DISCUSSION PAPER Nº IDB-DP-01027

Discrimination in Grading:

Evidence from Immigrants in Ecuador

Marcelo Drouet
Suzanne Duryea
Luana Marotta
Marcos A. Rangel
Lucina Rodríguez Guillén

Inter-American Development Bank
Migration Unit

August 2023

Discrimination in Grading:

Evidence from Immigrants in Ecuador

Marcelo Drouet
Suzanne Duryea
Luana Marotta
Marcos A. Rangel
Lucina Rodríguez Guillén

Inter-American Development Bank
Migration Unit

August 2023

JEL Codes: J24, J15

Keywords: teacher grading discrimination, immigrant students, education inequalities

Discrimination in Grading: Evidence from immigrants in Ecuador[*]

Marcelo Drouet
IADB[‡]

Suzanne Duryea
IADB[§]

Luana Marotta
IADB[¶]

Marcos A. Rangel
Duke University[‖]

Lucina Rodríguez Guillén
IADB[**]

Cynthia Van Der Werf
IADB[††]

August 2023

[*]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 low-cost interventions. We are grateful for the guidance and methodological input received by Eliana La Ferrara, Ana María Ibañez, and the participants of the Initial Workshop of 2022. We are grateful for the comments received from Felipe Muñoz. This study was made possible thanks to the Ministry of Education of Ecuador, which gave us access to administrative records and provided invaluable support.

[‡]Inter-American Development Bank. E-mail: msdrouet01@gmail.

[§]Inter-American Development Bank. E-mail: suzanned@iadb.org.

[¶]Inter-American Development Bank. E-mail: luanac@iadb.org.

[‖]Sanford School of Public Policy, Duke University; 262 Rubenstein Hall, Box 90312, Durham, NC 27708; Email: marcos.rangel@duke.edu; Phone: (919) 613-7340.

[**]Inter-American Development Bank. E-mail: lucinar@iadb.org.

[††]Inter-American Development Bank. E-mail: cvanderwerf@iadb.org.

Abstract

This article investigates whether discrimination taking the form of biased assessment of students by teachers is prevalent within Ecuadorian schools serving immigrants. Robust evidence is drawn from unique data pertaining to high-school students and educators. After holding constant performance in blindly scored tests of proficiency, we find that teacher-assigned Mathematics and Language grades suffer from cardinal and ordinal grading biases *against* children from immigrant households. We show that these results are robust with respect to the omission of socio-emotional traits that are valued by teachers. Heterogeneity analyses indicate key differences by the gender of the students and perceptions of teacher engagement.

JEL: I24, J15

1 Introduction

Children and adolescents are a particularly vulnerable group of migrants and face multiple barriers to education. Among those barriers is the lack of proper documentation that impede enrollment at destination and consequent disruptions in school attendance. When immigrant children manage to attend schools, they still have to go over a few additional hoops, including adaptation to different curricula, grade-level misplacement and, potentially, discrimination by peers and teachers. In this article, we take a closer and detailed look at the latter. While we propose discrimination within schools as a potential impediment for the assimilation of immigrants, we recognize that such a phenomenon may manifest itself in many different ways. Our paper focuses on a specific and relatively understudied form of teacher discrimination: a teacher's biased evaluation of students with respect to their scholastic proficiency and aptitude (i.e.: grading).

In this study, we examine teacher grading discrimination against students from immigrant families using data from Ecuadorian high schools. We employ detailed administrative data covering approximately 173,000 high-school students spread across nearly 6,000 classrooms in the 2018-2019 academic year. Our inference is based on the juxtaposition of teachers' subject-specific grades and scores from end-of-year standardized (and blindly marked) proficiency test (*Ser Bachiller*) covering the same official curriculum delivered in regular classes.

Ecuador has traditionally been an emigrant country and recent migration flows have included mostly Venezuelans escaping political and economic turmoils in their homeland. Unlike the setting explored by existing studies on teacher discrimination against immigrants, Ecuador provides a unique context, in which immigrants and natives share a relatively similar social and cultural background. For example, Alesina et al. (2018) examine bias in teacher grading in Italy, where most immigrants are noticeably different in skin tone, language, and culture from the majority of Italians. In our study, we examine whether teacher discrimination prevails in a context where immigrant families are not very different (and, in some aspects, more privileged) than natives.

The analyses show that portions of teachers' assessments in Mathematics and Language not explained by proficiency scores are associated with pupils' immigration status. Our most conservative estimates indicate that there is statistically significant underscoring and under-ranking of immigrants relative to native students. The measured gap in promotion rates between equivalently proficient and well-behaved students corresponds to an increase of 7.6% for Math (5.5% for Language) in the retention probability for the average immigrant student. Focusing exclusively on the ordinality aspect, we also uncover a gap that translates into a reduction of 1.8% for Math (2.5% for Language) in the probability of immigrants being graded above the classroom median. In practice, these gaps work as if teachers were "taxing" the average immigrant students' performance in proficiency tests by 0.12 in Math (0.11 in Language) of one standard deviation at the time her competence is being assessed. These results are shown robust to possible omissions of behavioral attributes and to the likely incidence of measurement error on scores from standardized tests used as covariates in our estimations. They are also very much in line with the expected subtlety of this particular form of discrimination, including with the prevalence of implicit association biases.

