I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11981

Revealing Stereotypes: Evidence from Immigrants in Schools

Alberto Alesina
Michela Carlana
Eliana La Ferrara
Paolo Pinotti

NOVEMBER 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11981

Revealing Stereotypes: Evidence from Immigrants in Schools

Alberto Alesina
Harvard University, IGIER Bocconi University, NBER and CEPR

Michela Carlana
Harvard Kennedy School and IZA

Eliana La Ferrara
IGIER and LEAP, Bocconi University

Paolo Pinotti
Bocconi University, DONDENA and Fondazione Rodolfo Debenedetti

NOVEMBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Revealing Stereotypes: Evidence from Immigrants in Schools*

If individuals become aware of their stereotypes, do they change their behavior? We study this question in the context of teachers' bias in grading immigrants and native children in middle schools. Teachers give lower grades to immigrant students compared to natives who have the same performance on standardized, blindly-graded tests. We then relate differences in grading to teachers' stereotypes, elicited through an Implicit Association Test (IAT). We find that math teachers with stronger stereotypes give lower grades to immigrants compared to natives with the same performance. Literature teachers do not differentially grade immigrants based on their own stereotypes. Finally, we share teachers' own IAT score with them, randomizing the timing of disclosure around the date on which they assign term grades. All teachers informed of their stereotypes before term grading increase grades assigned to immigrants. Revealing stereotypes may be a powerful intervention to decrease discrimination, but it may also induce a reaction from individuals who were not acting in a biased way.

JEL Classification: I24, J15

Keywords: immigrants, teachers, implicit stereotypes, IAT, bias in grading

Corresponding author:
Michela Carlana
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138
United States

E-mail: michela_carlana@hks.harvard.edu

* We acknowledge useful comments from seminar participants at Harvard University, BRIQ-Bonn, Mannheim University, Collegio Carlo Alberto, Uppsala University, University of Vienna, and University of Milan. We are grateful to the schools and teachers that took part in our project and to Gianna Barbieri and Lucia De Fabrizio from MIUR and Patrizia Falzetti and Paola Giangiacomo from INVALSI for giving us access to the administrative data used in this paper. Elena De Gioannis and Giulia Tomaselli provided invaluable help with data collection. Carlana acknowledges financial support from the "Policy Design and Evaluation Research in Developing Countries" Initial Training Network (PODER), which is financed under the Marie Curie Actions of the EU's Seventh Framework Programme (Contract no. 608109). La Ferrara acknowledges financial support from the ERC Advanced Grant "Aspirations, Social Norms and Development"(ASNODEV, Contract no. 694882).

1 Introduction

Stereotypes are over-generalized representations of characteristics of certain groups (Bordalo et al., 2016). They allow for easier and efficient processing of information, but they may cause biased judgment or even discrimination against particular groups. In addition, discrimination may lead to self-fulfilling prophecies by influencing the behavior of discriminated groups in the direction of the stereotypes. For example, individuals exposed to bias toward their own group may reduce effort, self-confidence, and productivity (Carlana, 2018; Bordalo et al., 2018; Glover et al., 2017).

We consider the case of immigrants in contemporary societies. Alesina et al. (2018) report robust evidence of anti-immigrant stereotypes in several receiving countries, including Italy. The problem is especially important in the case of possible discrimination experienced by young generations in schools. In the short-run, young immigrants who feel discriminated against may become discouraged and work less, with consequences for their careers and job placement. Eventually, they may also develop a belief that effort does not pay off for an immigrant, not only in school but also in the workplace and in society in general.[1] Thus, if negative stereotypes are internalized by individuals at a time when they are making long-term investments in their future, they may make suboptimal decisions that will impact their future careers and well-being. Making stereotypes known to those who interact with young generations may help address these problems.

In this paper, we study the behavior of teachers toward immigrant and native students in Italian middle schools.[2] This is an interesting context for two reasons. First, mass immigration is a relatively recent phenomenon in Italy. In fact, after Spain, Italy has experienced the second largest increase in the share of immigrants received among European countries over the last decade. Second, in the Italian education system -like in most European countries- middle school is a critical juncture at the end of which students get tracked into different education paths that affect their future education and work prospects.

We begin by showing that, holding constant performance on standardized, blindly-graded tests, immigrant students receive lower grades when graded by their teachers in a non-blind way compared to natives. Lower grades may of course reflect factors other than stereotypes, such as statistical discrimination or differential performance of non-native speakers on multiple choice tests. In order to isolate the role of stereotypes, we administer a survey to teachers including an Implicit Association Tests (IAT) to capture stereotypes against immigrants. The IAT is a computer-based tool developed by social psychologists, designed to minimize the risk of social desirability bias in

[1] In turn, the perception of effort put in by the needy significantly affects the propensity of society to help the poor (Alesina and Glaeser, 2004; Fong, 2001; Alesina et al., 2018).

[2] Immigrant students are defined according to their citizenship. They include first generation students born abroad and second generation students born in Italy from both parents who are not Italian citizens.

self-reported answers, and it is widely used in several contexts (e.g., Greenwald et al., 2009). As we explain in detail in Section 3.1, our IAT measures stereotypes using the difference in reaction times when teachers are asked to associate positive and negative attributes with immigrant-sounding and native-sounding names.

The results of our IAT show that teachers generally hold strong, negative stereotypes towards immigrant students. According to the metrics for IAT scores proposed by Greenwald et al. (2009), two thirds of teachers in our sample exhibit 'moderate to severe' stereotypes, and as much as 91 percent exhibit some stereotypes against immigrants. Additional evidence on the correlates of IAT score suggests that teachers' stereotypes against immigrants – and the resulting bias in grading – mainly reflects discriminatory beliefs. Negative stereotypes about immigrants do not reflect worse past experiences with this group, as measured by the immigrant-native gap in standardized test scores across previous cohorts of students met by each teacher. Furthermore, teachers' IAT scores correlate with their own self-reported prejudice against immigrants in other contexts, as measured by specific questions we adapted from the World Values Survey.

We find that math teachers with higher IAT score (indicating more negative stereotypes against immigrants) give lower grades to immigrant students compared to native ones with the same performance on blindly-graded tests. This provides suggestive evidence that at least part of the grade penalty in math reflects bias, and is not purely due to statistical discrimination or unobserved student characteristics. Stereotypes of literature teachers do not affect grades assigned to immigrants on average, potentially reflecting different teachers' expectations and standards applied in literature for immigrants who are less familiar with Italian.[3]

After showing that there is a link between discrimination in grading and teachers' stereotypes, we evaluate the effectiveness of a simple intervention aimed at reducing bias. The intervention consists of revealing own stereotypes to teachers. Specifically, we allow a randomly selected sample of teachers to see their IAT score shortly before the end of semester grading, while the other teachers were allowed to do so only shortly after. We find that both math and literature teachers who received their IAT score before the date when they submitted grades gave higher grades to immigrants relative to native students, compared to teachers who received their IAT scores afterwards. This result is driven by teachers who do not report explicit views against immigrants in answer to the World Values Survey question on labor market access for immigrants. In other words, the effect of revealing own IAT score holds only for those teachers who did not appear as biased in explicit

[3]In fact, higher implicit stereotypes of literature teachers have negative effects on grading for second generation immigrants and positive effects for first generation immigrants, with the latter being the least familiar with the language. In other words, literature teachers appear to internalize the need to help first generation immigrants because of language difficulties, regardless of their stereotypes. That is, even a biased teacher may acknowledge language problems of recent immigrants and be lenient although still biased.

survey questions.

The results of our experiment tell us three things. The first is that teachers who did not express explicit anti-immigrant views were not aware of (or had not fully internalized) their stereotypes, otherwise our communication would not have contained any new information based on which they could update their behavior. The second is that the intervention was effective for the set of teachers whose grading varied systematically with their IAT score (i.e., math teachers). The last is that this intervention may also induce reactions by people who may not act upon their stereotypes in a discriminatory way: literature teachers' grading of immigrants does not vary with their IAT score, yet they respond to receiving a message that tells them that they hold negative implicit stereotypes against immigrants.

Our work is related to several strands of literature. Within the broad area of empirical studies of discrimination, we contribute to the recent economics literature emphasizing the benefits of considering implicit bias (Van den Bergh et al., 2010; Guryan and Charles, 2013; Lowes et al., 2015; Burns et al., 2016; Bertrand and Duflo, 2017).[4] Exploiting data from French grocery stores, Glover et al. (2017) provide evidence that exposure to managers with stronger implicit bias, as measured by an IAT similar to ours, negatively affects performance of minorities in the workplace. In the context of gender bias, Reuben et al. (2014) show in a lab experiment that the gender IAT predicts employers' biased expectations against females and a suboptimal update of expectations after ability is revealed. Carlana (2018) shows that teachers' stereotypes affect the gender gap in math, track choice, and self-confidence in own mathematical abilities for girls in middle school.

Research in social psychology and medicine has examined individuals' emotional responses when provided feedback about their own implicit associations, showing that people tend to react defensively – for instance, by questioning the validity of the IAT – when provided with evidence about tensions between their own explicit and implicit bias (O'Brien et al., 2010; Howell et al., 2015; Sukhera et al., 2018). However, none of these papers investigates whether revealing own stereotypes to people has an impact on discriminatory behavior toward others. This is precisely the contribution of the present paper.

We also add to the literature that analyzes bias of teachers against minority children. This literature finds that teacher expectations are often biased against minority students. This may lead to a self-fulfilling prophecy with students ultimately behaving in the direction predicted by biased expectations (Papageorge et al., 2018; Jussim and Harber, 2005; Rosenthal and Jacobson, 1968). In a

[4]Economists have widely studied discrimination at least since Becker et al. (1957) (for a review of theoretical and empirical issues, see Altonji and Blank, 1999). Evidence of discrimination include that by employers (Bertrand and Mullainathan, 2004), police officers (Fryer Jr, 2016; Coviello and Persico, 2015; Knowles et al., 2001), courts (Dobbie et al., 2018; Alesina and La Ferrara, 2014), and teachers (Figlio, 2005).

lab experiment, Gilliam et al. (2016) track teachers' eye gazes while watching a video and find that, when expecting challenging behaviors, teachers gazed longer at black children, even if all children were behaving similarly. A few previous papers compare teacher-assigned (non-blind) grades and standardized (blind) test scores across minority and non-minority students (Botelho et al., 2015; Burgess and Greaves, 2013; Hanna and Linden, 2012; Van Ewijk, 2011) and across genders (Lavy and Sand, 2018; Lavy, 2008). Those papers cannot distinguish between the role of teachers' biases and unobserved student characteristics, which we are able to do using the IAT as a direct measure of teachers' stereotypes. Furthermore, none of the above papers test the effectiveness of remedial interventions.[5]

Finally, our results speak to recent studies that investigate how to reduce bias towards immigrants and affect people's inclination towards immigration policy. A first group of studies has analyzed the impact of providing information about immigrants on attitudes towards them (Grigorieff et al., 2017; Hopkins et al., 2018) and towards immigration policy (Facchini et al., 2016). A second group has focused on the contact hypothesis, i.e., the idea that promoting inter-group contact may help reduce prejudice (Allport, 1958). For instance, Burns et al. (2016) show that exposure to a roommate of a different race affects stereotypes (measured by the IAT), attitudes, and academic performance of students. As suggested also by a recent meta-analysis by Paluck et al. (2018), contact "typically reduces prejudice [but] (...) the absence of studies addressing adults' racial or ethnic prejudices [is] an important limitation for both theory and policy".

We consider a novel intervention that targets adults and consists of providing information to teachers on their own stereotypes. Our intervention has clear policy relevance given the increasing trend among institutions and corporations, especially in the U.S., to administer the IAT to their employees as an educational tool to increase awareness about race and gender implicit associations.[6] Our paper contributes to understanding the causal effect of providing this information to people.

The remainder of the paper is organized as follows. In section 2 we provide some background information about the grading system in Italian middle schools. Section 3 describes our data and the experimental design. In section 4 we present our results, and the last section concludes.

[5] Figlio et al. (2018) study the performance of immigrant children in Florida and relate it to some aspects of their home country cultural traits, in particular long-term orientation.

[6] Among others, Harvard University strongly encourages "every search committee member to take at least one Implicit Association Test (IAT)" (https://faculty.harvard.edu/recruitment-best-practices) and Starbucks has recently promoted a "racial bias training" for all employees (https://twitter.com/starbucks/status/997528229593280513).

