PhD THESIS DECLARATION

The undersigned

SURNAME *Carlana*
FIRST NAME *Michela*
PhD Registration Number *1691294*

Thesis title:
Essays on Gender and Immigration Economics

PhD in *Economics & Finance*
Cycle *28th*
Candidate's tutor *Professor Eliana La Ferrara*
Year of thesis defence *2018*

DECLARES

Under his responsibility:

1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the "Bilioteche Nazionali Centrali" (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary embargo protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

3) that the Bocconi Library will file the thesis in its "Archivio istituzionale ad accesso aperto" (institutional registry) which permits online consultation of the complete text (except in cases of a temporary embargo);

4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the candidate in unalterable format to Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:

   - thesis *Essays on Gender and Immigration Economics*;
   - by *Carlana Michela*;
   - defended at Università Commerciale "Luigi Bocconi" – Milano in *2018*;
   - the thesis is protected by the regulations governing copyright (Italian law no. 633, 22th April 1941 and subsequent modifications). The exception is the right of Università Commerciale "Luigi Bocconi" to reproduce the same for research and teaching purposes, quoting the source;

5) that the copy of the thesis submitted online to NORMADEC is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;

7) that the PhD thesis is subject to "embargo" as per the separate undersigned "PhD Thesis Temporary Embargo Request".

Date *31 January 2018*

SURNAME *Carlana*
FIRST NAME *Michela*

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to express my deepest and sincere gratitude to my advisor Eliana La Ferrara. I thank her for the enlightening guidance and inspiring instruction. She is an incredibly passionate researcher and teacher that I look up to immensely. I sincerely could not have imagined having a better mentor and role model for my early career as a researcher. I am immensely indebted to Alberto Alesina, Nicola Gennaioli, and Paolo Pinotti for providing deep insights, thoughtful conversations, and amazing research opportunities that inspired me to improve and expand my work. I am also very grateful to several professor that filled Bocconi University with stimulating discussions and helpful comments – in particular Jerome Adda, Thomas Le Barbanchon, Pamela Giustinelli, Selim Gulesci, Massimo Morelli, Nicola Pavoni, Guido Tabellini, Diego Ubfal, and Fernando Vega-Redondo. Finally, I would like to express my gratitude to Erich Battistin, my advisor during the Master at University of Padua: without his encouragement and initial guidance I would have not enrolled in a Ph.D. program.

I would like to thank my fellow graduate students from Bocconi University, for the stressful days and nights before exams and deadlines, interesting discussions, and for the fun – running and especially eating – during these years. Thanks Alessandro, Alexandra, Giacomo, Giampaolo, Laura, Tommaso, and (of course) Valerio!

My work immensely benefitted from the interaction with researchers from Harvard and IIES (Stockholm University) during my visiting periods in the academic year 2015-16 and 2016-17, respectively. I am particularly grateful to David Autor, Roland Fryer, Claudia Goldin, Larry Katz, Nathan Nunn, Andreas Madestam, Amanda Pallais, Jonathan de Quidt, David Stromberg, Jackob Svensson, and all graduate students that I met in Boston and Stockholm who made me feel at home while visiting their cities. Their support has been so important to enjoying grad school. A special thank to Marco, who has been a great friend – especially during the intense job market period – and an amazing coauthor.

The first chapter of my dissertation is funded under the grant "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 Number: 608109) and received financial support from the Laboratory for Effective Anti-poverty Policies (LEAP-Bocconi). The second chapter is funded by Fondazione CARIPLO, Compagnia di San Paolo and Fondazione Cassa di Risparmio di Padova e Rovigo and by DONDENA (Bocconi University). I am also thankful for financial support from the PhD school at Bocconi University as well as the *Research Visiting Grant* of the Cariplo foundation for the support while visiting Harvard University.

I thank for outstanding research assistance Cristina Clerici, Rosa De Vivo, Elena De Gioannis, and Giulia Tomaselli. I am indebted to Gianna Barbieri and Lucia De Fabrizio (Italian Ministry of Education, Statistics), Patrizia Falzetti, Michele Cardone, and Paola Giangiacomo (Invalsi) for generous support in providing the data used in the first two chapters. I am grateful to all principals and teachers of schools involved in my research work for their collaboration in data collection. This thesis would have not be possible without your invaluable help in data collection and extraction!

These years were brighter thanks to my friends since childhood, *"amiche da zero anni"*, Anna, Carla, Giorgia, Giulia, and Laura, and our days together with the same laughs, love, and enthusiasm. I want to thank also my friends from University of Padua, Anna, Daniele, Elisa, Francesca, Giulia, and Martina. Although space distance has often kept us apart, you are such an important part of my life and I look forward to sharing many more of life's lovely moments with you.

I have been blessed with the encouragement of a wonderful family, who has always supported me with love, even when my choices lead me far away from my hometown, Sandrigo. Throughout these years, the genuine support of my parents Susy and Gianni, the restless encouragement from my brother Alberto, and the invaluable hugs from my nephews Edoardo and Francesco have given me the strength to face all challenges and struggles that accompany Ph.D. life. I own them much more than words can describe.

Last but not least, I have been fortunate to have Valerio in my life during these years. You are the best gift I could wish for. I don't think I can thank you enough for your restless support, for believing in me, and for making me laugh every single day. Thanks for reminding me what matters in life. I am lucky to have you by my side and I dedicate my thesis to you.

I filled my doctoral studies with perseverance and passion. I have cultivated my enthusiasm for knowledge and the idea that research in economics can help improving our society. The Ph.D. chapter of my live is over. Now, I am looking forward to what comes next.

Abstract

This PhD thesis is composed of three chapters on gender and immigration economics. In the first chapter, I study the impact of exposure to gender-biased teachers on student achievement and self-confidence. The gender gap in math performance substantially increases when students are quasi-randomly assigned to teachers with stronger stereotypes (as measured by an implicit association test). The effect is driven by lower performance of female students, while there is no impact on males. Those lagging behind the most when assigned to biased teachers are girls from disadvantaged backgrounds and with lower initial level of achievements. Teacher bias induces females to self-select into less demanding high-schools, following the track recommendation of their teachers. Finally, teacher bias has a substantial negative impact on females' assessment of their own math ability. These findings are consistent with the hypothesis that ability-stigmatized groups underperform and fail to achieve their potential.

The second chapter, co-autored with E. La Ferrara and P. Pinotti, focuses on the educational choices of immigrant children in Italy. We estimate the impact of an innovative program that provided tutoring and career counseling to a random sample of immigrant children displaying high academic potential. We find that the program was successful in reducing educational segregation of male students. The effects are in the same direction but smaller and not significant for girls. Male treated students display an improvement in cognitive and soft skills (academic motivation and perceived environmental barriers). Both effects seem to have been internalized by teachers, who recommended them for a more demanding high-school. We show that changes in academic motivation induced by the treatment explain a sizable portion of the effect on the high-school choice, while the effect of increases in cognitive skills is negligible. Finally, we find evidence of positive spillovers of the intervention on immigrant peers of treated.

In the third chapter, I study the effects of immigration on natives' marriage, fertility, and family formation across US cities between 1910 and 1930, in a joint work with M. Tabellini. Instrumenting immigrants' location decision by interacting pre-existing ethnic settlements with aggregate migration flows, we find that immigration raised marriage rates, fertility, and the propensity to leave the parental house for young native men and women. We show that these effects were driven by the large and positive impact of immigration on native men's employment and occupational standing, which increased the supply of "marriageable men". We also explore alternative mechanisms - changes in sex ratios, natives' cultural responses, and displacement effects of immigrants on female employment - and provide evidence that none of them can account for a quantitatively relevant fraction of our results.

Introduction

This thesis investigates how people make important decisions, such as those regarding accumulation of human capital and family structure, and how individuals respond to public policies and cultural norms that may affect their life chances. I explore questions related to gender and immigration economics, with an emphasis on the role played by psychological factors. I focus on policy relevant issues, such as gender differences in math performance and the aspiration trap that may lead immigrant children to underinvest in their education. I have created innovative datasets by merging administrative information with newly collected or digitalized data and I exploit careful econometric identification to provide credible and rigorous answers.

The chapter *"Stereotypes and Self-Stereotypes: Evidence from Teachers' Gender Bias"* analysis whether exposure to gender bias influences human capital investments of children. Despite the impressive narrowing of gender differences in labor market outcomes and educational achievements, girls underperform in math compared to boys and women are still strongly underrepresented in highly profitable fields, such as science, technology, math, and engineering (STEM). Gender stereotypical beliefs about the lower math ability of women with respect to men are widespread and deeply-held (Guiso et al., 2008; Reuben et al., 2014). In the first chapter, I focus on the impact of exposure to teachers' gender bias on math performance, educational choices, and on self-confidence on own ability.

I build a unique dataset, by merging a newly collected questionnaire from students and more than 1.400 teachers together with administrative data from the Italian Ministry of Education and INVALSI (the National Institute for the Educational Evaluation of Instruction and Training).[1] I administered to each teacher a Gender-Science Implicit Association Test (IAT), a tablet-based test that exploits the reaction time to associations to calculate implicit bias. It was created by social psychologists (Greenwald et al., 1998) and recently used by economists when studying discrimination (Bertrand and Duflo, 2017; Glover et al., 2017). In order to identify the causal effect of exposure to teachers' bias, I exploit quasi-random assignment of students to teachers with different level of bias, within the same school and cohort.

First, this chapter shows that girls lag behind in math performance when assigned

[1] This dataset was created for this work and for Alesina et al. (2018).

to teachers with higher implicit bias, while there is no impact on boys. The gender gap in math performance generated during middle school increases by 34% if students are assigned to a teacher with one standard deviation higher implicit bias. Second, I provide evidence that teacher bias has a negative impact on the high-school track choice of female students, inducing them to undertake less demanding tracks. Lastly, girls are less likely to consider themselves "good at math" when exposed to teachers with stronger implicit associations between males and scientific field, even controlling for their standardized test scores. All results are robust to the inclusion of controls at both teacher and student level, interacted with pupils' gender. The findings are consistent with a theoretical framework whereby ability-stigmatized groups underperform failing to achieve their potential.

The last two chapters of my dissertation are focused on immigration. The growing number of immigrant students has profoundly changed the challenges that schooling systems have to face in order to ensure skill development in a diverse student population and promote social cohesion. Students from disadvantaged backgrounds, especially in schooling systems characterized by stratification in high school tracks, are substantially lagging behind in their cognitive and socio-emotional abilities (Fryer Jr and Levitt, 2004; Heckman et al., 2006; Cunha and Heckman, 2007). This could ultimately have long term effects on the educational and occupational careers of children from immigrant families, reducing social mobility and creating unequal opportunities (Guyon et al., 2012; Brunello and Checchi, 2007).

In the chapter *"Goals and Gaps: Educational Careers of Immigrant Children,"*, coauthored with E. La Ferrara and P. Pinotti, we document that immigrant children in Italy, especially males, tend to choose vocational training over more academic curricula relative to native students with similar ability, as measured by standardized test scores. Furthermore, we estimates the impact of a tutoring and career counseling program offered during middle school to a random sample of immigrant children displaying high academic potential. We find that the program was successful in reducing educational segregation of male students. Boys who participated in the intervention display an improvement in cognitive and soft skills (academic motivation and perceived environmental barriers). Both effects seem to have been internalized by teachers, who recommended them for a more demanding high-school. We show that changes in academic motivation induced by the treatment explain a sizable portion of the effect on the high-school choice, while the effect of increases in cognitive skills is negligible. Finally, we find positive spillovers on immigrant classmates of treated students, while there is no effect on native classmates.

Today, immigration is at the forefront of the political debate, and there are increasing concerns over its economic and social consequences. If we look at American history, however, this is not the first time that immigration is such a relevant and controversial issue. In fact, at the beginning of the twentieth century, following the inflow of more than 30 million Europeans, the share of foreign born in the US population was even higher than it is today, and opposition towards immigration was widespread. The last chapter of my dissertation analyzes the effect of exposure to immigration on fertility, family structure, education, and employment of *natives*.

In the chapter *"Happily Ever After: Immigration, Natives' Marriage, and Fertility"*, coauthored with M. Tabellini, we exploit plausibly exogenous variation in the number of European immigrants to US cities between 1910 and 1930 induced by WWI and the Immigration Acts to study the impact of immigration on marriage rates, fertility, and the propensity to leave the parental house for young native men and women. We find that, by promoting industrial expansion and economic activity, immigration increased the supply of native "marriageable" men who, because of their better employment prospects and occupational standing, became more attractive spouses. This, in turn, fostered natives' marriage rates for both men and women, and induced young adults to leave their parents' house earlier in their life. Higher marriage rates, in a period when oral contraception was not yet available, raised natives' fertility, mainly by increasing the number of women with at least one child (extensive margin).

Findings in this dissertation provide motivation for future work in several directions, on both gender and immigration economics. In my ongoing research, I plan to investigate concrete policies that could be implemented in order to alleviate the effect of gender and race stereotypes on students. Finally, I will expand my research to study the dynamics of cultural assimilation - between immigrants and natives as well as between different ethnic groups - using intermarriage as a proxy for the latter.

1. Stereotypes and Self-Stereotypes: Evidence from Teachers' Gender Bias

Abstract
I study the impact of exposure to gender-biased teachers on student achievement and self-confidence. The gender gap in math performance substantially increases when students are quasi-randomly assigned to teachers with stronger stereotypes (as measured by an implicit association test). The effect is driven by lower performance of female students, while there is no impact on males. Teacher bias induces females to self-select into less demanding high-schools, following the track recommendation of their teachers. Finally, teacher bias has a substantial negative impact on females' assessment of their own math ability. These findings are consistent with the hypothesis that ability-stigmatized groups underperform and fail to achieve their potential. *JEL:* J16, J24, I24.

1.1 Introduction

Over the last century, the narrowing of gender differences in labor market participation and educational outcomes has been impressive, up to a reversal of the gap in school attainment in many contexts (Goldin et al., 2006). In spite of this, boys outperform girls in math in most countries and the gender gap in favour of boys is even wider among the highest-achieving students (OECD, 2014). Several studies have shown that math test scores are good predictors of future occupation and earnings (Altonji and Blank, 1999). Gaining a better understanding of the reasons behind the emergence of the gap in math skills is of first-order importance to explain the enduring gender differences in readiness for science, technology, engineering, and math (STEM) universities and the underrepresentation of women in these highly profitable fields (Card and Payne, 2017).

The gender gap in math performance is generally attributed to either biologically based explanations in brain functioning or social conditioning.[1]. In this paper, I focus on the latter and I study whether exposure to gender stereotypes of teachers during middle school can affect math achievement, track choice, and self-confidence of boys and girls. According to social psychology literature, teachers believe math is more difficult for girls than equally achieving boys (Riegle-Crumb and Humphries, 2012; Tiedemann, 2002). Gender stereotypical beliefs are pervasive and deeply-held in most societies: women are believed to be worse than men in mathematics and arithmetic, even in tasks in which both genders perform equally well on average (Bordalo et al., 2016; Reuben et al., 2014).[2] However, our understanding of the role of gender stereotypes on educational outcomes is limited by the difficulty in measuring stereotypes. Also, no evidence exists on the effect of gender bias on students' self-confidence. This paper addresses both of these gaps.

Analyzing the role of teacher stereotypes on student outcomes presents two main challenges: identification and the measurement of stereotypes. I tackle the former by exploiting quasi-random assignment of students to teachers with different level of bias, within the same school. I measure stereotypes by collecting teacher bias using an Implicit Association Test (IAT). This is a computer-based tool developed by social psychologists (Greenwald et al., 1998) and recently used by economists when studying discrimination in the context of gender and race bias (Reuben et al., 2014; Glover et al., 2017; Lowes et al., 2015; Burns et al., 2016).

[1]For instance, Baron-Cohen (2003) elaborates the "empathizing-systemizing theory" whereby there are evolutionary differences among genders: females are stronger empathizers and males are stronger systemizers.

[2]Stereotypes are overgeneralized and simplified representation of differences between groups, which may hold a *kernel-of-truth* (Bordalo et al., 2017). For instance, the belief that *women are worse than men in math* is based on the empirical evidence that girls lag behind in math test-scores in most countries by the age of 14.

I find that the effect of teachers' gender stereotypes is negative and quantitativly significant. First, I show that the gender gap in math performance during middle school increases by 34 percent when students are assigned to teachers with one standard deviation higher bias. The gender gap in math improvements almost triples in classes where the math teacher has a "pro boys" attitude compared to classes in which she or he has a "pro girls" attitude.[3] The effect is driven by lower performance of females when assigned to biased teachers, while males are not affected by exposure to gender bias. Those lagging behind the most when assigned to biased teachers are girls from disadvantaged backgrounds and with lower initial level of achievement. Second, I provide evidence that teacher bias is correlated with their high-school recommendation to pupils and it induces females to undertake less demanding tracks. Finally, I discuss two mechanisms behind the negative impact of teacher bias on student achievements: self-stereotypes and pupil-teacher interaction. I show that teacher stereotypes have a substantial negative impact on girls' self-confidence in math. The findings are consistent with a model whereby ability-stigmatized groups underperform failing to achieve their potential.

To perform the analysis, I build a unique dataset, combining administrative information on pupils from the Italian Ministry of Education and the National Institute for the Evaluation of the Italian Education System (INVALSI) with a newly collected questionnaire on students and teachers in Italy. I survey more than 1,400 math and literature teachers, working in 103 schools in the North of Italy. As measure of gender bias, I collect Gender-Science IAT. The test exploits the reaction time to associations among male or female names and scientific or humanistic fields. The underlying assumption is that responses are faster and more accurate when gender and field subjects are more closely associated by the brain (Lane et al., 2007). Implicit bias has been found to correlate with many outcomes in the real world and in laboratory experiments, related for instance to hiring decisions (Reuben et al., 2014; Rooth, 2010). In addition to IAT scores, I have collected detailed information on teacher characteristics, such as family background, teaching experience and explicit gender beliefs. These data are matched with student performance in math and reading standardized test scores, family background, high-school track choice and teachers' track recommendation. Finally, the dataset is complemented by original information on self-confidence for a sub-sample of students.

I present evidence from two empirical strategies. The first one investigates the impact of teacher bias on the gender gap within the class. I include class fixed effects, which

[3] I consider the thresholds defined by Greenwald et al. (2003) to identify teachers with "pro boys", "pro girls", "without bias" attitude (the latter is a IAT score between -0.15 and 0.15). The gender gap in math performance is -0.035 standard deviations in "pro girls" classes and -0.10 standard deviations in "pro boys" classes. The increase by 34 percent in the gender gap when students are assigned to teachers with one standard deviation higher bias corresponds to an increase of 0.03 standard deviations with respect to an average gap in test scores generated during middle school of 0.08 standard deviations.

absorb all characteristics of peers, school environment, and teachers, including the level of gender stereotypes. I exploit variations in performance and track choice between boys and girls enrolled in the same class.[4] In the second empirical strategy, I compare students of the same gender, enrolled in the same school and cohort, but assigned to teachers with different level of bias. Both identification strategies rely on the quasi-random assignment of students to teachers with different level of bias. I provide supporting evidence showing that baseline characteristics of students are not systematically correlated with teacher bias.