Once the existence of evaluation gaps in teacher assessments is established, we rely on economic theory to examine its likely source in our context. We draw from a rich literature on statistical and screening discrimination.[1] We map our setting into these studies by focusing on two 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 graduation and ranking decisions. Second, due to the rapid inflow on Venezuelan immigrants, teachers may have perceived that immigration authorities employed lenient standards for the admission of recently-arrived students into Ecuadoran high-schools.[2] Twelfth-grade teachers are well aware of the implications of such admission policy, and priors regarding their students' proficiency may be downgraded as a result. Therefore, we hypothesize that when teachers issue report cards assessing the competence of their students, subtle biases are generated by the weighted combination of noisy information extracted from their own screening exams and stereotyped priors.

[1] 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).

[2] Venezuelans account for the largest percentage of immigrants living in Ecuador. See Section 2.

We then present evidence on the heterogeneity of these biases. In particular, we provide evidence of biases being pertinent on the teacher evaluation of immigrant and native boys but not for female students. We also document that these biased evaluation of boys are stronger in schools where teachers are perceived as “less engaged” with students. These patterns are compatible with informational issues, but we cannot rule out the possibility that such biases resulting from animus or anti-immigrant sentiment (taste discrimination), which would be in line with the observed reduction in the Ecuadoran Migrant Acceptance Index, as discussed in section 2.

The implication of our findings can be far reaching, and certainly go beyond differentials on high-school graduation and on ranking of students. There is an enormous potential for feedback effects in our context. This is the case because we detect discrimination in grading during the transition between high-schools 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. For our purposes, the key element of this reasoning is that teacher communications may steer investment decisions in one way or the other.[3] That is to say that parents (and teens themselves) likely update investment (and effort) decisions after extracting information from report cards issued by teachers. Therefore, if children’s perceived competence increases the returns or reduces the costs of investments, as in the traditional *Beckerian* human-capital framework, this mechanism can reinforce migrant gaps in the accumulation of human capital. 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.[4]

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

[4] See Mechtenberg (2009) for a formalization of this argument, and Lundberg and Startz (1983), who are explicit in modeling human capital investments’ response to the presence of discrimination. See also Lavy and Sand (2018) and Terrier (2020) for some important empirical results supporting this reasoning.

2 Institutional background

In the last few years, Ecuador has witnessed a considerable growth of their immigrant population, which, in great part, is due to the Venezuelan migration crisis. Between 2010 and 2020, the number of migrants in this country has more than doubled, going from 375 to 785 thousand.[5]

While in 2010 the migrant population in Ecuador was mostly made up of Colombians (59 percent), in 2020, 51 percent of the migrants in the country were from Venezuela and 25 percent from Colombia (DataMIG, 2023). Ecuador has become an attraction for Venezuelans because of its relatively liberal laws and formal institutions for migrants and refugees. The dollarization of the Ecuadorian economy has also contributed for the country to become a popular destination for some migratory flows in the region.

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

Yet, the proximity between immigrants and Ecuadorians has not prevented discrimination from happening. Rather, data show that attitudes and beliefs towards migrants in Ecuador have worsened considerably over the years. Gallup's Migrant Acceptance Index reveals, for example, that while in 2016 72% of the population perceived that having immigrants was a good thing for the country, only 27% did so by 2019. Equivalently, 84% said it was a good thing to have migrants moving into their neighborhood in 2016, but only 48% had the same view in 2019. Moreover, according to the

[5]As Ecuador permits all migrants, even those without a permit, to enroll in schools, the share of foreign-born students enrolled in the Ecuadorian school system increased accordingly. The percentage of students who are immigrants in the Ecuadorian educational system was 0.5% for the 2009-2010 period, and increased to 2.1% for the 2021-2022 period.

Latinobarómetro, an annual public opinion survey carried out with respondents from diverse Latin American countries, in 2020, 26% Ecuadorians agreed that immigrants were good for the country's economy, while the average for the region was 43%. Moreover, 79% Ecuadorians believed that immigration cause an increase in crime, whereas only 56% of the population in the region thought the same thing.

In school settings, results from standardized assessments show that immigrant and native students have similar performance levels. For example, Figure 1 shows the distribution of math test scores for immigrant and native students in the 2018 (*Ser Bachiller*), the test data used in our analysis. It reveals that students from immigrant families have a slightly higher performance in math than Ecuadorian students. Despite this, teachers believe that immigrant students have lower academic performance. In a survey conducted with teachers in Ecuador, we asked "Who, in your opinion, tend to achieve better academic results?"[6]. 17 percent of the surveyed elementary teachers said that non-migrants achieve better results, 4 percent said that immigrant students perform better, and 79 percent said both perform the same. The belief of *secondary school* teachers was even more pessimistic towards the performance of immigrant students: 21 percent said that non-migrant students have better academic performance. We also found that teachers believe immigrant students to have less motivation to learn, lower levels of effort to perform, less parent support, and lower preparation to do well in school.

On one hand, based on survey results of attitudes and beliefs of Ecuadorians towards immigrants, we could hypothesize that teacher biases may result from animus or anti-immigrant sentiment (taste discrimination). On the other hand, considering the beliefs of teachers towards the academic performance of immigrant students, potential grading discrimination might reflect lack of information. That is, in the absence of reliable knowledge about students' performance, teachers might underestimate the grades of immigrants, believing that they are less academically prepared (statistical discrimination).

