examine if such improved information also translates into increased weight given to proficiency signals when end-of-year Math evaluations are issued. If such coefficients are shown to be sensitive we can be more confident that statistical discrimination is at play in our study's environment.

6 Results

6.1 General results

Our initial estimations are derived from the specification in (10). Table I presents the results, illustrating the effect of the addition of controls over racial differentials in our two main dependent variables.[23] Panel A focuses on the Black-White gaps in final grades (0-100 scale). Group averages are presented in column 1. Considering all of the students in our sample, Whites are graded at 61.4

[22] Below we also examine the possibility of δ_1 being a function of race as predicted by screening discrimination arguments.

[23] 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.

on average while grades among Blacks average 55.7. 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) the polynomial for Math and Language contemporaneous standardized scores, and past performance in essays are included. Measured racial gaps are, not suprisingly, significantly reduced. Indeed, a large share of the competence differences seen by teachers is captured by performance in standardized tests of proficiency.

[Table I here]

In column 5 we resort to family background and information on past year's Math grades as additional control variables, with the hope of capturing both abilities and behavioral aspects relevant to the teacher that were not previously controlled for. Proxies for a child's behavioral attributes (self-reported, parent-reported, school-reported), over and above those indirectly captured by family socio-economic background, are included in column 6. 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.[24] Despite the reduction in size, estimated racial gaps are still statistically significant. At this point, conditional gaps are approximately 11 to 14% of the unconditional ones.

Finally, in columns 7 and 8 we tackle the robustness of our findings to the presence of measurement error on the proficiency score variables. As discussed above, because these are used as covariates in our analysis, biases on the estimation of *all* parameters are expected. We therefore employ polynomials of lagged 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 (see first-stage summary statistics in the appendix Table A.II). Moreover, over-identification tests indicate the validity of the sets of instru-

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

ments employed. 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 (marginal effects at the average proficiency level). The racial gaps are still significant after this correction. Indeed, they are statistically significant even when we employ the more stringent Schwarz criterion.[25]

Ultimately we find that Blacks' average Math grades are 0.35 points below those of equally proficient and well-behaved Whites, or that the former are regularly ranked 0.7 percentiles behind the latter. These amount to 6 and 7% of the unconditional gaps, respectivelly.[26] Importantly, the Black-White differentials in teacher-assigned grades we uncover are equivalent to the marginal effect of a reduction of 0.03 to 0.04 standard deviations in proficiency scores.

Table II reproduces these exercises with a focus on meaningful binary variables that summarize cardinal and ordinal gaps. According to these exercises, the measured racial gap in promotion rates between equivalently proficient and well-behaved students corresponds to a 4.1% increase in the retention probability for the average Black (Panel A). Focusing exclusively on the ordinality aspect (Panel B) we also estimate a gap that translates into a 4.5% reduction on the probability of Blacks being graded above the classroom median. These small (yet meaningfull) effects are very much in line with the subtleties we expect to permeate racial discrimination in grading.

[Table II here]

6.2 Robustness of main findings and modeling choices

We also explore expected heterogeneity in the size of racial differentials and its relation to some teacher characteristics (grading practices) to further examine the robustness of our findings to the omission of behavioral characteristics. In Table III (columns 1 and 2), before moving into the comparison across different data strata, we present a summary of the main effects under the full sample and under the subsample for which we have additional teacher characteristics (from survey

[25] 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 statistical significance of results.

[26] 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.

questionnaires). The contrast between these indicate that we should not expect selection biases when dealing with the smaller sample.

In the first set of stratifications (columns 3 to 5) we examine if the gaps in evaluation we measure are not generated by unobserved heterogeneity biases. 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. 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.[27]

[Table III here]

Columns 6 and 7 are solely based on teacher demographics (obtained combining official assignment records and survey questionnaires). We re-estimate our model using fixed-effects instrumental variables techniques for different strata according to teacher's race, which is examined here to investigate in-group biases. We see that no clear pattern emerges from these. Despite significant results among Whites and not among Black and mixed-race teachers, we cannot rule that point estimates are the same. These findings are incompatible with both the idea of taste discrimination (at least in its simplest format) against members of the out-group and the idea that teachers have more information about pupils of their own racial group.

In Table IV we investigate the robustness of our formulation by examining if the marginal impact of proficiency tests over grades are different for Black and White students, as predicted by

[27] In another exercise that examines unobserved heterogeneity biases we estimate if the difference in future drop-out rates between retained and non-retained Blacks were larger than among Whites. If they were it could mean that teachers observe other productive aspects on retained Blacks that they do not see in retained Whites. We find that this is not the case in our data covering 2011 and 2012, however.

the screening discrimination version of the model above. From the estimates presented, we have no evidence to support the idea that slope coefficients should be student-race specific.[28]

[Table IV here]

In principle, however, we cannot rule out 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.

6.3 Learning by grading?

In order to more directly examine the role of imperfect information we explore information on pupil-teacher matches, by utilizing 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. In this case we emulate a student-specific change in grading-effort (T) exerted by her current teacher. Simply put, larger T's should increase signal to noise ratios, increasing the marginal effect of (posterior) proficiency measures at the same time it reduces the one related to characteristics used to construct priors.

[Table V here]

It is clear from estimates in Table V that longer-term teacher-student interactions produce smaller grading gaps associated with racial identity. In other words, this empirical exercise reveals that while Black-White gaps in grades and rakings are salient for students attending classes with a teacher for the first time, no significant disparities are found among those that have already had classroom interactions with that instructor before eighth grade. It is also the case that acquaitance between teacher and students increase the weight given to proficiency scores on the determination of grades or rankings (steeper relation). Both these differences (in intercept and in slope) are shown statistically significant. In practice, Black students that have not interacted with their current teacher before eighth grade have their grades diminished by what is equivalent to a taxation of 0.06

[28] The conclusion remains unchanged if we restrict this analysis to classrooms with White teachers.

standard-deviations in proficiency tests. Those that are known to the teacher are not "taxed" at all. This is our main indication that imperfect information lies at the heart of the discrimination results we estimate.