This paper makes three contributions. First, I show that implicit bias correlates with several “expected" observable characteristics, as gender, field of study and cultural stereotypes in the place of birth of individuals (as measured by the World Value Survey and female labor force participation).[5] IAT does not correlate with variables such as gender of own children, teacher quality and experience.[6] Second, the paper provides evidence on the relevance of social conditioning in affecting the gender gap in math achievement and high-school track choice. More precisely, it uncovers the role of implicit bias in the context of education economics and pupil-teacher interactions. Third, it shows the influence of teachers on self-stereotypes and self-assessment of own math ability. This is a crucial channel to explain the underperformance of girls in math when assigned to more biased teachers.

This study adds to the recent literature in economics that has underlined the benefits from interacting with social psychologists and considering *implicit bias* in studying discrimination (Guryan and Charles, 2013; Bertrand and Duflo, 2017). Implicit stereotypes can operate even without awareness or intention to harm the stigmatized-group (Bertrand et al., 2005; Nosek et al., 2002). In particular, we may expect that teachers do not explicitly endorse gender stereotypes, but their implicit bias, embedded in their own experiences since childhood, affects their interaction with pupils. I collect IAT scores and I further examine the determinants captured by this test and the reaction time to stimuli.

Some interesting cross-countries evidence shows a correlation between gender gap in mathematics and gender equality. Guiso et al. (2008) and Nosek et al. (2009) provide evidence that gender gap in math performance is wider in those countries with low women

[4]Students are assigned to the same group of peers from grade 6 to grade 8. Teachers are assigned to classes and follow students during all years of middle school, with few exception due, for instance, to retirement or transfer to a different school.

[5]Thanks to the data used in Campa et al. (2010), I have access to the answers at province level of the following World Value Survey question: "When jobs are scarce, men have more right to a job than women".

[6]This has important implications for the estimation: when teachers controls or class fixed effects are not added in the regression, we need to consider teacher bias as including also characteristics correlated with IAT scores.

empowerment and higher implicit gender bias measured by IAT, respectively.[7] The economics literature analyzing the impact of gender stereotypes of teachers on student outcomes has mainly focused on either self-reported measures (Alan et al., 2017) or bias in grading, i.e. the gender differences in grades given in blind vs. open evaluations (Lavy and Sand, 2015; Terrier, 2015).[8] Compared to other measures of teacher bias, the Implicit Association Test has two main advantages. First, it does not suffer from social desirability bias that may be an issue in self-reported measures. Second, the measure of teacher bias is created without relying on data on student performance, which may capture random variation in unobservable characteristics of boys and girls, potentially correlated with future outcomes of pupils.

Finally, I contribute to understanding the importance of gender-biased environments in explaining the under-confidence of females in STEM fields. Gender differences in confidence and competitiveness have negative consequences for women's performance, scientific educational and occupational choices (Kugler et al., 2017; Reuben et al., 2015; Coffman, 2014).[9] Exposure to biased teachers activates negative self-stereotypes on female students. The results are consistent with the predictions of the stereotype threat theory (Steele and Aronson, 1995), according to which individuals at risk of confirming widely-known negative stereotypes reduce their confidence and underperform in fields in which their group is ability-stigmatized (Spencer et al., 1999).

This paper is organized as follows. Section 1.2 explains the setting analyzed, providing information on the Italian institutional background. Section 2.4 describes the data available on both students and teachers. Section 3.4 presents the estimation and identification challenges. The main results of the paper are presented in Section 1.5 and mechanisms are discussed in Section 2.6. Finally, Section 1.7 concludes. All supplementary material is provided in the Appendices.

1.2 Setting

In the Italian educational system, middle school lasts three years from grade 6 to 8. Students are assigned to classes at the beginning of grade 6 and they stay with the same

[7]Guiso et al. (2008) use four measures of gender equality: World Economic Forum's Gender Gap Index (GGI), World Values Surveys (WVSs), labor force participation of women and women's political participation measured by World Economic Forum. They find consistent results with all these measures.

[8]Lavy and Megalokonomou (2017), using a panel dataset, show that gender bias in grading of teachers is persistent over time and it influences students' university choice.

[9]Niederle and Vesterlund (2010) point out that gender differences in competitiveness may have some distortionary effects and exaggerate the advantage of males in math, especially in the right tail of the distribution of test scores.

peers for three years.[10] The general class formation criteria are established by an Italian law and details are specified by each school council in a formal document available on the website of the institution.[11] The general criteria mentioned by most schools are equal allocation of students across classes according to gender, disability, socio-economic status and ability level (as reported by the elementary school). I also collect additional information directly from the principal on how classes are formed, which is described in details in Appendix A.2. School principals report that the most important aspect in the class formation process is the comparability across classes and heterogeneity within class in the same school.[12] What is important for my analysis is that I can also test whether this intention of the principals is confirmed by the allocation of students to classes in my sample (see section 1.4.3).

Teachers are assigned to schools by the Italian Ministry of Education and their salary is determined by experience and rank in a centralized system. Teachers' allocation across school is determined by seniority: when they accumulate years of experience, they tend to move close to their home town and away from disadvantaged backgrounds (Barbieri et al., 2011). Each class is assigned by the principal to a math and Italian teacher among those available in the school and they usually follow students from grade 6 to grade 8. Every week, students spend at least 6 hours with the math teacher and 5 hours with the Italian teacher.[13] Students receive grades by teachers at the end of each semester, which may be affected not only by performance, but also by other factors as diligence, effort and improvements over time. Grades are given in a scale up to 10, where the pass grade is 6.

Standardized test score in math and reading are administered in grade 2, 5, 6, 8 and 10 by the National Institute for the Evaluation of the Italian Education System (INVALSI).[14] The tests are presented to all students as ability tests, thus making the gender stereotype in math potentially relevant. They are graded anonymously following a precise evaluation grid and by a different teacher than the one instructing students in the specific subject. Students are not informed about their performance on the test, except for the one in grade 8. The achievement test score of grade 8 is the highest stakes among these test scores, since it will affect 1/6 of the final score of students at the end of middle school. However,

[10]There are only few exceptions: students may be transfered to a different school or be required to repeat a grade. This affects less than 5% of students.

[11]The D.P.R. 20 marzo 2009 n.81 establishes, for instance, that the number of students per class in middle school should be between 18 and 27. Further information at school level is provided on the "Plan of Education Offer" ("Piano dell'Offerta Formativa").

[12]An analysis of Ferrer-Esteban (2011) shows that ability grouping across classes within schools occurs almost exclusively in the South of Italy, while all schools in my sample are from the North.

[13]Students can be enrolled in school from 30 to 43 hours per week and therefore the amount of time they spend with teachers vary. For instance, they spend from 6 to 9 hours with the math teacher. In some classes, Italian teachers also teach history and geography so they spend more time with students. The amount of hours per week spent with the Italian teacher therefore varies from 5 to 10

[14]The test score in grade 6 was administered only up to the school year 2012-13.

this final grade has no formal and direct impact for the enrollment in high-school or for the future educational career of students.

After middle school, students self-select into three different tracks: academic oriented (“liceo”), technical and vocational high-school. Each type of school is divided in several subtracks: the academic oriented track can be specialized in either scientific, humanistic, languages, human sciences, artistic or musical subjects, the technical track can be focused on technological or economic subjects, while the vocational track can have different core subjects, for instance hospitality training, cosmetics and mechanical workshop. Students are free to choose a high-school with no restriction on the track based on grades or ability. Giustinelli (2016) has shown that child’s enjoyment of the curriculum is one of the most important determinants of high school choice. Teachers give a non-binding track recommendation to families with an official letter sent to children’s home, which is also reported to the Ministry of Education.

The choice of high-school is strongly correlated with the university choice: 80% of graduates in STEM universities in 2015 did a scientific academic or a technical track during high-school (62% did the scientific academic high-school track). Among students enrolled in vocational track, only 1.7% of the cohort graduating in 2016 enrolled in university, while the percentage increases to 73.7% and 32.3% in the academic and technical track respectively. Interestingly, among students of the technical track the majority enrolls in either STEM or economics degrees: 62.5% vs. 52.4% of the academic track students.[15]

1.3 Data

During September 2016, I invited 156 middle schools to take part in a research project regarding “The role of teachers in high-school track choice,” out of which 91 accepted and provided all information necessary for my study. The sample was designed including all schools of the provinces of Milan, Brescia, Padua, Genoa and Turin with more than 20 immigrants in the school year 2011-12 enrolled in grade 6.[16]

I use four sources of data: teacher survey data, student survey data, administrative information from the Italian Ministry of Education (MIUR) and from the National Center for the Evaluation of the Italian Educational System (INVALSI). I collected directly

[15] Author’s calculation on MIUR data.

[16] More precisely, in 103 schools we obtain the authorization of the principal to administer the survey to teachers, but only 91 principals completed (without mistakes) the formal authorization to give me access to data from the National Institute for the Evaluation of the Italian Education System (INVALSI). In 2 cases, the principal explicitly stated they did not want to give access to INVALSI data. In most of the cases, the authorization (with all correct data) was not sent in time for the extraction of data from INVALSI. Finally, the number of schools according with 2011 data were 145. However, some of them where divided in different institutions and we follow all of 156 of them over time.

detailed information on teachers, including *implicit bias* measured by the Gender-Science Implicit Association Test (IAT), and on students' self-assessment of own ability in different subjects. Administrative data from MIUR contained information on gender, place of birth, high-school track choice, and their track recommendation to students. INVALSI provides information on standardized test scores and family background.

1.3.1 Teachers: Gender Stereotypes and Other Characteristics

From October 2016 to March 2017, I conducted a survey of around 1.400 math and Italian teachers. The questionnaire was administered directly one-to-one by enumerators using tablets in a meeting held in school buildings, in most of the cases in the early afternoon. Participants agreed to take part in the survey and signed an informed consent, in which it was explained that the survey was part of a research project aimed at analyzing the role of teachers in affecting students' track choice.[17] There was no direct reference to gender bias. The time to complete the survey was around 30 minutes and participating teachers did not receive compensation. Among all math and Italian teachers working in the schools involved in this research, around 80 percent completed our survey.[18] The survey is divided into two parts: the Implicit Association Test (IAT) and a questionnaire.

Gender-Science Implicit Association Test

In this research, the main focus is on *implicit* gender bias, using a measurement tool developed by social psychology called Implicit Association Test (IAT) (Greenwald et al., 1998; Lane et al., 2007). The idea underlying the test is that the easier the mental task, the faster the response production and the fewer the errors made in the process.[19] The IAT requires the categorization of words to the left or to the right of a computer or tablet screen and it provides a measurement of the strength of the association between two concepts (specifically, gender and scientific/humanistic subjects). Enumerators administered the test using touch screen tablets and they interact directly one-to-one with teachers. Subjects were presented with two sets of stimuli. The first set of stimuli were typical Italian names of females (e.g. Anna) and males (e.g. Luca), and the second set were subjects related to scientific fields (e.g., Calculus) and humanistic fields (e.g., Literature).

[17]The data collection was conducted for a broad research project involving also an ongoing work in which we study teachers' racial bias (Alesina et al., 2018).

[18]Only 4 math teachers, started the questionnaire and then did not finish it since they claimed either that they were not expecting such a long survey or that they could not understand the scope of the Implicit Association Test.

[19]This concept was initially developed by Donders (1868). Donders was very optimistic about the possibility of quantifying how mind works using the "time required for simple mental processes" and performed some of the first experiments making participants respond with the right hand to stimuli on the right side and with the left hand to stimuli on the left side.

One word at a time appears at the center of the screen and individuals are instructed to categorize them to the left or the right according with different labels displayed on the top of the screen (for instance on the right the label "Females" and on the left the label "Males"). Subjects are required to categorize the words as quickly as possible for seven blocks, i.e. seven rounds. To calculate the score, two types of blocks are used: in the first type, individuals are instructed to categorize to one side of the screen male names and scientific subjects and to the opposite side of the screen female names and humanistic subjects ("order compatible" blocks), while in the second type of blocks, individuals are instructed to categorize to one side of the screen female names and scientific subjects and to the opposite side of the screen male names and humanistic subjects ("order incompatible" blocks).[20] The order of the two types of blocks is randomly selected at individual level. Appendix Table A.1 presents the correlation between IAT score and whether the first task was order compatible or incompatible. The effect is small in magnitude and it disappears when controlling for school fixed effects.[21]

A broad strand of literature in social psychology and an increasing number of papers in economics have provided evidence on the validity of IAT scores in predicting relevant choices and behaviors (Nosek et al., 2007; Greenwald et al., 2009). For example, Reuben et al. (2014) shows in a lab experiment that higher stereotypes (measured by gender IAT) predict employers' bias expectations against female math performance and also suboptimal update of expectations after ability is revealed. Higher implicit gender bias is acquired at the beginning of elementary school and is generally associated with lower performance of females in math during college, lower desire to pursue STEM-based careers and lower association of math with self, even for women who had selected math-intensive majors (Cvencek et al., 2011; Nosek et al., 2002; Kiefer and Sekaquaptewa, 2007). Also in the context of race implicit bias, studies have shown the relevance of IAT scores in call-back rates of minority job applicants (Rooth, 2010) and in affecting job performance of minorities (Glover et al., 2017).

There is a lively debate in the literature on how to interpret IAT scores and to what extent they are capturing stable characteristics that do not vary over time (Banaji et al., 2004; Greenwald et al., 2009).[22] Thanks to a broad set of individual level information on

[20]The number of words that appear in the two types of evaluation blocks are 120. As in the standard IAT with a seven-block structure, individuals are asked to categorize only female and male words in the first block, only scientific and humanistic subjects in the second and fifth, while blocks three/ four and six/seven are those described in details and used for evaluation. Detailed explanation is provided in Appendix A.3.

[21]Each teacher performs both gender and race IAT. The order was randomized at individual level. In the Appendix Table A.3, I show the impact of the order of IATs on the score. The correlation is low and indistinguishable from zero. However, in all regressions I will control for ordering factors (even if they have no impact on the estimates).

[22]In particular, it has been shown that race bias (as measured by IAT) decreases after subjects viewed

teachers, I will contribute to this debate by analyzing the correlation between observables and IAT score in Section 1.4.2.

One of the critiques of implicit bias is that people may have a specific context in mind when completing the IAT, which may differ from the context the researcher wants to analyze. If anything, it would increase the noise of the measurement inducing an attenuation bias. However, in my case, this is not a relevant issue. The measure of bias I collect is strongly related to the schooling context and teachers are interviewed directly inside the school building.[23] Furthermore, individuals complete the survey in the presence of an enumerator and therefore I am sure of the identity of who completed the survey.[24]

Teachers' Questionnaire

After the Implicit Association Tests, enumerators invited teachers to complete a questionnaire with detailed information about family background of teachers (age, parents' education, place of birth, age and sex of children, etc) and career related aspects (type of contract, years of experience, whether they are involved in the management of the school or in the organization of Math Olympics Games, etc). Furthermore, they were also asked questions about explicit forms of bias, as for instance beliefs about gender differences in innate math ability and the standard Word Value Survey question: *"When jobs are scarce, men should have more right to a job than women"*.[25] Participants are in general reluctant to explicitly endorse gender stereotypes about differences in innate ability and employment (Nosek et al., 2002) due to social desirability bias in the responses. These aspects are potentially emphasized by the awareness of being interviewed as teachers. Enumerators collected the allocation of teachers to classes from the school year 2011-12 to the school year 2016-17, in order to merge teacher and student data. I confirmed all this information using data provided directly by schools and their websites.

---

pictures of admired African Americans and disliked White Americans (Dasgupta and Greenwald, 2001).

[23] An example in which this may be an issue is the following. Assume I was interested in evaluating the bias toward obese people in the work environment and I collected IAT associating "obese people" and "thin people" with "good" vs. "bad". The positive attitude captured by IAT of a person toward obese people may be due to the fact that his/her mother is obese and he/she loves her. In the job environment, however, the same person may have a neutral attitude toward obese people. This would induce a bias in our measure of attitude toward obese people in the workplace. The context the person has in mind when completing the IAT may have an important effect on the result. In our case, the context of IAT is the same as the outcome I want to evaluate.

[24] A less-expensive and time-consuming alternative could have been sending the survey by email. However, the potential drawbacks were low response rate and uncertain identity of the individual completing the survey.

[25] I also collected information about potential factors that may influence females' scientific track choice (interest for STEM, ability in math, low self-esteem, parents' influence toward different tracks, cultural stereotypes) using a scale of 1 to 5. Finally, I asked teachers to state the expected student performance in standardized test scores by gender. The response rate to the latter question was low and teachers were highly unsure of the answer they provided.

Descriptive Statistics on Math Teachers

The dataset includes 537 math teachers, 855 Italian teachers and 31 teachers of other subjects. The main focus of this paper is on the impact of math teachers gender stereotypes on the performance in the subject they teach. Among these 537 teachers, we restrict the main analysis to 301 teachers (“matched sample") who were working for a school in my sample in the school year 2014-15 and for which we have student data [26]. Appendix Table A.2 shows the balance table of the differences between the sample of teachers matched (301 teachers) and the other 236 math teachers who completed the IAT. As expected, teachers not matched are around 9 years younger, 40 percent less likely to have full-time contract and they have 12 years less of experience in teaching. However, as it can be clearly seen also from Appendix Figure A.1, not only the average, but also the entire distribution of implicit gender bias of the matched and not-matched teachers is extremely close (exact p-value of Kolmogorov-Smirnov: 0.946).

Table 1.1 reports descriptive statistics on math teachers. Most teachers are females (84%), they are on average 52 years old with 23 years of experience in teaching and 92% hold a full-time contract. The majority (65%) of math teachers are born in a city in the North of Italy where the study took place, but a substantial share is born in the Center or South of Italy and then migrated to the North to work. Most teachers graduated from programs in biology, natural sciences and other related subjects: 24% studied math, physics and engineering. At the bottom of Table 1.1, I report the summary statistics of explicit bias questions described in details in Appendix A.3. The variation in the answers on the equality of access to labor market of men and women and about innate gender difference is ability is low, potentially also due to social desirability bias: for instance, less than 2% of the interviewed teachers respond that they agree with the statement that women have less right to jobs than men when opportunities are low. It may be difficult to obtain revealed bias, given the widespread explicit rejection of stereotypes and a related reluctance of participants in revealing their bias, especially if interviewed as “teachers” in the presence of enumerators.

Based on IAT scores, math teachers are slightly gender biased (Figure 1.1): indeed, a positive IAT score indicates a stronger association between males with scientific subjects and female with humanistic subjects. For ease of interpretation of our results, I standardize the IAT score to have mean zero and variance one throughout the paper. Considering the thresholds typically used in the social psychological literature, 25% of teachers are slightly or moderately in favor of girls, 30% present little to no bias, 19% show slight

[26] As specified in section 2.4, for 12 schools we did not obtain this authorization on time or there was a mistake in the authorization form. Furthermore, we lost some observations because some schools changed the official code (called “meccanografico”) over the years of our sample and INVALSI guarantees access to data only for school codes whose principal has signed the authorization.