[6]This survey was applied as part of a follow up study we carried out in 2021 and 2022. The study included the design of two simple interventions aimed at reducing the teachers' bias in Ecuador. The baseline survey gathered information on the teachers' socio-demographic characteristics, current position, previous experience as a teacher, and perception about minority students of a sample of 2,092 teachers

3 Related literature

The question of whether teachers treat children of different backgrounds 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 do not detect significant 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.[7]

Our work complements more recent studies from the education and economics literature. Shay and Jones (2006) and Dorsey and Colliver (1995) examine quasi-experimental variation provided by institution-level policy changes regarding anonymity in the grading processes applied to college/graduate students and do not detect significant racial differentials. Figlio (2005) examines whether teachers' overall perception of a given student is affected by the "Blackness" of her first name, even after controlling for performance in standardized examinations. Using data from one school district in Florida, he uncovers evidence of lower teacher expectations for those perceived to have African American ancestry. Burgess and Greaves (2013) investigate differences in teacher grading according to ethnic background using observational data from England, finding significant underassessment of Black Caribbean and Black African pupils. Finally, Hinnerich et al (2011a, 2011b) conduct audit-like studies by transcribing and blindly re-grading tests assessed by teachers in Sweden and estimate gender (insignificant) and nationality (significant) gaps. Similar exercise conducted in Germany by Sprietsma (2013) also uncovers biases against exam solutions which had Turkish-sounding names randomly allocated to them (relative German-sounding names).

A common approach in this literature is to juxtapose subjective teacher evaluations with blind

[7] 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.

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 happens under different anonymity regimes. Using the blind score as the counterfactual to the non-blind teacher score, Lavy (2008) finds evidence of discrimination against males. Teacher biases based on class-level gender differences furthermore have both short and long-term consequences for boys' and girls' human capital accumulation (Lavy and Sand, 2018; Lavy and Megalokonomou, 2019).[8] Blind/non-blind contrasts 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 is clear evidence of discrimination. Finally, Burgess and Greaves (2013) and Botelho et al. (2015) use large-scale observational data in the UK and in Brazil, respectively, to investigate differences in teacher grading according to ethnic/racial background. They juxtapose objective tests with subjective teacher assessments and document significant underassessment of Black pupils (Black Caribbean and Black African in the case of the UK).

Our study builds on this literature by employing both blind and non-blind assessments of student mastery over the same skill set. In light of previous discussions, we underscore the contributions of our study context. First, we use large-scale observational data from the Ecuador that provides plausibly objective measures of student math and reading mastery alongside 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 biased behaviors of teachers embedded in the very test scores that anchor our models. These may include teachers' varied treatment or evaluation of students across nativity groups in a manner that differentially influences students'

[8]Terrier (2020) similarly shows teacher favoritism towards girls using blind and non-blind test scores, and finds that, as a result, girls 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.

end-of-year test scores.

The discussion presented here plays on four contributions provided by our context with respect to other studies in the literature. First, we examine discrimination within high-schools, a section of the educational trajectory when awareness of stereotypes and reactions to their prevalence is higher. This is the case due to the importance of identity formation and social connections among teenagers (Seider et al., 2019, 2022; Altschul et al., 2006; Elenbaas and Killen, 2017). In particular, McLoyd et al (2009) indicates that teenagers acquire more perspective-taking skills and that in itself contributes to a heightened ability to consciously assess racist (or xenophobic) priors, motivations, and decisions from people they interact with.[9] We argue that this patterns can only increase the relevance of feedback effects induced by teachers' assessment biases.

Second, we investigate discrimination regarding national origins at a time in which Ecuador and its neighbors were living a global economic downturn. Ecuador was on the receiving end of the mass exodus of Venezuelans fleeing the humanitarian crisis in their country. This was likely consequential for 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 & 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 data presented in section 4 suggest, growth in immigrant population in Ecuador during recent years seems to have boosted anti-immigration beliefs among natives. We argue that this "change of heart" should also be pertinent for Ecuadoran teachers who started interacting with a larger population of immigrants in their classrooms.

Third, we are well positioned to employ rich administrative data sources (more on this below) and confirm the results of smaller scale studies on xenophobia within schools. This includes the work of Alesina et al. (2018), Alan et al. (2020), Glock et al. (2013), and Carlana et al. (2022). This is important because the sheer size of and level of detail in our data base allows us to convey a

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

complete portrait of teacher and student-body characteristics associated with discrimination in actual classroom environments. Meanwhile, 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). Moreover, in our context there are both weak regulation of grading and non-disclosure of information regarding standardized test performance to acting parties (teachers) 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 or students themselves, and ii) last minute reactions to performance information are not likely sought either by evaluators or by those being evaluated.

Finally, the context analyzed in this paper is unique compared to existing research on teacher discrimination as cultural and religious differences between Ecuadorians and immigrants are minimal (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 natives and immigrants are more conspicuous. As noted by Carlana et al. (2022), most immigrants in Italy come from low and middle-income countries and come from lower socioeconomic backgrounds than native households. Immigrants are also more likely to work at low-skilled jobs (Mariani et al., 2020). In that setting, the authors found that immigrant children receive lower teacher-assigned grades than natives after controlling for their performance on standardized tests. Moreover, they found that grading gaps between natives and immigrants is correlated with teachers' stereotypes against immigrants. In our paper, we will be able to explore a similar question under a context with greater cultural, linguistic, and socioeconomic proximity between natives and immigrants.