Further experimenting with these ideas we examine stratifications based on different levels of detail in the information teachers have about their students. In Table VI we start by reproducing in column 1 the differences estimated in Table V. Columns 2, 3 and 4 focus on the proportion of students in a classroom that are "known" to the current teacher. The idea here is that by knowing a suficient number of students, teachers are able to employ relative references to grade their pupils. We do observed that to be the case, particulalry regarding the ranking measures (which we would expect to be more prone to the use of relative referencing). In practice, racial gaps are not observed in classrooms where teachers have had past interactions with at least 50% of the students, while they are in case teachers know relatively less students. The former also give more weight to proficiency scores when assessing competence and defining end-of-year grades.

[Table VI here]

In columns 5, 6 and 7 we turn to the idea that information flows result from teachers' tenure in a given school (and with a given population of students). When we estimate racial gaps employing this idea we find that indeed gaps are larger among teachers that have shorter tenure in the school. Differences in slopes are less precisely estimated but still support the role of information flows.[29] Finally, in columns 8 to 10 we examine if information is also spread via social interactions within the schools' neighborhood. Interestingly, in this case those that would supposedly know more about the student population seem to discriminate more. This makes us believe that the information flows that translate into reduced discrimintion need to be somewhat related to Math abilities, something that teachers do capture within classroom/school settings, but that neighbohors cannot easily estimate.

The robustness of our learning argument can be further put to the test by examining an alternative explanation for the findings above. In particular, we investigate if the assignment of teachers to students captured in our proposed measure of knowledge above is not simply revealing that

[29] We have also estimated models with teacher-student-specific interactions among longer tenure teachers and found room for "learning a student's type" even among those. This most likely means that the level of detail regarding a student's competence is finer when classroom interactions do occur than when information is provided via interactions with other teachers in the same school.

schools that assign teachers to the same students are culturally different (say, in terms of taste for discrimination) or even that students with specific behavioral characteristics select into longer-term interactions with teachers. In Table VII we falsify both these views by presenting evidence that neither Math-grade racial gaps nor its relation to proficiency are a function of length of interaction time between a student and her Language teacher. The exact same conclusion comes about when we use the identity of future Math teachers to measure acquaintance levels. Teachers that *will* spend time with a given student do not discriminate more or less today than those that will not. Alternatively, students that will spend more time interacting with their current Math teachers in the future do not have their racial identification playing a role on evaluations that is different than for those that that will not.

[Table VII here]

We conclude our analysis in Table VIII by verifying that our measures of learning and reduction in racial gaps are not a result of the inclusion of omited interactions in our econometric specification. In particular, we examine if when including interactions between teacher-student relation indicators and other control variables we do not eliminate the differences observed in the race coefficient. In fact we see no reason to believe this is the case. From what we can tell from the estimates, a student racial identity is likely used in our context as an indicator of lower proficiency. Its impact over grades is remediated if teachers get to know (and test) students for longer than an academic year. From this exercise we incidentally uncover some other interesting (yet less significant) patterns: i) social economic background variables seem to have a role similar ro race, with teachers potentially looking at indicators of those (which are not observed directly) to draw their priors regarding a students' competence, and; ii) behavioral traits have a role on the evaluation that is independent of how much a teacher knows about a specific student. Both findings further substantiate the formulation of the conceptual framework as we proposed above.

[Table VIII here]

Finally, in order to gather a sense of the size of these effects and (possibly) the mechancs behind grading discrimination, we explore a simple simulation exercise. We start by converting the blindly graded proficiency scores into a classroom-specific 0-100 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.[30] White students in the exact same situation have $h + 1$ as their assigned grades, following unbiased rounding rules. We then compare the average racial gap in the biased and unbiased conditions. We find that this biased rounding generates racial gaps of 0.94 (in end-of-year grades) and 1.94 (in percentile rankings). Notice, therefore, that the differences between Blacks and Whites we estimate above (particulalrly among teachers that know less about students) are aproximately half the size of 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 over a significant share of the teachers' population, for example. This exercise provides additional confidence that the results we uncover are not ignorable.

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 less likely to be deemed non-competent (below passing grade) than their equally proficient and equivalently well-behaved Black classmates. The former are also relatively more likely to be graded above their classroom median. Quantitatively, these correspond to a 4.1% increase in the retention probability and a 4.5% reduction in the probability of Blacks being at the top of their class grade distribution. Such effects are equivalent to "taxing" Blacks' performance in proficiency tests by 0.03 to 0.04 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

[30] 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.

well intentioned teachers issue report cards for their students with subtle biases (possibly incurred when rounding continuous marks into a discrete scale, for example) and end up adding obstacles to the acquisition educational credentials by Blacks. These are meaningfull effects resulting from racial discrimination in grading.

We find that these biases most likely result from imperfect information and statistical discrimination or, in other words, from the weighted combination of noisy proficiency signals extracted from teacher-designed exams and stereotyped priors. In the 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 is clearly shown to be the case in our data, particularly when we use the length of classroom-interaction time between a teacher and a given student.

Our findings lead to important education 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 for screening purposes the latter's race becomes. 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 grading discrimination within schools. We leave this for our future research on the topic.

University of Sao Paulo

University of Sao Paulo

Sanford School of Public Policy - Duke University and BREAD

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 Guimarāes; 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] Autor, David and David Scarborough (2008); "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments", *The Quarterly Journal of Economics*, 123 (1): 219-277.

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

[11] Ben-Zeev, Avi; Tara C. Dennehy; Robin I. Goodrich; Branden S. Kolarik; and Mark W. Geisler (2014); "When an âĂIJEducatedâĂİ Black Man Becomes Lighter in the MindâĂŹs Eye Evidence for a Skin Tone Memory Bias", *SAGE Open*, 4(1), January.

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

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

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

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

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

[22] 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.

[23] 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.

[24] Farber, Henry and Robert Gibbons (1996); "Learning and Wage Dynamics" *The Quarterly Journal of Economics*, 111 (4): 1007-1047.