Table 1.1: Summary Statistics from Math Teachers' Questionnaire

	Count	Mean	SD	Min	Max
Family and education					
Female	301	0.84	0.37	0.00	1.00
Born in the North	291	0.65	0.48	0.00	1.00
Age	290	51.90	8.38	31.00	66.00
Children	301	0.74	0.44	0.00	1.00
Number of children	215	1.84	0.80	0.00	5.00
Number of daughters	215	0.85	0.76	0.00	3.00
Low edu Mother	278	0.58	0.49	0.00	1.00
Middle edu Mother	278	0.29	0.46	0.00	1.00
High edu Mother	278	0.13	0.34	0.00	1.00
Advanced STEM	292	0.24	0.43	0.00	1.00
Degree Laude	256	0.17	0.37	0.00	1.00
Job characteristics					
Full time contract	285	0.92	0.28	0.00	1.00
Years of experience	287	22.94	10.79	3.00	48.00
Math Olympiad	292	0.19	0.39	0.00	1.00
Update Courses	292	0.94	0.24	0.00	1.00
Satisfy with teacher job	287	3.69	0.84	2.00	5.00
Implicit bias					
IAT Gender	301	0.09	0.37	-1.03	1.08
Self-reported explicit bias					
WVS Gender Equality	290	0.17	0.37	0.00	1.00
Gender Dif Innate Ability	280	1.51	0.76	1.00	3.00
Reason GenderGap: Interest for STEM	256	2.58	0.98	1.00	4.00
Reason GenderGap: Predisposition for STEM	241	2.12	1.03	1.00	5.00
Reason GenderGap: Low self-esteem	278	2.64	1.05	1.00	5.00
Reason GenderGap: Family support	278	3.14	1.08	1.00	5.00
Reason GenderGap: Cultural Stereotypes	279	2.15	1.16	1.00	5.00
Boys better in Invalsi	233	0.20	0.40	0.00	1.00
Girls better in Invalsi	233	0.32	0.47	0.00	1.00
Gender Equal in Invalsi	233	0.48	0.50	0.00	1.00
Observations	301				

Notes: First-hand data from teachers' questionnaire. We restrict the sample to teachers matched to students and therefore used in the main analysis of this paper. The balance table with the difference between teachers' matched and not matched with students' data is presented in Table A.2. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

bias against girls and 26% show moderate to severe bias against girls. [27] The sample of 1164 Italians used by Nosek et al. (2009) that decided to take the IAT online in a similar Gender-Science test have an average score of 0.40 (SD 0.40): the score of math teachers is on average lower than this sample (mean 0.09, SD 0.37, as shown in Table 1.1), while Italian teachers are very close to it (mean 0.39, SD 0.39, as shown in Table A.20). [28] Interestingly, the great majority of math teachers are women and this may have important implications for the association of scientific subjects with gender.

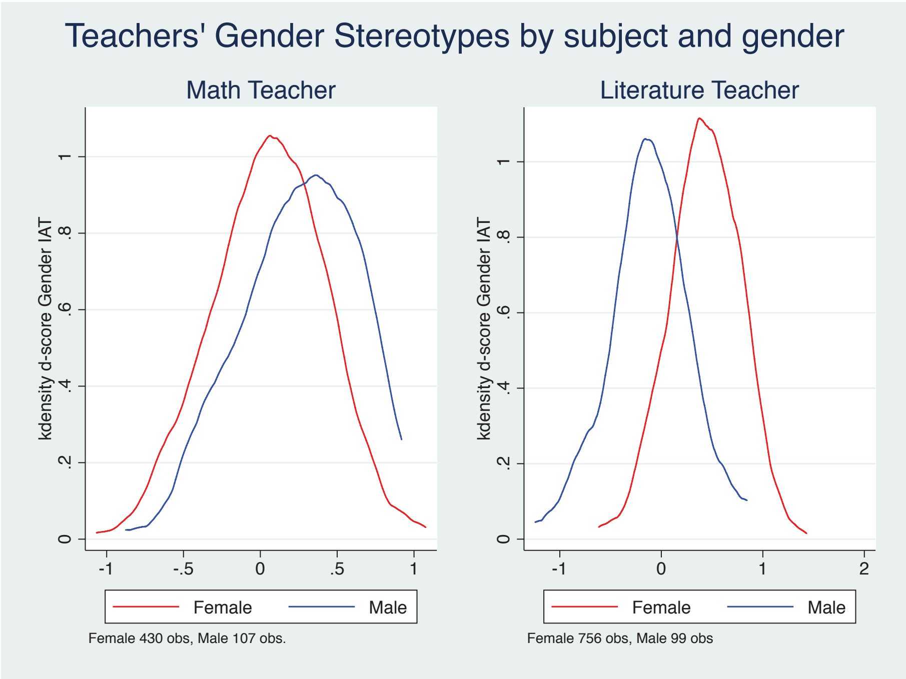


Figure 1.1: Teachers' Implicit Gender Bias (IAT measure) by gender and subject they teach

Notes: This graph shows the distribution of Gender-Science IAT scores for math and literature teachers, separated by gender. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes. The graph provides evidence that teachers in gender-incompatible fields have stereotypes closer to zero.

1.3.2 Students: Self-Stereotypes and Administrative Data

I use individual level information from the Italian Ministry of Education and from INVALSI for three cohort of students enrolled in grade 6 between school year 2010-11 and

[27] Greenwald et al. (2003) suggests that a raw IAT score below -0.15 show bias in favor of the stigmatized group, between -0.15 and 0.15 little to no bias, from 0.15 to 0.35 slight bias against the stigmatized group and a value higher than 0.35 as moderate to severe bias against the stigmatized group.

[28] In the paper by Nosek et al. (2009), individuals completed the IAT online in the *Implicit Project* website.

2012-13. [29] The data available include math and reading standardized test score in grade 6 and 8, parents' education and occupation, baseline individual information (date and place of birth, gender, citizenship), high-school track choice and official teachers' recommendation. Students in grade 8 in 2014 of 24 schools in this sample, around two months before the end of middle schools, are asked to complete a survey about their track choice. In particular, they need to mention all subjects they will learn during high-school and to report their belief about their own ability in each subject. The potential choices to that answer were: "good", "mediocre", "scarce".[30]

Table 1.2 reports summary statistics on students' information. I restrict the sample to students with information available on the standardized test score in grade 6 and 8 and for whom I have the implicit association test of their math teacher in grade 6. This is the sample that will be used in the empirical analysis of this paper. Appendix A.4 describes in details the sample selection and potential attrition issues. In our sample, 50% of students are males and boys and girls are balanced in terms of baseline characteristics related to place of birth, generation of immigration, parents' education and occupation. Test scores are standardized to have mean zero and standard deviation one per subject and year in which the test was taken. Females at the beginning of middle school are lagging behind of 0.19 standard deviations in math and ahead of 0.13 standard deviations in reading, with respect to males. In the same table, I also report the raw gender differences in outcomes. The high-school track choice in this sample is comparable to the average national choices in those years: females are almost 10 percentage points less likely to choose an academic scientific track and almost 25 percentage points less likely to enroll in a technical technological track. Girls are more likely to choose an academic track than boys, but not a top-tier academic track (which include classical and scientific tracks). Indeed, one third of females choose a social, linguistic or artistic academic tracks. Vocational school is chosen at an equal rate by both genders. However, teachers recommend 36% of males toward vocational track and 30% of females, while the scientific track is recommended only to 16% of males and 11% of females.[31] Finally, from the original information available for a sample of students, I observe that on average there are no gender differences in assessment of ability, but females are 9 percentage points less likely than boys to consider themselves good at math and boys are 5 percentage points less likely to consider themselves good at Italian compared to girls.

[29] Individual level data are anonymous and I obtained the authorization from each school principal to access data from their school.

[30] These information were collected by Carlana et al. (2017) in a random sample of 47 control schools to evaluate soft skills of students. The specific question exploited in this paper was not used in the paper by Carlana et al. (2017).

[31] In some schools, more than one recommendation is given to students. Here, I report summary statistics only for the first recommendation.

Table 1.2: Summary Statistics of students by gender

	Males	Females	Diff.	se
Baseline characteristics				
Std Math grade 6	0.233	0.038	0.195***	(0.020)
Std Ita grade 6	0.085	0.218	-0.133***	(0.019)
Born in the North	0.849	0.854	-0.005	(0.007)
Born in the Center/South	0.027	0.030	-0.003	(0.003)
Immigrant	0.189	0.173	0.016	(0.008)
Second Gen. Immigrant	0.080	0.074	0.006	(0.006)
HighEduMother	0.456	0.453	0.003	(0.010)
Missing Edu Mother	0.212	0.211	0.002	(0.008)
High Occupation Father	0.169	0.174	-0.005	(0.008)
Medium Occupation Father	0.321	0.303	0.017	(0.010)
Missing Occupation Father	0.206	0.214	-0.008	(0.008)
Outcomes				
Std Math grade 8	0.194	-0.021	0.214***	(0.020)
Std Ita grade 8	-0.006	0.176	-0.182***	(0.020)
High-school Track: Scientific	0.304	0.208	0.096***	(0.010)
High-school Track: Classic	0.043	0.079	-0.036***	(0.005)
High-school Track: Other Academic	0.097	0.336	-0.239***	(0.009)
High-school Track: Technical Technological	0.311	0.067	0.244***	(0.008)
High-school Track: Technical Economic	0.113	0.163	-0.050***	(0.008)
High-school Track: Vocational	0.132	0.148	-0.015*	(0.008)
Track recommendation: Scientific	0.164	0.110	0.054***	(0.008)
Track recommendation: Vocational	0.362	0.298	0.064***	(0.011)
Own ability: all subjects	0.656	0.646	0.010	(0.012)
Own ability: math	0.833	0.747	0.087**	(0.030)
Own ability: Italian	0.917	0.968	-0.051**	(0.018)
Observations	4698	4611		

Notes: This table reports the summary statistics and the difference between the two genders in outcomes and baseline characteristics. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

1.4 Empirical Strategy

1.4.1 Estimating Equation

The main purpose of this paper is to investigate the impact of teachers' gender stereotypes on student outcomes. I exploit two identification strategies. The first is aimed at investigating the impact of teacher bias on the gender gap within a class, estimating the following equation:

$$\begin{aligned} y_{ic} = \alpha_0 + \alpha_1(Female_i \times bias_c) + \alpha_2 Female_i + \eta_c + \\ + \mathbf{X}_i\rho_1 + (Female_i \times \mathbf{X}_i)\rho_2 + (Female_i \times \mathbf{Z}_c)\rho_4 + \epsilon_{ic} \end{aligned} \tag{1.1}$$

where y_{ic} is the outcome (math standardized test score, track choice, and self-confidence) of student i in class c taught in grade 8 by teacher with stereotype level $bias_c$. $Female_i$ is a dummy variable which assumes value 1 if the student i is a girl and $bias_c$ is the standardized value of the gender implicit bias of the teacher assigned to class c in grade 8.[32] I include fixed effects at class level η_c, which absorb the average effect of teacher bias in class c. Furthermore, for robustness, I include student characteristics $\mathbf{X}_i$ (parents education and occupation, immigration status and generation of immigration), and teacher characteristics $\mathbf{Z}_c$ (as gender, place of birth, age, teacher "quality"[33], type of contract, type of degree achieved and self-reported gender bias) interacted with the gender of student i. Standard errors are clustered at teacher level.

Crucially, in this identification strategy, class, teacher, and school level characteristics are absorbed by class fixed effects. Indeed, as described in Section 1.2, students are assigned to a class in grade 6 and attend all lectures with the same classmates until grade 8. We can only identify the impact of teacher bias on the gender gap in the dependent variable, i.e. the interaction between the gender of students and implicit stereotypes of teachers. The coefficient of interest, α_1, measures how the gender gap in the class changes when assigned to teachers with one standard deviation higher bias.[34] I expect the estimate of α_1 to be attenuated for the measurement error in the gender IAT score. Indeed, occasion-specific noise may introduce an attenuation bias, as suggested by Glover et al. (2017).[35] For robustness, I include controls for student characteristics $\mathbf{X}_i$ interacted with

[32] On average in 70% of the cases professors have been teaching to the same class from grade 6 to grade 8, in 11% of the cases from grade 7 and in 19% only for grade 8. Teachers are teaching on average in three math classes per year. For simplicity, I omit the subscript referring to teachers in equation 1.1. However, two different classes can be assigned to the same teacher.

[33] Teacher "quality" is proxied by being the teacher in charge of math Olympics in the school, updarefresherte courses and other observables. Appendix Table A.4 shows that being the teacher in charge of math Olympics in the school is correlated with the value added, especially for females.

[34] I discuss the exogeneity of student assignment to teachers in Section 1.4.3.

[35] Glover et al. (2017), while analyzing the impact on manager implicit bias on minority workers, suggest

the gender of the pupil. The regression also controls for the gender of students interacted with teacher characteristics $\mathbf{Z}_c$. This is potentially important to partial out differential impact by gender of sex, background, and experiences of teachers. Furthermore, this allows to establish whether the impact of teacher stereotypes on gender gap among classmates can be explained (or attenuated) by teachers' observables, as clarified in Section 1.4.2.

The second identification strategy relies on the comparison of students of the same gender enrolled in the same school, but assigned to teachers with different bias level. I investigate whether the impact of teacher stereotypes on gender gap is due to higher performance of boys, lower performance of girls or a combination of both. I estimate the following equation:

$$\begin{aligned} y_{icsy} = \beta_0 + \beta_1(Female_i \times bias_c) + \beta_2 Female_i + \beta_3 bias_c + \eta_{sy} + \\ + \mathbf{X}_i\rho_1 + (Female_i \times \mathbf{X}_i)\rho_2 + \mathbf{Z}_c\rho_3 + (Female_i \times \mathbf{Z}_c)\rho_4 + \epsilon_{icsy} \end{aligned} \tag{1.2}$$

where η_{sy} are school s by cohort y fixed effects and standard errors are clustered at teacher level. All other variables are defined as in equation (1.1).

Institution level characteristics are absorbed by school by cohort fixed effects. The advantage with respect to specification (1.1) is that we can analyze the impact of teacher stereotypes separately on male students (β_3) and on female students ($\beta_1 + \beta_3$). The drawback is that we cannot control for unobservable characteristics at the teacher or class level: this specification exploits variation in the level of teacher bias to which students of the same gender in the same school and cohort are exposed.

1.4.2 Correlation between implicit bias and individual characteristics

In this Section, I present evidence on the correlation between observable characteristics of teachers and IAT scores. This has important implications for the estimation. When teachers controls are not included, we need to consider teacher stereotype as including also characteristics correlated with teacher bias. This point will be carefully stressed analyzing the results.

Figure 1.1 plots the entire distribution of implicit bias for math and Italian teachers by gender: interestingly, individuals teaching a subject which is stereotypically associated with their gender (i.e. males teaching math and females teaching Italian) are more gender biased. Teachers are more likely to associate own gender with the subject they teach.

that we may expect an attenuation bias of approximately a factor of 1.8 due to measurement error in the IAT score.

This result is coherent with findings of Rudman et al. (2001) according to which individuals possess implicit gender stereotypes in self-favorable form because of the tendency to associate self with desirable traits.

The richness of the data collected allows me to associate individual level characteristics of teachers with the results from the Implicit Association Test (IAT) in order to dig deeper into the determinants captured by reaction time to stimuli. Table 1.3 shows the correlation between math teacher IAT score and their characteristics. Women teaching math are significantly less biased in associating gender with STEM and this explains a substantial portion of the low average IAT score for math compared to Italian teacher. In columns 2-5 (Panel A), I show the association with age, education of teachers' mother, and whether teachers have children. Among this group of comparable adults, implicit stereotypes is not affected by age. Teachers with mothers that graduated from high-school seems to be slightly less biased, even if the effect is imprecisely estimated. Finally, having children, and in particular daughters, do not significantly impact on gender stereotypes.[36]

Gender stereotypical beliefs are rooted in cultural traits, transmitted from generation to generation (Guiso et al., 2006). Indeed, I find that exposure to cultural norms is strongly associated with the IAT score. In column 1 of Table 1.3 (Panel B), I correlate the implicit bias with the place of birth of teachers. Around 35 percent of math teachers in this sample are born in the South where gender norms are stronger, as shown for instance by Campa et al. (2010) using World Value Survey data at Italian provincial level.[37] I further investigate how implicit associations are correlated with individual level beliefs and cultural norms in the place of birth. As shown in column 2 of Panel B, women labor force participation in the province of origin of teachers is negatively correlated with the IAT score.[38] In column 3 and 4 of Panel B, I use, as proxy of gender cultural stereotypes in the province of birth, the answers to the World Value Survey question on the relative rights of men and women to paid jobs when the latter are scarce. I find a positive correlation between lower gender stereotypes measured by this question and IAT scores. During the survey I administered, I asked the same question to teachers themselves and I find a low and indistinguishable from zero correlation. We may suspect that there is a social desirability bias in the self-reported measure when professors are interviewed in the school. In column 5 of Panel B, I correlate implicit bias and explicit beliefs about

[36]I also check whether the Gender-Science IAT score is correlated with the race IAT score. In the same regression as in Table 1.3, I find that the correlation is -0.068 (standard error 0.123). Hence, math teachers more biased in one sphere are not more biased also in the other sphere. The IAT score does not seem to capture a general "ability" in doing this type of test for math teachers.

[37]Italy is a country with low labor market participation of women, but substantial geographic variation across regions. In 2016, only 31 percent of women in the South of Italy were employed, while in the North around 58 percent were working, similarly to the average of OECD.

[38]The correlation between labor force participation of women and geographical regions is indeed extremely strong in Italy.

innate differences in ability between men and women and I find a weak positive correlation (not statistically significant). This result is not surprising in light of social psychology literature, where implicit often differ from explicit and self-reported stereotypes (Lane et al., 2007; Nosek et al., 2002).

In Panel C, columns 1 and 2, I correlate the IAT score with qualifications of the teacher (type of degree and whether the degree was achieved with honor), finding negative point estimates despite high standard errors. Another rough proxy of potential quality of teachers is related to having tenure (which is associated with higher experience in teaching), and being the professor in charge of math Olympiads in the school.[39] Also in these cases, point estimates are small and indistinguishable from zero.

Appendix Table A.3 shows jointly all correlation presented in separate regressions in Table 1.3. Interestingly, the results are substantially invariant: gender and place of birth of teachers are the two most relevant aspects in affecting IAT scores in all specifications.

1.4.3 Exogeneity Assumption

Next, I present evidence regarding the absence of a systematic correlation between gender bias of teachers and student characteristics and the absence of systematic grouping of students by socio-economic background and initial ability.

If parents are able to guess who the teacher is with higher stereotyping behaviour, they may try to place their daughter in a different class. Although this seems unlikely, it is also possible that they try to select teachers according to observables which are correlated with IAT score, as gender, place of birth of teachers, years of experience and tenure.[40] In Table 1.4, I provide evidence that student characteristics are not systematically correlated with the implicit bias of teachers. I would not be able to obtain causal estimates if teachers with higher gender bias are systematically more/less likely to be assigned to females or to females with specific characteristics in terms of parents' education and occupation, place of birth and ability. I might expect that if parents had control over assignment of their children to teachers, daughters of highly educated mothers would have been less likely to be assigned to more biased teachers, within school. Instead, I see that the difference is not statistically significant and the point estimate goes in the opposite direction. In columns 3, 4, 5 and 6, I analyze the correlation respectively with father occupation,

[39] In each school, usually only one professor is in charge of math Olympiad and anecdotally she is highly motivated and passionate teacher. Indeed, as shown in Appendix Table A.4, teachers in charge of math Olympics induce higher improvements in test scores of their students.

