WORKING PAPER Nº IDB-WP- 01586

Barriers to Immigrant Assimilation Evidence on Grading Bias in Ecuadorian High Schools

Marcos A. Rangel
Luana Marotta
Cynthia van der Werf
Suzanne Duryea
Marcelo Drouet
Lucina Rodríguez Guillén

Inter-American Development Bank
Migration Unit
Education Division
Gender and Diversity Division

March 2024

Barriers to Immigrant Assimilation Evidence on Grading Bias in Ecuadorian High Schools

Marcos A. Rangel *
Luana Marotta **
Cynthia van der Werf **
Suzanne Duryea **
Marcelo Drouet **
Lucina Rodríguez Guillén****

* Sanford School of Public Policy, Duke University
** Inter-American Development Bank
*** Vancouver School of Economics, The University of British Columbia

Inter-American Development Bank
Education Division
Gender and Diversity Division

March 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Barriers to immigrant assimilation: evidence on grading bias in Ecuadorian high schools / Marcos A. Rangel, Luana Marotta, Cynthia Van Der Werf, Suzanne Duryea, Marcelo Drouet, Lucina Rodríguez Guillén.
p. cm. - (IDB Working Papers Series ; 1586)
Includes bibliographic references.
1. Human capital-Ecuador. 2. Immigrant youth-Education (Higher)-Ecuador. 3. Discrimination in education-Ecuador. I. Rangel, Marcos. II. Marotta, Luana. III. Werf, Cynthia van der. IV. Duryea, Suzanne. V. Drouet, Marcelo. VI. Rodríguez Guillén, Lucina. VII. Inter-American Development Bank. Migration Unit. VIII. Inter-American Development Bank. Gender and Diversity Division. IX. Inter-American Development Bank. Education Division. X. Series.
IDB-WP-1586

http://www.iadb.org

Copyright © 2024 Inter-American Development Bank ("IDB"). This work is subject to a Creative Commons license CC BY 3.0 IGO (https://creativecommons.org/licenses/by/3.0/igo/legalcode). The terms and conditions indicated in the URL link must be met and the respective recognition must be granted to the IDB.

Further to section 8 of the above license, any mediation relating to disputes arising under such license shall be conducted in accordance with the WIPO Mediation Rules. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the United Nations Commission on International Trade Law (UNCITRAL) rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this license.

Note that the URL link includes terms and conditions that are an integral part of this license.

The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

Barriers to immigrant assimilation: Evidence on grading bias in Ecuadorian high schools[*]

Marcos A. Rangel
Duke University[‡]

Luana Marotta
IADB[§]

Cynthia Van Der Werf
IADB[¶]

Suzanne Duryea
IADB[‖]

Marcelo Drouet
IADB[**]

Lucina Rodríguez Guillén
UBC [††]

March 2024

[*]This document was conducted by the Migration, Education and Gender, Diversity and Inclusion specialists at the Inter-American Development Bank. This study falls under the Perception Initiative of the Migration Unit from the IADB, which aims to measure opinions and attitudes towards migrant populations and generate evidence that reduces prejudice through cost effective interventions. We are grateful for the suggestions by Eliana La Ferrara, Ana María Ibañez, Felipe Muñoz as well as by the participants of the 2022 IDB Initial Workshop, the 2023 HUMANS LACEA Network, the Association for Education Finance and Policy (AEFP) Education Development meeting, and the 2023 LACEA meeting. This study was made possible thanks to the Ministry of Education of Ecuador, which gave us access to administrative records and provided relevant information that contributed to the interpretation of our results.

[‡]Sanford School of Public Policy, Duke University; 262 Rubenstein Hall, Box 90312, Durham, NC 27708; Email: marcos.rangel@duke.edu.

[§]Inter-American Development Bank; 1300 New York Ave NW, Washington, DC 20577; E-mail: luanac@iadb.org.

[¶]Inter-American Development Bank; 1300 New York Ave NW, Washington, DC 20577; E-mail: cvanderwerf@iadb.org.

[‖]Inter-American Development Bank; 1300 New York Ave NW, Washington, DC 20577; E-mail: suzanned@iadb.org.

[**]Inter-American Development Bank; 1300 New York Ave NW, Washington, DC 20577; E-mail: msdrouet01@gmail.

[††]The University of British Columbia; Vancouver, BC V6T 1Z4; E-mail: lucina24@student.ubc.ca.

Abstract

We investigate the assimilation of immigrant youth in Ecuador. Focusing on formal schooling and employing administrative data from high schools, we document subtle ways by which assessment biases against students with an immigrant background play a significant role in this assimilation process. We find that, after holding constant performance on blindly scored proficiency tests, teacher-assigned grades in Mathematics and Spanish are consistently lower for students from immigrant families. We show that these results are robust with respect to the omission of socio-emotional and behavioral traits that are likely valued by teachers. These differentials are larger for male students and those attending urban schools. While these grading differentials have direct impact over high school graduation rates, they may also discourage future human capital investments, potentially leading to lower college attendance, distorted choice of major, and suboptimal labor market outcomes, which are all well know elements for the economic assimilation of immigrants.

JEL: I24, J15
Keywords: Immigration assimilation; human capital; teacher discrimination; grading bias.

1 Introduction

Children and adolescents are a particularly vulnerable group of immigrants and face multiple barriers to gaining an education. Among those barriers is the lack of proper documentation that may impede school enrollment after arriving in the destination country, resulting in disruptions in school attendance. When they are enrolled in school, immigrant students may encounter additional hurdles, including grade-level misplacement, the need to adapt to different curricula, and, potential discrimination by peers and teachers. This article focuses on specific aspects of the latter while recognizing that such phenomenon manifest itself in many different ways. In particular, we focus on a relatively understudied form of teacher discrimination: a teacher's biased evaluation of students' proficiency and aptitude.

Employing data from Ecuadorian high schools we examine whether teachers may show a propensity to discriminate against students from immigrant families when assigning grades. We employ detailed administrative data covering approximately 180,000 high school students spread across nearly 6,500 classrooms in the 2018-2019 academic year. Our inference is based on juxtaposing teacher-assigned grades with students' scores from the *Ser Bachiller*, a nationally administered standardized, and blindly scored proficiency test taken at the end of the academic year and covering the same official curriculum delivered in regular classes. Our results indicate that relative to native students with identical academic performance, there is statistically significant underscoring and under-ranking of immigrant students according to the subjective assessment of teachers.[1]

We find that immigrants have 10% higher chance of being retained for insuf-

[1] As more precisely described in section 4, our definition of "immigrants" refers to students with at least one household member, including themselves, who are in a situation of human mobility. We only focus on international migrants and refugees.

ficient performance on math and language when compared with equivalently proficient native students with similar socio-economic and socio-emotional characteristics. In general, when compared with native students who perform similarly on standardized academic tests, teacher-assigned grades for immigrant students are significantly lower. These differentials are equivalent to "taxing" the performance of immigrant students in standardized tests on Math and Language by 0.1 to 0.15 standard-deviations, respectively. Throughout, we show that results are shown to be robust to possible omissions of students' behavioral attributes, which have been found to influence teachers' grades (Ferman and Fontes, 2022).

The Ecuadorian context we study is particularly rich due to recent migration flows from Venezuelans seeking to escape the political and economic crisis in their homeland. In fact, Figure 1 shows the number of immigrants arriving in Ecuador has increased considerably over the last ten years, mostly fueled by the Venezuelan crisis. Importantly and unlike previous literature which explored settings where immigrants and natives are noticeably different in terms of race, language, economic background, and culture (Alesina et al., 2018; Carlana et al., 2022), Ecuador provides a unique context because most migrant families come from neighboring countries with similar social and cultural backgrounds. Our study, therefore, adds to the literature by documenting that grading discrimination prevails even in a setting where students from immigrant families—hereafter, immigrant students—are similar to (and, in some aspects, more privileged than) their native-born peers.

We also present evidence on the heterogeneity of these biases. There are clear indications that grading biases against immigrants are higher among male students, a finding that seems to be consistent with the broader existing literature: generally, there is more discrimination against immigrant men compared to native men than against immigrant women compared to native women (Gereke et al., 2020; Ji et al., 2021). Moreover, we also find that biased teacher assessments are more harmful for students at the top of the socio-economic and performance distribution. Finally, we find that grading bias is higher among instructors teaching in urban schools.

After establishing the presence of bias in teacher assessments in Ecuadorian high schools, our remaining analyses rely on economic theory to examine its likely source in that context. As discussed in Section 5, we characterize teachers as relatively sophisticated agents who evaluate students employing observed characteristics that are thought to be correlated with competence. In this case, the characteristics themselves convey information and can "help" teachers generate better assessments. This source of bias is often referred to as statistical discrimination (Altonji and Pierret, 2001; Blume, 2006; Bjerk, 2008; Lehmann, 2011). In our context, there are some reasons why teachers may associate migrant status with their students' level of ability. First, due to the rapid influx of Venezuelan immigrants, teachers may incorrectly or erroneously perceive that immigration authorities employed lenient standards for the admission of recently-arrived students into Ecuadoran high schools. Teachers are well aware of the implications of such potentially altered admission policies, and initial expectations regarding their students' proficiency may be lower as a result. Second, the vulnerable condition of many immigrants when entering the receiving country may lead teachers to believe that these immigrant families come from unstable backgrounds, characteristics which are correlated with poor academic performance. Therefore, we hypothesize that when teachers issue reports assessing the competence of their students, subtle biases may be generated by the weighted combination of noisy information extracted from the instructors' own screening exams and stereotyped beliefs or priors.