4 Data

Information on high-school graduating students comes from two major administrative data bases. The first comes from detailed students' transcript information from the Ministry of Education. The second is the 2018-2019 edition of *Ser Bachiller* standardized tests and its companion questionnaires (covering socio-emotional traits and socio-economic/demographic profiles of students and teachers). The *Ser Bachiller* was an exam administered to all students in Ecuador at the end of 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.[10]

The *Ser Bachiller* was a high stakes evaluation since students' performance on this test contributed to 30 percent of their graduation grade in high school. Their test scores were also weighed in college admission decisions. It is important to note that graduation test scores were computed centrally by a system from the Ministry of Education and high school teachers did not have access to standardized scores of their students. Therefore, teachers' evaluations could not be influenced by the performance of students on the *Ser Bachiller*.

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).

Lastly, our measure of immigration status comes from the following questions of 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?”*. A student is considered to have an immigrant background if she and any other family member (e.g. mother, father, brother and partner) have been in a situation of human mobility and that, based on the type of mobility, she is considered ”international migrant” or ”refugee”.[11] Due to a substantive rate of missingness on this variable we carefully construct our analysis also identifying non-responding students in order to guarantee comparison between self-declared immigrants and natives.

[10]Ecuador has four types of schools: fiscal (77%), private (18%), fiscal commissioners–or charters–(4%), and municipal (1%) schools. This paper uses data only from fiscal public schools.

[11]Our analysis includes all students with immigrant backgrounds regardless of their nationality or the family member's nationality (e.g., Venezuela, Colombia, etc.)

Data on covariates–students' socioeconomic background and socioemotional characteristics, as well as teacher characteristics, also come from the *Ser Bachiller* contextual questionnaires. We describe the entire data set on Appendix Tables A1 through A3 and also highlight the selected subsample containing only classrooms where we can identify at least one immigrant and one native student for models that control for classroom fixed effects.

5 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 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. 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 second, concentrating sole attention on the screening role of twelfth-grade instructors.

The basic intuition is that teachers have access to noisy signals of the students' proficiency in Math and Language, and observe both their behavior in class and their national identities. We define an objective function for graders of school work by assuming they operate as statisticians compelled to maximize the power of the hypothesis test embedded in the evaluation of a student's competence. In addition, we impose that teachers weight Type I and Type II errors symmetrically (i.e.: excessive lenience and excessive rigor are equally unwelcome). Evaluation errors can be reduced by exerting more screening effort, something we implicitly assume teachers either dislike (utility costs) or have limited access to due to high monetary/opportunity costs, or even that school authorities set the number of tests that can be applied to students in a given year (costs of effort could then be modeled as a function of distance to the "norm", such as in Holmstrom and Milgrom, 1991).[12]

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), they 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 personality traits deemed *productive* by teachers.[13] 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 ($t = 1, 2, ..., T$).[14] The higher

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

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

[14]For clarity of exposition, measurement error in teacher's tests is considered classical. We acknowledge that, due the bounded nature of grading scales in most of these classroom tests, errors would be negatively correlated with the 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

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 grading:

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

From this formulation there are two ways in which statistical nationality differentiation can be depicted. The first, rational stereotyping, is based on the idea that attributes including nationality ($\vec{b_i}$) can be informative in the computation of proficiency's best linear projection $E\left[p_i^*|s_i^1,...,s_i^T,\vec{b_i},\vec{a_i}\right]$.[15] In other words, the formulation of priors regarding group's average proficiency encompasses the use of other individual characteristics.[16]

The case of nationality discrimination at hand 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 immigrants' proficiency levels. 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:

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

way the main messages of the model.

[15]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.

[16]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.

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 they 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 proficiency signals (slopes).

Notice that in any of these representations, nationality bias is derived from the imprecision on the information about proficiency contained in the signals. It follows that improvements in the signal-extraction technology should make immigrant status 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, or if tests were made less noisy. We take this simple model to the data, emphasizing its prediction regarding learning a child's true proficiency. 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, 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 another class B if they have different grading standards in the spread of grades.[17] 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.

In order to facilitate 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 establishment of a common passing grade across the Ecuadoran system. 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 evaluated above or at grade 7 (seven) are deemed competent while those below are not. This cardinal notion ought to be common across all classrooms, even if in different levels of stringency (captured by a class fixed-effect). The second altenative is solely based on the ordinal dimension, being constructed as an indicator for grades above the classroom's median grade.

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. Since teachers receive a uniform curriculum, textbooks and practice exercises from the external examiner, their evaluations regarding proficiency should reflect the same skills and cognitive abilities as the standardized exam. 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 multiple choice exam. We, therefore, expect the objectivity inherent in the material to translate itself 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

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

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 own 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 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 $\vec{b_{ir}}$ lists elements affecting teachers' priors with regard to proficiency. Meanwhile, in order to make explicit further challenges to our empirical exercise, the elements in the vector of scholastic attributes ($\vec{a_i}$) were also be 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.

Given that our central objective is to consistently estimate δ_1 and $\vec{\delta_3}$, this simple empirical representation highlights the two main econometric problems we face: a) measurement error in proficiency scores, and b) unobserved heterogeneity.[18] 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 "draw" observations 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.