[40] Anecdotally, parents dislike being assigned to a teacher with a temporary contract that may change during the middle school years and has little experience. This paper focuses on variation of exposure to a sample of teachers that has been teaching in the same school since at least 2014. They have a lot of experience (on average 23 years) and almost all have a full-time contract. Almost all teachers included in my analysis have tenure. Hence, among these teachers, the selection on experience is unlikely.

Table 1.3: Correlation between teachers' characteristics and Gender IAT Score

	Panel A: Independent variables (background teachers' characteristics)				
	Female (1)	Age (2)	HighMotherEdu (3)	Children (4)	Daughters (5)
Dep. Var.: **Raw IAT** **score**	-0.188** (0.083)	0.016 (0.060)	-0.053 (0.060)	0.069 (0.145)	0.047 (0.075)
Obs.	301	301	301	301	301
R^2	0.347	0.327	0.337	0.330	0.331
	Panel B: Independent variables (cultural traits and beliefs)				
	BornNorth (1)	WomenLFP (2)	WVSCityBorn (3)	WVSIndiv (4)	InnateAbility (5)
Dep. Var.: **Raw IAT**	-0.154** (0.064)	-0.499** (0.247)	0.399* (0.211)	0.007 (0.086)	0.016 (0.041)
Obs.	301	286	261	301	301
R^2	0.348	0.361	0.399	0.325	0.328
	Panel C: Independent variables (education and teacher experience)				
	Ad.STEM (1)	Laude (2)	FullContract (3)	Olympiad (4)	JobSatisfy (5)
Dep. Var.: **Raw IAT** **score**	-0.092 (0.076)	-0.034 (0.075)	-0.049 (0.153)	0.059 (0.087)	0.054* (0.032)
Obs.	301	301	301	301	301
R^2	0.332	0.326	0.327	0.311	0.336
School FE	Yes	Yes	Yes	Yes	Yes

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and own teacher characteristics; the unit of observation is teacher t in school s. Standard errors are robust and clustered at school level in parentheses; the number of clusters is 90. School fixed effects are included in all regressions. The significance and magnitude of coefficients are not significantly impacted by the inclusion of FE. The variable "Female" indicates the gender of the teacher, "Born in the North " assumes value 1 if the teacher was born in the North of Italy, "HighMotherEdu" is a dummy which assumes value 1 is the mother of the teacher has at least a diploma,"Children" and "Daughters" are dummies which assumes value 1 if the teacher has children/daughters. The variable "Ad.STEM" assumes value 1 if the teacher has a degree in math, engineering and physics, "Laude" is a dummy which assumes value 1 if the degree was achieved with laude, "Full Contract" assumes value 1 is the teacher has tenure, "Olympiad" is 1 for teachers in charge of math Olympiad in the school, "JobSatisfy" is a categorical variable from 1 to 5 which captures self-reported job satisfaction of teachers, "Updates" captures whether teachers followed update courses in teaching during the academic year, "WomenLFP" is the labor force participation of women in the province of birth, "WVSCityBorn" is the WVS answer to the relative rights of men and women to paid jobs when the latter are scarce, "WVSIndiv" is the answer to the same question at individual level, "InnateAbility" regards the teacher belief about innate differences in math abilities between men and women,"ExplicitBias" is an index that summarizes explicit gender bias of teachers. We include the order of IATs for math teachers (if the first one was the gender IAT and if the first associations were order compatible or not) and missing categories if the information is not available. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

immigration background and for the proxy of ability using standardized test scores in reading in grade 6 and I do not find statistically significant correlation. Furthermore, the point estimates are small in magnitude as well. Finally, in the last column, I also include the standardized test score in math in grade 5, before entering middle school, despite the sample size is substantially reduced for data availability issues.[41] The assumption of quasi-random assignment of students in the sample to teachers with different level of gender bias, as measured by the Implicit Association Test, within a school, seems to be supported in the context under analysis.

The result is identical when observations are collapsed at teacher level, as shown in Appendix Table A.5. I also verify that teachers with higher bias are not systematically associated with fewer females in the top of the distribution. I find that this is not the case and, if anything, the sign of the correlation goes in the opposite direction. The results considering the share of female students in the top 10, 20 and 40 percent of the distribution in the standardized test score in grade 6 are shown in Appendix Table A.6.

Furthermore, even if some parents manage to allocate their children to teacher with higher "quality", it does not necessarily mean that they are less gender biased. For instance, the teacher in charge of math olympics in the school is usually considered as one of the best math teacher. It seems reasonable since, as shown in Appendix Table A.4 for the sample under analysis, his or her students improve the most their math performance in terms of value added, especially females. However, if anything, teachers in charge of math olympics have slightly more gender biased than others, as measured by IAT score (Table 1.3).

Finally, principals must assigned all teachers to a class since they have an exact number of teacher. Hence, they cannot avoid assigning a teacher to a class, even if he or she can guess who is the teacher with higher stereotypes.[42]

The second aspect regards the absence of systematic grouping of students by socio-economic background and initial ability. Within schools, classes are formed by the principal with the main objective of creating comparable groups in terms of gender, ability and socio-economic background across classes and therefore to guarantee heterogeneity within each class. This objective is spelled out in the official documents on the school websites of most schools and also emerges from self-reported information from principals discussed in Appendix A.2. It is important to stress that middle school teachers do not teach in elementary school as well. I have information about the observable characteristics of students

[41] I required standardized test score in math in grade 5. Unfortunately, for reasons related to confidentiality, I have obtained them only for those students that did not change school code between elementary and middle school. There are only few students for which I have this information.

[42] Principals do not have more math teachers available than classes in the school. Since each teacher with a full-time contract teaches three classes, teachers can be assigned to more than one school to cover all their required hours.

that are used to create classes (gender, education and occupation of parents, immigration status and generation of immigration). Plausibly, unobservable student characteristics are also unknown to school principals at the moment of class formation. I check whether class assignments are statistically independent with a series of Pearson Chi-Square tests (Lavy and Sand, 2015). First, I consider the assignment of individual level characteristics (gender, education and occupation of parents, immigration status and generation of immigration). Then, I also check that within each characteristic, class assignment is statistically independent from gender. I find that in less than 7.8% of the tests performed, the p-value is lower or equal than 5%[43]. This implies that for only 7.8% of the classes we cannot reject that there is non-random assignment of one characteristic of students. Hence, there is no strong evidence of systematic non-random formation of classrooms with respect to students' characteristics.

1.4.4 Reverse Causality

The measure of teacher gender stereotypes was collected between October 2016 and March 2017. Teacher data are matched with students who graduated from middle school between June 2013 and June 2015, as clarified in Figure 2.2. Similarly to the study of Glover et al. (2017), teacher bias is collected after students in the sample graduated from middle school and therefore after outcomes are realized. The main potential concern is that IAT scores are affected by exposure to students. Indeed, the IAT is expected to be the combination of two aspects: the former is a trait stable over time capturing the influence of cultural norms and experience, while the latter is occasion-specific variation and noise that may be affected by conditions while taking the test and stimuli received by the subject in the period right before the test.[44] Our potential concern given the timeline of the analysis is that the three cohort of pupils affect the stable trait of teachers' gender bias.

[43]Given the size of the Table, it is not reported in the paper but it is available upon request to the author.

[44]The test-retest reliability of IAT is generally considered as satisfactory by social psychology, with a correlation of 0.56 that does not change with the length of time between testing (despite being usually of less than one month in most studies) (Nosek et al., 2007).

Table 1.4: Exogeneity of assignment of students to math teachers with different bias

Dependent Variable: Math Teacher implicit gender bias (standardized)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fem	0.007	-0.011	0.004	0.015	0.008	-0.018	0.220
	(0.013)	(0.022)	(0.025)	(0.016)	(0.013)	(0.132)	(0.239)
Fem*HighEduMother		0.036				0.044	-0.002
		(0.034)				(0.031)	(0.045)
HighEduMother		0.018				0.005	-0.009
		(0.027)				(0.025)	(0.029)
Medium Occupation Father			0.013			0.007	0.038
			(0.024)			(0.022)	(0.035)
Fem*Medium Occupation Father			0.020			0.008	0.076
			(0.036)			(0.033)	(0.060)
High Occupation Father			0.015			0.018	0.005
			(0.032)			(0.027)	(0.041)
Fem*High Occupation Father			0.006			-0.012	-0.032
			(0.041)			(0.038)	(0.059)
Fem*Immigrant				-0.035		0.005	0.097
				(0.038)		(0.040)	(0.076)
Immigrant				0.059**		0.049*	0.045
				(0.029)		(0.029)	(0.056)
Fem* Std Ita grade 6					0.005	-0.005	-0.005
					(0.015)	(0.015)	(0.026)
Std Ita grade 6					-0.009	-0.009	-0.016
					(0.013)	(0.013)	(0.017)
Fem*Std Mat grade 5							-0.002
							(0.025)
Std Mat grade 5							-0.005
							(0.016)
School,year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Control	No	No	No	No	No	Yes	Yes
Obs.	9309	9309	9309	9309	9280	9280	1649
R^2	0.412	0.412	0.412	0.412	0.419	0.489	0.723

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and students' characteristics; 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 math teacher level in parentheses; the number of clusters is 301 in columns 1-6 and 131 in column 7. The variable "Fem" indicates the gender of the student, "HighEduMother" assumes value 1 if the mother has at least a 5 years diploma, "Medium Occupation Father" assumes value 1 if the father is a teacher or office worker, while "High Occupation Father" is 1 if the father is manager, university professor or an executive. "Immigrant" assumes value 1 is the student is not an Italian citizen, while "Std Mat grade 5" and "Std Ita grade 6" are the standardized test score in grade 5 in math and grade 6 in Italian respectively. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for the order of IAT in the questionnaire administered. For 29 students we do not observe the test score in Italian in grade 6. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

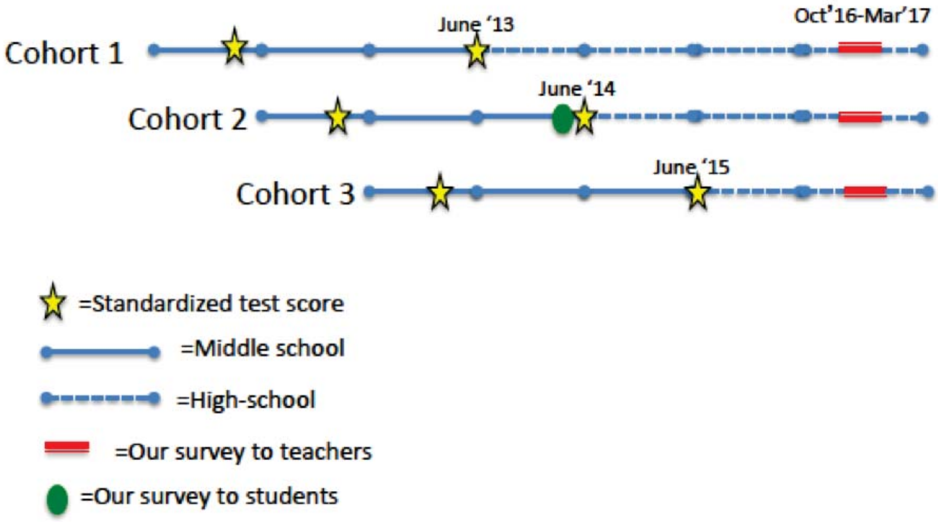


Figure 1.2: Timeline of main data available for students and teachers

Notes: This graph shows the timeline of data collected for the three cohorts of students. They graduated from middle school between 2013 and 2015. Teachers were surveyed between October 2016 and March 2017. Standardized test scores are administered at the end of grade 6 and 8.

Reverse causality seems unlikely for several reasons. First, as shown in Section 1.4.3, teachers with higher bias were not assigned to a differential treatment in terms of student characteristics. I control for student family background, ability in math and reading as measured by standardized test score (see Tables 1.4 and Appendix Table A.5) and share of females in the top of the math ability distribution (see Appendix Table A.6). Second, under the assumption of monotonic decay of the influence of students to teachers, I would expect a higher effect for the most recent cohort of student. However, results are stable in all three cohorts, as shown in the robustness analysis (Appendix Table A.8). Third, math teachers included in our analysis have been teaching on average for 23 years (with a median of 25 years) and therefore over time they were exposed to hundreds of females and males students. Furthermore, for data availability issues, we do not include in the sample the cohort of student graduating right before the school year in which the test was administered. Each teacher has been exposed on average to 4 classes (around 100 students) after those included in our analysis.[45]

In fact, there is a main advantage from exploiting this timing choice: taking the IAT

[45]Students who were enrolled in middle school in the school year 2015-2016 and 2016-2017 are not included in the sample. Usually, math teachers teach three classes per year (one in grade 6, one in grade 7 and one in grade 8). Hence, teachers are exposed to around 4 different classes and therefore around 100 students after the last cohort of students I analyze and before taking the IAT (i.e. the class in grade 8 in 2015-16 and classes in grade 6, 7, and 8 in 2016-17.

or knowledge about this study could not have affected students' performance nor teachers' or parents' attention to the issue of gender stereotypes for cohorts of boys and girls in this dataset.

1.5 The Impact of Teachers' Implicit Bias

In this section, I present the main results of the paper. I focus on the impact of teacher bias on student performance as measured by the standardized test scores in math (Section 1.5.1) and high school track choice (Section 1.5.2). Finally, I present some robustness checks and additional outcomes in Section 1.5.3 before analyzing the mechanisms behind the treatment (self-stereotypes and pupil-teacher interaction).

1.5.1 Performance in math

By the age of 14, girls are lagging behind in math compared to their male classmates by around 0.22 standard deviations, a result comparable to several other countries (Fryer Jr and Levitt, 2010; Bharadwaj et al., 2016).[46] As children complete more years of education, the differences between boys and girls gets bigger. The additional gender gap in math generated during the last two years of middle school is around 0.08 standard deviations, as shown in column 2 of Table 1.5. This paper analyzes what happens to the gender gap when students are quasi-randomly assigned to biased teachers.

Before moving to the causal estimates, Appendix Figure A.3 plots the relationship between teacher bias and math performance of male and female students. Each circle plots the average improvement in math test scores of students assigned to a math teacher with the indicated level of bias, aggregated into bins. The size of the circle indicates the number of observations per bin. These graphs plot the raw data, without removing fixed effect at class or school level. Nonetheless this figure tells a similar story compare to our regression analysis: female students are lagging behind when assigned to math teachers with higher implicit bias.

Table 1.5 shows the effect of teacher bias on gender gap in math performance within the class, presenting the results of estimating equation (1.1). Classes that are assigned to teachers with one standard deviation higher bias have 0.027 standard deviations higher gender gap in math performance. Considering an average gap of 0.08 standard deviations,

[46]In Appendix Figure A.2, I show the average gap in PISA test scores across countries. According to a meta-analysis performed on 100 studies in several countries, gender gaps in mathematics are around 0.29 standard deviations in high-school (Hyde et al. (1990), two years after the end of middle school. The average gender gap without controlling for class fixed effects is substantially invariant (0.21 standard deviations as shown in Table 1.2). Most of the variation in math performance is within classes, coherently with the target in class formation of heterogeneity within class and homogeneity across classes.

it corresponds to an increase of 34% of the gender difference in performance generated during middle school. Column 4 includes student characteristics $\mathbf{X}_i$ and their interaction with gender of the children. Adding these controls does not change the coefficient of interest.

Although the level of teacher bias and all characteristics are absorbed by the class fixed effect, as clarified describing equation (1.1), column 5 includes the interaction between student gender and teacher characteristics $\mathbf{Z}_{tc}$. If anything, the coefficient of interest *"Fem*Bias Teacher"* slightly increases in magnitude when all these interaction effects are absorbed. Observable characteristics of teachers, interacted with students' gender, are not driving the relation between gender gap and teacher bias. I report the coefficients only for the main characteristics of teachers interacted with students' gender, but the effects are mainly small and insignificant for all variables, including age, parents' education, whether he or she has children or daughters, whether he or she achieved the degree with *laude*, the type of teaching contract, refresher courses and appointment as teacher in charge of math Olympics. The latter controls are crude proxies for teacher quality in terms of improvements in standardized test scores, as shown analyzing the relation between value added and teachers observables the Appendix Table A.4. Finally, in the Appendix Table A.7, I split the sample among student of male and female teachers. Although the effect is not statistically significant for pupils assigned to male teachers due to a small sample size, the point estimates show that the impact of teachers' implicit bias on students achievement is comparable for male and female teachers.

As it can be seen in column 5 of Table 1.5, ceteris paribus, female students assigned to female teachers or to teachers with an advanced STEM degree have slightly lower, albeit insignificantly so, math achievement test scores in grade 8 compared to their classmates[47]. The impact of teacher gender is coherent with the result of Bharadwaj et al. (2016). However, other studies find that having a teacher of own gender helps improve performance, especially at college level (Dee, 2005; Carrell et al., 2010). Finally, teachers born in the North of the country do not have an heterogeneous effect on boys and girls. The results are robust to potential confounding aspects considering all information available on professors from their family background to their professional career.

To give a clearer interpretation, Figure 1.3 reports the same estimates using a categorical variable instead of the continuous one. I consider the thresholds defined by Greenwald et al. (2003), where no bias is the interval of IAT raw score between -0.15 and +0.15 and "pro boys"("pro girls") assumes value 1 when implicit bias is higher than 0.15 (lower than -0.15). Being assigned to a teacher with a "pro boys" attitude (45% of teachers) in STEM

[47]It should be noticed, however, that most of teachers in Italian middle schools are females, also in math. There is little variation on the gender of teachers.

Table 1.5: Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

Dependent Variable: Math standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	-0.222***	-0.078***	-0.080***	-0.036	-0.024
	(0.019)	(0.014)	(0.014)	(0.032)	(0.103)
Fem*Teacher Bias			-0.027**	-0.028**	-0.037***
			(0.013)	(0.013)	(0.014)
Fem*Teacher Fem					-0.056
					(0.037)
Fem*North Math Teacher					0.008
					(0.030)
Fem*Advanced STEM Teacher					-0.041
					(0.031)
Std Math grade 6		0.723***	0.723***	0.697***	0.699***
		(0.012)	(0.012)	(0.013)	(0.013)
Constant	0.198***	0.028***	0.028***	-0.112***	-0.112***
	(0.009)	(0.007)	(0.007)	(0.023)	(0.023)
Gender Gap	-0.222	-0.078	-0.078	-0.082	-0.082
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9309	9309	9309	9309	9309
R^2	0.209	0.618	0.618	0.625	0.625

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is math standardized test score in grade 8; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

compared to a teacher with a "pro girls" attitude (24% of teachers) leads to triple the gender gap in math improvements within the class (from -0.035 standard deviations to -0.10 standard deviations). The same results are reported in Appendix Table A.9, considering the thresholds defined by Greenwald et al. (2003) and also whether IAT score is positive or negative. As in Table 1.5.1, the effect is stronger when controlling for student and teacher characteristics interacted with pupil gender. In columns 4-6, we consider whether IAT score has a positive or negative sign finding similar results.