We examine this hypothesis by estimating heterogeneous effects by teachers' level of interest, class size, and the total number of students taught by an instructor. We assume that more interested teachers and those who teach fewer classes and smaller class sizes will have more opportunities to interact with each individual student and form a better judgment about his or her true level of ability, thus relying less on stereotypical priors about immigrants. We do find that grading bias is smaller for more interested instructors, and teachers with smaller-sized classes and teach fewer students. However, heterogeneous effects on the use of signals are not significant so that we cannot rule out the possibility that the assessment bias we uncover is resulting solely from animus or anti-immigrant sentiment of teachers (taste discrimination).

We call attention to the implication of our findings can be far reaching given the enormous potential for feedback effects between assessment biases and human capital investments. This is the case because we detect discrimination in grading during the transition between high school and either college or the labor market, at a time when students and parents invariably find themselves in the position of investors relying on the asset- return evaluations of more informed experts. In this case, intra-classroom evaluation biases may very well lead to gaps in college attendance and, ultimately, labor market outcomes, which are clearly essential elements of the immigrant assimilation process.

2 Context

Over the last decade, Ecuador has experienced considerable growth in its immigrant population, which is largely due to the Venezuelan migration crisis. As figure 1 shows, between 2010 and 2020, the number of migrants has more than doubled, going from 375,000 to 785,000 (International Migrant Stock, 2020). The composition of migrants in the country also changed. In 2010, Colombians accounted for 59 percent of Ecuador's migrant population, but, in 2020, 51 percent of the country's migrants came from Venezuela, and only 25 percent came from Colombia (DataMIG, 2023). As Ecuador allows all migrants, even those without a permit, to enroll in K-12 school, the share of foreign-born students enrolled in the Ecuadorian school system increased accordingly. The percentage of immigrant students in the Ecuadorian educational system was 0.5% for the 2009-2010 school year, but increased to 2.1% for the 2021-2022 academic year (Administrative Records - Ministry of Education Ecuador).

Since most immigrant families living in Ecuador come from neighboring countries that share common cultural and linguistic traits, its migrants usually share similar characteristics, including observable traits, with the native population. Like Ecuadorians, most immigrants speak Spanish and, in some regions, they have a similar accent–for instance, the accent of people from Ecuador's coastal region is similar to the Venezuelan accent (Palacios, 2017). Moreover, on average immigrants have higher levels of education: according to the National Employment Survey, 29 percent of the immigrants in Ecuador have a college degree, as opposed to 18 percent of natives (Encuesta Nacional de Empleo, Desempleo y Subempleo-ENEMDU, 2021).

Despite sharing a largely common linguistic and cultural heritage, tension between immigrants and Ecuadorians has grown; data show that in Ecuador, attitudes and beliefs towards migrants have worsened considerably in recent years. Gallup's Migrant Acceptance Index reveals that, while in 2016 72 percent of the Ecuadorian population believed that immigrants were good for the country, only 27 percent felt this way by 2019. Along the same lines, in 2016, 84 percent of Ecuadorians said it was a good thing to have migrants moving into their neighborhood, but only 48 percent held the same view in 2019. Moreover, according to the Latinobarómetro, an annual public opinion survey with respondents from diverse Latin American countries, in 2020, only 26 percent of Ecuadorians agreed that immigrants were good for the country's economy, while the regional average was 43 percent. Similarly, in 2020, 79 percent of Ecuadorians believed that immigrants cause an increase in crime, whereas only 56 percent of the region's population thought immigrants are associated with a rise in criminal activity.

In school settings, results from standardized skill assessments show that immigrant and native students perform at similar levels. Figure 2 shows the distribution of math and language test scores for immigrant and native students from the 2018 Ser Bachiller, the test data used in our study. The distributions in this chart reveal that immigrant students have slightly higher test scores compared to Ecuadorian students. Despite this evidence, teachers believe that immigrant students have poorer academic results than native pupils. In a survey conducted with almost 1,800 K-12 teachers in Ecuador between 2021 and 2022, we asked: "Who, in your opinion, tends to achieve better academic results?".[2] Among elementary school teachers, 17 percent said that native students achieve better results, 4 percent said that immigrant students perform better, and 79 percent said both perform at the same level. High school teachers are even more pessimistic regarding the performance of immigrant students: 21 percent said that native students perform better academically than immigrant students. As figure 3 shows, we also found that Ecuadorian teachers at different grade levels believe that native students are slightly more motivated, work harder, have greater family support, and are better prepared academically than immigrant students.

[2]We cannot verify if the teachers who participated in this 2021-2022 survey were included in the current study's sample, from the 2018-2019 school year, because the datasets do not have common identifiers for the teachers.

We also administered an Implicit Association Test (IAT) to a group of teachers who participated in this 2021-2022 survey and found that they displayed a moderate to strong implicit preference for natives over Venezuelans. The IAT measures the strength of associations between concepts (e.g., black people, LGBTQ people) and adjectives (e.g., good, bad) or stereotypes (e.g., athletic, clumsy), considering the time it takes for individuals to make these associations. As part of the 2021-2022 survey study, the IAT was applied so that teachers had to associate concepts related to Ecuadorians/Venezuelans with good/bad connotations. Figure 4 shows the distribution of the IAT scores, indicating that teachers tend to make associations that favor Ecuadorians over immigrant students.

3 Related Literature

The question of whether teachers treat children of different backgrounds differently is well-established. There is a tradition within the sociology literature of directly examining whether in the United States, teacher-related bias is a factor in assigning grades for courses (Bowles and Gintis, 1976; Farkas et al., 1990; Rist, 1973; Rosenthal and Jacobson, 1968; Sexton, 1961). Earlier work has only found modest biasing effects on teachers' grades (Sewell and Hauser, 1980; Williams, 1976). There also are a considerable number of contributions from the social psychology literature focusing on teacher perceptions of Black and White children (see Ferguson (1998) and references therein), which again only find weak relationships between Black stereotypes and measures of discriminatory actions.[3]

Unlike earlier studies, more recent literature, with more robust methods, shows that teachers' evaluations are not free from bias. Figlio (2005) examines whether

[3] See review of studies in Dovidio et al. (1996).

teachers' overall perception of a given student is affected by the "Blackness" of his or her first name, even after controlling for the student's performance on standardized examinations. Using data from one school district in Florida, he uncovers evidence that teachers have lower expectations for those students perceived to have African American ancestry. Burgess and Greaves (2013) investigate differences in teacher-assigned grades according to a student's ethnic background using observational data from England; they find significant underassessment of pupils with Black Caribbean and Black African ancestry. Finally, Hinnerich et al. (2011a,b) conduct audit-like studies by transcribing and blindly re-grading tests assessed by teachers in Sweden and estimate gaps based on gender (insignificant) and nationality (significant). A similar exercise conducted in Germany by Sprietsma (2013) also uncovers biases against exam results for students who had Turkish-sounding names randomly allocated to them (relative German-sounding names).

Another approach to assess teachers' bias is to juxtapose their subjective evaluations (i.e. grades) with blind assessments of student performance. One set of papers capitalizes on the fact that students in Israeli high schools take two examinations covering the same material with the same format during senior year, and that the grading of each exam occurs under different anonymity regimes. Using the blind score as the counterfactual to the non-blind score issued by the teacher's assessment, Lavy (2008) finds evidence of discrimination against male high school students. Furthermore teacher biases based on class-level gender differences have both short and long-term consequences for male and female human capital accumulation (Lavy and Sand, 2018; Lavy and Megalokonomou, 2019).[4] Blind and non-blind contrasts in assessing academic

[4] Terrier (2020) similarly shows teacher favoritism towards girls using blind and non-blind test scores, and finds that, as a result, females are more likely to choose a high school science track. Avitzour et al. (2020) probed the origins of these biases and document a correlation between *implicit* gender stereotypes and teacher assessment behavior.

performance are also explored in a randomized control trial designed and implemented by Hanna and Linden (2012). The authors identify statistically significant positive differences between blinded and non-blinded scores for members of lower castes in India, relative to upper castes, which offer clear evidence of discrimination. Finally, Burgess and Greaves (2013) and Botelho et al. (2015) use large-scale observational data from England and in Brazil, respectively, to investigate differences in teacher-assigned grades according to their students' ethnic/racial backgrounds. Both studies juxtapose objective tests with subjective teacher assessments and document significant underassessment of Black pupils.

Our study builds on this approach by employing both blind and non-blind assessments of student performance over the same skill set being assessed. We use large-scale observational data from Ecuador that provides plausibly objective measures of student math and reading mastery along with more subjective teacher evaluations of the same underlying skillset. Therefore, our blind and non-blind measures are well-suited for the task at hand, as both measures are taken contemporaneously. While our juxtaposition of teacher assessments and standardized test scores aims to capture evaluation bias, we acknowledge that this measure stops short of fully exploring the biased teacher behaviors embedded in the very test scores that anchor our models. These biased teacher assessments we measure hold constant teacher effort that may directly influence students' end-of-year test scores, for example.[5] We argue that negative feedback induced by biased assessments of students' academic aptitude and achievement can still be harmful, above and beyond the knowledge acquired during the school year. For instance, students who fail a math class may be less likely to choose a math-intensive major or decide not to attend college, perma-

[5] See also Brown and Bigler (2005) for a discussion regarding the United States context.

nently altering their earnings trajectory.