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

Unobserved heterogeneity adds another layer of complications because even in the absence of measurement error in scores, 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.[19] 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 self-reported perceptions of student engagement, behavior, and effort in school-related activities.

Ultimately, our main empirical model consists of regressing grades on nationality, race, gender, age, essay scores, parental socio-demographics, and our proxies for 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. $f(scores_{ir})$ is estimated as fourth-order polynomials of subject-specific test scores.[20]

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 her pupils. In this way we examine the central prediction from our statistical discrimination conceptual framework: learning of students' true types should preclude the use of national identity as an indicator of scholastic competence.

In practice, and in the spirit of Altonji and Pierret (2001), we test whether nationality differentials in teacher-assigned grades diminish as a teacher's information regarding students improves. 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

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

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

in our study's environment. In the exercises attempted below we proxied information with the intensity of engagement of teachers within a school (as reported by students). In practice we looked at stratified samples based on teachers' level of "interest on students learning" and absenteeism. We argue that the second in particular should reflect the amount of interaction and the potential for learning a students "true type".

7 Results

7.1 General results

Figure 1 shows the relationship between teacher-assigned grades and students' standardized test scores for migrant and non-migrant students. In most cases, for language and math, students with a migrant background receive lower grades from their teachers despite having a similar performance on the standardized exam to non-migrant students. The econometric strategy employed in this paper attempt to verify if this gap persists after controlling for attributes of migrant and non-migrant students that may influence teacher evaluations, such as their socioeconomic status and behavioral traits.

Table 1 shows the first results of this exercise and reports significant gaps in teacher-assigned grades between migrant and non-migrant students, conditional on a set of basic covariates (demographics and socio-economic variables, described in details on Table A2). Also, results are based only on data from classrooms that contain at least one immigrant student. The first column shows that there is a difference of 0.049 grade points in math and 0.058 in language between immigrants and natives, favoring the latter. Columns 2 controls for classroom fixed effects, while columns 3 and 4 add students' test scores and demographic and socioeconomic characteristics, respectively. According to the model with all controls (column 4), teacher-assigned grades for immigrant students are significantly lower by 0.043 points in math and 0.038 points in language. Moreover, we found that, for immigrants students, the likelihood of scoring equal to or higher than the passing grade–that is "7"–and the classroom median was, respectively, 0.5 percentage points (p.p.) and 2 p.p. lower then native students.

Table 2 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 by teachers' subjective evaluations (Ferman and Fontes, 2020), the omission of these traits will likely bias the results if immigrants are perceived to have different behavior in class. The proxies for working habits and socio-emotional traits are fully described in Table A3. Although the grading differential between immigrant and native students decrease after controlling for these socio-emotional traits, teacher biases persist.

7.2 Heterogeneity

In Table 3, we examine heterogeneity in grading differentials between immigrants and natives by gender. Results indicate that evaluation bias against immigrants are only pertinent among boys. Amongst girls, teachers seem to not be penalizing immigrants when evaluating scholastic competence. One potential explanation for this finding is that negative priors associated with immigrants in the region (e.g. incidence of violence) are typical of males. These stereotyped priors are, therefor, being used as a signal when teachers estimate the level of proficiency of immigrant boys (but not girls) .

7.3 Learning

Albeit imperfectly, we approximate the quality of a teacher's grading technology by looking at strata based on their engagement with students (as reported by the latter). Under statistical discrimination, the longer pupils and teachers interact, the less teachers will rely on stereotyped priors to grade their students, and, therefore, biased grading is less likely to happen.

To explore potential statistical discrimination, in Table 4 we split the samples according to the median values (across classrooms) of the classroom-level reported engagement levels. If learning occurs, the grading differential between immigrants and natives would be closer to zero for highly interested and less absent teachers. At the same time, teachers' grades would reflect to a greater extent students' true proficiency level, which is captured without bias by blindly-scored tests–in other words, the slope of students' proficiency level on the standardized tests would increase.

While the patterns are interesting and point estimates for the immigrant coefficient seem to indicate stronger biases among less engaged and highly absent teachers, the absence of variation in the coefficient attached to test scores does not conform to predictions of the learning model.

8 Conclusions

We explore potential biases in teacher grading against immigrant students by employing uniquely detailed administrative data from Ecuador and juxtaposing teachers' subject-specific grades and scores from end-of-year standardized (and blindly marked) 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 portions of teachers' assessments in Mathematics and Language not explained by proficiency scores and yet related to a pupil's immigration status. We find statistically significant underscoring and under-ranking of immigrants relative to native high-school graduating students.

The measured gaps are sizeable and very much in line with the expected subtlety of this particular form of discrimination, including with the prevalence of implicit association biases. Despite their subtlety, there are a number of reasons to believe they are quite relevant. They are, for example, equivalent to 42% of the raw (within-classroom) grade differential associated with having a mother with a college degree or more versus a mother with a high-school degree only. Most importantly, we believe that the implication of our findings can be far reaching given the enormous potential for feedback effects in our context. This is the case because we detect discrimination in grading during the transition between high-schools 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.

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 noise 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 evaluation purposes. Second, direct investment in teacher training with regard to the design of exams and tests may be warranted. Well-designed questions are easier to grade and more likely to differentiate students on the most relevant dimensions of proficiency. Finally, because blindly graded proficiency tests can be taken by immigrant students at time of admission and curved with the overall student population, the generation of individual report cards 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 schools, education and immigration authorities could do a better job on their use of performance information in order to maximize teaching efficiency. The reduction of grading discrimination of the sort we uncover would be an added bonus.