[25] 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).

[26] 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.

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

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

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

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

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

[32] 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.

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

[34] 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.

[35] Lang, Kevin (1986); "A Language Theory of Discrimination", *The Quarterly Journal of Economics* 101 (2): 363-382.

[36] 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.

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

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

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

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

[41] Lundberg, S. and R. Startz (2007); "Information and racial exclusion". *Journal of Population Economics*, vol. 20(3), pages 621-642, July

[42] Madeira, R. and M. A. Rangel (2013); "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.

[43] 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.

[44] 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.

[45] Neal, D. (2006); "Why has the Black-White Skill Convergence Stopped?" Handbook of the Economics of Education, Volume 1.
bibitem Ost, Ben (2014); "How Do Teachers Improve? The Relative Importance of Specific and General Human Capital." *American Economic Journal: Applied Economics*, 6(2): 127-51.

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

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

[48] 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.

[49] 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.

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

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

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

[53] 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.

[54] 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.

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

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

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

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

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

Table I

Unconditional and Conditional Racial Differentials in Grading - OLS and IV Estimations

	Averages	Black-White raw gaps	Black-White conditional gaps FE	FE	FE	FE	IV-FE	IV-FE
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: End-of-year assessment by teacher (0-100 scale)								
White	61.4							
Black	55.7	-5.664***	-4.966***	-1.581***	-0.698***	-0.646***	-0.349***	-0.347***
		(0.115)	(0.105)	(0.089)	(0.079)	(0.074)	(0.083)	(0.083)
Proficiency in Math				3.806***	2.025***	1.823***	9.238***	8.475***
				(0.058)	(0.052)	(0.048)	(2.779)	(0.782)
Over-ID test (J-statistic [p-value])							1.488 [.2225]	1.478 [.2240]
Panel B: Intra-classroom percentile rank of end-of-year assessment by teacher (0-100)								
White	41.9							
Black	32.1	-9.813***	-9.887***	-3.250***	-1.522***	-1.369***	-0.735***	-0.721***
		(0.200)	(0.205)	(0.177)	(0.161)	(0.153)	(0.175)	(0.172)
Proficiency in Math				8.124***	4.650***	4.268***	24.780***	20.442***
				(0.117)	(0.107)	(0.102)	(5.501)	(1.540)
Over-ID test (J-statistic [p-value])							1.659 [.1977]	1.971 [.1604]
Controls								
Classroom fixed-effects		No	Yes	Yes	Yes	Yes	Yes	Yes
Child demographics		No	No	Yes	Yes	Yes	Yes	Yes
Performance in standardized tests		No	No	Yes	Yes	Yes	Yes	Yes
Family background + 2009 Math gı		No	No	No	Yes	Yes	Yes	Yes
Behavioral traits		No	No	No	No	Yes	Yes	Yes
Order of scores' polynomial		-	-	4th	4th	4th	4th	3rd

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. Sample consists of 277,444 students in 10,614 classrooms. Marginal effect of proficiency scores evaluated at the mean proficiency level (for the population) are presented. Controls consist of classroom fixed-efefcts, child's gender and age polynomial (second order), a 4th-order (3rd order) polynomial function of concurrent Math z-scores interacted with Language z-scores and past performace in essays. Family background includes maternal education, age, region of birth (in or out of state), home ownership, ownership of automobiles, and number of wc's in the household. Behavioral traits include reports of parents regarding child's interest for school work, effort regarding studies and overall behavior. They also include Physical Education grades and Languague classes attendance rates for the first half of the school year. Finally, self-reported measures of behavior are included with indicators of procrastination with homework, class-skiping and interest in extra-curricular Math activities.

Table II

Unconditional and Conditional Racial Differentials in Grading - OLS and IV Estimations

	Averages	Black-White raw gaps	FE	FE	FE	FE	IV-FE	IV-FE
				Black-White conditional gaps				
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Above or at passing grade indicator (x 100)								
White	91.6							
Black	86.3	-5.246***	-4.922***	-1.613***	-0.771***	-0.875***	-0.568***	-0.570***
		(0.225)	(0.214)	(0.204)	(0.198)	(0.192)	(0.205)	(0.202)
Proficiency in Math				3.348***	1.642***	1.379***	3.243	3.412*
				(0.127)	(0.122)	(0.118)	(7.146)	(1.890)
Over-ID test (J-statistic [p-value])							0.576 [.4478]	0.495 [.4817]
Panel B: Above classroom median grade indicator (x 100)								
White	39.6							
Black	26.2	-13.486***	-13.510***	-4.637***	-2.296***	-2.060***	-1.200***	-1.177***
		(0.304)	(0.310)	(0.278)	(0.259)	(0.253)	(0.279)	(0.275)
Proficiency in Math				11.381***	6.656***	6.151***	33.425***	27.088***
				(0.187)	(0.176)	(0.170)	(8.715)	(2.456)
Over-ID test (J-statistic [p-value])							1.603 [.2055]	1.998 [.1575]
Controls								
Classroom fixed-effects		No	Yes	Yes	Yes	Yes	Yes	Yes
Child demographics		No	No	Yes	Yes	Yes	Yes	Yes
Performance in standardized tests		No	No	Yes	Yes	Yes	Yes	Yes
Family background + 2009 Math gr		No	No	No	Yes	Yes	Yes	Yes
Behavioral traits		No	No	No	No	Yes	Yes	Yes
Order of scores' polynomial		-	-	4th	4th	4th	4th	3rd

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. Sample consists of 277,444 students in 10,614 classrooms. Marginal effect of proficiency scores evaluated at the mean proficiency level (for the population) are presented. Controls consist of classroom fixed-efefcts, child's gender and age polynomial (second order), a 4th-order (3rd order) polynomial function of concurrent Math z-scores interacted with Language z-scores and past performace in essays. Family background includes maternal education, age, region of birth (in or out of state), home ownership, ownership of automobiles, and number of wc's in the household. Behavioral traits include reports of parents regarding child's interest for school work, effort regarding studies and overall behavior. They also include Physical Education grades and Languague classes attendance rates for the first half of the school year. Finally, self-reported measures of behavior are included with indicators of procrastination with homework, class-skiping and interest in extra-curricular Math activities.