The discussion presented here focuses on four contributions provided by our context with respect to other studies in the literature. First, we examine discrimination within the high school setting, a period of the educational trajectory when a student's awareness of prevailing stereotypes is higher. This is the case because the stakes are higher, due to the importance of teenager's formation of identity and social connections as they navigate the path to adulthood (Seider et al., 2019, 2022; Altschul et al., 2006; Elenbaas and Killen, 2017). In particular, (McLoyd et al., 2009) indicates that teenagers acquire more skills accounting for self-awareness and the perspectives of others, which in itself contributes to a heightened ability to consciously assess racist (or xenophobic) prior assumptions, motivations, and decisions rendered by people they interact with. We argue that the high school student's ability to detect these behavioral patterns increases the impact of negative feedback effects that may be induced by teachers' biased assessments.

Second, we investigated discrimination related to national origins at a time during which Ecuador was on the receiving end of the mass exodus of Venezuelans fleeing the humanitarian crisis in their country. The migratory crisis was likely consequential for influencing Ecuadorians' attitudes and beliefs towards immigrants, as sociotropic concerns related to the impact of migration on the natives' own economic opportunities (Adida et al., 2019; Bansak et al., 2016; Hainmueller and Hiscox, 2010; Hainmueller and Hopkins, 2014, 2015; Valentino et al., 2019) and concerns about how migration changes local customs and traditions (Adida et al., 2019; Bansak et al., 2016; Hainmueller and Hopkins, 2014; Hopkins, 2010) often drive opposition to immigrants. Indeed, as discussed in section 2, the rapid growth in Ecuador's immigrant population since 2015, and especially after 2017, seems to have boosted anti-immigration beliefs among natives. We argue that this significant "change of heart" should also be pertinent for understanding the attitudes of Ecuadoran teachers, who within a fairly short period, started interacting with a larger population of immigrant students.

Third, we are well positioned to employ rich administrative data sources (more on this below) to complement previous studies finding evidence of discrimination within schools. This prior work includes Alesina et al. (2018), Alan et al. (2020), Glock et al. (2013) and Carlana et al. (2022) who examine this question in experimental settings where evidence of discriminatory behavior may be due to the one-shot nature of the event (even when incentivizing schemes curb hypothetical biases). In contrast, the sheer size and level of detail in our data base allow us to convey a complete portrait of teacher and student-body characteristics associated with discrimination in actual classroom environments. Two features of the setting provide additional advantages. First, in our context grading is weakly monitored so subtle discriminatory behavior, in the form of biased grades issued by classroom teachers, is rarely detected by the school authorities or students themselves. In addition, information regarding standardized test performance is not disclosed to teachers. Therefore, teachers' assessment is not influenced by more objective information about students' performance.

Finally, the context analyzed in this paper is unique compared to previous studies of discrimination against migrants within schools. The cultural and religious differences between Ecuadorians and immigrants are minimal compared to those differences among natives and immigrants in contexts analyzed by existing research (Ibanez et al., 2022; Olivieri et al., 2022). For example, Alesina et al. (2018), who also explore teacher grading discrimination against immigrant students using a similar methodological approach, use data from Italy, where the differences between the native-born and the immigrant population are more conspicuous. As noted by Carlana et al. (2022), most of the immigrants living in Italy come from low and middle-income countries and have more deprived socioeconomic backgrounds compared to native households. In Italy, immigrants are also more likely to work at low-skilled jobs (Mariani et al., 2020). In the Italian setting, where the immigrant population is markedly different compared to the native population, Alesina et al. (2018) found that immigrant children receive lower teacher-assigned grades than natives after controlling for their performance on standardized tests. Moreover, they found that such gaps in assessments between natives and immigrants are correlated with teachers' harboring negative stereotypes toward immigrants. Given the setting and data used in this paper, we are able to explore a similar question regarding grade discrimination under a context with greater cultural, linguistic, and socioeconomic proximity between natives and immigrants.

4 Data

Analysis is based on data of public school students from Ecuador who were in the last year of high school in the 2018-19 academic year. Information on these students comes from two data bases. First, the Ministry of Education's records of student transcripts in all subject areas. The second is the 2018-2019 version of the *Ser Bachiller* standardized tests and its companion questionnaires (covering socio-emotional traits and socio-economic/demographic profiles of students). The *Ser Bachiller* was an exam administered to all Ecuadorian students finishing high school between the years of 2014 and 2019. However, in this study, we only use *Ser Bachiller* data from public school students, for whom transcript records were available.[6]

[6] Ecuador has four types of schools: publicly-funded (77%), private (18%), publicly funded charter schools (4%), and municipal (1%) schools. This paper uses data only from publicly-funded schools.

The *Ser Bachiller* was a high-stakes evaluation, since students' performance determined whether they could graduate from high school and weighed in college admission decisions. Specifically, the score on the test contributed 30 percent to a final high school score, calculated by the Ministry of Education, that determines whether or not each student may graduate [7]. It is important to note that the *Ser Bachiller* scores of individual students were not available to schools. Therefore, teachers' evaluations could not be influenced by the performance of students on this standardized exam.

Although students' test scores and grades are available for different subject areas, this paper only focuses on math and language, for which the underlying skillset measured by the standardized test and teacher evaluation are more closely aligned, as proposed by Alesina et al. (2018). The grades range from 0 to 10, and students are required to reach a minimum of 7 to pass the school year [8].

Lastly, our measure of a student's immigration status comes from the following answers to the *Ser Bachiller* contextual questionnaire: "Has any member of your household, including you, been in a situation of human mobility (migration, return, seeking refuge)?"–and in its original language, *"¿Algún miembro de tu hogar, incluyéndote a ti, ha estado en situación de movilidad humana (migración, retorno, refugio)?"*–and "Which type of mobility?"–*"¿Qué tipo de movilidad?"*. Based on these two questions, a pupil is considered to have an immigrant background if s/he or any other family member (e.g. mother, father,

[7] After finishing the third and last year of high school, a score was calculated centrally by the Ministry of Education that took into account students' grades in all three years of high school and their scores on the *Ser Bachiller* test. This combined score ranged from 0 to 10 and students were required to reach a minimum of 7 to graduate. This study only uses students' grades in the third and last year of high school as well as their *Ser Bachiller* scores.

[8] Note that students may reach a minimum of 7 in all courses during the last year of high school and, yet, are not able to graduate. As described in the previous footnote, to graduate from secondary education, they need to reach a minimum of 7 in the combined score that takes into account grades from all high school years and performance in the *Ser Bachiller* test

brother or sister) have been in a situation of human mobility classified as an “international migrant” or “refugee.” [9]

It is important to note that there are two caveats about how the *Ser Bachiller* measures human mobility. First, it does not have information on the students' nationality. However, as described in section 2, we know that, during the 2018-2019 school year, most immigrant students in Ecuadorian high schools came from Venezuela and Colombia. Second, this measure is incomplete, as 51 percent of the data is missing. Because the *Ser Bachiller* contextual questionnaire is large (320 questions), students were required to answer different blocks of questions, with some common items, such as gender and race. Nevertheless, the assignment to these different blocks of questions was random. Table 1 shows that, as expected, the group of students with and without missing the measure of human mobility do share similar characteristics.

Data on covariates describing students' socioeconomic background and socioemotional characteristics also come from the *Ser Bachiller* contextual questionnaires. A summary of the data and description of each variable are provided on Table 2. This summary only includes students from classrooms where we can identify at least one immigrant and one native student because our main strategy uses classroom fixed effects to detect discrimination in grading.[10]

5 Conceptual Framework

We focus our attention on a stylized description of how grade discrimination may arise, and that leads directly into our empirical specifications. The model

[9]We argue that Ecuadorian students with immigrant relatives may also experience discrimination that is similar to what their foreign-born peers experience. As research shows, children of a foreign-born family suffer the negative effects of immigration regardless of their own citizenship status–see Heinrich et al. (2023) for a review of this literature.

[10]In our analysis, we also include students from classrooms without variation in migrant status to improve the precision of our estimates.

is by no means intended to be general, but rather used as a descriptive device to emphasize a particular source of differentiation in teachers' assessments. In principle, there are two basic reasons why teachers might systematically misevaluate the competence of students with certain individual 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 evaluate (hard to measure) competencies by also using observed characteristics perceived to be correlated with academic mastery. In this case, the outward characteristics themselves convey information, and can "help" teachers generate biased assessments. These alternative sources of discrimination are well-known in the economics literature. The like/dislike rationale is a loose representation of taste discrimination (Becker, 1957), whereas the second falls under the realm of statistical discrimination (Arrow et al., 1971; Phelps, 1972; Aigner and Cain, 1977). In our model we highlight the operation of the second, concentrating sole attention on the screening role of twelfth-grade instructors. We model how statistical discrimination operates, by focusing on the screening role played by twelfth-grade instructors, and how their priors/beliefs may affect these assessments, and contribute to grade discrimination.

The basic intuition governing our model of grade discrimination is that teachers have access to noisy signals of the students' proficiency in math and language, and observe both an individual student's behavior in class and his or her national identity. We define an objective function for the task of grading by assuming that teachers essentially operate as statisticians, and are compelled to maximize the power of the hypothesis test embedded in the evaluation of a student's competence. In addition, we impose the condition that teachers weight Type I and Type II errors symmetrically (i.e.: excessive lenience and excessive rigor are equally unwelcome). Evaluation errors can be reduced by exerting more screening effort, something we implicitly assume teachers either dislike to perform (utility costs), have limited access to better screening technologies due to high monetary/opportunity costs, or even that school authorities limit the number of tests that can be administered to students in a given year (the costs of effort could then be modeled as a function of distance to the "norm", such as in Holmstrom and Milgrom (1991)).[11]

This descriptive model of grade discrimination, based on statistical discrimination, can be quantified. 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:

$$\min_{g_i} E\left[\sum_{i=1}^{n} \frac{1}{2}\left(g_i - g_i^*\right)^2\right], \tag{1}$$

where we omit teacher-level subscripts for clarity of exposition and impose symmetry and tractability by adopting a simple quadratic function for the disutility generated by evaluation errors.