2 Institutional background

In the Italian schooling system, the share of immigrant children, defined as children without Italian citizenship, has substantially increased in the last two decades, moving from less than 1 percent in 1998 to 10 percent in 2018. Immigrant students come from diverse geographic backgrounds, with the most represented nationalities being Romanian, Albanian, Moroccan, Chinese, Filipino, and Indian (see Appendix Table A.1).

Middle school is the analog of grades 6 to 8 in the US system. Students during those years are assigned to the same class for all subjects, with an average class size of 24. Students are usually taught by the same teachers for all three years of middle school, and they spend at least 6 hours with the math teacher and 5 hours with the literature teacher per week. Teachers are assigned to schools by the Italian Ministry of Education and their allocation is determined by seniority: teachers with more experience can get schools that are higher in their preference ranking, and tend to move close to their home town and away from disadvantaged areas (Barbieri et al., 2011).

Grades received by students in Italian middle schools typically range from 3 to 10; 6 is the minimum passing grade. Students are assessed continuously, but they receive an official grade for each subject twice a year: at the end of the first semester (in January) and at the end of the second (in June). These two sets of 'final' grades mostly reflect their performance on exams, but also incorporate a broader evaluation of their diligence in doing homework and attention during lectures. In the questionnaire we administered to teachers, we ask them to describe the relative weights they assign to these various factors. Both math and literature teachers reported that they assign significant importance to attention in the classroom. Thus, end of semester grades may include a non-negligible component of subjective evaluation by teachers. While the final (possibly subjective) grades are given each year, standardized test scores in math and reading are administered only at the end of middle school (grade 8) to all Italian students. These are mainly multiple choice exams prepared by the National Institute for the Evaluation of the Italian Education System (INVALSI). Importantly, for the purposes of our study, INVALSI tests are blind graded following a precise evaluation grid, and the grading is not done by students' own teachers.

The Italian schooling system is characterized by tracking at the end of grade 8. Students are free to choose between academic high-school (*liceo*), technical high-school, and vocational high-school. Academic and technical schools offer much better educational and employment prospects than vocational schools (Carlana et al., 2018).

3 Data and empirical strategy

3.1 Teachers survey and the IAT

Our sample includes 65 schools in five cities of Northern Italy: Milan, Brescia, Padua, Genoa and Turin. We invited these schools to take part in our research project, describing the scope of our research as being about "The role of teachers in high-school track choice". A sizeable part of the questionnaire was indeed devoted to understanding the criteria used by teachers to advise students on high school choice for a different project. We intentionally avoided mentioning immigrants when describing our research, because we did not want teachers to select into participating based on their sensitivity to the topic of immigration or their bias. From September 2016 to March 2017, we administered the survey to all math and literature teachers in the 102 schools in our sample. However, only 65 schools were surveyed before the end of January 2017, which is when the mid-term grades are assigned, so these are the schools included in the main analysis of this paper. We show below that the remaining 38 schools are not different in terms of most observable characteristics.

The survey was conducted by enumerators using tablets during meetings held in school buildings with all math and literature teachers of each school. Enumerators gave each teacher one tablet to complete the survey independently and were available in the room to answer questions or help with the tablet if requested. Teachers who agreed to take part in the survey gave written informed consent. The time to complete the survey was around 30 minutes, and participants did not receive any compensation. Among all math and literature teachers working in the schools included in the final sample, around 80 percent completed our survey, yielding a sample of 1384 teachers for our analysis.

The first part of our survey involved an IAT aimed at measuring implicit stereotypes toward immigrants. This was followed by a series of questions on teachers' demographic characteristics, teaching experience, and explicit beliefs about immigrants. The idea underlying the IAT is that the easier the mental task, the faster the response production. Therefore, the test requires categorizing words to the left or to the right of a computer or tablet screen and measures the strength of the association between two concepts based on response times.[7] Specifically, the IAT that we administered associates immigrant/native children with words related to good/bad in the specific schooling context. Teachers were presented with two sets of stimuli. The first set included typical Italian names (e.g., Francesca or Luca) and common names among immigrant children in Italy (e.g., Fatima or Alejandro). The second set consisted of positive adjectives (e.g., smart) and negative ones

[7] This concept was initially developed by Donders (1868).

(e.g., lazy). One word at a time (either a name or an adjective) appears at the center of the screen, and individuals are instructed to categorize it to the left or to the right according to different labels displayed on the top of the screen. For instance, the right label might say "Immigrant" and the left one might say "Italian". Names and adjectives randomly appear at the center of the screen, and subjects are asked to categorize the words as quickly as possible. In one type of round, subjects are asked to categorize native-sounding names and negative adjectives to the same side of the screen, whereas in another, they are asked to categorize immigrant-sounding names and negative adjectives to the same side. The order of the two types of rounds is randomly selected at the individual level.

To the extent that teachers are biased against immigrant students, they should react more slowly when the label "Immigrant" ("Italian") is associated with positive (negative) adjectives, because those associations are more difficult for their minds to process. The IAT thus measures stereotypes by the difference in reaction times between rounds in which native-sounding names and negative adjectives appear on the same side of the screen and rounds in which immigrant-sounding names and negative adjectives are on the same side. It is hard to fake the IAT, as this would require strategically speeding up or slowing down in certain blocks, and the improved scoring algorithm (Greenwald et al., 2003) discards observations in which reaction times are abnormally slow or fast. Also, faking is quite unlikely within our sample of teachers, as the IAT is not widely known in Italy.

IATs have the great advantage of (i) avoiding social-desirability bias in the response and (ii) capturing implicit associations that may be unknown to the individual, but may nevertheless affect his or her interaction with the stigmatized group. For this reason, the IAT is a widely-used tool in social psychology. Although there is some mixed evidence on its predictive validity (Blanton et al., 2009; Oswald et al., 2013; Olson and Fazio, 2004), implicit bias has been found to correlate with outcomes in the real world and in laboratory experiments (Nosek et al., 2007; Greenwald et al., 2009; Burns et al., 2016), including, e.g., call-back rates of job applicants (Rooth, 2010).

[Insert Figure 1]

Figure 1 shows the distribution of IAT scores for the teachers in our sample. A positive score means a stronger association between foreign-sounding names and negative attributes, while a negative score suggests a stronger association between foreign-sounding names and positive attributes. Using the typical thresholds in the literature (Greenwald et al., 2009), an IAT score between -0.15 and 0.15 indicates no bias, a score between 0.15 and 0.35 (in absolute value) a slight association between the two concepts, and a score higher than 0.35 (in absolute value) a moderate to severe association. According to this metric, the IAT scores in Figure 1 suggest that teachers are generally biased against immigrants. The mean IAT score is 0.47, which is slightly higher

than the mean of 0.41 in the sample of Italians who decided to take the IAT online on the website https://implicit.harvard.edu, and substantially higher than the mean of most European countries (see Figure A.1). Over 67 percent of the teachers exhibit moderate to severe bias, i.e., a score greater than 0.35.

[Insert Table 1]

Table 1 (columns 1 to 5) shows the correlation between a number of teachers' characteristics and their IAT scores. Female teachers are less biased than male teachers, and so are teachers born in northern Italy compared to those born in southern Italy, although in both cases the difference is very small. We find no relevant differences along other characteristics such as having children or parental background. On the other hand, IAT scores are correlated with beliefs about immigrants, as elicited by a question we included in the survey. In particular, column 5 shows that teachers who believe that immigrants and natives should have equal opportunities of accessing available jobs on average have weaker implicit stereotypes against immigrants.[8]

In columns 6 and 7 of Table 1, we test whether teachers' stereotypes reflect the relative average ability or variation in the ability of the native and immigrant students to whom teachers were previously exposed. We collected the standardized test scores (INVALSI) of the students taught by the teachers in our sample during the five years prior to our analysis. We could recover previous students' test scores for 779 out of 1384 teachers, which explains the reduction in the sample in columns 6 to 9.[9] We find no meaningful correlation between teachers' IAT and the relative average test scores (column 6) or the standard deviation difference in test score (column 7) of native and immigrant students they taught in the past. Therefore, stronger stereotypes toward immigrant students do *not* seem to reflect statistical discrimination based on objective information on average group ability.

The last two columns show that the results remain basically unchanged when we introduce all regressors at the same time (column 8) and when we include school fixed effects (column 9). Overall, the evidence in Table 1 is consistent with the IAT capturing negative beliefs against immigrants.

In the Appendix Table A.2, we show the correlation between teachers' implicit stereotypes and beliefs about factors affecting the different high-school choice of native and immigrant children (the detailed questions are reported in Appendix B.2). Teachers with a higher IAT score are more likely to believe that the lower ability is a crucial explanation of why immigrants are more likely to go to vocational high schools, rather than academic oriented ones, while belief in the importance

[8] We took the wording of this question from the World Value Survey.

[9] We include teachers who had at least three immigrant (and native) students.

of other factors that may determine the choice of high-school track, such as economic reasons, are not correlated with IAT scores.

3.2 Student performance

To measure students' academic performance, we obtained administrative data on the grades received by students at the end of grade 8 for schooling years between 2011-12 and 2015-16 from all schools in our sample. We have two separate measures of performance for each student: grades received from teachers on non-blindly graded tests, and blindly-graded, standardized test scores from INVALSI. Appendix Figure A.2 shows the distribution of grades assigned by teachers (Panel A) and of standardized test scores (Panel B) separately for native and immigrant students. In both cases, the distribution for natives first order stochastically dominates that for immigrants. This fact may reflect differences in ability between native and immigrant students as well as differences in the grading policy of teachers toward the two groups. As a first step to isolate the latter, we compare grades received from teachers by native and immigrant students with comparable standardized test score.

[Insert Figure 2]

Figure 2 plots the average grades given by teachers to immigrant and native students (on the vertical axis) by quintiles of the standardized test score (on the horizontal axis), with the associated 95 percent confidence intervals. Conditional on obtaining the same standardized test score, immigrant students receive significantly lower grades from teachers along the whole distribution. The average gap is 0.13 points for math and 0.20 points for literature, comparable to the difference in grades by mother's education. In fact, controlling for the quintiles of the standardized test score, students whose mothers have a university degree receive a grade that is on average 0.15 points higher for math and 0.21 for literature, compared to children of mothers with a high-school diploma.

While this gap in grades could be due to bias against immigrants, it may also be explained by the ability of teachers to assess aspects of multidimensional competence (e.g., oral expression, behavior in class, etc.) that are not easily captured by multiple choice, standardized tests. We address this point in section 4.1.

3.3 The experiment

We administered the survey with the IAT test in 102 schools, but, due to logistical constraints, we only managed to complete the data collection for 65 of them before the end of semester grading

in January.[10] Only these 65 schools are thus included in our experiment, so that our sample for the evaluation of the experimental intervention contains 6,031 students in grade 8 in the school year 2016-2017 and their 533 teachers, 262 in math and 271 in literature. We have information on student grades and their math teachers for 5,141 pupils, and on student grades and their literature teachers for 5,138.

Balance in the characteristics of teachers surveyed before the end of January (hence included in our analysis) and after the end of January (not included) is shown in Appendix Table A.3. Balance in the characteristics of their students is shown in Appendix Table A.4. In both cases the two groups are comparable in terms of the characteristics we care about. Importantly, the sample of teachers is perfectly balanced in terms of implicit bias against immigrants (first row in Appendix Table A.3). For the few characteristics that differ across samples (e.g., gender or place of birth), the normalized difference never exceeds the cutoff of 0.25 suggested by Imbens and Rubin (2015).[11] Overall, the sample of schools used for the intervention is representative of the original population we started from.

We offered the possibility of receiving feedback on the IAT score to all teachers in our sample and more than 80 percent of teachers chose to have it. The timing of feedback was randomized across schools. Teachers in half of the schools (the treated group) received the feedback before the end-of-semester grading, which took place at the end of January 2017. Teachers in the remaining schools (the control group) received the feedback within two weeks. We chose to randomize at the school level rather than at the teacher level in order to avoid contamination between teachers who received the early feedback and those who received the feedback after term grading. The teachers were not aware that we could observe the grades they gave to their students. We obtained this information directly from the National Evaluation Center, substantially reducing the risk of experimenter demand effects in teachers' grading (De Quidt et al., 2018).