Table III

Conditional Racial Differentials in Grading by Teacher's Evaluation Practices and Race - IV Estimations

			Teacher's Grading Practices			Teacher's Race	
	Full sample [1]	Responding quests. [2]	Objective grader [3]	Subjective grader [4]	Non-Objective grader [5]	White grader [6]	Black + Mixed grader [7]
Panel A: End-of-year assessment by teacher (0-100 scale)							
Black	-0.347***	-0.340***	-0.426***	-0.380***	-0.283**	-0.362***	-0.265
	(0.083)	(0.089)	(0.145)	(0.105)	(0.114)	(0.092)	(0.194)
Proficiency in Math	8.475***	8.153***	8.687***	6.732***	7.987***	8.809***	7.021***
	(0.782)	(0.834)	(1.560)	(1.008)	(0.994)	(0.855)	(1.948)
Panel B: Intra-classroom percentile rank of end-of-year assessment by teacher (0-100)							
Black	-0.721***	-0.743***	-0.957***	-0.884***	-0.627***	-0.741***	-0.582
	(0.172)	(0.187)	(0.305)	(0.220)	(0.236)	(0.191)	(0.414)
Proficiency in Math	20.442***	20.412***	18.323***	17.748***	21.597***	21.277***	16.666***
	(1.540)	(1.642)	(2.958)	(1.967)	(1.978)	(1.666)	(4.170)
Sample students	*277,444*	*233,750*	*86,485*	*171,727*	*147,846*	*224,936*	*52,198*
Sample teachers	*10,614*	*8,925*	*3,305*	*6,548*	*5,641*	*8,596*	*2,006*

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. See notes in Table 1.

Table IV
Marginal Effects of Proficiency over Grades by Student's Race

	Base model [1]	Interacted model [2]	Differential [2]
Panel A: End-of-year assessment by teacher (0-100 scale)			
Proficiency in Math	8.475***		
	(0.782)		
Proficiency in Math x White		8.290***	
		(0.344)	
Proficiency in Math x Black		8.549***	0.259
		(0.751)	(0.786)
Panel B: Intra-classroom percentile rank of end-of-year assessment by teacher (0-100)			
Proficiency in Math	20.442***		
	(1.540)		
Proficiency in Math x White		18.611***	
		(0.744)	
Proficiency in Math x Black		19.048***	0.437
		(1.501)	(1.593)
Sample students	*277,444*	*277,444*	*277,444*
Sample teachers	*10,614*	*10,614*	*10,614*

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. Sample consists of 277,444 students in 10,614 classrooms. See notes in Table 1.

Table V

Conditional Racial Differentials in Grading and Learning Students' Types - IV Estimations

	Math teacher knows student [1]	Math teacher does not know [2]	Difference [3] = [2] - [1]
Panel A: End-of-year assessment by teacher (0-100 scale)			
Black	-0.092	-0.427***	-0.335*
	(0.172)	(0.095)	(0.197)
Proficiency in Math	11.664***	7.517***	-4.147**
	(1.612)	(0.925)	(1.862)
Panel B: Intra-classroom percentile rank of end-of-year assessment by teacher (0-100)			
Black	0.079	-0.960***	-1.038**
	(0.364)	(0.197)	(0.414)
Proficiency in Math	24.999***	19.119***	-5.880*
	(3.063)	(1.826)	(3.572)

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. Sample consists of 277,444 students in 10,614 classrooms. See notes in Table 1. Teachers are identified as knowing a given student if they have taught in classes to which the student was assigned between 2007 and 2009.

Table VI
Conditional Racial Differentials in Grading and Learning Students' Types - IV Estimations by Information Level

	Difference when knowledge specific to student	Classroom-level acquaintance rate: Math teacher knows 50% or more of class	Math teacher knows less than 50% of class	Difference	Tenure in school: Math teacher in school 3 years or more	Math teacher in school less than 3 years	Difference	Knowledge of neigborhood: Math teacher from school neighborhood	Math teacher NOT from school neighborhood	Difference
	[1]	[2]	[3]	[4] = [3] - [2]	[5]	[6]	[7] = [6] - [5]	[8]	[9]	[10] = [9] - [8]
Panel A: End-of-year assessment by teacher (0-100 scale)										
Black	-0.335*	-0.113	-0.422***	-0.309	-0.220**	-0.617***	-0.397**	-0.390***	-0.198	0.193
	(0.197)	(0.168)	(0.098)	(0.194)	(0.110)	(0.162)	(0.195)	(0.101)	(0.196)	(0.220)
Proficiency in Math	-4.147**	10.736***	7.493***	-3.243**	9.128***	5.053***	-4.075*	8.105***	8.219***	0.114
	(1.862)	(1.251)	(0.999)	(1.601)	(0.920)	(1.935)	(2.144)	(0.907)	(2.350)	(2.516)
Panel B: Intra-classroom percentile rank of end-of-year assessment by teacher (0-100)										
Black	-1.038**	0.088	-1.006***	-1.093***	-0.507**	-1.249***	-0.741*	-0.926***	-0.169	0.757*
	(0.414)	(0.353)	(0.203)	(0.408)	(0.229)	(0.334)	(0.404)	(0.213)	(0.396)	(0.450)
Proficiency in Math	-5.880*	23.823***	19.176***	-4.648	21.478***	17.378***	-4.100	19.759***	23.375***	3.617
	(3.572)	(2.461)	(1.965)	(3.144)	(1.813)	(3.728)	(4.145)	(1.778)	(4.770)	(5.083)
Sample students	*277,444*		*277,444*			*233,750*			*233,750*	
Sample teachers	*10,614*		*10,614*			*8,925*			*8,925*	

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. See notes in Table 1. Teachers are identified as knowing a given student if they have taught in classes to which the student was assigned between 2007 and 2009. This info is aggregated at the classroom level in columns (2) and (3). Tenured is defined from responses to teacher questionnaires. The same is the case for neighborhood, which is a function of how far the teacher has to travel to teach in a given school.