Importantly, we allow teachers to broadly define competence. As in Mechtenberg (2009), teachers acknowledge true proficiency (p^*_i) and other directly observed scholastic attributes ($\vec{a_i}$) as elements to be rewarded. Mechtenberg (2009) refers to the latter as *attitudes*, which we envision as a broad concept that includes habits, styles, behavior, and any other socio-emotional or person-

[11]One could also envision a technological constraint that limits the choices of teaching and testing effort.

ality traits deemed *productive* by teachers.[12] That is to say:

$$g_i^* = \alpha_1 p_i^* + \vec{a_i'}\vec{\alpha_2} \tag{2}$$

Teachers do not observe true proficiency directly, so we further assume that they collect a sequence of noisy (yet unbiased) signals $s_i^t = p_i^* + u_i^t$. Signals result from formulating and grading tests/exams, and hence we associate them with evaluation effort ($t = 1, 2, ..., T$).[13] The higher the 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:

$$\hat{p}_i^* = \frac{\sigma_{p*}}{\sigma_{p*} + \sigma_{\bar{u}}}\bar{s}_i + \frac{\sigma_{\bar{u}}}{\sigma_{p*} + \sigma_{\bar{u}}}\beta_1, \tag{3}$$

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 teachers to follow when assigning grades:

$$g_i = \theta\alpha_1\beta_1 + (1 - \theta)\alpha_1\bar{s}_i + \vec{a_i}'\vec{\alpha_2}. \tag{4}$$

From this formulation, there are two ways to depict statistical differentiation using a student's nationality. The first, rational stereotyping, is based on the idea

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

[13] For clarity of exposition, measurement error in teacher's tests is considered classical. We acknowledge that, due to the bounded nature of grading scales in most of these classroom tests, errors would be negatively correlated with the student's true proficiency level. As long as the absolute value of the covariance between the error and the true proficiency is smaller than the noise variance (Black et al., 2000), introducing non-classical measurement error does not alter in any way the main tenets of the model.

that attributes, including nationality, ($\vec{b_i}$) can be informative in the computation of proficiency's best linear projection $E\left[p_i^*|s_i^1,...,s_i^T,\vec{b_i},\vec{a_i}\right]$.[14] In other words, the formulation of prior expectations regarding a group's average proficiency levels encompasses the use of other individual characteristics.[15]

The case at hand, discrimination based on national origin, can be illustrated within our context. Due to the rapid inflow of immigrant students, twelfth-grade teachers now may assume that a particularly lenient rule for grade-promoting students was used. In the absence of any other information, teachers will therefore have lower expectations regarding the proficiency levels of immigrant students. If we let $\vec{b_i}$ be a scalar corresponding to an indicator $Immigrant_i$ not included in $\vec{a_i}$, we can amend the optimal grading equation to:

$$g_i = \theta\alpha_1\beta_1 + (1-\theta)\alpha_1\bar{s}_i + \vec{a_i}'\vec{\alpha_2} + \theta\alpha_1\beta_2 Immigrant_i. \qquad (5)$$

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 a function of nationality or cultural background. Lang (1986) raised this as a possible result of communication difficulties between Whites (teachers) and Blacks (students), while Lundberg and Startz (2007) suggest that such biases are the outcome of differential rates of social interaction. In our model screening discrimination would be embedded on nationality-specific signal-to-noise ratios: θ_1 and $\theta_1 + \theta_2 Immigrant_i$. Under these circumstances, the practical distinction with respect to Equation (5) would solely come from the inclusion of nationality-specific effects of average

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

[15] 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-stereotypical stimulus (regarding educational attainment). We see this as a version of implicit association biases.

proficiency signals (slopes).

Notice that in all of these representations, nationality bias is derived from the imprecise information about proficiency contained in the signals. It follows that improvements in the signal-extraction technology should make a student's immigrant status a less relevant element of the grade assignment process. At the same time, the relationship between teacher-issued grades and individual test scores results on blindly scored exams should be strengthened. This would be the case if teachers were to (exogenously) increase grading effort, if new information were distributed, or if the tests administered by the teachers were made less noisy. We use this simple model to test the data, emphasizing its prediction regarding learning a student's true proficiency level. Further discussions on alternative specifications and identification challenges are presented in the empirical section below.

6 Empirical Strategy

6.1 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, how to compare the grades assigned by different teachers is not straightforward. 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 (1) point gain in class *A* can hardly be compared to the same absolute gain in another class *B* if the spread of grades is due to teachers using different grading standards.[16] At first, we simply put aside this concern and use grades as our dependent variable, but we do so recognizing that (within this scale) measured gaps have both cardinal and ordinal meanings.

To facilitate the interpretation of the practical impacts of our main results we also present an alternative binary dependent variable: an indicator of minimum competence. This was made common among teachers by the establishment of a common passing grade across the Ecuadoran school system. So, independently of a teacher's choices regarding the 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 students who are evaluated as having skill level at or above grade 7 (seven) are deemed competent while those below are not. This cardinal notion ought to be common across all classrooms, even if there are different levels of stringency (captured by a class fixed-effect), depending on the individual teacher.

A second practical concern is the different natures of the exams administered by teachers within the school context and the external standardized tests used for monitoring a student's mastery of skills. Teachers' evaluations of students' proficiency should reflect the same skills and cognitive abilities as measured by the standardized exam because both teachers and exam follow the same curriculum. Yet, it is plausible that proficiency 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 only capture reading skills via a multiple- choice exam. We, therefore, expect the objectivity inherent in the material to translate itself

[16] In other words, the non-additive nature of this grading heterogeneity implies that linear fixed-effects will not wash them out.

into skills more easily measured in a test-like format. However, as previously discussed in the literature (Bettinger, 2012; Alesina et al., 2018), this problem is less severe for math as the skills assessed by standardized tests in this subject area tend to be more aligned with those in teacher evaluations.

6.2 Econometric Issues

In essence, 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 classroom examinations. Meanwhile, other skills also considered relevant by teachers are factored into the *productive attributes* term ($\vec{a_i}$). Therefore, we propose the following empirical representation that incorporates teacher/classroom fixed-effects (η_r) and a pupil-level disturbance term (ϵ_{ir}):

$$g_{ir} = \delta_1 f(scores_{ir}) + \vec{x_{ir}}'\vec{\delta_{21}} + \vec{z_{ir}}'\vec{\delta_{22}} + \vec{b_{ir}}'\vec{\delta_3} + \eta_r + \epsilon_{ir}, \tag{6}$$

where $f(scores_{ir})$ is a function of a student's performance on the objective test results available in our data that replaces the "theoretical" average level of proficiency captured in teacher-designed examinations ($\bar{s}_{ir}$), and once again $\vec{b_{ir}}$ lists elements affecting teachers' priors with regard to proficiency. Meanwhile, to make explicit the further challenges to our empirical exercise, the elements in the vector of scholastic attributes ($\vec{a_i}$) are also decomposed into observed and unobserved components, with $\vec{x_{ir}}$ representing elements observed both by teachers and the econometricians and $\vec{z_{ir}}$ standing for those only observed by the former. The outcome (g_{ir}) represents students' grade, which ranges from 0 to 10.

Given that our central objective is to consistently estimate δ_1 and $\vec{\delta_3}$, this simple empirical representation highlights a potential econometric problems: unobserved heterogeneity. Unobserved heterogeneity adds a layer of complications because elements of $\vec{b_{ir}}$ may very well be related to elements of $\vec{z_{ir}}$. In particular, we worry about behavioral indicators that are available to teachers during classroom interactions and are correlated with nationality.[17] 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 student's self-reported perceptions of their work habits, socio-emotional characteristics, and behavior, all of which have been found to be strong predictors of how teachers assign grades (Ferman and Fontes, 2022).

Ultimately, our main empirical model consists of regressing students' grades on their migration status while controlling for their standardized test scores, race, gender, whether they repeated a grade, mother's education (as a proxy of socio-economic background), and indicators of work habits, socio-emotional characteristics and behavior. These are all considered elements of the vector $\vec{x_{ir}}$ while the remaining elements of $\vec{z_{ir}}$ not observed by the econometrician are either absorbed by the classroom fixed-effects or by the disturbance term.

6.3 Learning

We also extend the analysis to explore the heterogeneity of the parameters according to teacher and student-body characteristics. In particular, we pay attention to the amount of knowledge a given teacher has about each of his or her students. In this way, we examine the central prediction from our statistical discrimination conceptual framework: learning about a student's true ability should preclude the use of national identity as a biased indicator of scholastic competence.

[17] Cornwell et al. (2013) face a similar issue in the case of gender differentials in grading.

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

We use three proxies of teachers' opportunity to interact with each individual student and learn her or his true ability level: a) teachers' interest, b) class size, and c) total number of students taught by teachers. For the first measure, we use the following proxy from the *Ser Bachiller* student questionnaire to gauge teacher's interest: "At your school, are most of your teachers interested in their students doing well?" (an in its original language, *En tu escuela, ¿La mayoría de tus maestros está interesado en que los estudiantes estén bien*?). We calculated the proportion of students in each classroom who answered "yes" to this question and divided the sample into "less interested teachers" (below the median) and "more interested teachers" (above the median). For the second measure, we use administrate data to see how many students are in each classroom [18]. For the third measure, we calculate the total number of students taught by each teacher considering all high school classrooms where she or he works [19]. Teachers have more opportunity to learn about the true ability of each one of their students when they have smaller sized classes and when they teach fewer students overall (Marotta, 2019; Elacqua and Marotta, 2020). For these measures, we also divided the sample in half to determine "smaller" versus "larger"

[18] In Ecuador, students attend all courses in an academic year with the same group of students. Therefore, the class size in Lenguage, for example, is the same as the class size in Math.