The feedback was provided over e-mail. Each teacher received his/her IAT score and a brief description of the test explaining whether in their case the association between immigrant names

[10] The difference in the times when the survey was administered depended on logistical constraints on our side (e.g., availability of tablets and enumerators) and on the schools' side. In particular, the principals needed to agree on the data collection and needed to consent to share administrative data on teachers' grades and standardized test scores of all their students.

[11] The normalized difference we show in column 4 is the formula recommended by Imbens and Wooldridge (2009):

$$\Delta = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2 + S_2^2}}$$

where $\overline{X}_1$ and $\overline{X}_2$ are the means of covariate X in the two sub-groups that are being compared, and S_1^2 and S_2^2 are the corresponding sample variances of X. Imbens and Rubin (2015) recommend as a rule of thumb that Δ should not exceed 0.25.

and good/bad adjectives was "slight", "moderate" or "strong"based on the thresholds typically used in the literature (Greenwald et al., 2009) and represented also in Figure 1. They were also reassured that these results would not be shared with anyone. The detailed text of the e-mail is reported in Appendix B.3. Appendix Table A.5 provides evidence on the correlation between several teacher characteristics and the decision to receive the feedback. Interestingly, there is no significant correlation with either implicit or explicit biases against immigrants – respectively, IAT scores and answers to the World Value Survey question on immigrant rights to jobs – nor are there significant correlations with the subject taught or own gender.[12]

[Insert Table 2 and 3]

Tables 2 and 3, respectively, show the average characteristics of teachers and students in treated and control schools. As should be expected by virtue of randomization, teachers and students in the two groups are balanced in terms of observable characteristics. None of the variables are statistically different at conventional levels between the two groups. The share of female teachers is slightly higher in the treatment group (only significant at the 10 percent level), but even in this case the standardized difference is below the critical threshold of 0.25 suggested by Imbens and Rubin (2015).

[Insert Figure 3]

Figure 3 shows the timeline of the survey and experiment, as well as the periods covered by the data on standardized test scores and teacher-assigned grades. The top part of the figure shows the sample of students used to analyze whether teachers with more implicit bias actually discriminate against immigrants. For this exercise, we exploit cohorts of students graduating from middle school between June 2012 and June 2016 for whom we have the standardized test score (INVALSI) in June of grade 8 and teacher-assigned grades in exactly the same period. In other words, the analysis of bias in grading is done using *end-of-year* grades because this is the only time at which two independent measurements of achievement exist. Importantly, knowledge of our study could not affect the behavior of teachers toward these cohorts of children given that they graduate from middle school before our data collection.

[12]Instead, there is a significant correlation with the time employed to complete the survey: completion time one standard deviation or more than the mean value is associated with a 5 percentage point higher probability of consenting to receive the feedback. In addition, those who completed only the IAT and not the survey were almost 10 percentage points less likely to consent to receive the email about own stereotypes. These correlations do not survive when including school fixed effects, which explain a substantial share of the variation in the choice of receiving feedback as shown by the R-squared.

The bottom part of the figure shows the timeline for the survey and experiment. During the first semester of school year 2016/17 (October-January), we administered the survey and the IAT to all teachers in our sample. During the last week of January 2017, before end of semester grading, we sent feedback about own IAT score to teachers in a random group of schools. All other teachers were allowed to see their score after the mid-term grading (i.e., first week of February 2018). The impact of the intervention is tested by looking at the *end-of-semester* grades given by teachers to the cohort of students in grade 8, who eventually graduated from middle school in June 2017. No standardized test scores exist for evaluations at the end of the first semester, which implies that in analyzing the impact of our experiment we cannot control for the INVALSI score (but this does not affect our ability to estimate the impact of the intervention, given randomization). Data on teacher-assigned grades at the end of the first semester are collected by INVALSI only for students in grade 8. This is why we restrict our analysis to students in grade 8.

4 Results

4.1 Implicit biases and grading

Table 4 investigates the relationship between teachers' implicit bias and grading of immigrant vs. native students. The sample includes all students enrolled in grade 8 in our schools between school years 2011-12 and 2015-16. The dependent variable is end-of-year grade received from the math (Panel A) and literature (Panel B) teacher. Note that this grade differs from the mid-term grade given at the end of January, which we use to evaluate the impact of our experiment. The advantage of considering end-of-year grades for this part of the analysis is that students take the INVALSI standardized test at the same time, which allows us to compare students' performance on blindly and non-blindly graded tests. All specifications in columns 1-6 of Table 4 include teacher fixed effects. We include class fixed effects as a robustness check in the last column.

[Insert Table 4]

Column 1 shows that immigrant children receive on average half a grade less than natives in both math and literature (on a scale going from 3 to 10, where 6 is the passing grade). The gap is slightly larger for literature compared to math: indeed, we should expect language to be a stronger barrier for immigrants in the former subject than in the latter. After controlling for student ability, as measured by a cubic polynomial of the standardized test score (INVALSI), the gap decreases to about 0.1 points, or a tenth of a standard deviation, in both math and literature (column 2).

In column 3 of Panel A, we interact the dummy for immigrant student with math teacher's IAT. We find that math teachers with stronger implicit stereotypes against immigrants give them lower grades, holding constant students' ability (as measured by the standardized test score). A one standard deviation increase in teacher's IAT score (0.26, equivalent to moving from an unbiased teacher to a slightly biased one) is associated with a 0.033 decrease in immigrants' grades in math on average, or half of the average gap between immigrants and natives with identical performance on standardized tests. Therefore, negative (implicit) attitudes towards immigrants systematically correlate with the way in which math teachers evaluate student performance. These biases in teachers' minds may operate in an unconscious way, without any intention to directly harm the stigmatized group (Gilliam et al., 2016; Van den Bergh et al., 2010).

The remaining columns of Table 4, Panel A, show that the relationship between teachers' IAT scores and immigrants' grades in math is robust to including additional covariates. In column 4, we control for student characteristics (gender, first vs. second generation immigrant, and mother's education), and in column 5 we add teacher characteristics (age, gender, place of birth, and whether the teacher holds an advanced STEM degree). We also interact all these variables with the immigrant dummy to make sure that our main explanatory variable of interest does not capture the effect of other teacher characteristics correlated with the IAT score. Note that after including all interactions of the immigrant dummy with student and teacher characteristics (i.e., in columns 4 to 7 of Table 4), we can no longer interpret the coefficient of the standalone immigrant dummy.[13] We find that the estimated effect of the IAT score on the grading of immigrant students remains unaffected. In column 6, we add the grade received by students at the end middle school for their behavior in the classroom. This grade is decided jointly by all teachers of the class, with a stronger weight given to the math and literature teachers. Although the variable *Behavior* may be endogenous, so we cannot interpret the resulting coefficients in a causal way, it is reassuring that the inclusion of the variable does not affect the coefficient on *Immigrant*Teacher's IAT*, suggesting that our results are not driven by worse behavior in the classroom of immigrants compared to natives. Our coefficient of interest is also unchanged by replacing teacher fixed effects with class fixed effects (last column of the table).

The evidence in Panel A of Table 4 confirms that math teachers with stronger implicit stereotypes, as measured by higher race IAT scores, give lower grades on average to immigrant students than to native students with the same performance on blindly-graded, standardized tests. Panel B of Table 4 shows that this is not true for literature teachers, as there is no relationship between IAT scores and immigrants' grades in literature (Columns 3-7). We propose two non-mutually exclu-

[13]For this reason, at the bottom of each panel we report the mean of the native-immigrant gap in grades when including all control variables but not their interaction with the immigrant dummy.

sive explanations for the different results for math and literature teachers. First, multiple choice, standardized tests in literature may be ill-suited to measure skills that are considered important by literature teachers (e.g., language proficiency), while the problem may be less severe in math (Bettinger, 2012). If this is the case, misalignment between the skills measured by standardized tests and by teacher evaluations may hamper the identification of subjective biases in the latter. Second, taking into account the additional difficulties faced by non-native speakers in their subject, literature teachers may impose lower standards on immigrant than on native students. Indeed, literature teachers with stronger stereotypes may expect less from non-native speakers and this effect may offset any negative impact of (implicit or explicit) biases on grading of immigrant students. Appendix Table A.6 provides supporting evidence for the latter explanation using the generation of immigration of children. First generation immigrants are less likely to be proficient in the native language than second generation immigrants. In line with our proposed explanation, we find that literature teachers with stronger stereotypes give relatively higher grades to first generation than to second generation immigrants (coefficients on *First Generation*Teacher's IAT* in columns 8-10), although statistical power is an issue when detecting effects within specific subgroups. By contrast, the effect of teacher stereotypes on grades in math, for which language proficiency is less of an issue, does not vary between first and second generation immigrants (coefficients on the same variable in columns 3-5).

4.2 Revealing implicit stereotypes

Figure 4 shows the effect of our treatment on the grades given by math (top panel) and literature (bottom panel) teachers, to immigrant (left) and native (right) students in January 2017. The sample includes 8th graders during the school year 2016-2017. Teachers are not aware that we observe the grades they give to their students, substantially reducing the risk of experimenter demand effects. The grade distribution, on the horizontal axis, goes from 3 to 10. Colored bars represent the grade distribution of students whose teachers received the feedback before the end-of-semester grading, while black lines represent the grades of students whose teachers received the feedback after. The figure shows that receiving the feedback on the IAT before grading shifts the grade distribution to the right for immigrant students and to the left for native ones.

[Insert Figure 4]

Table 5 quantifies the average effects in Figure 4. The variable '*Early Feedback*' in Panel A is an indicator for whether the school was randomized into the treated group. We are thus comparing grading by teachers eligible for receiving the feedback in time to adjust their grades (i.e., before

end of January) and teachers who could receive the feedback only after end-of-semester grading. The coefficient in the first row of Panel A is therefore an intention-to-treat effect. Math teachers eligible for treatment (i.e., '*Early Feedback*' taking value 1) give on average 0.4 points more to immigrants and 0.15 points less to natives compared to teachers randomized into the control group (columns 1-3). The effect on grading of immigrant students in literature is qualitatively similar, but only about 3/4 as large in magnitude (columns 4-6).

[Insert Table 5]

In Panel B of Table 5 we rescale the intention-to-treat effect by the take-up rate of the early feedback, which was above 80 percent, in order to compute the treatment effect of bias revelation. The variable '*Email*' in Table 5 takes value 1 if the teacher actually received the feedback and 0 if he/she did not receive any feedback. The effect on the interaction between the treatment and immigrant students' grades increases in magnitude to about $+0.5$ for math and $+0.4$ for literature, while the effect for natives is around -0.2 points for both subjects.

It should be stressed that the magnitude of the treatment effect in Table 5 is not comparable to the magnitude of the bias in grading (i.e., the difference between teacher grades and standardized test scores) shown in Table 4. The experiment was done at the end of the first semester (when no standardized test scores exist), while the bias in grading is measured at the end of the second semester (when we have information on both standardized test scores and teacher-assigned grades). Also, the grading policy of teachers likely differs between the first and the second semester, especially around the pass grade. In both semesters, students fail if they obtain a grade lower than 6, but failing has very different consequences in the first and second semester. Failing in the first semester represents a 'warning' with no immediate consequences, while students may be retained if they fail more than one subject in the second semester. For this reason, teachers may be more hesitant to fail students in the second than in the first semester. Indeed, the average fraction of students failing either literature or mathematics (or both) is 21 percent in the first semester, but only 2 percent in the second semester. Among immigrant students, failure rates in the first and second semester reach 31 and 4 percent, respectively. Lower propensity to fail students in the second semester sets a floor to teacher grades – and, possibly, to penalties for immigrant students – compared to the first semester. For this reason, stereotypes likely induce larger effects on relative grades of immigrant and native students in the first than in the second semester.

[Insert Table 6]

Visual evidence in Figure 4 suggests that treatment effects may be particularly large around the margin between passing and failing students (i.e., between grades 6 and 5), especially for math

teachers. This is confirmed in Table 6, in which we estimate the effect of the early feedback on the probability of failing immigrant and native students, respectively. For math teachers (columns 1-3), early IAT feedback decreases the probability of failing immigrants by around 10 percentage points, whereas failing rates of native students remain unaffected. There is no significant effect for literature teachers.

Overall, we document large effects of revealing own stereotypes to math teachers on both grades and failure rates of immigrant students in math. As for literature teachers, the effect on grading is smaller, and there is no effect on the probability of failing immigrant students in math. These findings dovetail nicely with the evidence in Table 4 that teacher stereotypes are less relevant for grading in literature than for grading in math – with the caveat, discussed above, that standardized test scores may be a noisy measure of student ability in literature.