References

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.

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.

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

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

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.

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

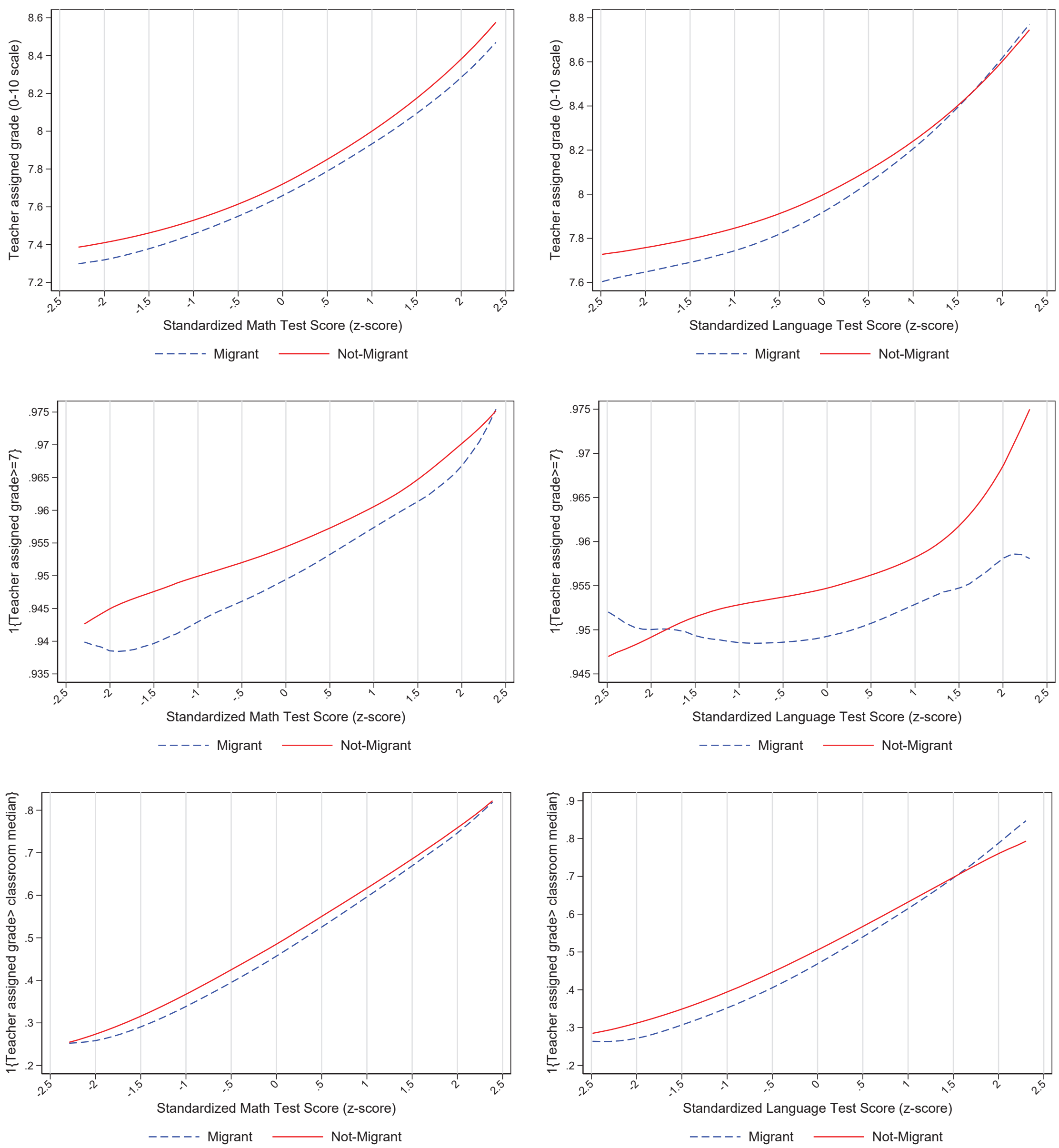

Figure 1: Teacher evaluation differentials: Math (Left) and Language (Right)

Source: *Ser Bachiller*.
Note: Based on 2018-2019 microdata.

Table 1: Teacher Grading and Test Performance: By number of immigrants in the school

	School with few number of immigrants (1)	School with many number of immigrants (2)
PANEL A: *Mathematics*		
Under "human mobility" condition	-0.008	-0.033***
	(0.015)	(0.010)
Math Test score	0.192***	0.221***
	(0.010)	(0.007)
PANEL B: *Spanish Language*		
Under "human mobility" condition	-0.006	-0.025***
	(0.014)	(0.010)
Language Test score	0.192***	0.177***
	(0.009)	(0.006)
Observations	35,878	70,172
Average number immigrants per school	0.47	14.69

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts. The average number of immigrant students per school is 6.65.