[19] We are only able to observe teachers and students in the last year of high school. Therefore, the third measure does not count the number of students taught by teachers in other grade levels.

class sizes, and teachers who teach "fewer" and "more" students. Because class size and number of students taught are associated with school location, which in turn was found to be related to teacher bias, we also estimate these heterogeneous effects controlling by school fixed-effects.[20]

7 Results

7.1 General Results

Figure 5 shows the unconditional relationship between teacher-assigned grades and students' standardized test scores for immigrant and native students. For language and math, immigrant students receive lower grades from their teachers despite similar performance on the standardized exam when compared to native students. The econometric strategy employed in this paper attempts to verify if this gap persists after controlling for attributes of migrant and native students such as their socioeconomic status and behavioral traits, that may influence teacher evaluations.

Table 3 shows the first results of this exercise, which finds significant gaps between migrant and native students in teacher-assigned grades, conditional on a set of basic characteristics (demographics and socio-economic variables, described on table 2). The first column shows that there is a difference of 0.074 grade points in math and 0.096 in language between immigrants and natives, favoring the latter [21]. Column 2 controls for classroom fixed effects, while columns 3 and 4 add students' test scores and socio-demographic characteristics, respectively. According to the model with all controls (column 4), teacher-assigned

[20] We found that, in our context, grading bias was much more prevalent among teachers larger in urban areas.

[21] The outcome, students' grade assigned teachers, ranges from 0 to 10 and has a mean score of 7.68 and standard deviation of 1.03 in Math and a mean score of 7.96 and standard deviation of 1.13 in Language.

grades for immigrant students are significantly lower by 0.043 points in math and 0.052 points in language when compared with native students with similar academic performance on standardized tests. Moreover, we found that, for immigrant students, the likelihood of reaching the passing grade (that is, a grade equal to or greater than 7) is 0.6 percentage points lower than native students for both math and language.

Table 4 expands the analysis and investigates the robustness of findings with respect to the omission of socio-emotional characteristics. Considering that students' socio-emotional traits are often accounted for in teachers' subjective evaluations (Ferman and Fontes, 2020), the omission of these traits will likely bias the results if immigrants are perceived to behave differently in class. We also control for parental involvement, which can also influence the grades that teachers assign to students. The proxies for work habits, socio-emotional traits, classroom behavior, and parental involvement are described in table 2. After controlling for these possible confounders, discrimination in teacher-assigned grades decreases, but the bias does not disappear. It is important to note that proxies of student socio-emotional characteristics might be also influenced by teacher (or school) bias. For example, the measure of behavior is based on whether students were suspended from school. If schools are more likely to suspend immigrant students, we are underestimating teacher grading bias by controlling for that proxy of behavior.

7.2 Heterogeneity

Table 5 shows heterogeneity in grading differentials between immigrants and natives based on a student's gender, socioeconomic background, and academic performance. Results indicate that evaluation bias against immigrants is only applicable to males. Teachers do not seem to penalize female immigrants when evaluating their scholastic competence. One potential explanation for this finding is that negative priors associated with immigrants in Latin America (e.g. incidence of violence) are typically ascribed to males (Gereke et al., 2020; Ji et al., 2021). These stereotypes may, therefore, be used as an informational signal when teachers estimate the level of proficiency of immigrant students who are male, but not female.

We also find that grading bias is larger among students at the top of the socio-economic and performance distributions. We conjecture that these are resulting from the heterogeneity of priors teachers form about individuals in the opposite ends of the income/performance distribution. This highlights the existence of a glass ceiling applicable to immigrants. Finally, in table 6, we show that teacher grading bias is higher in urban schools. This result is somewhat surprising, as in other contexts, the Observatory of Public Attitudes to Migration (OPAM) has found that people living in rural areas have been shown to hold more negative perceptions of immigrants, probably because migrants relocating to rural areas tend to have lower levels of schooling and income (another predictor of discrimination). However, anecdotal evidence from focus groups in Ecuador suggests that rural communities are more welcoming to immigrants and tend to treat them more equally. This can explain why grade discrimination may be less prevalent in rural locations, and more prevalent in urban locations.

7.3 Learning

Albeit imperfectly, we explore the quality of teachers' grading technology when looking at their opportunity to learn about their students' true ability. Under statistical discrimination, the longer pupils and teachers interact, the less teachers will rely on stereotyped priors to grade their students, behavior which lowers the potential for introducing biases into student assessments.

To explore potential statistical discrimination, in table 7 we see if grading bias is lower for more interested teachers, teachers with smaller class sizes, and instructors who teach fewer students overall (considering other classrooms). As discussed previously, these attributed should be associated with teachers' opportunity to interact with each individual student and get to know his or her true ability level, precluding the use of national identity as an indicator of academic competence. If learning a student's true ability occurs, the grading differential between immigrant and native students would be closer to zero for more interested teachers, and teachers with smaller class sizes, and those who teach fewer students. At the same time, to a greater extent grades would reflect a student's true proficiency level–in other words, the slope of student scores on the standardized tests would increase.

The point estimates for the immigrant coefficient seem to indicate stronger biases among less interested teachers, those with larger class sizes, and instructors in charge of more students. However, these heterogeneous effects are not significant. Moreover, the absence of variation in the coefficient attached to test scores does not conform to the predictions of the learning model. Table 8 also shows the heterogeneous effects by class size and number of students taught when controlling for school fixed effects, given that both measures are correlated with whether schools are located in urban or rural areas, a potential cofounder, as shown by 6. According to this model, teacher grading bias against immigrant students is also stronger when class sizes are larger and among teachers with more students. However, the point estimates are not precise enough to make definite conclusions about the potential explanations for these differences.

8 Conclusions

We explore potential biases in how teacher evaluate immigrant students by employing uniquely detailed administrative data from Ecuador. The information allows us to juxtapose subject-specific grades issued by teachers with scores from end-of-year standardized (and blindly scored) proficiency tests covering the same official curriculum delivered in regular classes. This exercise, enriched by a detailed socio-emotional profile of students, allows us to precisely estimate the portions of teacher-based assessments in mathematics and language that are not explained by proficiency scores and yet are related to a pupil's immigration status. We find statistically significant underscoring and under-ranking of immigrant students relative to native students when these students are graduating from high school. Given the particular setting that we use in this study, in which immigrants share similar characteristics with Ecuadorians, and by some metrics are more privileged, our results add complexity to the results of studies that focus on immigrants with characteristics that sharply contrast with the native population. In other words, our results suggest that assessment biases prevail even in a place where immigrants are incredibly similar to natives.

There is also enormous potential for feedback effects in our context. The implications of our findings can be far reaching on both the individual level and the macroeconomic level, and certainly go beyond differentials that rank student performance in high school and secondary graduation rates. We detect discrimination in grading during the crucial transition period between finishing high school and deciding whether to enter college or the labor market; for many people, the latter choice means an end to their educational attainment. This juncture is a time when students and parents invariably find themselves in a position akin to one of investors relying on the asset-return evaluations of more informed experts to make long-term decisions. In the context here, the decision is one of whether to further invest in human capital. For our purposes, the key element of this reasoning is that teacher evaluations made over the course of a student's final year in high school may steer decisions in one way or the other regarding whether to invest in further education. Stated more explicitly, parents (and teens themselves) likely update investment (and effort) decisions after extracting information from assessment reports issued by teachers. Therefore, if a student's perceived competence, or lack thereof, increases the returns associated with the costs of investments, or makes such investment returns seem more risky, as in the traditional Beckerian human-capital framework, teacher-based discrimination in grading may act as a mechanism that reinforces immigrant gaps in the accumulation of human capital. In this case, intra-classroom evaluation biases may very well lead to higher gaps in rates of college attendance and, ultimately, in long-term reduced labor market outcomes for immigrants. Educational attainment and labor market performance are essential elements of the immigrant assimilation process.

Considering the role played by misinformation in the results presented here, and beyond its scientific interest, we draw three lessons for education and immigration policies from our analysis. First, curbing teacher rotation can be particularly important for immigrant students (over and beyond any effect on learning *per se*) because increasing interactions between a group of students and a given teacher would diminish the influence of noisy information on the evaluation of scholastic proficiency. The more a teacher gets acquainted with a given student, the less relevant the pupil's immigration status becomes for making assessments. 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 can be taken by immigrant students at the time of school admission, and graded on a curve with the overall student population, the generation of individual reports with resulting scores could aid teachers in their competence evaluations. This additional information should make teachers better able to evaluate their students without resorting to biased priors. Above all, public educators and immigration authorities could do a better job on their use of performance information to maximize teaching efficiency. Reducing the type of discrimination in grading we have exposed in this study would be an added bonus.

References

Adida, C. L., Lo, A., and Platas, M. R. (2019). Americans preferred syrian refugees who are female, english-speaking, and christian on the eve of donald trump's election. *PLOS ONE*, 14(10):e0222504.

Aigner, D. J. and Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review*, 30(2):175–187.

Alan, S., Duysak, E., Kubilay, E., and Mumcu, I. (2020). Social exclusion and ethnic segregation in schools: The role of teacher's ethnic prejudice. *The Review of Economics and Statistics*, pages 1–45.