4.3 Heterogeneous effects

In Table 7, we explore how changes in teacher behavior induced by our intervention vary with teacher and student characteristics. For comparison, columns 1 and 5 report the baseline estimated effects in math and literature, respectively, according to the most stringent specifications in columns 3 and 6 of Table 5.

[Insert Table 7]

We first examine heterogeneity in the (intention-to-treat) effect by the initial level of teacher explicit bias against immigrants. If being revealed one's own (implicit) bias is more informative for teachers unaware of such bias, we should expect a greater reaction from teachers who reported no explicit bias in the initial survey. To test this hypothesis, in columns 2 and 6 of Table 7 we interact the indicator variables for immigrant students and teachers' early feedback with the dummy variable *WVS*, which is equal to 1 for teachers who in our survey agree that "immigrants and natives should have equal opportunities of accessing available jobs". Not only are these teachers slightly less biased to begin with (see the previous Table 1), they are also more responsive to our intervention, as demonstrated by the positive coefficient of the triple interaction term in columns 2 and 6. This suggests that teachers actually react to being revealed a bias they were unaware of.

In the remaining columns of Table 7, we turn to heterogeneity by student characteristics. In columns 3 and 7 we explore the role of the generation of immigration. Revealing stereotypes to teachers does not lead them to differentially adjust the grades they give to first or second generation immigrants. In column 4 and 8, we estimate the differential effect by area of origin of immigrant

students: Eastern Europe, Africa, Latin America, and Asia. Math teachers receiving the early feedback increase grades more for students who are geographically or linguistically closer to natives – respectively, Eastern Europeans (the excluded category) and Latin Americans. Math teachers do not respond to the treatment when they grade African or Asian students. The pattern is less clear for literature teachers. The heterogeneous effects by region of origin observed for math teachers are suggestive of a role of cultural or linguistic proximity as mediators in the impact of our treatment.

Finally, we also explored heterogeneity of the effect with the timing of when teachers took the survey or were informed of the IAT: before/after Christmas, and number of days before/after end-of-semester grading. However, we could not detect any significant differences along these dimensions.

5 Conclusions

Immigrant children receive lower teacher-assigned grades than natives after controlling for their performance on anonymously graded standardized tests. This result per se is not a proof of bias, since there may be characteristics that differentiate immigrants from natives that are observable to teachers, but unobservable to the econometrician (e.g., disciplinary problems or differences in performance on standardized multiple choice questions versus open ended ones). However, by administering an Implicit Association Test (IAT) to teachers, we are able to show that the difference in grading of natives and immigrants is correlated with teachers' stereotypes against immigrants for math, potentially the most objective among teacher-assigned grades. The effect is not statistically significant for literature teachers, although this masks interesting heterogeneous effects, with second (first) generation immigrants receiving lower (higher) grades from teachers with more stereotypes. Literature teachers with stronger stereotypes may expect less from children who are less familiar the language and, thus, they may be positively surprised when such students perform well.

We also show that informing the teachers about their own stereotypes may be an effective policy to reduce discrimination in grading. Both math and literature teachers randomized into a treatment that consists of receiving feedback on their IAT before end-of-term grading react to this information by increasing the grades they give to immigrants and decreasing the grades they give to natives. Only teachers with positive explicit views toward immigrants react to the treatment.

Our results speak to a new policy debate, particularly to recent efforts by corporations and institutions in the U.S. and Canada to increase awareness about implicit bias by encouraging every search committee member or new employee to take an IAT. In the context of schooling, the IAT

test is simple to implement and it would not cost much to ask every teacher to take it, say at the beginning of every academic year. Our results suggest that this may help counteract implicit bias that teachers hold towards negatively stereotyped groups. However, we also want to stress that the implications of such a policy are not straightforward. By making teachers aware of their 'implicit' biases, their evaluation of students becomes more fair if they were acting upon their stereotypes by giving lower grades to immigrants (i.e. math teachers). But it is possible that teachers whose negative stereotypes do not translate into discriminatory behavior may also react, thus inducing positive discrimination toward immigrant children. Further research is needed to assess efficient and equitable ways to de-bias individuals.

References

Alesina, A. and Glaeser, E. (2004). *Fighting poverty in the US and Europe: A world of difference*. Oxford University Press.

Alesina, A. and La Ferrara, E. (2014). A test of racial bias in capital sentencing. *The American Economic Review*, 104(11):3397–3433.

Alesina, A., Miano, A., and Stantcheva, S. (2018). Immigration and redistribution. *NBER Working Paper No. 24733*.

Allport, G. W. (1958). *The nature of prejudice: Abridged*. Doubleday.

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. *Handbook of labor economics*, 3:3143–3259.

Barbieri, G., Rossetti, C., and Sestito, P. (2011). The determinants of teacher mobility: Evidence using Italian teachers' transfer applications. *Economics of Education Review*, 30(6):1430–1444.

Becker, G. S. et al. (1957). *Economics of Discrimination*. University of Chicago Press.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. *Handbook of Economic Field Experiments*, pages Pages 309–393.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? a field experiment on labor market discrimination. *The American Economic Review*, 94(4):991–1013.

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.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., and Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94(3):567.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*.

Bordalo, P., Coffman, K. B., Gennaioli, N., and Shleifer, A. (2018). Beliefs about gender. *American Economic Review*.

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.

Burns, J., Corno, L., and La Ferrara, E. (2016). Interaction, stereotypes and performance. Evidence from South Africa. *Working Paper*.

Carlana, M. (2018). Implicit stereotypes: Evidence from teachers' gender bias. *Working Paper*.

Carlana, M., La Ferrara, E., and Pinotti, P. (2018). Goals and gaps: Educational careers of immigrant children. *Mimeo Bocconi Univeristy*.

Coviello, D. and Persico, N. (2015). An economic analysis of Black-White disparities in the New York Police Department's stop-and-frisk program. *The Journal of Legal Studies*.

De Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *The American Economic Review*.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *The American Economic Review*.

Donders, F. (1868). *On the speed of mental processes. Translation by WG Kostor in Attention and performance II, ed. WG Koster*. North Holland.

Facchini, G., Margalit, Y., and Nakata, H. (2016). Countering public opposition to immigration: The impact of information campaigns. *Working Paper*.

Figlio, D. (2005). Names, expectations and the Black-White test score gap. *NBER Working Paper*.

Figlio, D., Giuliano, P., Özek, U., and Sapienza, P. (2018). Long-term orientation and educational performance. *Working Paper*.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public economics*, 82(2):225–246.

Fryer Jr, R. G. (2016). An empirical analysis of racial differences in police use of force. *NBER Working Paper*.

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., and Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions. *Research Study Brief. Yale University, Yale Child Study Center, New Haven, CT*.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics*.

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of personality and social psychology*, 85(2):197.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, 97(1):17.

Grigorieff, A., Roth, C., and Ubfal, D. (2017). Does information change attitudes towards immigrants? representative evidence from survey experiments. *Working Paper*.

Guryan, J. and Charles, K. K. (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal*, 123(572):F417–F432.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–68.

Hopkins, D. J., Sides, J., and Citrin, J. (2018). The muted consequences of correct information about immigration. *Working Paper*.

Howell, J. L., Gaither, S. E., and Ratliff, K. A. (2015). Caught in the middle: Defensive responses to IAT feedback among whites, blacks, and biracial black/whites. *Social Psychological and Personality Science*, 6(4):373–381.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Jussim, L. and Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and social psychology review*, 9(2):131–155.

Knowles, J., Persico, N., and Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*.

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

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

Lowes, S., Nunn, N., Robinson, J. A., and Weigel, J. (2015). Understanding ethnic identity in Africa: Evidence from the Implicit Association Test (IAT). *The American Economic Review*, 105(5):340–45.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1):36–88.

O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., and Blodorn, A. (2010). But I'm no bigot: How prejudiced White Americans maintain unprejudiced self-images. *Journal of Applied Social Psychology*, 40(4):917–946.

Olson, M. A. and Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: personalizing the IAT. *Journal of Personality and Social Psychology*, 86(5):653.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., and Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2):171.

Paluck, E. L., Green, S. A., and Green, D. P. (2018). The contact hypothesis re-evaluated. *Behavioural Public Policy*, (1–30).

Papageorge, N. W., Gershenson, S., and Kang, K. (2018). Teacher expectations matter. *NBER Working Paper No. 25255*.

Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408.

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3):523–534.

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

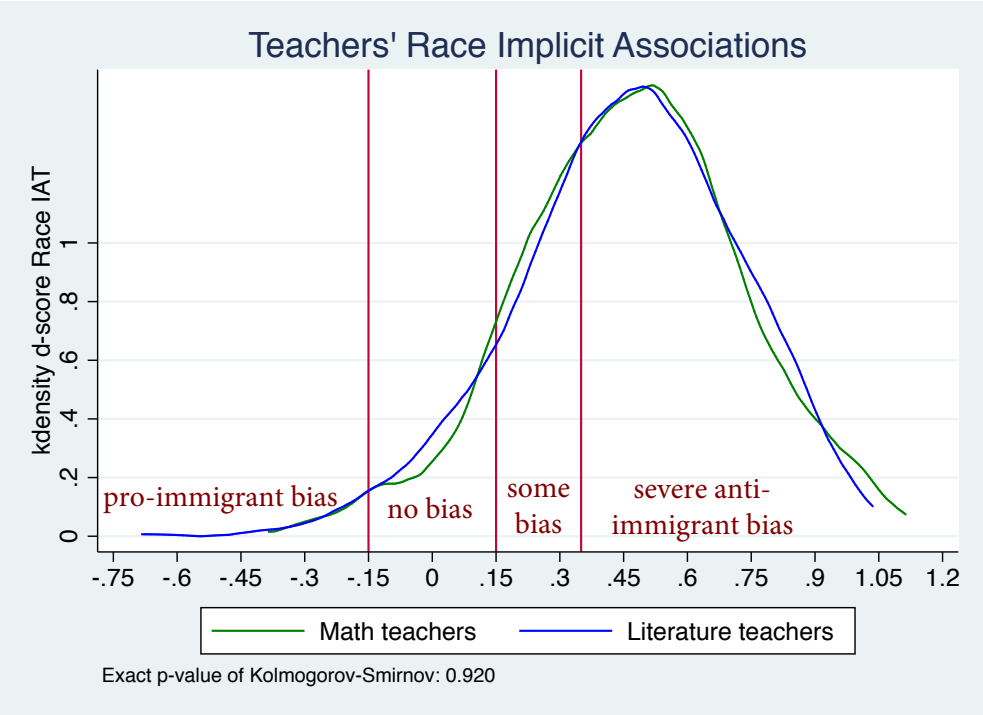
Sukhera, J., Milne, A., Teunissen, P. W., Lingard, L., and Watling, C. (2018). The actual versus idealized self: Exploring responses to feedback about implicit bias in health professionals. *Academic Medicine*, 93(4):623–629.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., and Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2):497–527.

Van Ewijk, R. (2011). Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review*, 30(5):1045–1058.