Table 2: Teacher Grading and Test Performance: By percentage of immigrants in the school

	School with low percentage of immigrants (1)	School with high percentage of immigrants (2)
PANEL A: *Mathematics*		
Under "human mobility" condition	-0.015	-0.030***
	(0.015)	(0.010)
Math Test score	0.182***	0.230***
	(0.010)	(0.007)
PANEL B: *Spanish Language*		
Under "human mobility" condition	-0.020	-0.018*
	(0.014)	(0.010)
Language Test score	0.193***	0.176***
	(0.009)	(0.006)
Observations	40,124	65,926
Average percentage of immigrants per school	0.0003	0.0125

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts. The average percentage of immigrant students per school is 0.0064.

Table 3: Teacher Grading and Test Performance: By number of immigrants in the school (our definition)

	School with few number of immigrants (1)	School with many number of immigrants (2)
PANEL A: *Mathematics*		
Under "human mobility" condition	-0.003	-0.029***
	(0.023)	(0.009)
Math Test score	0.210***	0.221***
	(0.015)	(0.006)
PANEL B: *Spanish Language*		
Under "human mobility" condition	-0.019	-0.020**
	(0.014)	(0.009)
Language Test score	0.198***	0.179***
	(0.013)	(0.005)
Observations	17,722	88,328
Average number immigrants per school	0.77	7.55

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts. The average number of immigrant students per school is 3.45.

Table 4: Teacher Grading and Test Performance

	Raw Difference	+Classroom FE	+Test Score	+Demographics and SES	– Alternative Outcomes – Passing Grade ×100	Above Median ×100
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: *Mathematics*						
Under “human mobility” condition	-0.049***	-0.033***	-0.046***	-0.043***	-0.450**	-1.999***
	(0.012)	(0.010)	(0.009)	(0.009)	(0.187)	(0.618)
Math Test score			0.341***	0.228***	0.802***	14.909***
			(0.006)	(0.006)	(0.122)	(0.388)
PANEL B: *Spanish Language*						
Under “human mobility” condition	-0.058***	-0.036***	-0.056***	-0.038***	-0.437**	-2.390***
	(0.013)	(0.009)	(0.009)	(0.008)	(0.173)	(0.626)
Language Test score			0.298***	0.203***	0.508***	13.123***
			(0.005)	(0.005)	(0.091)	(0.358)
Observations	106,050	106,050	106,050	106,050	106,050	106,050

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts.

Table 5: Teacher Grading and Test Performance: Robustness to socio-emotional characteristics

	Base Model	+Aspiration and working habits	+Socio-emotional profile	– Alternative Outcomes – Passing Grade $\times 100$	Above Median $\times 100$
	(1)	(2)	(3)	(4)	(5)
PANEL A: *Mathematics*					
Under "human mobility" condition	-0.043***	-0.032***	-0.025***	-0.390**	-0.812
	(0.009)	(0.008)	(0.008)	(0.187)	(0.608)
Math Test score	0.228***	0.213***	0.211***	0.730***	13.871***
	(0.006)	(0.006)	(0.006)	(0.121)	(0.379)
PANEL B: *Spanish Language*					
Under "human mobility" condition	-0.038***	-0.026***	-0.020**	-0.382**	-1.178*
	(0.008)	(0.008)	(0.008)	(0.173)	(0.615)
Language Test score	0.203***	0.183***	0.181***	0.452***	11.738***
	(0.005)	(0.005)	(0.005)	(0.090)	(0.351)
Observations	106,050	106,050	106,050	106,050	106,050

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts.

Table 6: Teacher Grading and Test Performance: By gender

	Amongst Girls only	Amongst Boys only
	(1)	(2)
PANEL A: *Mathematics*		
Under "human mobility" condition	-0.008	-0.044***
	(0.012)	(0.012)
Math Test score	0.203***	0.220***
	(0.009)	(0.009)
PANEL B: *Spanish Language*		
Under "human mobility" condition	-0.007	-0.034***
	(0.012)	(0.011)
Language Test score	0.195***	0.159***
	(0.008)	(0.008)
Observations	35,963	38,024

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts.

Table 7: Teacher Grading and Test Performance Amongst Boys: By teacher perceived involvement

	Highly interested teachers (1)	Less interested teachers (2)	Highly absent teachers (3)	Less absent teachers (4)
PANEL A: *Mathematics*				
Under "human mobility" condition	-0.009	-0.070***	-0.056***	-0.029*
	(0.018)	(0.016)	(0.016)	(0.017)
Math Test score	0.206***	0.234***	0.234***	0.205***
	(0.014)	(0.012)	(0.013)	(0.013)
PANEL B: *Spanish Language*				
Under "human mobility" condition	-0.036**	-0.031**	-0.037**	-0.029*
	(0.017)	(0.015)	(0.015)	(0.016)
Language Test score	0.154***	0.162***	0.165***	0.151***
	(0.013)	(0.011)	(0.011)	(0.012)
Observations	16,314	21,710	20,078	17,946

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts.