Alesina, A., Carlana, M., Ferrara, E. L., and Pinotti, P. (2018). Revealing Stereotypes: Evidence from Immigrants in Schools. Working Paper 25333, National Bureau of Economic Research.

Altonji, J. G. and Pierret, C. R. (2001). Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics*, 116(1):313–350.

Altschul, I., Oyserman, D., and Bybee, D. (2006). Racial-ethnic identity in mid-adolescence: Content and change as predictors of academic achievement. *Child development*, 77(5):1155–1169.

Arrow, K. J., Princeton University, Industrial Relations Section, Woodrow Wilson School of Public and International Affairs, Princeton University, Conference Office, and Princeton University Conference on Discrimination in Labor Markets, editors (1971). *The theory of discrimination.* Princeton N.J.

Avitzour, E., Choen, A., Joel, D., and Lavy, V. (2020). On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes. Technical Report w27818, National Bureau of Economic Research.

Bansak, K., Hainmueller, J., and Hangartner, D. (2016). How economic, humanitarian, and religious concerns shape european attitudes toward asylum seekers. *American Association for the Advancement of Science*, 354(6309):217–222.

Becker, G. S. (1957). *The Economics of Discrimination*.

Ben-Zeev, A., Dennehy, T. C., Goodrich, R. I., Kolarik, B. S., and Geisler, M. W. (2014). When an "Educated" Black Man Becomes Lighter in the Mind's Eye: Evidence for a Skin Tone Memory Bias. *SAGE Open*, 4(1):1–9.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics*, 94(3):686–698.

Bjerk, D. (2008). Glass Ceilings or Sticky Floors? Statistical Discrimination in a Dynamic Model of Hiring and Promotion*. *The Economic Journal*, 118(530):961–982.

Black, D. A., Berger, M. C., and Scott, F. A. (2000). Bounding Parameter Estimates with Nonclassical Measurement Error. *Journal of the American Statistical Association*, 95(451):739–748.

Blume, L. E. (2006). The Dynamics of Statistical Discrimination. *The Economic Journal*, 116(515):F480–F498.

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

Bowles, S. and Gintis, H. (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. Routledge & Kegan Paul.

Brown, C. and Bigler, R. (2005). Children's perceptions of discrimination: A developmental model. *Child development*, 76(3):533–553.

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

Carlana, M., La Ferrara, E., and Pinotti, P. (2022). Goals and gaps: Educational careers of immigrant children. *Econometrica*, 90(1):1–29.

Cornwell, D., Mustard, B., and Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264.

Dovidio, J. F., Brigham, J., Johnson, B., and Gaertner, S. (1996). Stereotyping, prejudice, and discrimination: Another look. *Stereotypes and Stereotyping, New York, Guilford*.

Elacqua, E. and Marotta, L. (2020). Is working one job better than many? assessing the impact of multiple school jobs on teacher performance in rio de janeiro. *Economics of education review*, 78.

Elenbaas, L. and Killen, M. (2017). Children's perceptions of social resource inequality. *Journal of Applied Developmental Psychology*, 48:49–58.

Farkas, G., Grobe, R. P., Sheehan, D., and Shuan, Y. (1990). Cultural Resources and School Success: Gender, Ethnicity, and Poverty Groups within an Urban School District. *American Sociological Review*, 55(1):127–142.

Ferguson, R. F. (1998). Can schools narrow the Black–White test score gap? In *The Black–White test score gap*, pages 318–374. Brookings Institution Press, Washington, DC, US.

Ferman, B. and Fontes, L. (2020). Discriminating behavior: Evidence from teachers' grading bias. *Munich Personal RePEc Archive - University Library of Munich*.

Ferman, B. and Fontes, L. (2022). Assessing knowledge or classroom behavior? evidence of teachers' grading bias. *Journal of Public Economics*, 216.

Figlio, D. N. (2005). Names, Expectations and the Black-White Test Score Gap. SSRN Scholarly Paper ID 684721, Social Science Research Network, Rochester, NY.

Gereke, J., Schaub, M., and Baldassarri, D. (2020). Gendered discrimination against immigrants: experimental evidence. *Frontiers in sociology*, 5(59).

Glock, S., Kneer, J., and Kovacs, C. (2013). Preservice teachers' implicit attitudes toward students with and without immigration background: A pilot study. *Studies in educational evaluation*, 39(4):204–210.

Hainmueller, J. and Hiscox, M. (2010). Attitudes toward highly skilled and low-skilled immigration: Evidence from a survey experiment. *American Political Science Review*, 104(1):61–84.

Hainmueller, J. and Hopkins, D. (2014). Public attitudes toward immigration. *Annual Review of Political Science*, 17:225–249.

Hainmueller, J. and Hopkins, D. (2015). The hidden american immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science*, 59(3):529–548.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in Grading. *American Economic Journal: Economic Policy*, 4(4):146–168.

Heinrich, C., Hernández, M., and Shero, M. (2023). Repercussions of a raid: Health and education outcomes of children entangled in immigration enforcement. *Journal of Policy Analysis and Management*, 42(2):350–392.

Hinnerich, B., Hoglin, E., and Johannesson, M. (2011a). Ethnic discrimination in high school grading: Evidence from a field experiment. *Working Paper Series in Economics and Finance*, (733).

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2011b). Are boys discriminated in Swedish high schools? *Economics of Education Review*, 30(4):682–690.

Holmstrom, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7:24–52.

Hopkins, D. (2010). Politicized places: Explaining where and when immigrants provoke local opposition. *American Political Science Review*, 104(1):40–60.

Ibanez, A., Moya, A., Ortega, M., Rozo, S., and Urbina, M. (2022). Life out of the shadows: Impacts of amnesties in the lives of migrants. *IZA Institute of Labor Economics, Discussion Paper Series*, IZA DP No. 15049.

Ji, T., Tybur, J. M., and van Vugt, M. (2021). Gendered outgroup prejudice: An evolutionary threat management perspective on anti-immigrant bias. *Group Processes Intergroup Relations*, 24(1):177–192.

Lang, K. (1986). A Language Theory of Discrimination. *The Quarterly Journal of Economics*, 101(2):363–382.

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

Lavy, V. and Megalokonomou, R. (2019). Persistency in Teachers' Grading Bias and Effects on Longer-Term Outcomes: University Admissions Exams and Choice of Field of Study. Technical Report w26021, National Bureau of Economic Research, Cambridge, MA.

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

Lehmann, J. K. (2011). Job assignment and promotion under statistical discrimination: Evidence from the early carrers of lawyers. *Unpublished manuscript, University of Houston*.

Lundberg, S. and Startz, R. (2007). Information and racial exclusion. *Journal of Population Economics*, 20(3):621–642.

Mariani, R., Pasquini, A., and Rosati, F. (2020). Elementary facts about immigration in italy. what do we know about immigration and its impact. *CEIS Working Paper*, No. 488.

Marotta, L. (2019). Teachers' contractual ties and student achievement: The effect of temporary and multiple-school teachers in brazil. *Comparative Education Review*, 63(3).

McLoyd, V. C., Kaplan, R., Purtell, K. M., Bagley, E., Hardaway, C. R., and Smalls, C. (2009). Poverty and socioeconomic disadvantage in adolescence.

Mechtenberg, 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(4):1431–1459.

Olivieri, S., Ortega, F., Rivadeneira, A., and Carranza, E. (2022). Shoring up economic refugees: Venezuelan migrants in the ecuadoran labor market. *Migration Studies*, 9(4):1590–1625.

Palacios, A. (2017). Variación y cambio lingüístico en situaciones de contacto. *Iberoamericana Vervuert*.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, 62(4):659–661.

Rist, R. C. (1973). *The Urban School: A Factory for Failure. A Study of Education in American Society*. MIT Press, Cambridge, Mass.

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

Seider, S., Clark, S., Graves, D., Kelly, L. L., Soutter, M., El-Amin, A., and Jennett, P. (2019). Black and latinx adolescents' developing beliefs about poverty and associations with their awareness of racism. *Developmental Psychology*, 55(3):509.

Seider, S., Graves, D., and Diaz, B. (2022). The development of awareness of racism in adolescents of color over 4 years of high school. *Journal of community psychology*, 50(1):64–89.

Sewell, W. H. and Hauser, R. M. (1980). *The Wisconsin Longitudinal Study of Social and Psychological Factors in Aspirations and Achievements**.

Sexton, P. C. (1961). *Education and income: inequalities of opportunity in our public schools*. Viking Press.

Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical Economics*, 45(1):523–538.

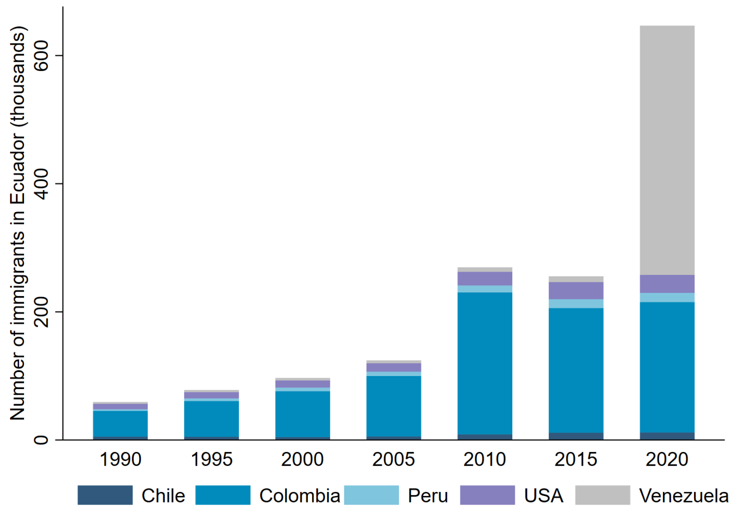
Terrier, C. (2020). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review*, 77:101981.