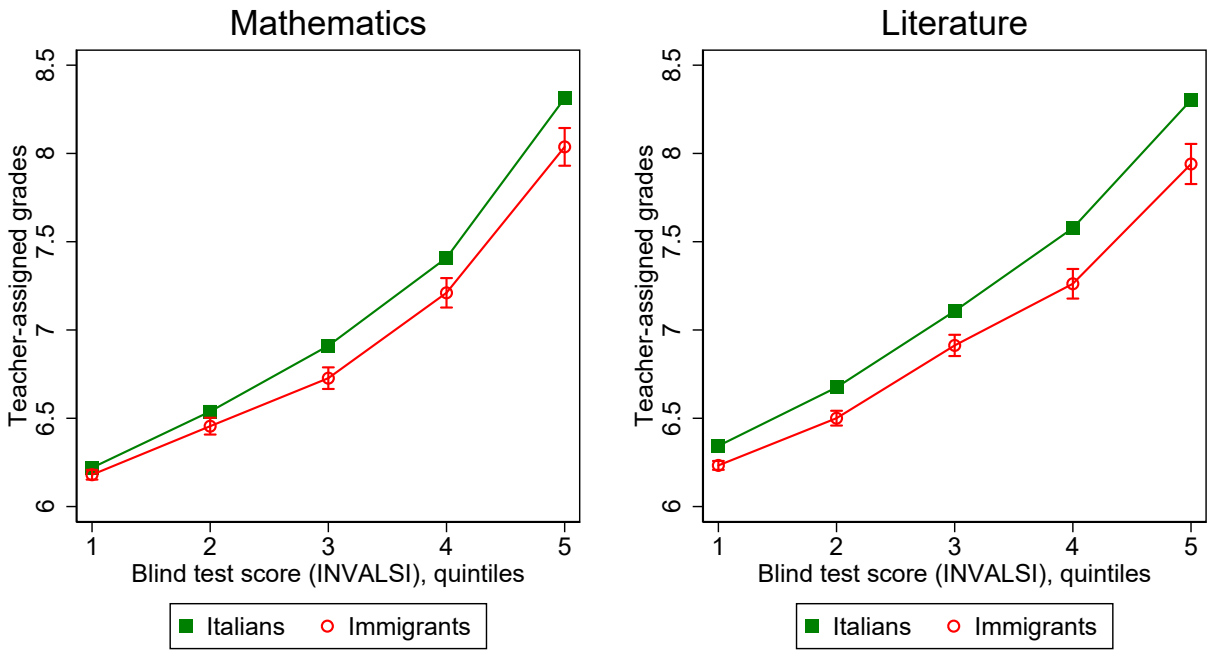
Tables and Figures

Figure 1: Distribution of the race IAT score across teachers



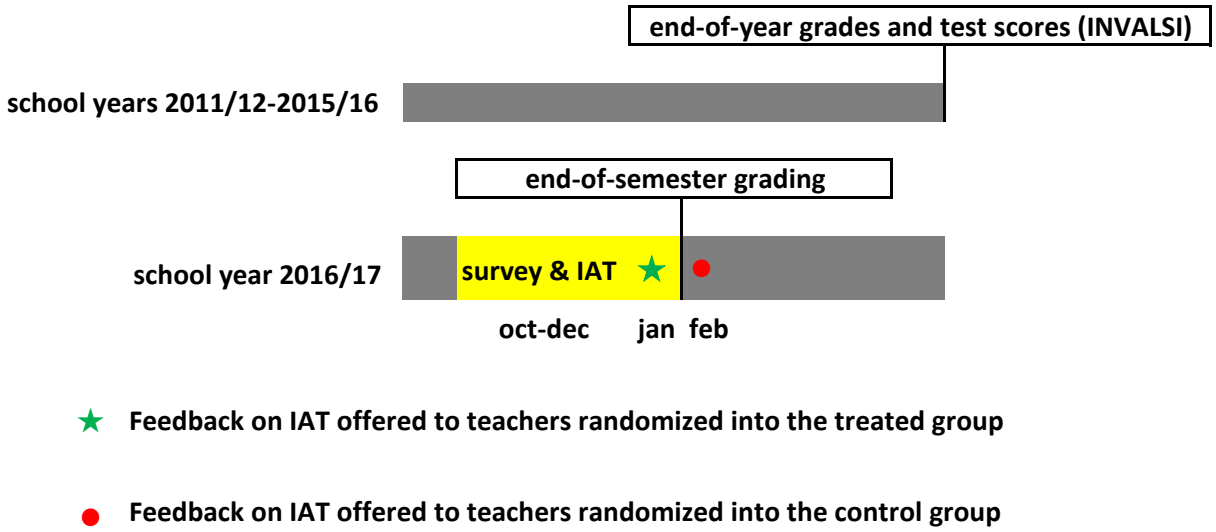
Notes: This graph shows the distribution of raw IAT scores for mathematics and for literature teachers. A positive value indicates a stronger association between "natives"-"good" and "immigrant"-"bad". The vertical lines indicate the critical thresholds suggested by Greenwald et al. (2009) for defining different levels of bias, also indicated in the graph.

Figure 2: Teacher-assigned grades vs. blindly-graded, standardized test scores



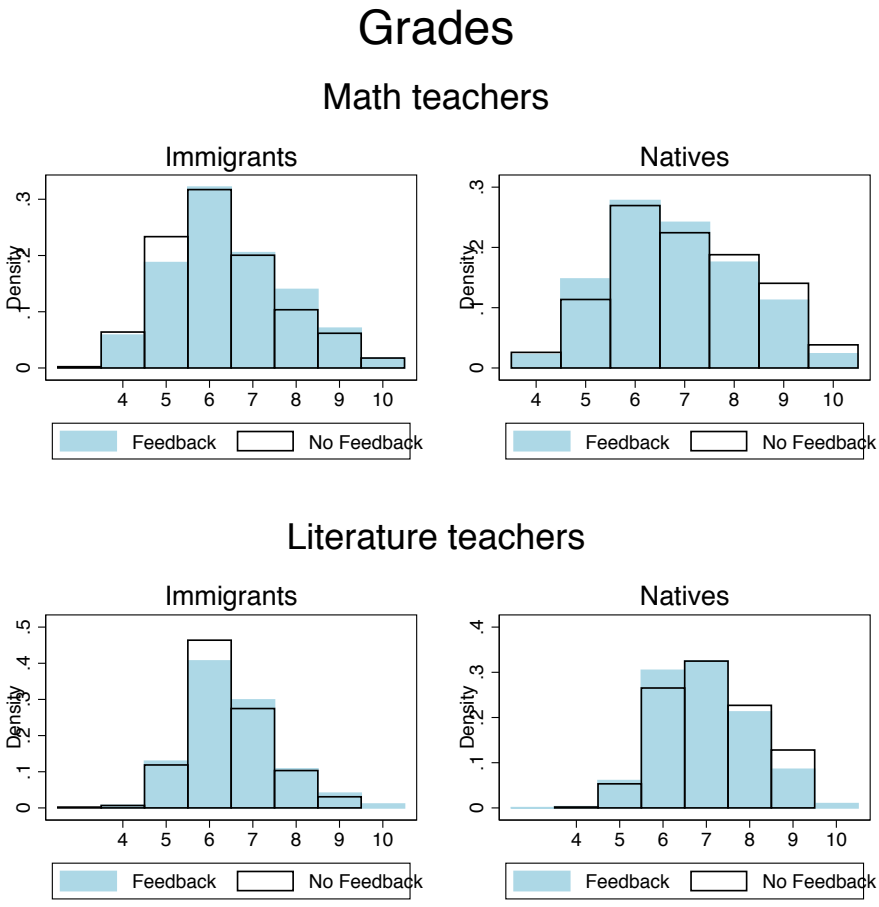
Notes: This graph shows teacher-assigned grades (non-blindly graded) on the vertical axis and quintiles of the standardized test score INVALSI (blindly-graded) on the horizontal axis at the end of grade 8. Teacher-assigned grades are on a scale from 3 to 10, with 6 as the pass grade. The green squares and lines are for native students, while the red circles and lines are for immigrant students. The left panel presents evidence from grades in mathematics, while the right panel presents grades in literature. Students in this sample completed grade 8 between school years 2011-2012 and 2015-2016.

Figure 3: Timeline



Notes: This figure shows the timeline of the data collection, survey, and experiment. As described at length in Section 3, we obtained administrative data on end-of-year teacher-assigned grades as well as on standardized, blindly graded test scores for school years 2012/13 through 2015/16. During the first semester of the 2016/17 school year (October-January), we administered the survey and the IAT to all teachers in our sample. On January 2017, before end of semester grading, we sent feedback about own IAT score to a random group of teachers. All other teachers were allowed to see their score after the end of semester grading (i.e., February 2018).

Figure 4: The impact of revealing bias to teachers on grading of immigrant and native students



Notes: This graph shows the distribution of grades given to native and immigrant children by math and literature teachers eligible (light blue bars) and non-eligible (striped bars) for receiving feedback about own IAT score before end of semester grading.

Table 1: Correlation between teachers characteristics and IAT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.: IAT score (stereotypes against immigrants)									
Female	-0.042**							-0.039	-0.046
	(0.020)							(0.027)	(0.032)
Born in the North		-0.026*						-0.056***	-0.042**
		(0.014)						(0.017)	(0.020)
Children			0.021					-0.014	-0.009
			(0.050)					(0.020)	(0.025)
Middle edu Mother				0.027				0.027	0.031
				(0.017)				(0.022)	(0.026)
High edu Mother				-0.022				-0.022	-0.008
				(0.021)				(0.028)	(0.035)
WVS Immigrants' Rights to Job					-0.058***			-0.045*	-0.043
					(0.017)			(0.025)	(0.031)
Native-Imm INVALSI(/100)						-0.040		0.010	0.025
						(0.085)		(0.092)	(0.120)
SD Native- SD Imm INVALSI(/100)							0.002	0.002	0.002
							(0.002)	(0.002)	(0.002)
Obs.	1384	1384	1384	1384	1384	779	779	779	779
R^2	0.063	0.063	0.060	0.065	0.066	0.092	0.094	0.123	0.203

Notes: This table reports OLS estimates, where the dependent variable is IAT score of teachers and the unit of observation is teacher *t* in grade 8 of school *s*. We include controls for the order of IATs and for whether the blocks were presented on a order compatible or incompatible way (which was randomized at the individual level). The variable "WVS Immigrants' Rights to Job" equals 1 for teachers believing that immigrants should have the same right to jobs as natives. "Native-Imm INVALSI(/100)" indicates the difference in average standardized test scores of native and immigrant students assigned to the teacher in the previous four years; "SD Native- SD Imm INVALSI(/100)" is the difference in standard deviations. In columns 6-9, the number of observations decreases because information on past students is not available for all teachers; in these columns, we control for the number of observations with information available for immigrants and native children. *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

Table 2: Balance table - Teacher characteristics

Variable	(1) Control	(2) Treatment	(3) Diff.	(4) Normalized difference
Panel A: Math teachers				
IAT Immigrants	0.481	0.455	-0.026	-0.070
	(0.278)	(0.253)	(0.032)	
Female	0.782	0.888	0.107	0.204
	(0.415)	(0.316)	(0.049)**	
Born in the North	0.641	0.579	-0.062	-0.090
	(0.482)	(0.496)	(0.070)	
Age	48.641	49.064	0.423	0.031
	(9.795)	(9.667)	(1.599)	
Years of experience	18.310	19.681	1.371	0.081
	(11.923)	(12.128)	(2.090)	
Children	0.664	0.720	0.056	0.086
	(0.474)	(0.450)	(0.061)	
Low edu Mother	0.541	0.466	-0.075	-0.106
	(0.501)	(0.501)	(0.063)	
Middle edu Mother	0.306	0.366	0.060	0.090
	(0.463)	(0.484)	(0.058)	
High edu Mother	0.153	0.168	0.015	0.028
	(0.362)	(0.375)	(0.047)	
Degree *cum Laude*	0.200	0.266	0.066	0.110
	(0.402)	(0.443)	(0.060)	
Advanced STEM	0.248	0.191	-0.056	-0.096
	(0.434)	(0.395)	(0.055)	
Observations	119	143	262	
Panel B: Literature teachers				
IAT Immigrants	0.493	0.472	-0.021	-0.057
	(0.275)	(0.254)	(0.028)	
Female	0.915	0.903	-0.012	-0.029
	(0.281)	(0.297)	(0.038)	
Born in the North	0.737	0.779	0.042	0.070
	(0.442)	(0.416)	(0.059)	
Age	49.737	48.979	-0.758	-0.059
	(8.720)	(9.305)	(1.359)	
Years of experience	22.035	20.755	-1.280	-0.082
	(11.146)	(10.793)	(1.570)	
Children	0.718	0.695	-0.023	-0.036
	(0.452)	(0.462)	(0.060)	
Low edu Mother	0.486	0.504	0.018	0.025
	(0.502)	(0.502)	(0.072)	
Middle edu Mother	0.355	0.321	-0.035	-0.051
	(0.481)	(0.469)	(0.068)	
High edu Mother	0.159	0.176	0.017	0.032
	(0.367)	(0.382)	(0.063)	
Degree *cum Laude*	0.340	0.323	-0.017	-0.025
	(0.476)	(0.469)	(0.060)	
Observations	117	154	271	

Notes: Standard errors are robust and clustered at the school level. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table 3: Balance table - Students' characteristics