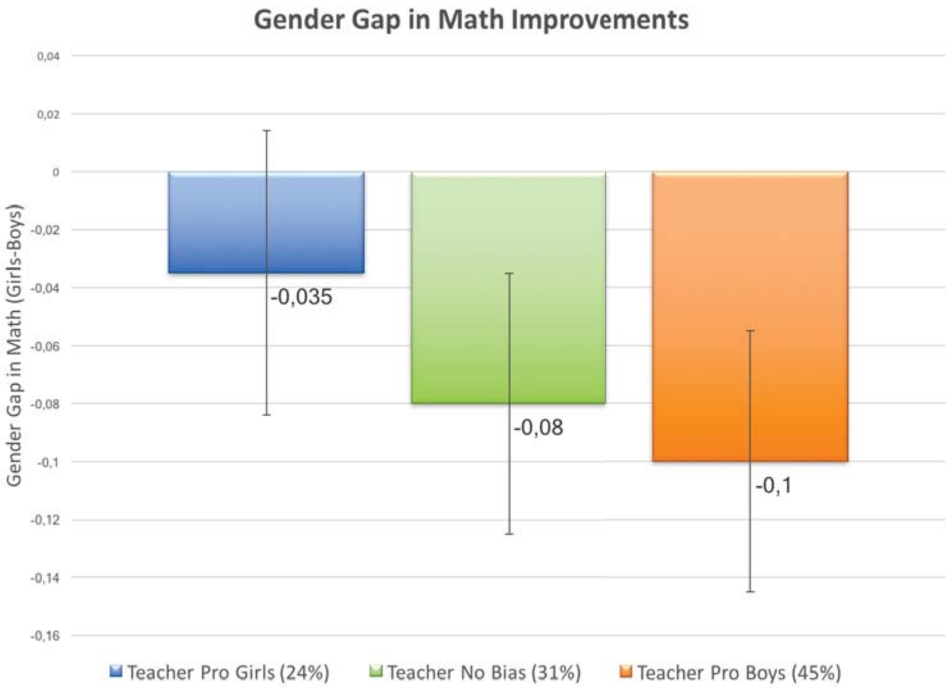


Figure 1.3: Effect of teacher bias on student math performance

Notes: This graph shows the effect of teacher stereotypes on student achievement. We consider the thresholds defined by Greenwald et al. (2003) where no bias is the interval of IAT raw score between -0.15 and +0.15. The attitude of the teacher in associating fields with gender is considered "pro girls" if the score is lower than -0.15 (24% of teachers) and "pro boys" if the score is higher than +0.15 (45% of teachers). The variable in the y axis is the gender gap in improvements in math between grade 6 and 8, when class fixed effects are absorbed.

Are biased teachers worse instructors or are they helping boys to learn math? I next investigate the effect of teacher bias from estimating directly equation (1.2), comparing students of the same gender within the same school and cohort, but assigned to different classes. Figure 1.4 shows that having a teacher with strong gender stereotypes has a negative impact on female students, while a bias in favour of girls has a positive impact in their math improvements. The linear approximation presented in this paper seems appropriate. There is no statistically significant impact on male students, throughout the whole distribution of teacher bias. Table 1.6 mirrors Figure 1.4: it presents the results of the regression analysis and shows that girls are lagging behind when assigned to more bias teacher, while boys are not affected by teacher stereotypes. The results are robust to the inclusion of the same controls as in Table 1.5. In this specification the characteristics of teachers are not absorbed by class fixed effects and therefore controls at teacher level, included in columns 5, are particularly relevant. Furthermore, controls for the amount of math hours per week are included in this specification and interacted with the student gender. Indeed, in almost all schools some classes have an extended school day and they spend more time with all teachers, including the math one. Adding all these controls does not significantly impact on the main results. A potential interpretation of this finding is that it is sufficient to be exposed for six hours per week for one school year for the effect to kick in.

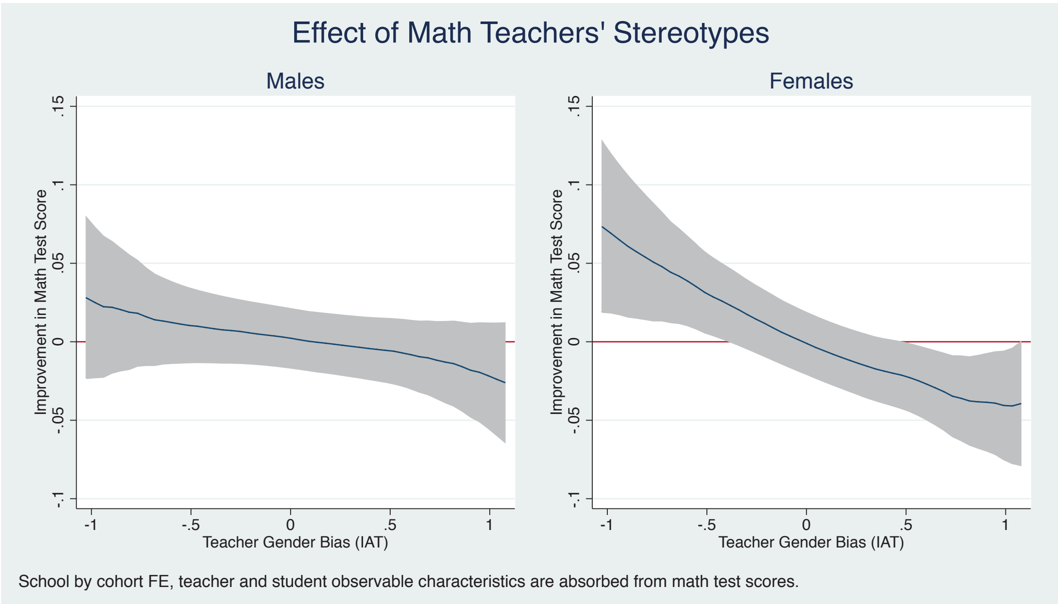


Figure 1.4: Effect of teacher bias on student math performance by gender

Notes: This graph shows the effect of teacher stereotypes on student achievement by gender. The variable in the y axis is the residualized standardized test score in grade 8, after controlling for school by cohort fixed effects, student and teacher level controls. The variable in the x axis is the raw IAT score. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females.

The differential response by gender is consistent with the previous results in the economic literature: females are negatively affected by teachers of male-typed domains, as math (Kugler et al., 2017). Coffman (2014) finds that individuals are significantly less likely to contribute with their ideas in gender incongruent fields and this is particularly strong for women, leading to more missed opportunities among female in male-typed categories than for males in female-typed categories. Furthermore, the type of task affects gender differences in the willingness to complete, with wider gaps in stereotypically male tasks (Niederle and Vesterlund, 2010; Große and Riener, 2010).

Heterogeneous effects

We now examine which students are most affected by teacher bias. Table 1.7 shows that the effect of implicit stereotypes is stronger for the most disadvantaged groups of female students, in term of background characteristics. Based on the estimates in column 2, a standard deviation increase in teacher bias leads to an increase of the gender gap of 0.049 standard deviations among students with low educated mothers and of 0.027 standard deviations among students with higher level of mother education (at least a diploma), although the difference is indistinguishable from zero at usual levels. In the following column, I analyze the impact of teacher bias in the three terciles of the distribution of the standardized test score in grade 6. The effect is stronger for students in the lowest tercile (-0.070, with standard error 0.027) and turning positive, but not statistically significant, only for students in the top of the initial ability distribution in grade 6. Finally, the effect if anything is slightly stronger among immigrants, even if the difference with natives is not statistically significant at usual levels.

Why do girls from more disadvantaged backgrounds suffer the most from the interaction with biased teachers? The empirical evidence presented is coherent with stereotype threat model (Steele and Aronson, 1995): individuals with higher risk of conforming to the predicament that *"women are bad at math"* are those more deeply affected. Indeed, male students are not influenced by teacher stereotypes and among females those strongly affected are from disadvantaged backgrounds, especially in terms of initial math achievements who are at higher risk of confirming the negative expectations on their group. Appendix A.7 presents a conceptual framework that illustrates how teacher stereotypes can differentially affect effort and outcomes of students in the bottom and the top of the ability distribution.[48] One complementary explanation, coherent with the interaction theory (McConnell and Leibold, 2001), is that female students with highly educated mothers or with higher initial level of math achievement may need less interaction with their math teacher in order to avoid lagging behind with their peers. They are more likely to have both additional support to believe in their own abilities and alternative role models.

In order to investigate further the second potential explanation, I analyze the heterogeneous effect according to the "quantity" of interaction time between teacher and students. The last two columns of Table 1.7 analyze whether there are heterogeneous effects in terms of years of exposure and hours per week. Indeed, around 75% of students

[48] This conceptual framework is an extension of the stereotype threat model presented by Dee (2014).

Table 1.6: Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - school FE regression

Dependent Variable: Math standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	-0.234***	-0.092***	-0.093***	-0.034	-0.020
	(0.022)	(0.015)	(0.015)	(0.033)	(0.107)
Fem*Teacher Bias			-0.022*	-0.024*	-0.032**
			(0.013)	(0.013)	(0.013)
Teacher Bias			-0.011	-0.011	-0.006
			(0.015)	(0.014)	(0.013)
Fem*Math Teacher Fem					-0.052
					(0.040)
Math Teacher Fem					0.061
					(0.041)
Fem*North Math Teacher					0.013
					(0.031)
Math Teacher born North					0.027
					(0.035)
Fem*Advanced STEM Teacher					-0.031
					(0.034)
Advanced STEM					0.026
					(0.034)
Std Math grade 6		0.716***	0.715***	0.687***	0.688***
		(0.011)	(0.011)	(0.012)	(0.012)
Gender Gap	-0.214	-0.077	-0.077	-0.081	-0.082
School, year FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9309	9309	9309	9309	9309
R^2	0.136	0.576	0.577	0.585	0.588

Notes: This table reports OLS estimates of equation 1.2, where the dependent variable is math standardized test score in grade 8; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (school by cohort) is 185. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

interact with the math teacher for six hours per week, while the rest for 9 hours per week. Furthermore, I exploit the fact that around 20% did not have the same teacher for all three years of middle school. However, for both variables, I do not see a statistically or economically significant pattern. Most likely the impact of teacher gender stereotypes begins at lower intensive margins and we do not have proxies of the "quality" of teacher-student interaction that would be necessary to further investigate this mechanism.

1.5.2 Choice of High-School Track and Teachers Recommendation

High-school track choice is the first crucial career decision in the Italian schooling system. Students and their families are free to choose their most-preferred track, with no constraints based on grades or teachers' official track recommendation. There are three main types of high-school: academic, technical and vocational. As shown in Table 1.2, there are substantial gender differences in the type of tracks selected: the preferred choice among females are academic track related to psychology, languages and art, while for males the preferred choices are academic scientific and technical technological tracks. Students in different tracks have in most cases little to no interaction during the school day since buildings are generally separated. Finally, the choice of high-school is strongly correlated with university choice, as discussed in Section 1.2. From a policy perspective, the scientific academic path is interesting since it easily opens up career opportunities in STEM related fields, while the vocational choice is highly correlated with almost no tertiary education. Hence, I explore the impact of teacher bias on the track choice at the end of middle school, with a focus on the choice of the scientific academic track and on the vocational track.

Table 1.8, Panel A, shows that girls are 9.4 percentage points less likely than boys to attend a scientific track and equally likely to attend a vocational track. Controlling for the standardized test score in math reduces half of the gap in the choice of scientific track, which is present also in track recommendations received from teachers (Panel B). However, I find a close to zero and insignificant effect of teacher bias on gender gap in scientific track choice (Panel A, columns 2-4) and in the recommendation of teachers toward a scientific track (Panel B, columns 2-4). The inclusion of controls at student and teacher level interacted with the pupil gender do not affect the point estimates of interest.

Recent work suggests that women are more responsive to negative feedback than men in STEM fields (Kugler et al., 2017). However, the scientific track is chosen by females with highly educated parents or with high achievement tests, whose performance was not affected by teacher bias, as shown by analyzing the heterogeneous effects in Section

Table 1.7: Estimation of the effet of teachers' gender stereotypes

Dependent Variable: Math standardized test score in grade 8						
Heterogeneous effects by	**Student Characteristics**				**Interaction time with teacher**	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.024	-0.020	0.033	-0.023	-0.050	-0.033
	(0.103)	(0.103)	(0.112)	(0.103)	(0.104)	(0.104)
Fem*Teacher Bias	-0.037***	-0.049**	-0.070***	-0.036**	-0.040**	-0.065**
	(0.014)	(0.021)	(0.027)	(0.015)	(0.016)	(0.031)
Fem*T Bias*HighEduM		0.022				
		(0.028)				
Fem*T Bias*Top tercile Math6			0.100***			
			(0.035)			
Fem*T Bias*Middle tercile Math6			0.011			
			(0.035)			
Fem*T Bias*Immigrant				-0.011		
				(0.038)		
Fem*T Bias*Extended School Day					0.012	
					(0.026)	
Fem*T Bias*Same Math Teacher						0.031
						(0.035)
Gender Gap	-0.082	-0.082	-0.082	-0.082	-0.082	-0.082
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	9309	9309	9309	9309	9309	9309
R^2	0.626	0.626	0.627	0.626	0.626	0.626

Notes: This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on math standardized test score in grade 8 by observable characteristics of the student and by interaction time with teacher; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student, "HighEduM" wether the mother has at least a diploma, "tercile Math6" is the tercile of standardized test score in math in grade 6 and "Immigrant" is a dummy equal to 1 if the student is not Italian citizen. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughthers, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, type of contract and education of the teacher' mother. Regressions are all fully saturated even if not all interactions are shown in the table. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table 1.8: Estimation of the effect of teachers' gender bias on track choice- class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A- Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.094***	-0.048***	0.025	0.171*	0.014	-0.009	0.020	0.016
	(0.012)	(0.011)	(0.019)	(0.092)	(0.009)	(0.010)	(0.023)	(0.071)
Fem*T Bias		0.009	0.008	0.001		0.023**	0.020**	0.020**
		(0.012)	(0.011)	(0.011)		(0.009)	(0.009)	(0.009)
Fem*T Fem				-0.032				0.034
				(0.029)				(0.022)
Std Math 6		0.178***	0.159***	0.159***		-0.104***	-0.091***	-0.091***
		(0.008)	(0.008)	(0.008)		(0.007)	(0.007)	(0.007)
Constant	0.299***	0.242***	0.106***	0.108***	0.141***	0.174***	0.207***	0.205***
	(0.006)	(0.006)	(0.015)	(0.015)	(0.005)	(0.006)	(0.016)	(0.016)
MeanYFem	0.205	0.205	0.205	0.205	0.155	0.155	0.155	0.155
Obs.	8463	8463	8463	8463	8463	8463	8463	8463
R^2	0.113	0.214	0.233	0.236	0.119	0.190	0.208	0.211
	Panel B- Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.045***	-0.019**	0.033**	0.016	-0.059***	-0.110***	-0.125***	-0.059
	(0.010)	(0.009)	(0.015)	(0.081)	(0.013)	(0.011)	(0.024)	(0.092)
Fem*T Bias		0.001	-0.000	-0.007		0.018*	0.018*	0.024**
		(0.009)	(0.009)	(0.009)		(0.010)	(0.010)	(0.011)
Fem*T Fem				-0.053**				0.024
				(0.025)				(0.036)
Std Math 6		0.126***	0.113***	0.113***		-0.246***	-0.217***	-0.217***
		(0.009)	(0.009)	(0.009)		(0.008)	(0.008)	(0.008)
Constant	0.156***	0.129***	0.059***	0.059***	0.376***	0.428***	0.518***	0.517***
	(0.005)	(0.004)	(0.011)	(0.011)	(0.006)	(0.006)	(0.017)	(0.017)
MeanYFem	0.110	0.110	0.110	0.110	0.317	0.317	0.317	0.317
Obs.	7086	7086	7086	7086	7086	7086	7086	7086
R^2	0.152	0.238	0.249	0.251	0.150	0.362	0.389	0.391
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S Controls	No	No	Yes	Yes	No	No	Yes	Yes
T Controls	No	No	No	Yes	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is the high-school track choice; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

1.5.1.[49] These female students are likely to have additional academic-oriented role models in addition to their math teacher and a lower vulnerability to the gender stereotypes.

Teacher stereotypes have stronger impact at the bottom of the ability distribution. Indeed, we can observe in columns 6 of Panel A that females, when assigned to a teacher with one standard deviation higher implicit bias, are more likely than their male classmates to attend vocational track by around 2 percentage points. This effect corresponds to an increase of 13% with respect to the mean probability of attending vocational training for girls. This result mirrors an analogous differential in teachers' track recommendation toward vocational school as shown by Panel B, columns 6. The subsequent two columns include characteristics of teachers and pupils and their interaction with the gender of the latter. Adding these controls does not change the coefficient of interest. When exposed to less gender-biased environment, female students are more likely to attend the technical track, instead of vocational (see Appendix Table A.11).

Figure 1.5 reports the same estimates using a categorical variable instead of the continuous one to offer a clearer representation of the results.[50] Girls assigned to a teacher with a "pro boys" attitude have a probability of 16.3% of attending the vocational track, while female students assigned to a teacher with a "pro girls" attitude have a 6 percentage points lower probability of attending the same track, which corresponds to a decrease by 37%.

[49] In the questionnaire administered to teachers, I ask them why girls, compared to boys with the same math performance, are less likely to attend the scientific track: the reason identified as the most important is the parental influence (for the summary statistics see Table 1.1).

[50] I use the thresholds defined by Greenwald et al. (2003) and exploited also in Figure 1.3. For more details, check Section 1.5.1.

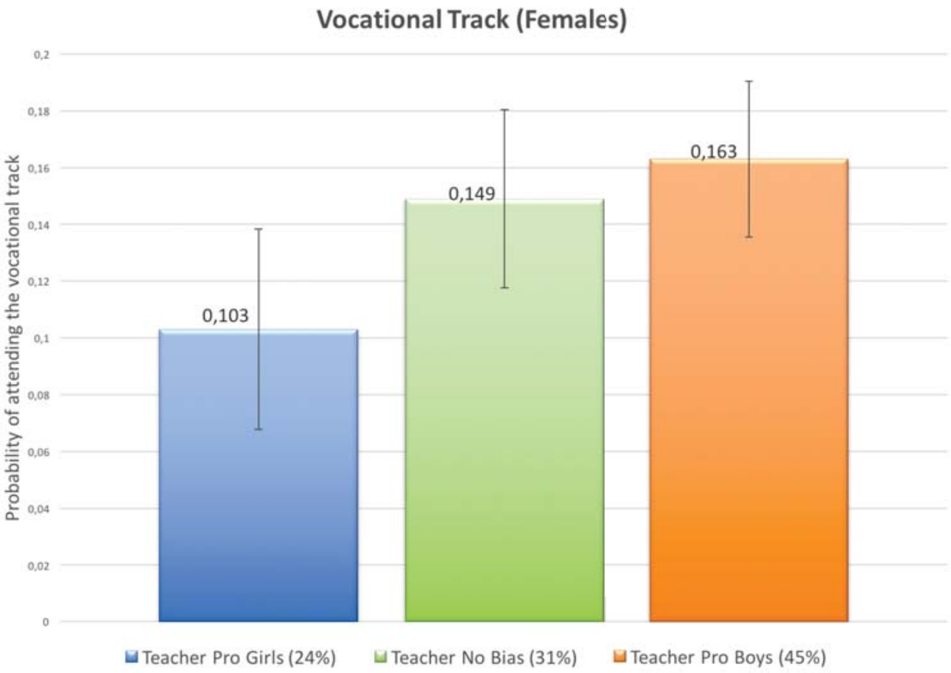


Figure 1.5: Effect of teacher bias on choice of vocational track of females

Notes: This graph shows the effect of teacher stereotypes on female students' track choice. We consider the thresholds defined by Greenwald et al. (2003) where no bias is the interval of IAT raw score between -0.15 and +0.15. The attitude of the teacher in associating fields with gender is considered "pro girls" if the score is lower than -0.15 (24% of teachers) and "pro boys" if the score is higher than +0.15 (45% of teachers). The variable in the y axis is the gender gap in improvements in math between grade 6 and 8, when class fixed effects are absorbed.