Trawalter, S., Hoffman, K. M., and Waytz, A. (2012). Racial Bias in Perceptions of Others' Pain. *PLOS ONE*, 7(11):e48546.

Valentino, N., Soroka, S., Iyengar, S., Aalberg, T., Duch, R., Faile, M., , Hahn, K., Hansen, K., Harell, A., Helbling, M., Jackman, S., and Kobayashi, T. (2019). Economic and cultural drivers of immigrant support worldwide. *British Journal of Political Science*, 49(4):1201–1226.

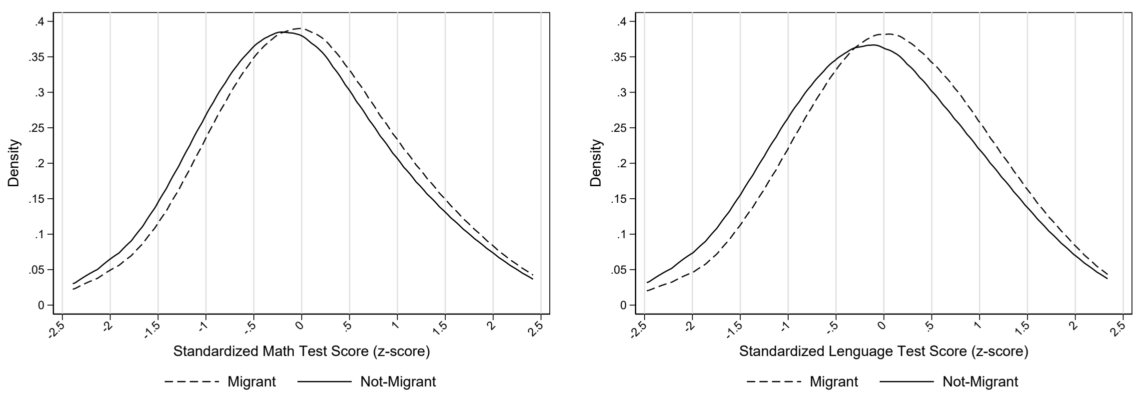
Williams, T. (1976). Teacher Prophecies and the Inheritance of Inequality. *Sociology of Education*, 49(3):223–236.

Figure 1: Number of Immigrants in Ecuador, 1990 to 2020



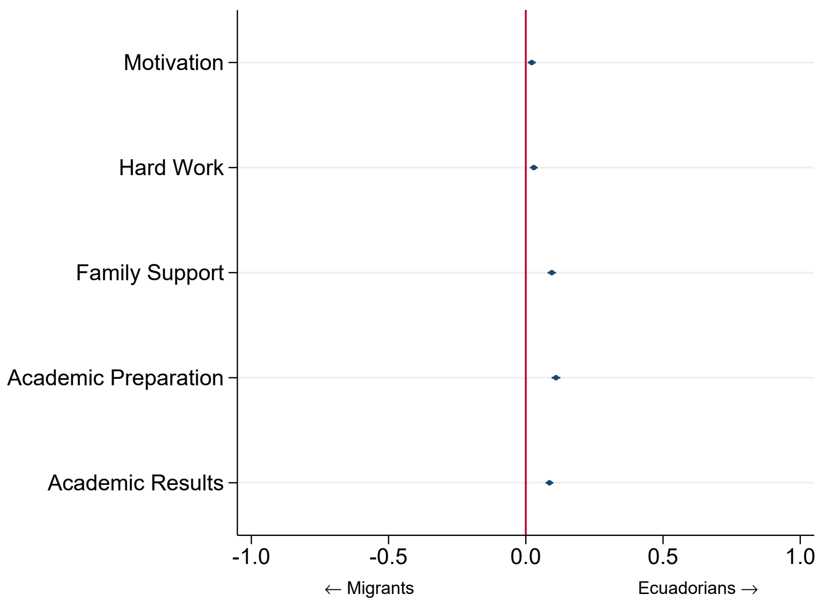
Source: International Migrant Stock 2020: Destination and Origin.

Figure 2: Test Score Differentials: Math (Left) and Language (Right)



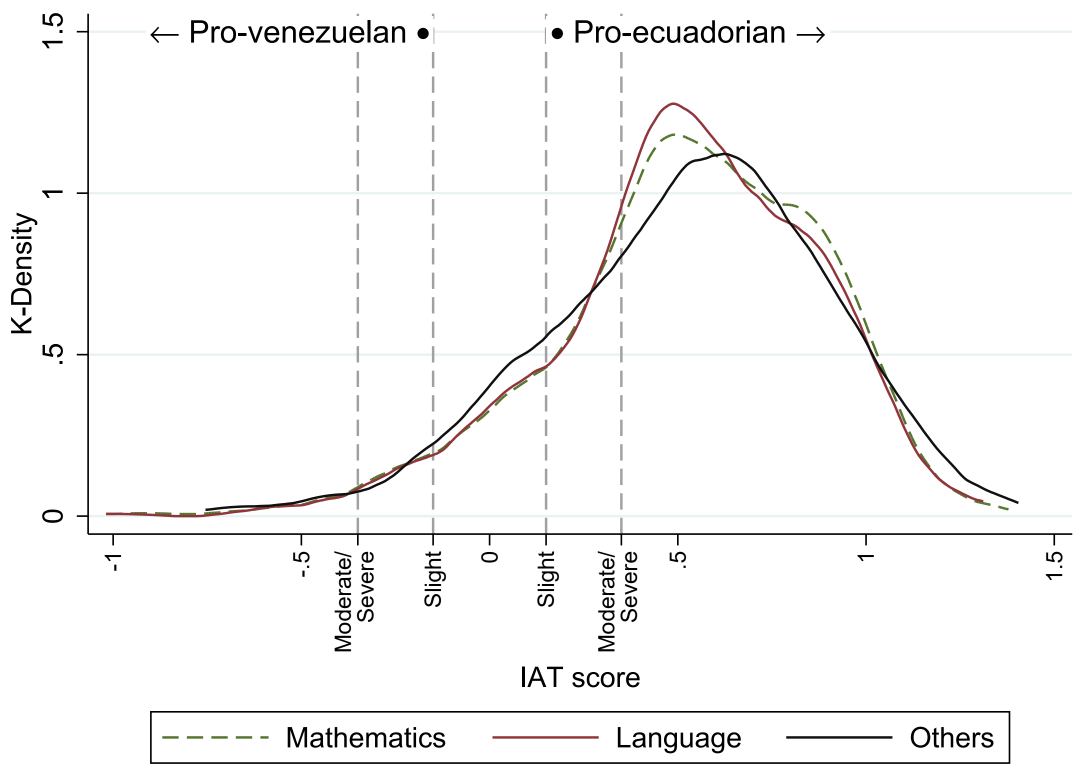
Source: *Ser Bachiller* 2018-2019 microdata.

Figure 3: Beliefs Teachers Have Regarding Immigrant Students



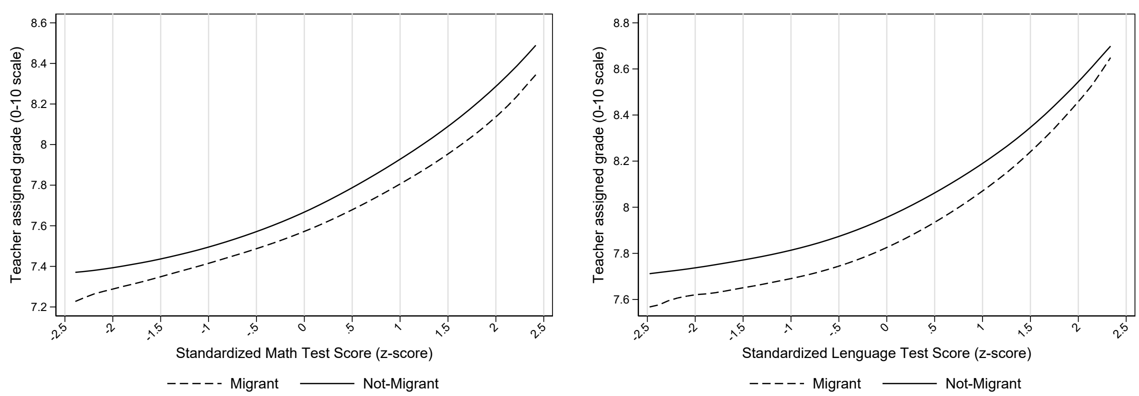
Note: These data come from a survey conducted with 1,773 teachers in Ecuador from different grade levels and subject areas that measured, among other things, their perception of immigrant students. In this survey, teachers were asked: "In your opinion, who are more motivated to learn," "work harder to reach their goals," "have more support from their family,", "start the school year more academically prepared," and "have better academic results." The teachers had to select an option from a 5-point scale, where the midpoint ("0" in this figure) represented a neutral position. Each question asked teachers about their perceptions about different groups of students (for example, "boys versus . girls", "high versus low socioeconomic students"), among which they express their beliefs about "immigrant versus native students.". The results for this survey are shown in this figure.

Figure 4: Unconscious Teacher Biases Against Venezuelans as measured by the Implicit Association Test



Note: These data come from an Implicit Association Test (IAT) applied to a sub-sample of teachers who participated in the study described in Figure 1. The IAT was conducted with 1.380 teachers. This test measures the strength of associations between concepts (e.g., black people, LGTBQ people) and evaluations (e.g., good, bad) or stereotypes (e.g., athletic, clumsy), considering the time it takes for an individual to categorize these concepts into two categories. In this study, the IAT was administered so that participants had to associate concepts related to Venezuelans / Ecuadorians with good / bad adjectives. This graph shows the IAT score distribution for teachers and positive values indicate a greater association between "natives" - "good" and "Venezuelans" - "bad".