Appendix

Table A1: Descriptive Statistics - Students (Ser Bachiller 2018-2019)

	Full sample	Classrooms w/o immigrant mix	Classrooms w immigrant mix	Classrooms w immigrant mix + missing<50%
	(1)	(2)	(3)	(4)
Immigration proxy				
"Human mobility" undetermined	0.480	0.492	0.472	0.409
Under "human mobility" condition	0.037	0.000	0.061	0.064
Geography				
Costa	0.605	0.646	0.579	0.648
Rural	0.209	0.272	0.169	0.152
Demographics				
Female	0.505	0.503	0.507	0.507
Disable	0.009	0.009	0.010	0.009
White	0.021	0.019	0.022	0.023
Mestizo	0.828	0.807	0.841	0.836
Afro-descendant	0.041	0.043	0.040	0.043
Indigenous/Montubio	0.106	0.127	0.093	0.094
Other race	0.004	0.004	0.004	0.004
Socio-economic status				
SES Index	-0.164	-0.304	-0.076	-0.059
Missing SES Index	0.003	0.003	0.003	0.002
Mother not educated	0.041	0.047	0.037	0.035
Mother basic education	0.521	0.554	0.500	0.484
Mother high-school education	0.292	0.265	0.309	0.320
Mother college education	0.077	0.065	0.085	0.090
Missing mother ed.	0.069	0.069	0.069	0.071
Home environment				
No dictionary	0.252	0.263	0.245	0.242
No books	0.101	0.106	0.098	0.101
Proficiency and school grades				
Standardized test score, Math	770.806	764.925	774.528	770.572
Standardized test score, Language	776.092	768.728	780.753	777.867
Math grade	7.704	7.704	7.704	7.728
Math grade>=7	0.952	0.950	0.953	0.955
Math grade>=class median	0.474	0.475	0.474	0.474
Lang. grade	7.974	7.983	7.968	8.010
Lang. grade>=7	0.954	0.955	0.954	0.959
Lang. grade>=class median	0.484	0.483	0.484	0.484
Observations (students)	173,174	67,124	106,050	62,802
Observations (classrooms)	6,053	2,504	3,549	2,072

Notes: Ser Bachiller 2018-2019 data.

Table A2: Descriptive Statistics - Students in Mixed-Immigration Status Classrooms

	Immigrant (1)	Non-Immigrant (2)	Undetermined status (3)
Immigration proxy			
"Human mobility" undetermined	0.000	0.000	1.000
Under "human mobility" condition	1.000	0.000	0.000
Geography			
Costa	0.527	0.607	0.557
Rural	0.188	0.163	0.174
Demographics			
Female	0.472	0.514	0.504
Disable	0.011	0.010	0.010
White	0.026	0.022	0.022
Mestizo	0.831	0.842	0.841
Afro-descendant	0.041	0.040	0.040
Indigenous/Montubio	0.098	0.092	0.093
Other race	0.004	0.004	0.004
Socio-economic status			
SES Index	0.040	-0.079	-0.088
Missing SES Index	0.001	0.001	0.005
Mother not educated	0.040	0.036	0.039
Mother basic education	0.476	0.497	0.506
Mother high-school education	0.321	0.313	0.304
Mother college education	0.090	0.087	0.082
Missing mother ed.	0.073	0.067	0.070
Home environment			
No dictionary	0.246	0.243	0.246
No books	0.082	0.098	0.099
Proficiency and school grades			
Standardized test score, Math	778.740	772.229	776.258
Standardized test score, Language	787.154	778.441	782.213
Math grade	7.671	7.720	7.693
Math grade>=7	0.949	0.954	0.951
Math grade>=class median	0.461	0.477	0.472
Lang. grade	7.931	7.990	7.951
Lang. grade>=7	0.950	0.955	0.953
Lang. grade>=class median	0.468	0.490	0.481
Observations	6,467	49,517	50,066

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts.

Table A3: Socio-Emotional Profile - Students in Mixed-Immigration Status Classrooms

	Immigrant (1)	Non-Immigrant (2)	Undetermined status (3)
Aspiration and working habits			
Aspire to go to college	0.861	0.861	0.852
Always attentive to student obligations	0.579	0.652	0.639
Study $<$ 1 hour a day	0.150	0.137	0.139
Study between 1 and 2 hours a day	0.396	0.388	0.391
Study between 2 and 4 hours a day	0.222	0.229	0.225
Study more than 4 hours a day	0.225	0.240	0.237
Socio-emotional profile			
Missing socio-emotional profile	0.030	0.055	1.000
– Attention to detail			
Almost never attentive to details	0.118	0.091	-
Almost always attentive to details	0.409	0.377	-
Always attentive to details	0.427	0.464	-
– Procrastination			
Almost never finishes what starts	0.058	0.049	-
Almost always finishes what starts	0.430	0.379	-
Always finishes what starts	0.471	0.508	-
– Perseverance (reverse)			
Almost never persevere on task	0.219	0.197	-
Almost always persevere on task	0.440	0.438	-
Always persevere on task	0.259	0.251	-
– Creativity			
Almost never comes up with new ideas	0.203	0.204	-
Almost always comes up with new ideas	0.449	0.441	-
Always comes up with new ideas	0.265	0.234	-
– Empathy			
Almost never empathize	0.079	0.076	-
Almost always empathize	0.456	0.446	-
Always empathize	0.418	0.406	-
– Kindness			
Almost never kind	0.031	0.026	-
Almost always kind	0.410	0.367	-
Always kind	0.523	0.548	-
– Extroversion			
Almost never extrovert	0.179	0.158	-
Almost always extrovert	0.362	0.374	-
Always extrovert	0.385	0.374	-
– Spontaneity			
Almost never spontaneous	0.133	0.123	-
Almost always spontaneous	0.398	0.393	-
Always spontaneous	0.413	0.404	-
Observations	6,467	49,517	50,066

Notes: Ser Bachiller 2018-2019 and administrative data on public-school transcripts.