(1)		(2)	(3)	(4)	
Variable		**Control**	**Treatment**	**Diff.**	**Normalized Diff.**
Female		0.503	0.490	-0.012	-0.018
		(0.500)	(0.500)	(0.014)	
First Gen Imm		0.079	0.088	0.008	0.021
		(0.271)	(0.283)	(0.015)	
Born before 2003		0.051	0.064	0.013	0.039
		(0.220)	(0.244)	(0.012)	
Grade Ita June '16		7.145	7.116	-0.028	-0.019
		(1.052)	(1.046)	(0.069)	
Grade Math June '16		7.199	7.135	-0.065	-0.037
		(1.250)	(1.227)	(0.074)	
Grade Ita June '15		7.232	7.180	-0.053	-0.035
		(1.054)	(1.054)	(0.064)	
Grade Math June '15		7.376	7.309	-0.067	-0.037
		(1.288)	(1.287)	(0.067)	
Mother	Less than high school	0.206	0.254	0.048	0.080
		(0.405)	(0.435)	(0.038)	
	High education	0.165	0.213	0.048	0.087
		(0.371)	(0.410)	(0.042)	
	Low Occupation	0.143	0.174	0.031	0.060
		(0.350)	(0.379)	(0.022)	
	Mid Occupation	0.340	0.353	0.014	0.020
		(0.474)	(0.478)	(0.035)	
	High Occupation	0.100	0.137	0.037	0.082
		(0.300)	(0.344)	(0.033)	
Father	Less than high school	0.257	0.302	0.045	0.071
		(0.437)	(0.459)	(0.044)	
	Low education	0.152	0.178	0.026	0.050
		(0.359)	(0.383)	(0.043)	
	Low Occupation	0.245	0.271	0.026	0.043
		(0.430)	(0.444)	(0.037)	
	Mid Occupation	0.340	0.359	0.019	0.028
		(0.474)	(0.480)	(0.036)	
	High Occupation	0.177	0.217	0.040	0.071
		(0.382)	(0.412)	(0.050)	
Observations		2,756	3,275	6,031	

Notes: Standard errors are robust and clustered at the school level. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table 4: Teachers' IAT scores and grades assigned to immigrant students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. **Dependent Variable: Grades given by Math Teacher**							
Immigrant	-0.439***	-0.091***	-0.032	0.049	1.151*	1.582**	0.928
	(0.025)	(0.018)	(0.038)	(0.046)	(0.637)	(0.619)	(0.626)
Imm*Teacher's IAT			-0.033*	-0.030*	-0.032*	-0.030*	-0.030*
			(0.019)	(0.018)	(0.018)	(0.017)	(0.018)
Behavior						0.485***	
						(0.012)	
Imm* Behavior						-0.101***	
						(0.017)	
Mean of Native-Imm Gap	-0.439	-0.091	-0.091	-0.075	-0.075	-0.076	-0.065
Obs.	21846	21846	21846	21846	21846	21353	21846
R^2	0.093	0.475	0.475	0.505	0.505	0.592	0.540
Panel B. **Dependent Variable: Grades given by Literature Teacher**							
Immigrant	-0.574***	-0.110***	-0.108***	-0.082*	0.334	1.141*	0.316
	(0.022)	(0.017)	(0.036)	(0.043)	(0.724)	(0.627)	(0.716)
Imm*Teacher's IAT			-0.001	-0.000	0.002	0.012	0.009
			(0.017)	(0.017)	(0.017)	(0.015)	(0.016)
Behavior						0.437***	
						(0.010)	
Imm*Behavior						-0.086***	
						(0.015)	
Mean of Native-Imm Gap	-0.574	-0.110	-0.110	-0.103	-0.103	-0.094	-0.111
Obs.	20457	20457	20457	20457	20457	20097	20457
R^2	0.149	0.524	0.524	0.558	0.558	0.643	0.595
Teacher FE	Yes	Yes	Yes	Yes	Yes	Yes	No
Class FE	No	No	No	No	No	No	Yes
INVALSI cubic	No	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes	Yes	Yes
Student Controls*Imm	No	No	No	Yes	Yes	Yes	Yes
Teacher Controls *Imm	No	No	No	No	Yes	Yes	Yes

Notes: This table reports OLS estimates, where the dependent variable is grade given by teachers in math (Panel A) and literature (Panel B) in grade 8, and the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. "Teacher's IAT" is the raw score of IAT divided by the standard deviation. We include a cohort dummy in all regressions and "INVALSI cubic" indicates the cubic polynomial of INVALSI test score in grade 8. Student controls include gender, generation of immigration, and mother's education. Teacher controls include gender, place of birth, age, and advanced STEM degree (i.e., physics, math, engineering). "Behavior" is a grade between 5 and 10 assigned jointly by all teachers to measure behavior in the classroom. "Mean of Native-Imm Gap" indicates the average difference in grades between immigrants and natives when including fixed effects and all control variables, but not their interaction with the "Immigrant" dummy. *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

Table 5: Impact of revealing stereotypes to teachers on grades

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var:	**Grades given by Math Teacher**			**Grades given by Literature Teacher**		
Panel A: Intention to Treat						
Early Feedback*Imm	0.392***	0.437***	0.439***	0.312***	0.302***	0.288***
	(0.142)	(0.129)	(0.126)	(0.103)	(0.083)	(0.087)
Early Feedback	-0.153	-0.176*	-0.155	-0.150*	-0.160**	-0.147*
	(0.100)	(0.094)	(0.095)	(0.084)	(0.072)	(0.075)
Immigrant	-0.713***	-0.687***	0.915	-0.697***	-0.685***	-0.107
	(0.091)	(0.222)	(1.274)	(0.055)	(0.131)	(1.361)
Obs.	5141	5141	5141	5138	5138	5138
R^2	0.023	0.108	0.118	0.037	0.167	0.174
Panel B: Local Average Treatment Effect						
Email*Imm	0.501***	0.552***	0.554***	0.403***	0.392***	0.366***
	(0.171)	(0.161)	(0.156)	(0.132)	(0.112)	(0.114)
Email	-0.206	-0.234*	-0.208	-0.202*	-0.214**	-0.194*
	(0.131)	(0.127)	(0.128)	(0.111)	(0.099)	(0.101)
Immigrant	-0.713***	-0.659***	0.998	-0.697***	-0.624***	-0.306
	(0.090)	(0.226)	(1.252)	(0.054)	(0.142)	(1.408)
Obs.	5141	5141	5141	5138	5138	5138
R^2	0.022	0.105	0.116	0.035	0.161	0.169
F- stat	84.5	94.2	101.1	106.8	125.9	138.8
Mean dep. var.	6.83	6.83	6.83	6.95	6.95	6.95
Student Controls	No	Yes	Yes	No	Yes	Yes
Student Controls*Imm	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Teacher Controls*Imm	No	No	Yes	No	No	Yes

Notes: This table reports OLS estimates (Panel A) and IV estimates (Panel B), where the dependent variable is grade given by teachers in math (columns 1-3) and literature (columns 4-6) at the end of the first semester of grade 8 (January); the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at the school level. "Early Feedback" is a dummy variable indicating whether the teacher was eligible for receiving the feedback before end of semester grading (January) or after end of semester grading (February). "Email" is a dummy variable indicating whether teachers eligible for receiving the feedback before end of semester grading actually requested it. The coefficients in Panel B are estimated by instrumental variables, using "Early Feedback" as an instrument for "Email". Student controls include gender, generation of immigration, and education of the mother, interacted with whether the student is an immigrant. Teacher controls include gender, place of birth, advanced STEM degree (as physics, math, engineering), and age, interacted with whether the student is an immigrant. *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

Table 6: Impact of revealing stereotypes to teachers on the probability of failing students

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var:	**Math Fail (Grade <6)**			**Literature Fail (Grade <6)**		
Panel A: Intention to Treat						
Early Feedback*Imm	-0.094**	-0.103**	-0.107***	-0.012	-0.014	-0.009
	(0.044)	(0.041)	(0.038)	(0.023)	(0.020)	(0.021)
Early Feedback	0.026	0.029	0.022	0.000	-0.002	-0.009
	(0.028)	(0.025)	(0.025)	(0.015)	(0.012)	(0.012)
Immigrant	0.161***	0.122*	-0.217	0.075***	0.056	-0.366
	(0.032)	(0.072)	(0.409)	(0.014)	(0.034)	(0.343)
Obs.	5141	5141	5141	5138	5138	5138
R^2	0.015	0.059	0.067	0.011	0.038	0.045
Panel B: Local Average Treatment Effect						
Email*Imm	-0.119**	-0.129***	-0.133***	0.001	-0.003	-0.011
	(0.052)	(0.049)	(0.046)	(0.019)	(0.017)	(0.016)
Email	0.035	0.038	0.030	-0.015	-0.017	-0.010
	(0.037)	(0.033)	(0.033)	(0.028)	(0.024)	(0.025)
Immigrant	0.161***	0.117*	-0.226	0.075***	0.053	-0.369
	(0.032)	(0.071)	(0.400)	(0.014)	(0.034)	(0.343)
Obs.	5141	5141	5141	5138	5138	5138
R^2	0.017	0.059	0.068	0.011	0.038	0.045
Mean dep. var.	0.173	0.173	0.173	0.067	0.067	0.067
Student Controls	No	Yes	Yes	No	Yes	Yes
Student Controls*Imm	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Teacher Controls*Imm	No	No	Yes	No	No	Yes

Notes: This table reports OLS estimates (Panel A) and IV estimates (Panel B), where the dependent variable is the probability of obtaining a grade lower than 6 (pass grade) in math (columns 1-3) and literature (columns 4-6) at the end of the first semester of grade 8 (January), and the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at the school level. "Early Feedback" is a dummy variable indicating whether the teacher was eligible for receiving the feedback before end of semester grading (January) or after end of semester grading (February). "Email" is a dummy variable indicating whether teachers eligible for receiving the feedback before end of semester grading actually requested it. The coefficients in Panel B are estimated by instrumental variables, using "Early Feedback" as an instrument for "Email". Student controls include gender, generation of immigration, and education of the mother, all interacted with whether the student is an immigrant. Teacher controls include gender, place of birth, advanced STEM degree (as physics, math, engineering), and age, interacted with whether the student is an immigrant. *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