Table VII

Conditional Racial Differentials in Grading and Learning Students' Types - IV Estimations for Falsification of Hypothesis

	Language teacher knowledge of student			Future Math teacher knowledge of student		
	Lang. teacher knows student	Lang. teacher does not know	Difference	Future Math teacher knows student	Future Math teacher does not know	Difference
	[1]	[2]	[3] = [2] - [1]	[4]	[5]	[6] = [5] - [4]
Panel A: End-of-year assessment by teacher (0-100 scale)						
Black	-0.347*	-0.359***	-0.012	-0.463*	-0.326***	0.137
	(0.186)	(0.093)	(0.208)	(0.247)	(0.088)	(0.263)
Proficiency in Math	9.151***	8.358***	-0.793	9.512**	8.146***	-1.366
	(1.424)	(0.939)	(1.705)	(4.131)	(0.783)	(4.219)
Panel B: Intra-classroom percentile rank of end-of-year assessment by teacher (0-100)						
Black	-0.267	-0.841***	-0.574	-0.754	-0.709***	0.046
	(0.405)	(0.191)	(0.446)	(0.510)	(0.186)	(0.546)
Proficiency in Math	20.533***	20.645***	0.112	18.323**	20.144***	1.821
	(2.857)	(1.842)	(3.402)	(7.918)	(1.562)	(8.101)

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. Sample consists of 277,444 students in 10,614 classrooms. Teachers are identified as knowing a given student if they have taught in classes to which the student was assigned between 2007 and 2009 in columns (1) and (2). Future teachers are identified as knowing a given student if they are teaching and will teach in classes to which the student is going to be assigned in 2010 and 2011 in columns (4) and (5). See notes in Table 1.

Table VIII

Conditional Racial Differentials in End-of-year assessment by teacher (0-100 scale) and Learning Students' Types - IV Estimations for Signals Beyond Race

	Base Model		Interactions with SES added		Interactions with behavior added	
	Math teacher knows student	Math teacher does not know	Math teacher knows student	Math teacher does not know	Math teacher knows student	Math teacher does not know
	[1]	[2]	[3]	[4]	[5]	[6]
Black	-0.092	-0.427***	-0.106	-0.423***	-0.109	-0.423***
	(0.172)	(0.095)	(0.186)	(0.093)	(0.216)	(0.105)
Proficiency in Math	11.664***	7.517***	11.690***	7.516***	11.691***	7.519***
	(1.612)	(0.925)	(1.619)	(0.925)	(1.614)	(0.925)
Male	-2.876***	-3.052***	-2.864***	-3.055***	-2.857***	-3.057***
	(0.152)	(0.087)	(0.152)	(0.087)	(0.153)	(0.087)
Family background (SES)						
Mom HS grad.			0.274**	0.309***	0.277**	0.308***
			(0.131)	(0.074)	(0.132)	(0.074)
Mom some college			0.120	0.385**	0.131	0.383**
			(0.316)	(0.177)	(0.316)	(0.177)
Mom college grad.			0.254	0.427***	0.261	0.425***
			(0.277)	(0.150)	(0.277)	(0.150)
Home ownership			0.074	0.160**	0.073	0.161**
			(0.124)	(0.067)	(0.124)	(0.067)
Behavioral traits						
Well behaved (parental report)					1.164***	1.267***
					(0.121)	(0.070)
Poorly behaved (parental report)					-0.813***	-0.553***
					(0.218)	(0.125)
High-effort behaved (parental report)					0.953***	0.951***
					(0.142)	(0.083)
Low effort (parental report)					-0.846***	-0.738***
					(0.149)	(0.083)
Level of interest in school work (parental report)					0.490***	0.462***
					(0.032)	(0.018)
PE grades					0.692***	0.763***
					(0.040)	(0.023)
School attendance (Language classes)					0.118***	0.109***
					(0.008)	(0.004)
School attendance (self-report)					1.184***	1.171***
					(0.185)	(0.101)
Don't procastinate (self-report)					1.741***	1.868***
					(0.128)	(0.073)
Enrolled in Math extra curricular (self-report)					0.967***	0.972***
					(0.123)	(0.067)

Notes: Standard-errors in parentheses are clustered at the classroom level. *** 1%, ** 5% and * 10% significance levels. Samples consist of 277,444 students in 10,614 classrooms. See notes in Table 1.