Appendix Table A.10 shows the results estimating equation (1.2), with school fixed effects instead of class. They confirm the previous evidence of a substantial impact on female students in terms of choice of vocational training. Finally, Appendix Table A.12 presents results from the heterogeneity analysis and, as expected, the impact of teacher bias has a stronger effect on the track choice of female students from disadvantaged background. The enrollment of females from the bottom tercile of the distribution increases by 4.3 percentage points for one standard deviation higher bias of the math teacher (which corresponds to a 15.8% increase with respect to the mean value for this group).

1.5.3 Additional Results and Robustness Checks

Appendix Table A.13 provides evidence of the impact of math teacher bias on reading standardized test scores, presenting the results of estimating equation (1.1). Although the effect is significant only including the controls, there are some negative cross-subject spillovers in performance. Additionally, Appendix Table A.14 shows estimates of the impact on math performance of the Italian teacher bias. The gender bias of Italian teacher does not affect the gender gap in math performance. The point estimates are small, indistinguishable from zero and not affected by inclusions of controls either at Italian teacher level or at pupil level.[51] Biased teachers in male-typed domains activate stereotypes on female students. Indeed, gender bias of Italian teachers has no statistically significant impact on reading and math performance of students, neither boys nor girls.

All results exploit information on three cohorts of students. In Appendix Table A.8, I show the effect of the main specification presented in Table 1.5 for the three different cohort of students separately. Reassuringly also for the potential reverse causality concerns expressed in Section 1.4.4, results are not statistically different in the three cohorts, even if, since the number of observation decreases splitting the sample, estimates are noisier.[52]

In the Italian schooling system, at the end of each academic year, teachers decide whether the student is admitted to the following grade. This decision is based on the overall assessment of students, including both performance and behavior in class. The retention rate of males is higher compared to the one of females. For instance, among students who attended the test score in grade 6 (9837 students), 6.0% of males and 3.3% of females are retained in (at least) one of the three years of middle school. In Table A.15, I check whether math teachers' bias has an impact on retention rate, but I do not find any significant impact, neither without nor with the inclusion of the controls at teacher and student level. Furthermore, I also check that teacher implicit stereotypes does not differentially impact the probability of taking the standardized test score in grade 8 (Table A.15, columns 5-8), conditional on taking the one in grade 6. These results suggest that the sample used in our main table on performance in math is not biased by differential attrition by gender, induced by teacher bias. Additional checks on potential sample selection issues are addressed in Appendix A.4.

Finally, in the Appendix Table A.16, I consider the impact of self-reported gender bias. The impact of self-reported bias on student performance is generally small and in most specifications indistinguishable from zero. However, the impact of IAT score on student achievement is not significantly affected when I control for reported bias. This evidence supports the distinctiveness of implicit and explicit cognition (Greenwald et al., 1998) in the context of gender stereotypes of teacher.

[51]In Appendix A.5, I show the summary statistics for Italian teachers and I delve deeper into the role of Italian teachers.

[52]For the first cohort, I have fewer observations because some schools change the code identifying the school that year for administrative reasons and I am not allowed to access data identified with the older codes.

1.6 Discussion of Potential Mechanisms

In this section, I discuss the mechanisms behind the negative impact of teacher bias on student achievement. I focus mainly on two aspects: self-stereotypes and interaction theory.[53] I use student survey data to analyze more deeply the former aspect, while for the latter I rely on the social psychology evidence on the interaction between teachers and pupils by gender. In Appendix A.7, I present a conceptual framework including both aspects.

Self-Stereotypes

Self-confidence plays a crucial role in affecting performance, especially in gender-incongruent areas, such as female performance in math (Coffman, 2014). According with social psychology, the development of academic self-concept begins since childhood and is strongly influenced in the period after elementary school by stereotypes communicated by significant others, such as parents and teachers (Ertl et al., 2017). Girls may believe that both own signal of ability and the signal received by teachers carry relevant information. However, if the signal received from teachers is biased by beliefs that women have lower ability than men in math or are less suitable for a STEM career, females will develop a lower self-assessment of own ability in the scientific field and potentially invest less in their STEM education. The idea is consistent with the stereotype threat theory developed in social psychological literature (Steele and Aronson, 1995), according to which individuals at risk of confirming widely-known negative stereotypes reduce their confidence and underperform in fields in which their group is ability- stigmatized (Spencer et al., 1999).[54]

I find that biased math teachers activate negative self-stereotypes and induce females to believe that they are *worse at math* than what would be expected given their achievements. This result is important for at least two reasons. First, it shows that self-confidence of women in math is affected by social conditioning from teachers. Second, this is an important mechanism to understand the effect of teacher bias on math performance of female students.

[53]There is a third theory that could be consistent with the negative impact of teacher bias on female student math performance. According with the *animus theory*, teachers may dislike female students, treating them badly or giving them more unpleasant assignments, causing girls to dislike math. In our context, it seems unlikely that teachers assign different tasks to students by gender in terms of exams or homework. Furthermore, in appendix A.6 we provide evidence that teachers favor female students in math grading, comparing blinded and no-blinded scores, as emerges in several other countries (Lavy and Sand, 2015; Terrier, 2015).

[54]Despite the rich literature in social psychology about stereotype threat since 1990s, only recently have economists directly analyzed this phenomenon, finding partially contradictory evidence. One of the first steps taken in this direction has been Fryer et al. (2008), which finds no evidence of stereotype threat behavior in influencing women's performance in math, while Dee (2014) shows a substantial impact of activating negatively stereotyped identity (i.e., student-athlete) on test score performance.

Table 1.9 assesses the extent to which bias of teachers affect one's own assessment of ability, for a sample of around 800 students for whom I collected self-confidence measures, as described in section 1.3.2. I present results for self-stereotypes in math in Panel A, in reading in Panel B and on average of all other subjects in Panel C. As shown in column 1, females are 9.4 percentage points less likely to consider themselves good at math (which corresponds to 11% percent lower probability than males). Female students are generally found to be more critical about their abilities in math than male students even if they have the same grade, as shown in PISA tests as well (OECD, 2015a). However, females are 5.2 percentage points more likely to consider themselves good in Italian (which corresponds to 6% percent higher probability than males), but on average both equally assess their own ability. In classes assigned to math teachers with higher bias, the gender gap in self-assessment of own ability in math is increasing. In particular, in classes assigned to teachers with one standard deviation higher bias, the gender gap in self-assessment increases by 4.5 percentage points, controlling for the test score in grade 6 as in our main specification in equation (1.1). Adding student and teacher level controls interacted with pupil gender do not substantially affect the point estimate of interest (columns 3 and 4).

In Section 1.5.1, I provide evidence that the gender gap in math achievement increases during middle school in classes assigned to a more biased teacher. Hence, in the last three columns of Table 1.9, I also control for the mediating role of performance measured at the end of middle school in order to analyze whether gender gap in own assessment is merely due to different achievements at the end of middle school. I find that gap in own assessment is reduced only by less than one third: teacher stereotypes have an additional impact on own assessment of math capabilities, on top of measured ability, that may have detrimental effects for investment choices in education and occupation.

In Appendix Table A.17, I show the result of the specification described with school fixed effects instead of class fixed effects (as in equation 1.2). Consistent with the results in Table 1.6, there is a negative impact of teacher bias on self-stereotypes of female students and no impact on male students. All results are robust to the inclusion of controls at pupil and teacher level and their interaction with student gender.

In Panel B and C of Tables 1.9 , I focus on the impact of math teacher bias on self-assessment respectively in Italian and all other subjects. Female students seem to compensate for the low confidence in math with higher self-assessment in Italian, the other main subject taught during middle school. There is no impact on other subjects. The effects are robust to the inclusion of controls at individual level (column 3 and 4) and at teacher level (column 4) and are coherent in both specification, including class and the school fixed effects (see Appendix Table A.17). Finally, in the last three columns of Panel B, I control for the standardized test score in Italian in grade 8: as expected, it does not

Table 1.9: Estimation of the effet of teachers' gender bias on self-stereotypes- class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A- Dependent Variable: Being good/mediocre at math (vs. being bad)							
Fem	-0.094***	-0.067**	-0.093	0.174	-0.053*	-0.074	0.195
	(0.029)	(0.028)	(0.065)	(0.188)	(0.028)	(0.065)	(0.200)
Fem*Teacher Bias		-0.045**	-0.049**	-0.066**	-0.030	-0.033	-0.052*
		(0.021)	(0.022)	(0.030)	(0.021)	(0.023)	(0.030)
Std Test Math		0.138***	0.135***	0.136***	0.157***	0.151***	0.148***
		(0.024)	(0.023)	(0.024)	(0.023)	(0.023)	(0.024)
Constant	0.837***	0.808***	0.809***	0.800***	0.810***	0.820***	0.812***
	(0.015)	(0.015)	(0.048)	(0.047)	(0.015)	(0.048)	(0.046)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	747	747	747	747	747	747	747
R^2	0.110	0.216	0.236	0.253	0.248	0.266	0.281
Panel B- Dependent Variable: Being good/mediocre at Italian (vs. being bad)							
Fem	0.052**	0.057**	0.045	0.166	0.047**	0.035	0.135
	(0.023)	(0.023)	(0.048)	(0.215)	(0.021)	(0.046)	(0.203)
Fem*Teacher Bias		0.038**	0.038**	0.026	0.038**	0.039**	0.029
		(0.018)	(0.019)	(0.022)	(0.017)	(0.019)	(0.021)
Constant	0.916***	0.908***	0.937***	0.946***	0.917***	0.953***	0.963***
	(0.012)	(0.012)	(0.034)	(0.035)	(0.011)	(0.034)	(0.035)
Std Test score Italian	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	664	664	664	664	664	664	664
R^2	0.115	0.134	0.148	0.175	0.148	0.161	0.189
Panel C- Dependent Variable: Average own ability in other subjects							
Fem	0.035	0.019	0.021	-0.213	0.016	0.018	-0.219
	(0.027)	(0.029)	(0.062)	(0.224)	(0.028)	(0.062)	(0.227)
Fem*Teacher Bias		-0.014	-0.015	-0.020	-0.018	-0.020	-0.025
		(0.023)	(0.024)	(0.027)	(0.024)	(0.024)	(0.027)
Constant	1.672***	1.689***	1.674***	1.681***	1.687***	1.670***	1.676***
	(0.014)	(0.016)	(0.041)	(0.041)	(0.015)	(0.040)	(0.040)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	802	802	802	802	802	802	802
R^2	0.096	0.125	0.137	0.157	0.130	0.141	0.161
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	Yes	Yes
Math Teacher Controls	No	No	No	Yes	No	No	Yes

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is self-stereotypes in grade 8; 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 math teacher level in parentheses; the number of clusters is 58. The number of fixed effects (classes) is 62. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

affect the estimate since math teacher stereotypes do not impact gender gap in reading performance. A deeper analysis of the impact on reading test scores or of the gender gap of the Italian teacher is presented in Appendix A.5.

Interaction Theory

A second potential mechanism is related to the *interaction theory* (McConnell and Leibold, 2001): math teachers with higher gender bias may spend less time (in terms of either quantity or quality) interacting with girls, especially those performing poorly. Biased teachers may choose to allocate more time or tailor math classes to the learning of boys and top-performing girls since they are more likely to attend a STEM track during high school. However, we do not find evidence of higher achievement of these groups of students when exposed to a gender-biased environment. Unfortunately, I do not have measures of the "quality of interaction" between teachers and student by gender to directly test this mechanism.

The social psychology literature provides evidence that math teachers interact differently with male and female students. It has been shown that they believe math is more difficult for girls than equally achieving boys (Riegle-Crumb and Humphries, 2012; Tiedemann, 2002).[55].

Hence, biased teachers are more likely to fail to recognize talent of some students in math related fields and set a lower bar for their learning. Teachers' erroneous expectations may lead to a self-fulfilling prophecy: they may fail to recognize students' talent and therefore not encourage them to fulfill their potential (Rosenthal and Jacobson, 1968; Cooper and Good, 1983). Furthermore, Sadker and Sadker (2010) document that teachers spend more time interacting with boys, while Hyde et al. (1990) suggests that math is taught as a set of computational methods to girls, while boys are encouraged to exert independence. Finally, Keller (2001) find that teachers convey their stereotyping of *mathematics as a male domain* through their classroom instruction and affect students' own association between math and males.

All these aspects suggest that gender-biased interaction between pupils and teachers is an important mechanisms behind the main results of this paper on the impact of teacher stereotypes on student achievements. They are also very important to understand the mechanisms through which self-stereotypes of students are activated when exposed to gender-biased teachers.

[55] Using Italian data from INVALSI, I show in Appendix A.8 that this perception of teachers mirrors a self-perception of students. Female students compared to boys with the same performance are more likely to believe their achievement is the result of effort and less likely to believe it is the result of ability.

1.7 Conclusion

In most OECD countries, women outnumber men in tertiary education, but they are by far a minority in highly paid fields such as science, technology, engineering and math, especially when excluding teaching careers. The prospects for change are not optimistic considering that on average in OECD countries less than 5 percent of 15-years-old girls are planning to pursue a career in these fields compared to around 20 percent of boys according to 2015 PISA data. Social conditioning has a strong impact on development of skills and educational choices. This paper shows that the gender gap in math performance can be partially explained by teacher implicit bias. Females, especially those from disadvantaged backgrounds, are lagging behind when assigned to teachers with higher implicit stereotypes (as measured by an Implicit Association Test). Males, the group not ability-stigmatized in terms of math performance, are not affected by teacher bias. Teacher stereotypes affect high-school track choice, leading female students assigned to a teacher with higher implicit bias to be more likely to attend a vocational school. Furthermore, they foster low expectations about own ability and lead to underperformance in male-typed domains. Indeed, females are more likely to consider themselves bad in math at the end of middle school if they are assigned to a biased teacher, even controlling for their ability measured by standardized test scores. These findings are consistent with a model whereby ability-stigmatized groups under-assess own ability and underperform fulfilling negative expectations about their achievements. Unconscious biases and implicit associations can form an unintended and often an invisible barrier to equal opportunity.

These results raise the question of which kind of policies should be implemented in order to alleviate the impact of gender stereotypes. The gap in math performance generated during middle school would be 35% smaller if no teachers had negative gender stereotypes (from 0.078 to 0.051 standard deviations). The implicit bias measured by IAT score at this stage of development should not be used to make decisions about others, as hiring or firing decisions. IAT scores are educational tools to develop awareness of implicit preferences and stereotypes. Hence, one set of potential policies may be aimed at informing people about own bias or training them in order to assure equal behavior toward individuals of ability-stigmatized groups and others. An alternative way to fight against the negative consequences of stereotypes is increasing self-confidence of female in math or providing alternative role models, as done in the context of Indian elections, where exposure to female leaders weakens gender stereotypes in the home and public spheres (Beaman et al., 2009). More research is needed to further investigate the impact of both type of policies.

A. Appendix

A.1 Additional Figures and Tables

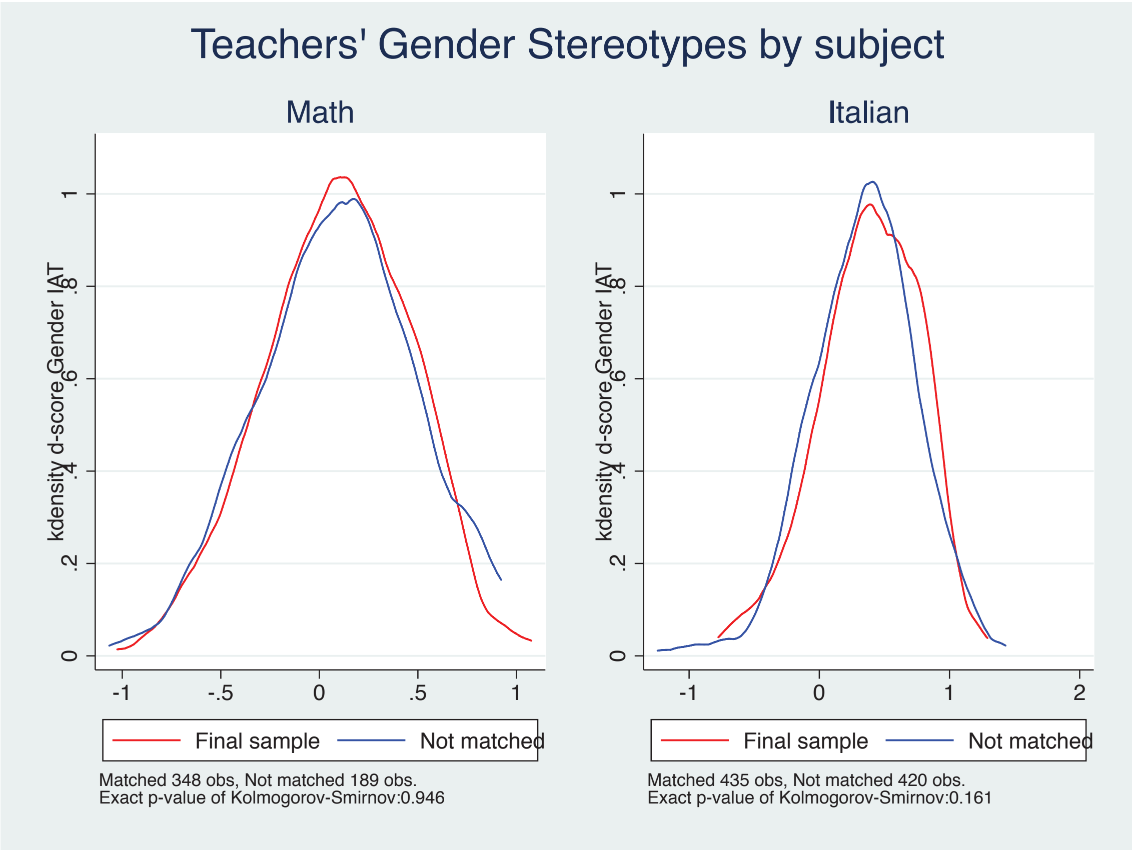


Figure A.1: Teachers' Implicit Gender Bias (IAT measure) by subject of matched and unmatched sample

Notes: This graph shows the distribution of Gender-Science IAT scores for math and literature teachers. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes. The graph provides evidence that data on teachers used in this papers (i.e. those teaching in the school before 2015) are similar in terms of gender stereotypes compared to those who completed the survey but are not included in the main analysis.