Figure 5: Unconditional Relationship Between Teacher-Assigned Grades and Blindly-Scored Tests



Source: *Ser Bachiller* 2018-2019 microdata and administrative records of students' transcripts.

Table 1: Balance Test: Sample With and Without Missing Information on Immigration Status

Variables		Without missing	With missing	Difference
Demographics				
Female		0.504	0.506	0.003
Race	White or mestizo	0.845	0.847	0.002
	Afro-descendant	0.042	0.041	-0.001
	Indigenous	0.061	0.059	-0.002
	Montubio	0.049	0.049	0.000
Socio-economic status				
Mother's education	Until primary school	0.479	0.477	-0.002
	Middle school	0.126	0.126	-0.000
	High-school	0.311	0.315	0.004*
	College or higher	0.084	0.082	-0.002
Location				
Urban		0.791	0.791	0.000
Coast		0.602	0.603	0.001
Observations		88,505	91,308	

Source: Ser Bachiller 2018-2019.

Note: Estimates are significant at $*p{<}0.10, **p{<}0.05, ***p{<}0.01$.

Table 2: Summary Statistics

Variables		Natives	Migrants	Difference
Demographics				
Female		0.510	0.473	-0.038***
Race	White or mestizo	0.872	0.870	-0.002
	Afro-descendant	0.040	0.041	0.002
	Indigenous	0.048	0.061	0.013***
	Montubio	0.036	0.023	-0.013***
Socio-economic status				
Mother's education	Until primary school	0.443	0.413	-0.030***
	Middle school	0.126	0.119	-0.007
	High-school	0.337	0.365	0.027***
	College or higher	0.094	0.104	0.010*
Work habits				
Hours studying at home	Less than 1 hour	0.138	0.151	0.012**
	From 1 to 2 hours	0.390	0.395	0.006
	3 hours	0.230	0.230	0.000
	4 hours or more	0.241	0.223	-0.018**
Work hard		0.728	0.725	-0.004
Socio-emotional				
Attentive to details		0.887	0.861	-0.026***
Persist on the task		0.938	0.933	-0.006
Kindness		0.967	0.963	-0.004
Empathy		0.904	0.908	0.005
Creativity		0.716	0.747	0.030***
Parental involvement				
Parents check homework		0.849	0.811	-0.039***
Parents check grades		0.910	0.883	-0.027***
Behavior				
No school suspension		0.980	0.975	-0.006
Observations		34,286	3,809	

Source: Ser Bachiller 2018-2019.

Note 1: Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. Because our main strategy relies on classroom fixed effects, this table only includes students from mixed classrooms, with both migrant and natives. Note 2: The questions used to identify the socioemotional characteristics from the Ser Bachiller contextual questionnaires are: i) attentive to details - When you carry out a task, are you very careful with the details?; ii) persist on the task - Do you finish what you start?; iii) kindness - Are you friendly with other people?; iv) empathy - When you try to understand others, do you try to put yourself in their position?; v) creativity - Do you have ideas that other people have not thought of before?

Table 3: Teacher Grading Bias

	Teacher assigned grade				Passing grade
	(1)	(2)	(3)	(4)	(5)
Panel A: Mathematics					
Migrant	-0.074***	-0.037***	-0.052***	-0.043***	-0.546**
	(0.020)	(0.013)	(0.012)	(0.012)	(0.241)
Test scores			0.333***	0.331***	0.899***
			(0.003)	(0.003)	(0.069)
Panel B: Spanish Language					
Migrant	-0.096***	-0.045***	-0.068***	-0.052***	-0.629***
	(0.022)	(0.013)	(0.013)	(0.012)	(0.222)
Test scores			0.266***	0.255***	0.665***
			(0.003)	(0.003)	(0.051)
Classroom FE	No	Yes	Yes	Yes	Yes
Socio-demographics	No	No	No	Yes	Yes
Observations	179813	179813	179813	179813	179813

Note: Standard errors are clustered at the classroom level. Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. Socio-demographic characteristics include students' sex, race, their mother's education, and whether they have repeated a grade level.

Table 4: Teacher Grading Bias: Robustness to Socio-Emotional Characteristics

	Teacher assigned grade					Pa gr
	(1)	(2)	(3)	(4)	(5)	
Panel A: Mathematics						
Migrant	-0.043***	-0.042***	-0.040***	-0.041***	-0.029**	-0.
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.
Test scores	0.331***	0.313***	0.313***	0.313***	0.304***	0.7
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.
Panel B: Spanish Language						
Migrant	-0.052***	-0.050***	-0.047***	-0.047***	-0.035***	-0.
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.
Test scores	0.255***	0.237***	0.237***	0.236***	0.228***	0.5
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.
Work habits	No	Yes	Yes	Yes	Yes	Y
Socio-emotional skills	No	No	Yes	Yes	Yes	Y
Parental involvement	No	No	No	Yes	Yes	Y
Behavior	No	No	No	No	Yes	Y
Observations	179813	179813	179813	179813	179813	17

Note: Standard errors are clustered at the classroom level. Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. All models control for classroom fixed effects and socio-demographic characteristics. Models gradually control for: students' work habits (how often they work hard and number of hours of study time per day); socio-emotional skills (attention to details, perseverance, kindness, empathy, and creativeness); parental involvement (how often parents ask about homework and grades); and behavior (never have been suspended from school for bad behavior).

Table 5: Teacher Grading Bias: By Students' Characteristics

	Teacher assigned grade					
	Male	Female	Low SES	High SES	Low achiever	Hig achie
	(1)	(2)	(3)	(4)	(5)	(6
Panel A: Mathematics						
Migrant	-0.044***	-0.003	-0.030	-0.041**	-0.016	-0.03
	(0.016)	(0.018)	(0.019)	(0.017)	(0.017)	(0.0
Test score	0.287***	0.314***	0.289***	0.327***	0.193***	0.38
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.0
Panel B: Spanish Language						
Migrant	-0.061***	-0.009	-0.024	-0.052***	-0.011	-0.05
	(0.017)	(0.016)	(0.019)	(0.017)	(0.017)	(0.0
Test score	0.206***	0.241***	0.217***	0.244***	0.150***	0.28
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.0
Observations	88531	90381	79839	87189	92389	874

Note: Standard errors are clustered at the classroom level. Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. All models control for classroom fixed effects, students' socio-demographic characteristics, work habits, socio-emotional skills, parental involvement, and behavior.

Table 6: Teacher Grading Bias: By Rural or Urban Location

	Teacher assigned grade	
	Rural	Urban
	(1)	(2)
Panel A: Mathematics		
Migrant	0.025	-0.040***
	(0.023)	(0.013)
Test scores	0.310***	0.303***
	(0.007)	(0.003)
Panel B: Spanish Language		
Migrant	0.006	-0.043***
	(0.025)	(0.013)
Test scores	0.236***	0.226***
	(0.006)	(0.003)
Observations	37523	142290

Note: Standard errors are clustered at the classroom level. Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. All models control for classroom fixed effects, students' socio-demographic characteristics, work habits, socio-emotional skills, parental involvement, and behavior.

Table 7: Teacher Grading Bias: Exploring Statistical Discrimination

	Teacher Assigned Grade					
	Less interested teachers	More interested teachers	Larger class sizes	Smaller class sizes	Teach more students	Teach few students
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Mathematics						
Migrant	-0.037**	-0.019	-0.033*	-0.025*	-0.045***	-0.010
	(0.016)	(0.017)	(0.019)	(0.015)	(0.017)	(0.016)
Test score	0.311***	0.298***	0.304***	0.304***	0.306***	0.303***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Panel B: Spanish Language						
Migrant	-0.051***	-0.016	-0.056***	-0.018	-0.056***	-0.012
	(0.016)	(0.017)	(0.019)	(0.015)	(0.018)	(0.015)
Test score	0.233***	0.223***	0.219***	0.237***	0.227***	0.229***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	90146	89610	83905	95908	89386	90427

Note: Standard errors are clustered at the classroom level. Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. All models control for classroom fixed effects, students' socio-demographic characteristics, work habits, socio-emotional skills, parental involvement, and behavior.

Table 8: Teacher Grading Bias: Exploring Statistical Discrimination with School Fixed Effects

	Class size		Number of students taught	
	Math	Language	Math	Language
	(1)	(2)	(3)	(4)
Migrant	-0.029	-0.036**	0.003	-0.024
	(0.019)	(0.018)	(0.020)	(0.020)
Test scores	0.319***	0.269***	0.311***	0.267***
	(0.008)	(0.008)	(0.009)	(0.009)
Larger class size	0.056*	-0.022		
	(0.032)	(0.034)		
Larger class size*Migrant	-0.003	-0.012		
	(0.028)	(0.028)		
Larger class size*Test scores	0.002	-0.037***		
	(0.012)	(0.012)		
Teach more students			-0.047	-0.075
			(0.055)	(0.057)
Teach more students*Migrant			-0.067**	-0.037
			(0.030)	(0.030)
Teach more students*Test scores			0.018	-0.032**
			(0.014)	(0.013)
Observations	179813	179813	179813	179813

Note: Standard errors are clustered at the classroom level. Estimates are significant at $*p<0.10, **p<0.05, ***p<0.01$. All models control for school fixed effects, students' socio-demographic characteristics, work habits, socio-emotional skills, parental involvement, and behavior.