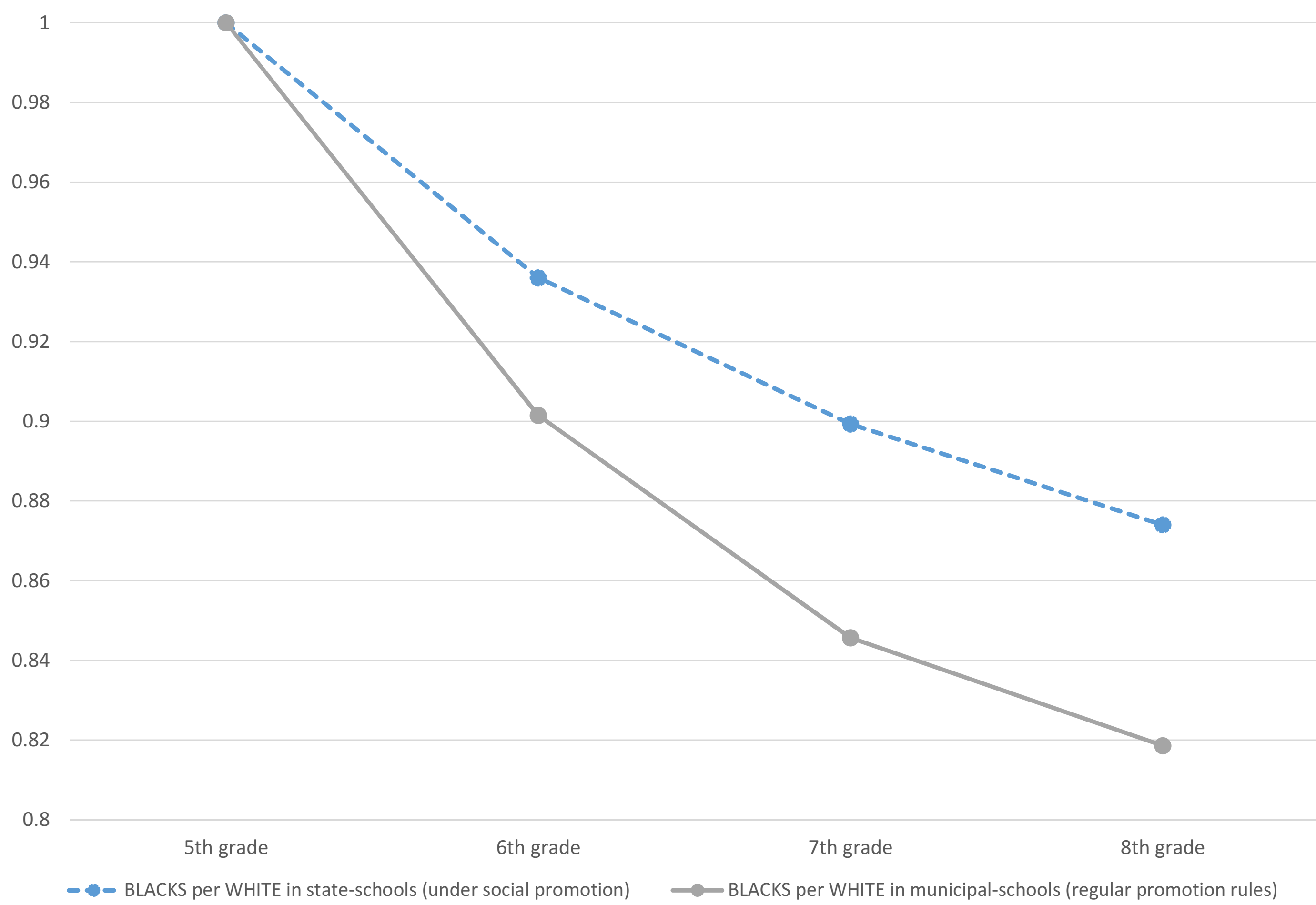

FIGURE I: DIFFERENTIAL ATTRITION (5TH TO 8TH GRADE) BY RACE AND SCHOOL SYSTEM'S PROMOTION RULES

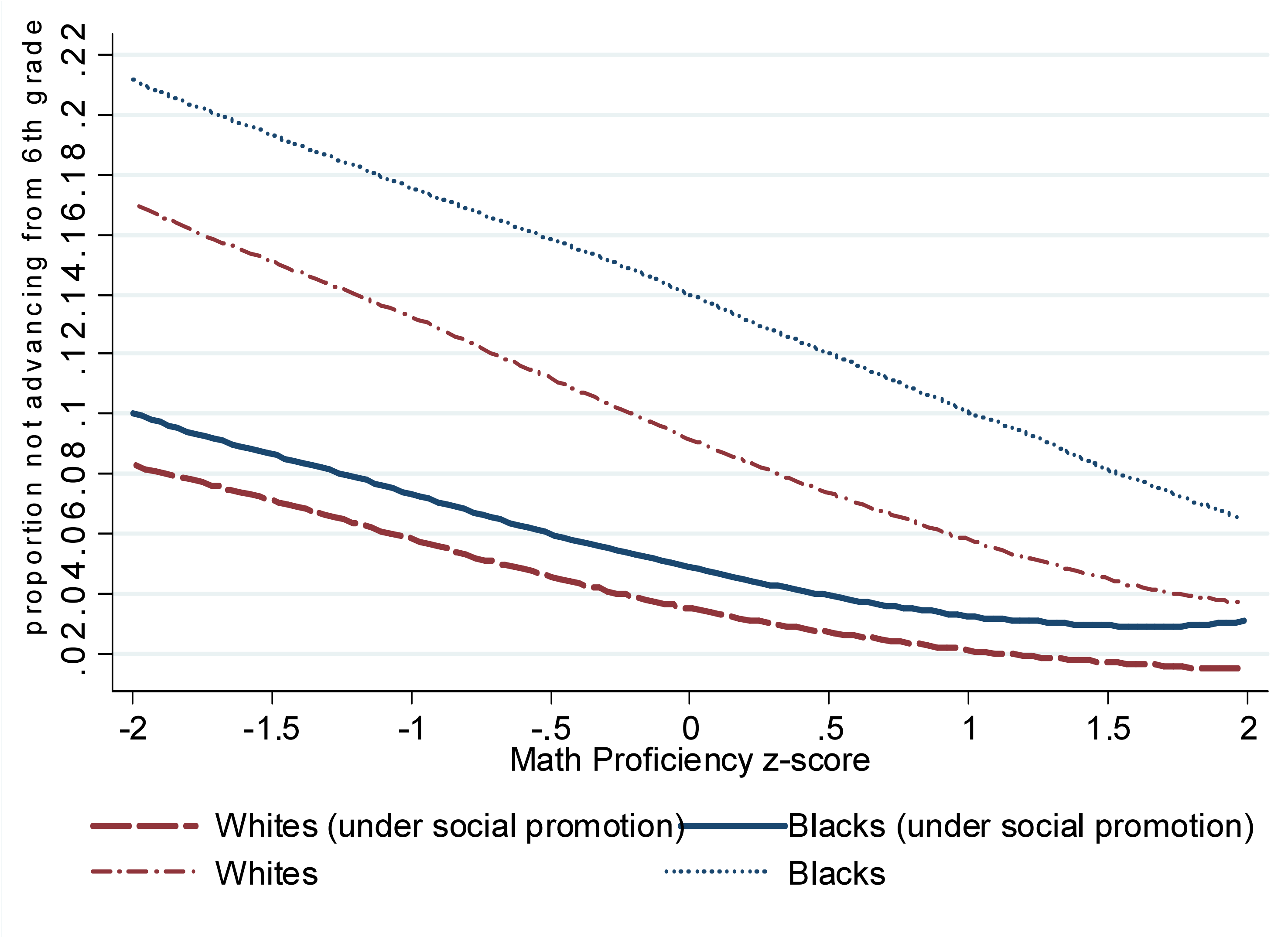

FIGURE II: DIFFERENTIAL ATTRITION FROM 6TH GRADE BY RACE, SCHOOL SYSTEM'S PROMOTION RULES AND PROFICIENCY LEVEL

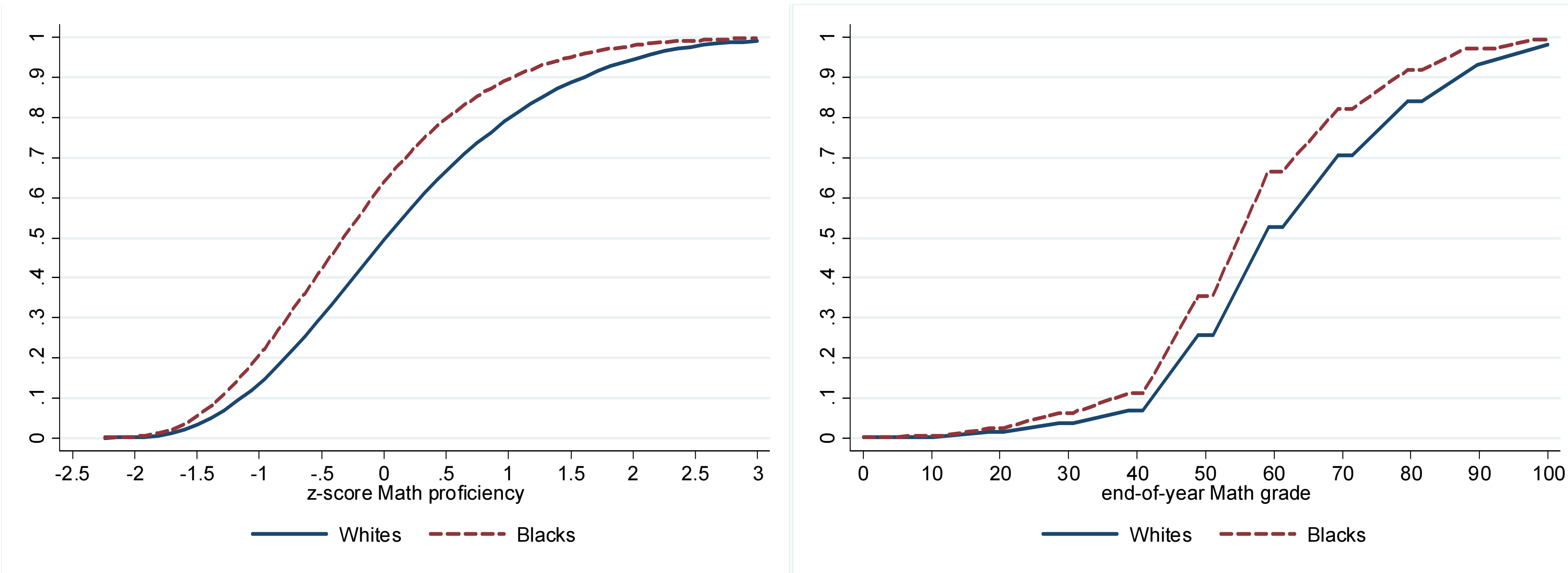

FIGURE III: CUMULATIVE DISTRIBUTION FUNCTIONS FOR PROFICIENCY SCORES AND TEACHER-ASSIGNED GRADES FOR 8TH GRADERS

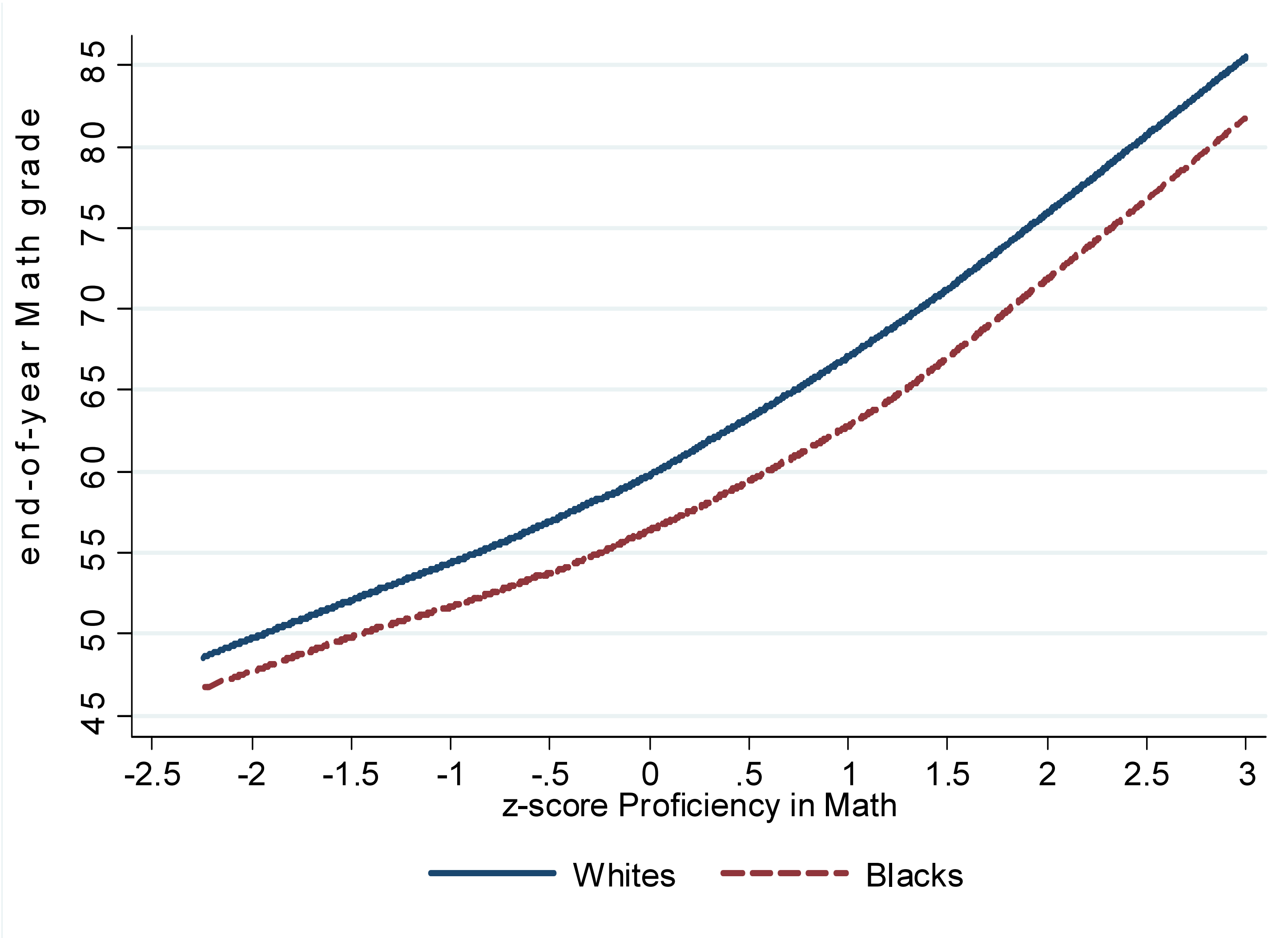

FIGURE IV: SMOOTHED RAW RELATION BETWEEN PROFICIENCY SCORES AND TEACHER-ASSIGNED GRADES FOR 8TH GRADERS

Table A.I

Descriptive statistics

	Whites mean (se)	Blacks mean (se)	Black-White (classroom FE) mean (se)
Proficiency and school performance			
8th grade Math scores (z-scores)	0.134 (0.0050)	-0.220 (0.006)	-0.249 (0.006)
8th grade Language scores (z-scores)	0.202 (0.0050)	-0.243 (0.006)	-0.343 (0.006)
Proficiency in past Math exam	217.755 (0.1800)	201.999 (0.242)	-11.521 (0.250)
Proficiency in past Language exam	215.870 (0.1920)	197.434 (0.262)	-14.018 (0.265)
Proficiency in past Sciences exam	238.236 (0.2330)	216.271 (0.321)	-16.501 (0.325)
Family background			
Mom HS grad	0.221 (0.002)	0.170 (0.002)	-0.037 (0.003)
Mom some college	0.030 (0.001)	0.024 (0.001)	-0.003 (0.001)