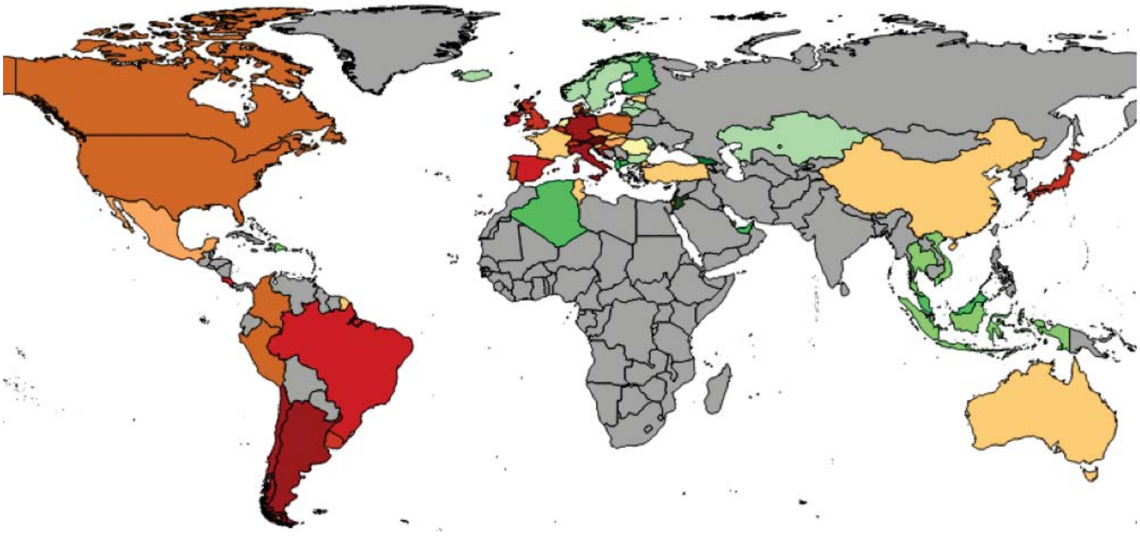


Figure A.2: Gender differences in math standardized test score PISA

Notes: This graph shows the difference in math PISA test scores between females and males. Countries in which girls lag behind are colored in red, while countries where boys lag behind are colored in green. *Source:* Author's calculation on PISA data (2015).

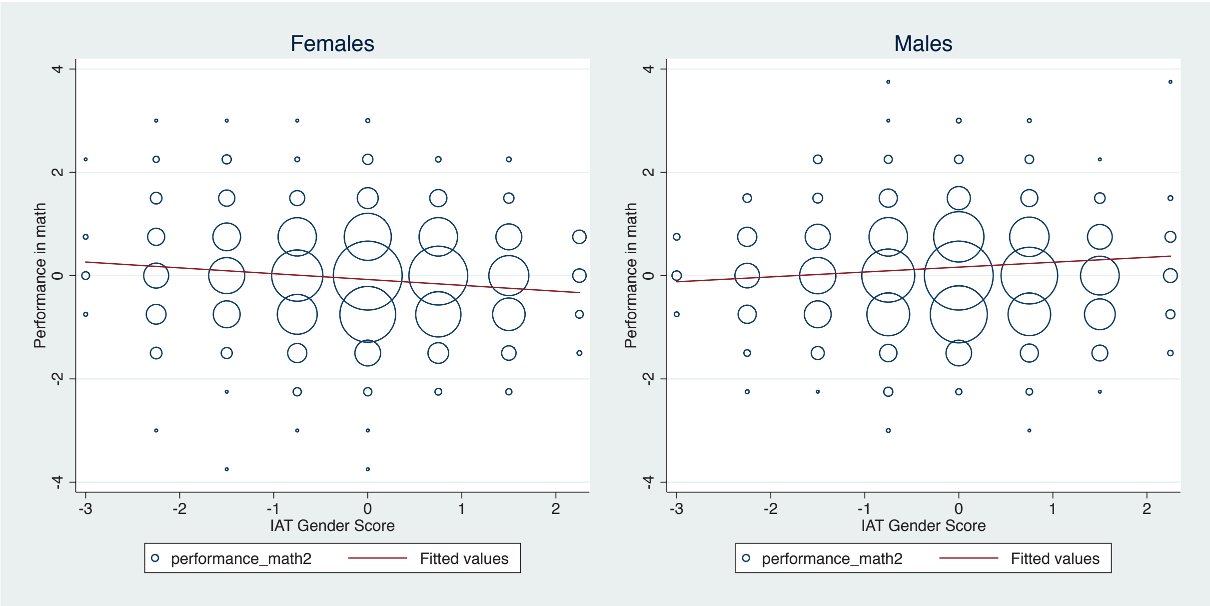


Figure A.3: Correlation between the performance in math and the implicit bias of math teachers (as measured by IAT)

Notes: This graph shows the effect of teacher stereotypes on student achievement by gender. The variable in the y axis is the improvement in standardized test score in middle school. The variable in the x axis is the raw IAT score. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females.

Table A.1: Correlation between implicit bias d-score and order of different parts of the survey

Dependent variable : d-score Implicit Bias Math Teachers	(1)	(2)	(3)	(4)	(5)
First IAT Gender	-0.035			-0.034	-0.040
	(0.037)			(0.037)	(0.047)
First Questionnaire, then IAT		-0.151		-0.166	0.017
		(0.127)		(0.126)	(0.185)
Order Compatible IAT Gender			-0.051*	-0.052*	-0.049
			(0.029)	(0.030)	(0.039)
Constant	0.106***	0.097***	0.119***	0.134***	0.131***
	(0.019)	(0.016)	(0.020)	(0.022)	(0.024)
School FE	No	No	No	No	Yes
Obs.	534	534	534	534	534
R^2	0.002	0.002	0.005	0.009	0.168

Notes: This table reports OLS estimates of the correlation between order of IAT and IAT score. A higher value of IAT score means stronger implicit association between Male-Science and Female-Literature. The dummy "First IAT Gender" captures the order of IATs (genderand race). The variable "Order Compatible IAT Gender" captures whether it was asked to associate together first more likely compatible categories (Male-Scientific vs. Female-Humanistic) or the opposite (Female-Scientific vs. Male-Humanistic). Finally, in 8 cases for the math teacher and 32 cases for the Italian teacher we asked to complete first a questionnaire and then the IATs. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.2: Balance table of the differences between teachers matched (not matched) with students in the sample

	Math Teachers				Italian Teachers			
	Not matched	Matched	Dif.	se	Not matched	Matched	Dif.	se
Female	0.763	0.837	-0.074*	(0.034)	0.858	0.910	-0.052*	(0.022)
Born in the North	0.545	0.653	-0.108*	(0.043)	0.606	0.737	-0.131***	(0.033)
Age	42.569	51.903	-9.335***	(0.767)	42.889	51.367	-8.477***	(0.594)
Full time contract	0.516	0.916	-0.400***	(0.035)	0.731	0.988	-0.257***	(0.022)
Yeas of experience	10.646	22.941	-12.295***	(0.938)	13.854	23.695	-9.842***	(0.691)
Teaching in 2015-16	0.560	0.987	-0.427***	(0.030)	0.617	1.000	-0.383***	(0.023)
Children	0.572	0.744	-0.172***	(0.040)	0.569	0.729	-0.160***	(0.032)
Number of children	1.858	1.842	0.016	(0.095)	1.674	1.786	-0.112	(0.074)
Number of daughters	0.890	0.851	0.039	(0.087)	0.916	0.815	0.101	(0.068)
Low edu Mother	0.415	0.579	-0.164***	(0.045)	0.387	0.516	-0.129***	(0.036)
Middle edu Mother	0.401	0.291	0.110*	(0.043)	0.403	0.376	0.027	(0.036)
High edu Mother	0.184	0.129	0.054	(0.033)	0.210	0.108	0.102***	(0.026)
Advanced STEM	0.237	0.240	-0.003	(0.038)	0.000	0.000	0.000	(0.000)
Math Olympiad	0.088	0.192	-0.104***	(0.031)	0.000	0.000	0.000	(0.000)
Update Courses	0.851	0.938	-0.087***	(0.026)	0.863	0.932	-0.069**	(0.021)
Degree Laude	0.276	0.168	0.108**	(0.039)	0.386	0.293	0.093**	(0.035)
IAT Gender	0.104	0.087	0.016	(0.033)	0.346	0.386	-0.040	(0.027)
IAT Race	0.472	0.458	0.014	(0.023)	0.450	0.464	-0.014	(0.018)
Boys better in Invalsi	0.241	0.202	0.039	(0.043)	0.095	0.099	-0.003	(0.026)
Girls better in Invalsi	0.304	0.322	-0.018	(0.048)	0.550	0.527	0.023	(0.043)
Gender Equal in Invalsi	0.456	0.476	-0.021	(0.052)	0.355	0.374	-0.020	(0.042)
Satisfy with teacher job	3.749	3.692	0.057	(0.079)	3.862	3.890	-0.028	(0.061)
WVS Gender Equality	0.150	0.166	-0.015	(0.032)	0.110	0.105	0.005	(0.022)
Reason Gender Gap:								
Interest for STEM	2.543	2.579	-0.036	(0.095)	2.869	2.648	0.221**	(0.076)
Predisposition for STEM	2.127	2.124	0.003	(0.103)	2.222	2.158	0.064	(0.081)
Low self-esteem	2.905	2.638	0.267**	(0.094)	2.694	2.540	0.154	(0.079)
Family support	3.155	3.144	0.011	(0.098)	3.118	2.941	0.176*	(0.076)
Cultural Stereotypes	2.461	2.147	0.314**	(0.105)	2.316	2.166	0.150	(0.086)
Gender Dif Innate Ability	1.558	1.514	0.043	(0.069)	1.490	1.371	0.119*	(0.049)
Reported gender bias	-0.003	0.002	-0.006	(0.106)	-0.060	0.058	-0.118	(0.083)
Observations	236	301			422	431		

Notes: First hand data from teachers' questionnaire. We compare teachers' matched with students' data with those not matched.

Table A.3: Correlation between teachers' characteristics and Gender IAT Score

Dependent variable: raw IAT score of math teachers				
	Restricted sample		All teachers	
	(1)	(2)	(3)	(4)
Female	-0.169***	-0.192**	-0.188***	-0.170***
	(0.056)	(0.088)	(0.044)	(0.058)
Born in the North	-0.054	-0.123*	-0.071**	-0.110**
	(0.047)	(0.072)	(0.033)	(0.044)
Age	0.040	0.047	-0.020	-0.035
	(0.039)	(0.063)	(0.018)	(0.024)
Age sq.	-0.000	-0.000	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)
High Edu Mother	-0.003	-0.003	0.010	0.006
	(0.047)	(0.063)	(0.037)	(0.044)
Children	0.195	0.031	0.028	0.111
	(0.123)	(0.095)	(0.054)	(0.309)
Daughters	0.071	0.048	0.013	-0.026
	(0.062)	(0.086)	(0.050)	(0.060)
Advanced STEM	-0.073	-0.105	-0.094**	-0.139***
	(0.052)	(0.077)	(0.042)	(0.050)
Degree Laude	-0.062	-0.033	-0.064	-0.053
	(0.059)	(0.084)	(0.042)	(0.051)
Full time contract	-0.063	-0.074	0.005	-0.005
	(0.095)	(0.141)	(0.057)	(0.075)
Math Olympiad	0.095	0.066	0.070	0.098
	(0.066)	(0.093)	(0.053)	(0.070)
Satisfy with teacher job	0.033	0.039	0.010	0.013
	(0.024)	(0.034)	(0.017)	(0.020)
Refresher courses	0.012	0.009	0.027	0.018
	(0.026)	(0.036)	(0.020)	(0.025)
Explicit Bias	0.009	0.020	0.019	0.038
	(0.024)	(0.038)	(0.020)	(0.026)
School FE	No	Yes	No	Yes
Obs.	301	301	534	534
R^2	0.116	0.417	0.096	0.285

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and own teacher characteristics; the unit of observation is teacher t in school s. Standard errors are robust and clustered at school level in parentheses; the number of clusters is 90. We include the order of IATs for math teachers and missing categories if the information is not available. The restrictd sample includes data expoits on teachers in the main regressions of this paper. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.4: The impact of math teacher characteristics on students' improvement in performance

	Dependent variable: Std Math test score in grade 8								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: X=	**Teacher Female**			**Teacher born North**			**Advanced STEM**		
X	0.075**	0.084**	0.079*	-0.008	-0.021	0.007	0.007	0.015	0.027
	(0.029)	(0.037)	(0.043)	(0.031)	(0.035)	(0.030)	(0.032)	(0.035)	(0.033)
Fem*X		-0.019	-0.033		0.026	0.019		-0.016	-0.009
		(0.039)	(0.038)		(0.030)	(0.028)		(0.033)	(0.031)
Obs.	9309	9309	9309	9309	9309	9309	9309	9309	9309
R^2	0.529	0.529	0.562	0.528	0.528	0.562	0.528	0.528	0.562
Panel A: X=	**Math Olympics**			**Full-time contract**			**Degree with laude**		
X	0.075*	0.039	0.031	0.056	0.042	0.044	0.042	0.044	0.074**
	(0.039)	(0.044)	(0.039)	(0.044)	(0.050)	(0.047)	(0.039)	(0.044)	(0.035)
Fem*X		0.073*	0.070*		0.028	0.018		-0.004	-0.002
		(0.041)	(0.038)		(0.075)	(0.068)		(0.041)	(0.040)
Obs.	9309	9309	9309	9309	9309	9309	9309	9309	9309
R^2	0.529	0.529	0.563	0.528	0.528	0.562	0.528	0.528	0.563
School FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table reports OLS estimates of the impact of math teachers' characteristics on improvements in math performance of their students; the unit of observation is the student. Standard errors are robust and clustered at teacher level in parentheses. All columns include a dummy for the gender of the student ("Fem"), standardized test score in grade 6 and the interaction with the gender of the student. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.5: Exogeneity of assignment of students to math teachers with different bias

Dependent Variable: Math Teacher implicit gender bias (standardized)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ShareF	0.883					0.739	0.506	-1.610
	(1.013)					(1.051)	(1.006)	(3.136)
ShareM HighEduMum		-0.117				-0.246	-0.173	0.793
		(0.708)				(0.719)	(0.752)	(2.625)
ShareF HighEduMum		0.573				0.449	0.502	-0.310
		(0.606)				(0.595)	(0.665)	(2.137)
ShareM HighOccDad			0.402			0.861	1.078	0.888
			(1.152)			(1.217)	(1.076)	(3.833)
ShareF HighOccDad			0.289			-0.074	-0.074	0.589
			(0.744)			(0.897)	(0.972)	(2.258)
ShareM MedOccDad			0.222			0.750	0.636	0.478
			(0.764)			(0.738)	(0.780)	(1.251)
ShareF MedOccDad			1.055			0.869	0.500	0.173
			(0.816)			(0.881)	(0.887)	(1.985)
ShareM Immigrant				0.851		1.113*	1.162*	-0.122
				(0.638)		(0.637)	(0.623)	(1.977)
ShareF Immigrant				-0.218		-0.080	-0.102	0.560
				(0.619)		(0.669)	(0.701)	(1.255)
Male Std Ita6					-0.145	0.031	-0.055	-0.500
					(0.273)	(0.301)	(0.314)	(1.333)
Fem Std Ita6					0.220	0.082	0.145	0.991
					(0.271)	(0.312)	(0.305)	(0.855)
Male Std Mat5								-0.187
								(0.593)
Fem Std Mat5								-0.458
								(0.422)
Constant	-0.368	-0.112	-0.572	-0.081	0.030	-1.400	-1.529	-0.910
	(0.516)	(0.444)	(0.697)	(0.163)	(0.103)	(0.941)	(1.371)	(3.250)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teach. Control	No	No	No	No	No	No	Yes	Yes
Obs.	301	301	301	301	301	301	301	110
R^2	0.329	0.332	0.341	0.337	0.328	0.366	0.444	0.590

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and students' characteristics; 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 school level in parentheses; the number of clusters is 90 in columns 1-6 and 40 in column 7. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for the order of IAT in the questionnaire administered. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.6: Correlation between IAT score of teacher and share of females among best performing students

Dependent variable: raw IAT score math teachers						
	(1)	(2)	(3)	(4)	(5)	(6)
Share Females Top10	0.360	0.586*				
	(0.219)	(0.312)				
Share Females Top20			0.068	0.489		
			(0.289)	(0.442)		
Share Females Top40					0.170	0.515
					(0.389)	(0.560)
School FE	No	Yes	No	Yes	No	Yes
Teacher Controls	No	Yes	No	Yes	No	Yes
Obs.	301	301	301	301	301	301
R^2	0.023	0.434	0.012	0.425	0.012	0.422

Notes: This table reports OLS estimates of the correlation between math teacher bias, measured by IAT score, and share of top performing females in grade 6; the unit of observation is teacher t. Standard errors are robust and clustered at school level in parentheses. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for the order of IAT in the questionnaire administered. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.7: Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

Dep. Variable	**Std Math Grade 8**					
	All		**Female Teachers**		**Male Teachers**	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.080***	-0.036	-0.086***	-0.052	-0.037	0.044
	(0.014)	(0.032)	(0.015)	(0.035)	(0.036)	(0.081)
Fem*Teacher Bias	-0.027**	-0.028**	-0.029**	-0.030**	-0.029	-0.022
	(0.013)	(0.013)	(0.014)	(0.014)	(0.040)	(0.041)
Std Math grade 6	0.723***	0.697***	0.724***	0.698***	0.719***	0.695***
	(0.012)	(0.013)	(0.013)	(0.013)	(0.034)	(0.036)
Student controls	No	Yes	No	Yes	No	Yes
Obs.	9309	9309	8006	8006	1303	1303
R^2	0.618	0.625	0.619	0.626	0.603	0.613

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is math standardized test score in grade 8; 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 math teacher level in parentheses. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.8: Estimation of the effect of teachers' gender stereotypes on standardized test score in math in grade 8 for the different cohorts

	Dependent Variable: Math standardized test score in grade 8			
	All Students (1)	First Cohort (2)	Second Cohort (3)	Third Cohort (4)
Fem*Teacher Bias	-0.037***	-0.045	-0.031	-0.041**
	(0.014)	(0.040)	(0.023)	(0.019)
Std Math grade 6	0.699***	0.671***	0.689***	0.722***
	(0.013)	(0.032)	(0.021)	(0.016)
Constant	-0.112***	0.051	-0.182***	-0.111***
	(0.023)	(0.053)	(0.037)	(0.031)
Class FE	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
Obs.	9309	1984	4143	3182
R^2	0.626	0.623	0.608	0.661

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is math standardized test score in grade 8; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.9: Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

	Std Math 8th grade					
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.035	0.014	0.053	-0.045**	0.002	0.017
	(0.025)	(0.036)	(0.104)	(0.020)	(0.035)	(0.103)
Fem *No Bias	-0.045	-0.048	-0.042			
	(0.034)	(0.033)	(0.033)			
Fem*Pro Boys	-0.065**	-0.073**	-0.093***			
	(0.032)	(0.031)	(0.032)			
Fem*Positive Bias				-0.056**	-0.059**	-0.076***
				(0.026)	(0.026)	(0.026)
Std Math grade 6	0.723***	0.697***	0.699***	0.724***	0.698***	0.699***
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Obs.	9309	9309	9309	9309	9309	9309
R^2	0.618	0.625	0.626	0.618	0.625	0.626