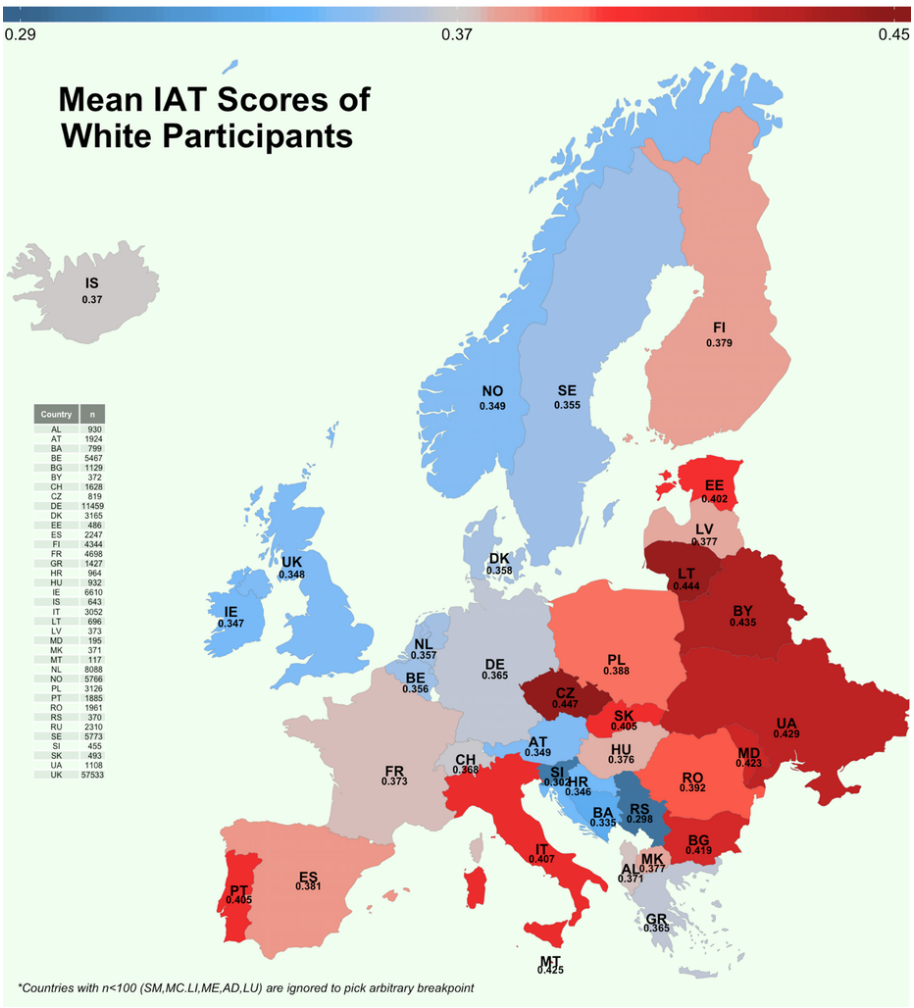
Table 7: Heterogeneous effects of revealing bias to teachers on grades, by teacher and student characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	**Grades given by Math Teacher**				**Grades given by Literature Teacher**			
Early Feedback*Imm	0.439***	0.051	0.415***	0.615***	0.288***	-0.200	0.245**	0.440***
	(0.126)	(0.236)	(0.141)	(0.136)	(0.087)	(0.185)	(0.106)	(0.140)
Early Feedback	-0.155	-0.131	-0.155	-0.155	-0.147*	0.080	-0.147*	-0.147*
	(0.095)	(0.180)	(0.095)	(0.095)	(0.075)	(0.159)	(0.075)	(0.075)
Early Feedback*Imm*WVS		0.545*				0.663***		
		(0.293)				(0.187)		
Early Feedback*WVS		-0.138				-0.288*		
		(0.225)				(0.168)		
Imm*WVS		-0.090				-0.135		
		(0.187)				(0.144)		
Early Feedback*First Gen Imm			0.052				0.091	
			(0.153)				(0.138)	
Early Feedback*Africa				-0.473*				-0.189
				(0.269)				(0.228)
Early Feedback*Latin				-0.105				-0.257
				(0.250)				(0.189)
Early Feedback*Asia				-0.441*				-0.260
				(0.251)				(0.220)
Obs.	5141	5141	5141	5141	5138	5138	5138	5138
R^2	0.118	0.120	0.118	0.123	0.174	0.179	0.174	0.176
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls*Imm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls*Imm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports OLS estimates, where the dependent variable is grade given by teachers in math (columns 1-4) and literature (columns 5-8) at the end of the I semester of grade 8 (January), and the unit of observation is student *i*, in class *c*, taught by teacher *t* in grade 8 of school *s*. The regressions are fully saturated: all interactions are included even if not reported in the table. "Early Feedback" is a dummy variable indicating whether the teacher was eligible for receiving the feedback before end of semester grading (January) or after end of semester grading (February)."WVS Immigrants' Rights to Job" equals 1 for teachers believing that immigrants should have the same right to jobs as natives. "First Gen Imm" assumes value 1 for first generation immigrants. "Africa", "Latin", and "Asia" indicates the place of origin of immigrants, where the excluded category is Eastern Europe. Standard errors are robust and clustered at the school level. Student controls include gender, generation of immigration, and mother's education, all interacted with whether the student is immigrant. Teacher controls include gender, place of birth, advanced STEM degree (physics, math, engineering), age, interacted with whether the student is animmigrant. *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

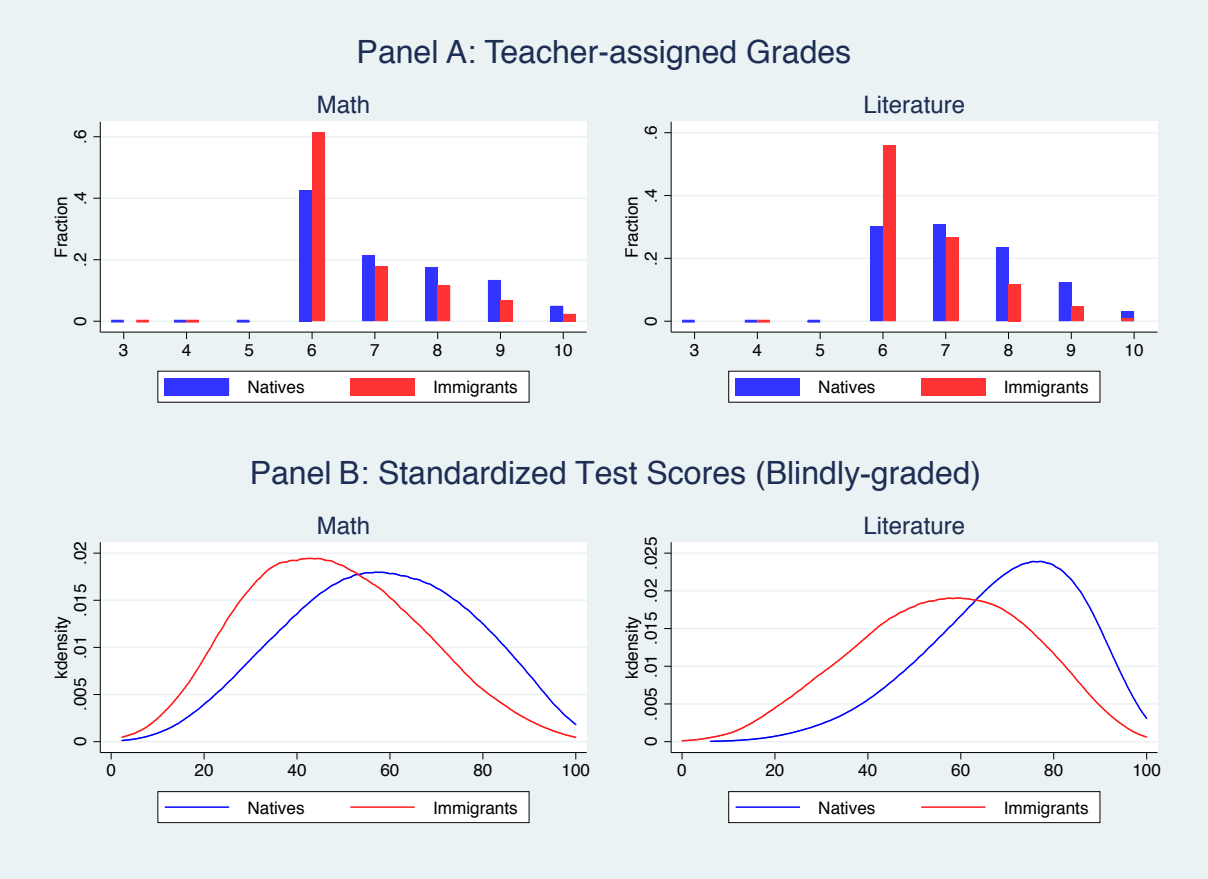
Appendix

Figure A.1: Average race IAT scores across European countries



Notes: This map shows the average race IAT scores across European countries. Red indicates a stronger association between "natives"-"good" and "immigrant"-"bad". *Source: Tom Stafford,* theconversation.com

Figure A.2: Distribution of grades



Notes: These graphs show the distribution of teacher-assigned grades (Panel A) and standardized test scores (Panel B) in math and literature across native and immigrant students.

Table A.1: Country of birth of immigrant students (school year 2016-2017)

Place of Birth	**Number of students**	**Share among immigrant children**
Romania	158,428	19.2%
Albania	112,171	13.6%
Morocco	102,121	12.4%
China	49,514	6.0%
Philippines	26,962	3.3%
India	25,851	3.1%
Moldavia	25,308	3.1%
Ukraine	19,956	2.4%
Pakistan	19,934	2.4%
Egypt	19,925	2.4%
Tunisia	18,613	2.3%
Peru	18,018	2.2%
Ecuador	16,153	2.0%
Macedonia	15,193	1.8%
Nigeria	14,853	1.8%

Source: Italian Ministry of Education. This table reports the total number of students by country of birth for the 15 most represented nationalities, and their share among all immigrant students.

Table A.2: Correlation between teacher characteristics and IAT

	(1)	(2)	(3)
Dep. Var.: Teacher's IAT			
Reason Immigrant-Native Gap High-School: economic	-0.015	-0.014	0.000
	(0.026)	(0.027)	(0.029)
Reason Immigrant-Native Gap High-School: behavior	-0.018	-0.021	-0.013
	(0.014)	(0.014)	(0.015)
Reason Immigrant-Native Gap High-School: ability	0.032**	0.031**	0.031**
	(0.014)	(0.014)	(0.015)
Reason Immigrant-Native Gap High-School: language	0.025	0.027	0.017
	(0.020)	(0.020)	(0.021)
Reason Immigrant-Native Gap High-School: information	-0.010	-0.010	-0.007
	(0.016)	(0.016)	(0.016)
Reason Immigrant-Native Gap High-School: prejudices	0.022	0.022	0.019
	(0.016)	(0.015)	(0.016)
Female		-0.040*	-0.045**
		(0.020)	(0.021)
Born in the North		-0.027*	-0.023
		(0.014)	(0.016)
WVS Immigrants' Rights to Job		-0.055***	-0.051**
		(0.017)	(0.020)
Children		0.094	0.085
		(0.071)	(0.080)
Middle edu Mother		0.028	0.030
		(0.017)	(0.019)
High edu Mother		-0.021	-0.029
		(0.021)	(0.023)
School FE	No	No	Yes
Obs.	1384	1384	1384
R^2	0.067	0.086	0.152

Notes: This table reports OLS estimates, where the dependent variable is IAT score of teachers, and the unit of observation is the teacher t in grade 8 of school s. We include controls for the order of IATs and for whether the blocks were presented on an order compatible or incompatible way (which was randomized at the individual level). The variables "Reason Immigrant-Native Gap High-School: X" equal 1 for teachers believing that X is a very important or important reason for the Immigrant-Native gap in high-school choice. *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

Table A.3: Teacher characteristics across schools included and not included in the experiment

Variable	(1) Not in the Sample	(2) Final Sample	(3) Diff.	(4) Normalized Difference
IAT Immigrant	0.457	0.474	0.017	0.046
	(0.257)	(0.264)	(0.020)	
Female	0.821	0.874	0.053	0.104
	(0.384)	(0.332)	(0.034)	
Teaching Math	0.464	0.492	0.027	0.039
	(0.500)	(0.500)	(0.023)	
Born in the North	0.636	0.684	0.049	0.073
	(0.482)	(0.465)	(0.045)	
Age	49.155	49.093	-0.062	-0.005
	(8.636)	(9.374)	(0.902)	
Full time contract	0.887	0.858	-0.029	-0.062
	(0.317)	(0.349)	(0.030)	
Years of experience	20.469	20.203	-0.265	-0.017
	(10.825)	(11.537)	(1.051)	
Children	0.634	0.700	0.066	0.099
	(0.483)	(0.459)	(0.044)	
Low edu Mother	0.539	0.498	-0.041	-0.058
	(0.500)	(0.501)	(0.047)	
Middle edu Mother	0.348	0.338	-0.011	-0.016
	(0.478)	(0.473)	(0.043)	
High edu Mother	0.113	0.165	0.052	0.106
	(0.317)	(0.371)	(0.031)*	
Degree Laude	0.230	0.284	0.054	0.088
	(0.422)	(0.452)	(0.031)*	
Reason Immigrant-Native Gap High-School: economic	3.844	3.945	0.101	0.066
	(1.088)	(1.063)	(0.081)	
Reason Immigrant-Native Gap High-School: behavior	2.661	2.575	-0.086	-0.057
	(1.040)	(1.098)	(0.079)	
Reason Immigrant-Native Gap High-School: ability	2.822	2.536	-0.286	-0.173
	(1.169)	(1.163)	(0.103)***	
Reason Immigrant-Native Gap High-School: language	3.552	3.533	-0.019	-0.014
	(0.984)	(0.991)	(0.084)	
Reason Immigrant-Native Gap High-School: information	2.741	2.697	-0.043	-0.028
	(1.064)	(1.091)	(0.090)	
Reason Immigrant-Native Gap High-School: prejudices	2.667	2.728	0.061	0.042
	(1.023)	(1.051)	(0.073)	
WVS Immigrants' Rights to Job	0.792	0.800	0.008	0.013
	(0.407)	(0.401)	(0.038)	
Observations	224	533		

Notes: This table compares the average characteristics of teachers across schools included and not included in the experiment. Standard errors are robust and clustered at the school level. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table A.4: Teacher characteristics across schools included and not included in the experiment