Mom college grad	0.044 (0.001)	0.030 (0.001)	-0.008 (0.001)
Mom aged 16 to 24	0.007 (0.000)	0.011 (0.001)	0.003 (0.001)
Mom aged 25 to 34	0.169 (0.001)	0.173 (0.002)	0.006 (0.003)
Mom aged 45 to 59	0.170 (0.001)	0.163 (0.002)	0.000 (0.002)
Mom 60 or older	0.007 (0.001)	0.010 (0.001)	0.003 (0.001)
Home owned	0.541 (0.002)	0.507 (0.004)	-0.030 (0.003)
Number of cars owned	0.583 (0.003)	0.406 (0.004)	-0.119 (0.004)
Number of exclusive-use wc's in household	1.069 (0.004)	0.899 (0.006)	-0.111 (0.005)
Behavioral traits			
Well behaved (parental report)	0.426 (0.002)	0.316 (0.003)	-0.092 (0.003)
High effort (parental report)	0.161 (0.001)	0.135 (0.002)	-0.023 (0.002)
Interest level 0-10, (parental report)	5.892 (0.020)	5.315 (0.027)	-0.391 (0.022)
PE grade in first bi-monthly evaluation	7.015 (0.012)	6.686 (0.016)	-0.160 (0.011)
Language classes attendance in first bi-monthly evaluation	91.705 (0.052)	90.639 (0.077)	-0.528 (0.058)
Does not skip classes often (self report)	0.689 (0.002)	0.635 (0.003)	-0.031 (0.003)
Does not procrastinate with homework (self report)	0.250 (0.002)	0.183 (0.002)	-0.054 (0.003)
Enrolled in extr-curricular Math activities (self report)	0.446 (0.003)	0.371 (0.003)	-0.045 (0.003)

Notes: Standard-errors in parentheses are clustered at the classroom level. Estimation of differences conducted including classroom fixed-effects. Samples consist of 277,444 students in 10,614 classrooms, of which 10.2% are Black and 44.5% are White. Remainder consists of mixed-race/Brown population.

Table A.II

First-Stage Regressions' Summary Statistics

	4th-order polynomial		3rd-order polynomial	
	F-test of instruments	P-values	F-test of instruments	P-values
Endogenous variables				
Proficiency score in Math (z-score)	734.89	0.000	863.64	0.000
Proficiency score in Math (z-score) squared	144.13	0.000	1645.96	0.000
Proficiency score in Math (z-score) to the third	348.09	0.000	666.91	0.000
Proficiency score in Math (z-score) to the fourth	152.27	0.000	-	0.000
Language score x Proficiency score in Math (z-score)	756.6	0.000	3465.75	0.000
Language score x Proficiency score in Math (z-score) squared	745.81	0.000	1343.68	0.000
Language score x Proficiency score in Math (z-score) to the third	280.32	0.000	1032.86	0.000
Language score x Proficiency score in Math (z-score) to the fourth	251.02	0.000	-	0.000
Proficiency score in Language (z-score)	3419.5	0.000	7994.92	0.000

Notes: Samples consist of 277,444 students in 10,614 classrooms. Instruments are polynomiasl of past test scores in Math and Language and past test scores in Natural Sciences.