Variable	(1) Not in the Sample	(2) Final Sample	(3) Diff.	(4) Normalized difference
Female	0.481	0.496	0.015	0.021
	(0.500)	(0.500)	(0.012)	
Immigrant	0.256	0.185	-0.072	-0.122
	(0.437)	(0.388)	(0.022)***	
High Edu Mother	0.139	0.191	0.052	0.100
	(0.346)	(0.393)	(0.035)	
High Edu Father	0.132	0.166	0.034	0.068
	(0.338)	(0.372)	(0.035)	
Mother High-skill	0.103	0.120	0.017	0.039
	(0.304)	(0.325)	(0.029)	
Mother Mid-Skill	0.289	0.347	0.058	0.088
	(0.453)	(0.476)	(0.024)**	
High Occupation Father	0.179	0.199	0.020	0.035
	(0.384)	(0.399)	(0.041)	
Medium Occupation Father	0.310	0.350	0.041	0.061
	(0.462)	(0.477)	(0.024)*	
Grade Math June '16	7.223	7.164	-0.059	-0.033
	(1.307)	(1.238)	(0.052)	
Grade Ita June '16	7.131	7.129	-0.001	-0.001
	(1.064)	(1.049)	(0.046)	
Observations	2,597	6,031	8,628	

Notes: This table compares the average characteristics of students across schools included and not included in the experiment. Standard errors are robust and clustered at the school level. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table A.5: Correlation between teachers characteristics and willingness to receive the feedback

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable: Dummy for whether the teacher wants to receive the feedback							
IAT Immigrant	0.004					0.000	0.031
	(0.032)					(0.033)	(0.032)
Teaching Math		0.026				0.021	0.029
		(0.022)				(0.022)	(0.023)
Female			0.003			0.004	0.020
			(0.031)			(0.031)	(0.028)
WVS				-0.036		-0.034	-0.002
				(0.029)		(0.030)	(0.029)
Time Survey: slow					0.053*	0.053*	0.014
					(0.031)	(0.031)	(0.032)
Time Survey: fast					-0.017	-0.015	0.004
					(0.053)	(0.054)	(0.047)
Time Survey: missing					-0.096**	-0.093**	-0.032
					(0.046)	(0.046)	(0.050)
FE school	No	No	No	No	No	No	Yes
Obs.	1384	1384	1384	1384	1384	1384	1384
R^2	0.000	0.001	0.000	0.001	0.004	0.006	0.247

Notes: Robust standard errors clustered at school level are in parentheses. All columns include dummy variables for missing characteristics (if any). "Time Survey: fast " equals 1 for teachers who took fewer than 11 minutes to complete the survey. "Time Survey: slow" equals 1 for teachers who took more than 20 minutes to complete the survey. The average completion time is around 15.5 minutes. "Time Survey: missing" indicates that a teacher did not complete the survey with the tablet and only did the IAT.

Table A.6: Teachers' IAT scores and grades assigned to immigrants, distinguishing between first and second generation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable	**Grades given by Math Teacher**					**Grades given by Literature Teacher**				
Immigrant	-0.357***	-0.111***	-0.046	0.053	1.159*	-0.505***	-0.117***	-0.074	-0.051	0.362
	(0.033)	(0.023)	(0.047)	(0.052)	(0.636)	(0.028)	(0.021)	(0.045)	(0.047)	(0.725)
First Gen Imm	-0.136***	0.035	0.024	0.044	0.045	-0.115***	0.012	-0.054	-0.023	-0.026
	(0.034)	(0.027)	(0.052)	(0.052)	(0.052)	(0.030)	(0.024)	(0.045)	(0.045)	(0.045)
Imm*Teacher IAT			-0.037	-0.032	-0.035			-0.024	-0.017	-0.016
			(0.023)	(0.022)	(0.023)			(0.023)	(0.022)	(0.022)
First Gen *Teacher IAT			0.007	0.003	0.006			0.038	0.028	0.029
			(0.025)	(0.025)	(0.025)			(0.023)	(0.023)	(0.023)
Obs.	21846	21846	21846	21846	21846	20457	20457	20457	20457	20457
R^2	0.094	0.475	0.475	0.505	0.505	0.150	0.524	0.524	0.558	0.558
Teacher FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
INVALSI cubic	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Student Controls*Imm	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Teacher Controls *Imm	No	No	No	No	Yes	No	No	No	No	Yes

Notes: This table reports OLS estimates, where the dependent variable is grade given by teachers in math (Columns 1-5) and literature (Columns 6-10) in grade 8, and the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. "'Teacher's IAT" is the raw score of IAT divided by the standard deviation. We include a cohort dummy in all regressions and "INVALSI cubic" indicates the cubic polynomial of INVALSI test score in grade 8. Student controls include gender, generation of immigration, and education of the mother. Teacher controls include gender, place of birth, age, and advanced STEM degree (physics, math, engineering). *, **, and *** indicate significance at the 10%, 5%, and 1% percent level respectively.

B Online Appendix

B.1 Description of Implicit Association Test

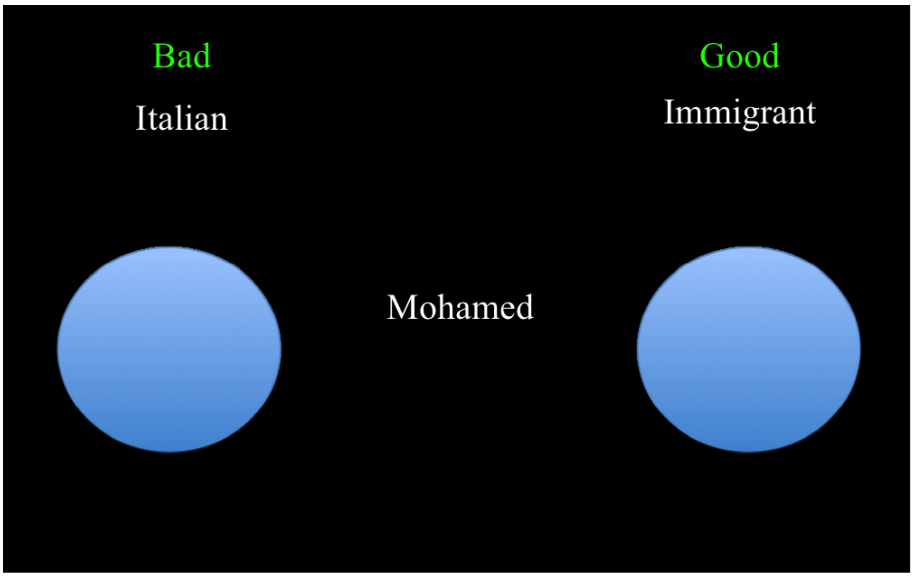
We invite teachers to complete a seven-block IAT following the schematic overview presented in Table B.1.1. Half of the teachers, randomly selected at the individual level, completed the IAT in the order as presented in Table B.1.1 ("order compatible" task first), while the other half completed the IAT with the blocks in the following order: 1, 5, 6, 7, 2, 3, 4 ("order incompatible" task first). An example screenshot of the latter task is presented in Figure B.1.1, while all the words presented to teachers are shown in the box below (with the original in Italian in parentheses). On average, there is a small difference in the IAT score between individuals who perform the order compatible task first vs. the order incompatible task first. Hence, in all regressions where there are no teacher fixed effects, we control for whether the first task was order compatible.

The blocks used to calculate the IAT score are blocks 3, 4, 6, and 7. The number of words that need to be categorized is 20 in blocks 3 and 6 and 40 in blocks 4 and 7, as in the standard IAT with 7 blocks. The scoring procedure follows the guidelines of the improved scoring algorithm defined by Greenwald et al. (2003).

Table B.1.1: Schematic overview of the Immigrant Implicit Association Test

Blocks	Left Categories	Right Categories
1	Italian	Immigrant
2	Good	Bad
3	Italian Good	Immigrant Bad
4	Italian Good	Immigrant Bad
5	Bad	Good
6	Italian Bad	Immigrant Good
7	Italian Bad	Immigrant Good

Figure B.1.1: Example of the screenshot of the tablet in the "order incompatible" task



- **IAT with male names of immigrants and natives**

  1. Immigrant *(Immigrato)*: Youssef, Mohamed, Gheorghe, Alejandro, Li Yi, Pascual
  2. Italian *(Italiano)*: Marco, Simone, Daniele, Francesco, Lorenzo, Mattia
  3. Good *(Bravo)*: Prepared *(Preparato)*, Intelligent *(Intelligente)*, Capable *(Capace)*, Studious *(Studioso)*, Able *(Abile)*, Precise *(Attento)*, Willing *(Volenteroso)*, Respectful *(Rispettoso)*
  4. Bad *(Impreparato)*: Disrespectful *(Irrispettoso)*, Slow *(Tardo)*, Incapable *(Incapace)*, Boisterous *(Irrequieto)*, Lazy *(Pigro)*, Distracted *(Distratto)*, Demotivated *(Demotivato)*, Insufficient *(Scarso)*

- **IAT with female names of immigrants and natives**

  1. Immigrant *(Immigrata)*: Fatima, Naila, Adina, Iryna, Jiaxin, Beatriz
  2. Italian *(Italiana)*: Valentina, Sara, Giorgia, Francesca, Elisa, Alice
  3. Good *(Brava)*: Prepared *(Preparata)*, Intelligent *(Intelligente)*, Capable *(Capace)*, Studious *(Studiosa)*, Able *(Abile)*, Precise *(Attenta)*, Willing *(Volenterosa)*, Respectful *(Rispettosa)*
  4. Bad *(Impreparata)*: Disrespectful *(Irrispettosa)*, Slow *(Tarda)*, Incapable *(Incapace)*, Boisterous *(Irrequieta)*, Lazy *(Pigra)*, Distracted *(Distratta)*, Demotivated *(Demotivata)*, Insufficient *(Scarsa)*

B.2 Teacher Questionnaire

1) *Immigrant children, with the same grades of natives, are more likely to choose a vocational track. According to your experience, how much do you think these factors affect the choice of immigrants? Answers on a scale from 1 to 5.*

1. *Economic reasons*
2. *Bad behaviour at school*
3. *Insufficient abilities for more demanding schools*
4. *Knowledge of the language*
5. *No information about educational and occupational careers*
6. *Perception of prejudices in school or at work*

2) Do you agree or disagree with the following statements? When jobs are scarce, employers should give priority to Italian people over immigrants. Possible answers: Agree, Neither Agree nor Disagree, Disagree, Don't know

B.3 Email with the feedback

The exact wording of the email with the feedback about own implicit bias is reported in this appendix translated in English. Instead of the XXX, teachers saw the precise score (for example, 0.25). We followed the standard categorization of IAT scores (Greenwald et al., 2009): no association if the score is between -0.15 and 0.15, slight association for values between $|0.15|$ and $|0.35|$, moderate association between $|0.35|$ and $|0.60|$, and strong association for scores higher than $|0.60|$.

Subject: Result of the Implicit Association Test – Research Project of Bocconi University

Dear teacher,

As per your request, we are writing you to let you know your result of the Implicit Association Test that you completed during the questionnaire administered by Bocconi University and related to the research titled "The role of teachers in high-school track choice". You did this test using a tablet in the school building where you work. The Implicit Association Test was administered to teachers in middle school to measure and increase the awareness of potential unconscious preferences or associations.

IMPLICIT ASSOCIATION TEST: this test investigates the automatic associations between immigrant and Italian names with positive associations (e.g., good) and negative associations (e.g., bad). You completed this test separately with male and female names.
Your Immigrant-Native Implicit Association Test score using male names of natives and immigrants is XXX, which suggests a (slight/moderate/strong) association between positive attributes and Italian/immigrant names, and between negative attributes and immigrant/Italian names (or no automatic associations between positive attributes and Italian or immigrant names).
Your Immigrant-Native Implicit Association Test score using female names of natives and immigrants is XXX, which suggests a (slight/moderate/strong) association between positive attributes and Italian/immigrant names, and between negative attributes and immigrant/Italian names (or no automatic associations between positive attributes and Italian or immigrant names).

We want to underscore that this test reveals implicit attitudes and not behaviours. Our attitudes may derive from the cultural and social context where we live, and it is not obvious that explicit and implicit behaviours coincide. All of your responses will be held in confidence: only the researchers involved in this study will have access to the information you provide. Your responses will not be shared with other people. Data collected will be published in aggregate form and it will not be possible to link them with the teacher or the school. We hope that you found this test useful. Thank you for the time you dedicated to our research.

The Research Team