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is math standardized test score in grade 8; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Instead of using a continuous variable as teacher bias we use categorical variables. In columns 1-3, we consider the thresholds defined by Greenwald et al. (2003) where no bias is the interval of IAT raw score between -0.15 and +0.15. In columns 4-6, we consider a positive or negative sign in the IAT score. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughthers, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.10: Estimation of the effet of teachers' gender stereotypes on track choice- school FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.093***	-0.049***	0.022	0.151*	0.004	-0.018*	0.010	0.005
	(0.014)	(0.013)	(0.019)	(0.089)	(0.010)	(0.011)	(0.022)	(0.067)
Fem*Teacher Bias		0.008	0.007	0.002		0.024***	0.022**	0.022**
		(0.011)	(0.011)	(0.011)		(0.009)	(0.009)	(0.009)
Teacher Bias		-0.007	-0.006	-0.005		-0.014**	-0.013**	-0.011
		(0.008)	(0.008)	(0.008)		(0.007)	(0.007)	(0.007)
Std Math grade 6		0.171***	0.151***	0.152***		-0.103***	-0.090***	-0.090***
		(0.007)	(0.008)	(0.008)		(0.006)	(0.006)	(0.006)
Constant	0.301***	0.248***	0.110***	0.021	0.141***	0.172***	0.209***	0.223***
	(0.011)	(0.010)	(0.015)	(0.074)	(0.008)	(0.008)	(0.017)	(0.052)
Obs.	8463	8463	8463	8463	8463	8463	8463	8463
R^2	0.067	0.171	0.193	0.197	0.068	0.144	0.164	0.169
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes

(Continues)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.058***	-0.031***	0.021	-0.002	-0.058***	-0.110***	-0.124***	-0.086
	(0.012)	(0.011)	(0.015)	(0.078)	(0.015)	(0.013)	(0.023)	(0.095)
Fem*Teacher Bias		0.000	-0.001	-0.007		0.018*	0.018*	0.022**
		(0.009)	(0.009)	(0.009)		(0.010)	(0.010)	(0.010)
Teacher Bias		-0.006	-0.006	-0.001		-0.015*	-0.016**	-0.016*
		(0.008)	(0.007)	(0.007)		(0.008)	(0.008)	(0.008)
Std Math grade 6		0.124***	0.110***	0.111***		-0.237***	-0.208***	-0.209***
		(0.009)	(0.008)	(0.008)		(0.008)	(0.008)	(0.008)
Constant	0.171***	0.147***	0.074***	0.060	0.366***	0.412***	0.509***	0.478***
	(0.011)	(0.010)	(0.013)	(0.066)	(0.015)	(0.013)	(0.019)	(0.081)
Obs.	7086	7086	7086	7086	7086	7086	7086	7086
R^2	0.097	0.187	0.198	0.203	0.103	0.321	0.351	0.355
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1.2, where the dependent variable is math standardized test score in grade 8; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 185. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract, education of the teacher' mother and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.11: Estimation of the effect of teachers' gender stereotypes on technical technological track

Dependent Variable: Track choice Technical Technological				
	(1)	(2)	(3)	(4)
Fem	-0.242***	-0.257***	-0.344***	-0.388***
	(0.011)	(0.011)	(0.024)	(0.087)
Fem*Teacher Bias		-0.022*	-0.019*	-0.015
		(0.012)	(0.011)	(0.012)
Std Math grade 6		-0.042***	-0.029***	-0.030***
		(0.007)	(0.007)	(0.007)
Fem*Math Teacher Fem				0.016
				(0.036)
Fem*North Math Teacher				0.008
				(0.023)
Fem*Advanced STEM Teacher				0.019
				(0.026)
Constant	0.310***	0.323***	0.409***	0.408***
	(0.006)	(0.006)	(0.018)	(0.018)
Class FE	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes
Obs.	8463	8463	8463	8463
R^2	0.199	0.205	0.218	0.220

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is the high-school track choice; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.12: Estimation of the effet of teachers' gender bias

Dependent Variable: Track choice Vocational						
Heterogeneous effects by	**Student Characteristics**				**Interaction time with teacher**	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	0.016	0.013	0.003	0.015	0.001	0.018
	(0.071)	(0.071)	(0.073)	(0.071)	(0.071)	(0.071)
Fem*Bias Teacher	0.020**	0.021	0.043**	0.017*	0.013	0.034*
	(0.009)	(0.015)	(0.021)	(0.009)	(0.011)	(0.020)
Fem*Bias T*HighEduM		-0.005				
		(0.017)				
Fem*Bias T*Top tercile Math6			-0.048**			
			(0.023)			
Fem*Bias T*Middle tercile Math6			-0.015			
			(0.026)			
Fem*Bias T*Immigrant				0.019		
				(0.027)		
Fem*Bias T*Extended School Day					0.032	
					(0.022)	
Fem*Bias T*Same Math Teacher						-0.016
						(0.023)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	8463	8463	8463	8463	8463	8463
R^2	0.211	0.211	0.216	0.211	0.211	0.211

Notes: This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on the choice of vocational high-school track by observable characteristics of the student and by interaction time with teacher; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student, "HighEduM" wether the mother has at least a diploma, "tercile Math6" is the tercile of standardized test score in math in grade 6 and "Immigrant" is a dummy equal to 1 if the student is not Italian citizen. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract and education of the teacher' mother. Regressions are all fully saturated even if not all interactions are shown in the table. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.13: Estimation of the effect of teachers' gender stereotypes on reading standardized test score in grade 8 - class FE regression

Dependent Variable: Reading standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	0.172***	0.076***	0.075***	0.079**	0.192*
	(0.020)	(0.015)	(0.015)	(0.034)	(0.113)
Fem*Teacher Bias			-0.021	-0.025*	-0.034**
			(0.015)	(0.014)	(0.015)
Fem*Teacher Fem					-0.055
					(0.047)
Fem*North Math Teacher					0.005
					(0.034)
Fem*Advanced STEM Teacher					-0.030
					(0.037)
Std Ita grade 6		0.738***	0.738***	0.702***	0.703***
		(0.012)	(0.012)	(0.013)	(0.013)
Constant	0.000	-0.064***	-0.064***	-0.203***	-0.202***
	(0.010)	(0.008)	(0.008)	(0.023)	(0.023)
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9285	9285	9285	9285	9285
R^2	0.190	0.591	0.592	0.601	0.602

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is reading standardized test score in grade 8; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.14: Estimation of the effect of Italian teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

Dependent Variable: Math standardized test score in grade 8			
	(1)	(2)	(3)
Female	-0.055***	-0.042	0.547***
	(0.016)	(0.033)	(0.142)
Fem*Ita Teacher Bias	0.005	0.001	0.004
	(0.016)	(0.016)	(0.018)
Fem*Ita Teacher Fem			-0.049
			(0.068)
Fem*Born North Ita Teacher			-0.045
			(0.037)
Std Math grade 6	0.722***	0.695***	0.695***
	(0.012)	(0.013)	(0.013)
Constant	0.016*	-0.105***	-0.104***
	(0.008)	(0.024)	(0.023)
Class FE	Yes	Yes	Yes
Student Controls	No	Yes	Yes
Teacher Controls	No	No	Yes
Obs.	8650	8650	8650
R^2	0.612	0.619	0.620

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is math standardized test score in grade 8; the unit of observation is student i, in class c taught by Italian teacher t in grade 8 of school s. Standard errors are robust and clustered at Italian teacher level in parentheses; the number of clusters is 352. The number of fixed effects (classes) is 504. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.15: Estimation of the effect of teachers' gender stereotypes on retention rate and on the probability of doing the standardized test score in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable	**Retention Rate**				**Doing Test in Grade 8**			
Fem	-0.027***	-0.037***	-0.048***	-0.046	0.014***	0.022***	0.006	-0.001
	(0.005)	(0.005)	(0.011)	(0.034)	(0.005)	(0.005)	(0.007)	(0.034)
Fem*Teacher Bias		0.004	0.004	0.003		0.002	0.004	0.005
		(0.005)	(0.005)	(0.005)		(0.005)	(0.005)	(0.005)
Std Math grade 6		-0.053***	-0.042***	-0.042***		0.040***	0.025***	0.024***
		(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)
Constant	0.060***	0.070***	0.064***	0.065***	0.939***	0.932***	1.051***	1.052***
	(0.002)	(0.003)	(0.010)	(0.010)	(0.002)	(0.003)	(0.007)	(0.007)
Obs.	9837	9837	9837	9837	9837	9837	9837	9837
R^2	0.099	0.136	0.153	0.154	0.175	0.200	0.390	0.391
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is the retention rate in columns 1-4 and the probability of doing the standardized test score in grade 8 in columns 5-8; 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 math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 551. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.16: Estimation of the effect of teachers' explicit and implicit bias on standardized test score in grade 8 - class FE regression

	Std Math 8th grade					
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.059***	0.040	0.026	-0.061**	0.005	-0.011
	(0.017)	(0.106)	(0.106)	(0.029)	(0.106)	(0.105)
Fem*Dif Innate Ability	0.033*	0.028	0.029*			
	(0.017)	(0.017)	(0.017)			
Fem*Teacher Bias			-0.036**			-0.041***
			(0.014)			(0.014)
Fem*WVS				0.020	0.014	0.012
				(0.033)	(0.032)	(0.031)
Std Math grade 6	0.719***	0.693***	0.693***	0.721***	0.695***	0.695***
	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	Yes	Yes	No	Yes	Yes
Obs.	8697	8697	8679	8952	8952	8934
R^2	0.618	0.626	0.627	0.619	0.626	0.627

Notes: This table reports OLS estimates, where the dependent variable is the standardized test score in grade 8; 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 math teacher level in parentheses. The variable "Fem" indicates the gender of the student, "Dif Innate Ability" assumes value -1 if the teacher think there are no innate differences in math abilities between men and women, 0 if there are few and 1 if there are up to several differences. "WVS" is the World Value Survey answer to the question on the right to access to jobs for men and women. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.17: Estimation of the effect of teachers' gender bias on self-confidence- school FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A- Dependent Variable: Being good/mediocre at math (vs. being bad)							
Fem	-0.124***	-0.092***	-0.102*	0.203	-0.073***	-0.075	0.181
	(0.028)	(0.027)	(0.061)	(0.185)	(0.027)	(0.061)	(0.179)
Fem*Teacher Bias		-0.035	-0.043*	-0.062**	-0.021	-0.028	-0.046
		(0.021)	(0.023)	(0.031)	(0.021)	(0.023)	(0.030)
Teacher Bias		0.009	0.016	0.030	0.015	0.022	0.017
		(0.016)	(0.016)	(0.020)	(0.016)	(0.016)	(0.018)
Std Test Math		0.128***	0.125***	0.131***	0.146***	0.139***	0.146***
		(0.021)	(0.020)	(0.022)	(0.022)	(0.021)	(0.023)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	747	747	747	747	747	747	747
R^2	0.076	0.188	0.208	0.242	0.220	0.237	0.269
Panel B- Dependent Variable: Being good/mediocre at Italian (vs. being bad)							
Fem	0.046*	0.045*	0.037	0.319*	0.037	0.028	0.141
	(0.025)	(0.025)	(0.049)	(0.188)	(0.023)	(0.047)	(0.192)
Fem*Teacher Bias		0.038**	0.040**	0.030*	0.039**	0.041**	0.035*
		(0.017)	(0.017)	(0.017)	(0.016)	(0.016)	(0.018)
Teacher Bias		-0.005	-0.006	0.002	-0.003	-0.004	0.003
		(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.016)
Std Test score Italian	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	664	664	664	664	664	664	664
R^2	0.043	0.068	0.085	0.153	0.080	0.092	0.149
Panel C- Dependent Variable: Average own ability in all other subjects							
Fem	0.038	0.025	0.020	-0.252	0.022	0.015	-0.227
	(0.029)	(0.031)	(0.062)	(0.221)	(0.031)	(0.063)	(0.221)
Fem*Teacher Bias		-0.011	-0.011	-0.021	-0.013	-0.014	-0.021
		(0.025)	(0.025)	(0.026)	(0.025)	(0.025)	(0.027)
Teacher Bias		0.021	0.022	0.030**	0.018	0.019	0.025*
		(0.019)	(0.019)	(0.013)	(0.019)	(0.019)	(0.013)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	802	802	802	802	802	802	802
R^2	0.030	0.059	0.072	0.140	0.063	0.075	0.144
School, Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	Yes	Yes
Math Teacher Controls	No	No	No	Yes	No	No	Yes

Notes: This table reports OLS estimates of equation 1.2, where the dependent variable is self-stereotypes in grade 8; 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 math teacher level in parentheses; the number of clusters is 58. The number of fixed effects (school, cohort) is 23. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother, and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

A.2 Survey to School Principals

School principals were asked to complete a paper questionnaire, including information about the career counseling service offered by the school to students, class formation criteria at the beginning of middle school and update courses offered to teachers. We received the questionnaire completed by 82 principals. Among them, 94% have a full-time contract, the average experience is 8.2 years and in 15% of schools the principal is mainly ruling a different complex of school and she has been assigned the direction of the institution in our sample temporarily until new staff is hired. This practice is wide-spread in Italy.

Among 81 principals who completed the questions on class formation, 64 percent consider the heterogeneity across classes in the ability level as “Extremely important” and 33 percent as “Important”. The heterogeneity across classes in the socio-economic status is considered as “Extremely important” by 60 percent of principals and “Important” by 29 percent. The equal allocation of immigrant across classes is considered “Extremely Important” by 25 percent of principals and as “Important” by 38 percent. The summary statistics on the importance given to the different criteria in class formation are summarized in Figure A.4.

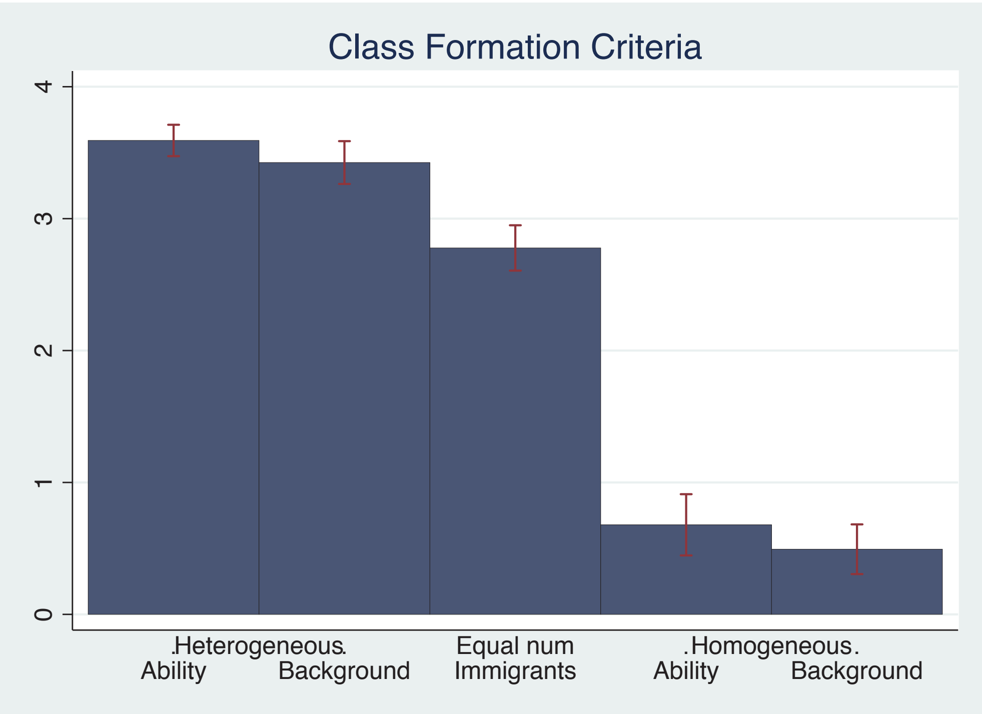


Figure A.4: Class formation criteria according with principals

An entire section of the questionnaire is dedicated to career counseling practices done in the school and the share of schools offering the different services is reported in Figure

A.5. Most institutions declare to offer information on high-schools curricula and attitudinal tests that help high-school choice. Around one third of schools organize meetings with psychologists at individual or group level to induce students to reflect on this important choice. However, very few institutions try to sensitize females toward STEM education.

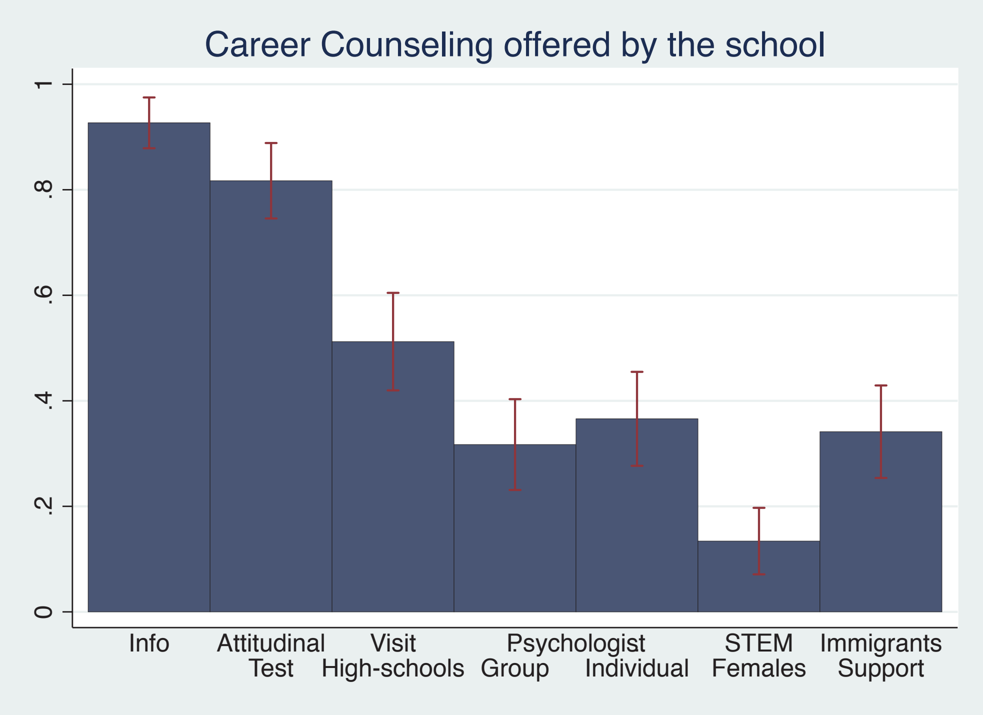


Figure A.5: Share of schools offering different career counseling services

Principals are asked to assign a score from 1 to 10 to the dedication and competence of math and Italian teachers in the middle school: generally principals are generous with the evaluation, but if anything, math teachers are considered as marginally less committed to their work and with lower level of competence than their Italian colleagues, as shown in Figure A.6.

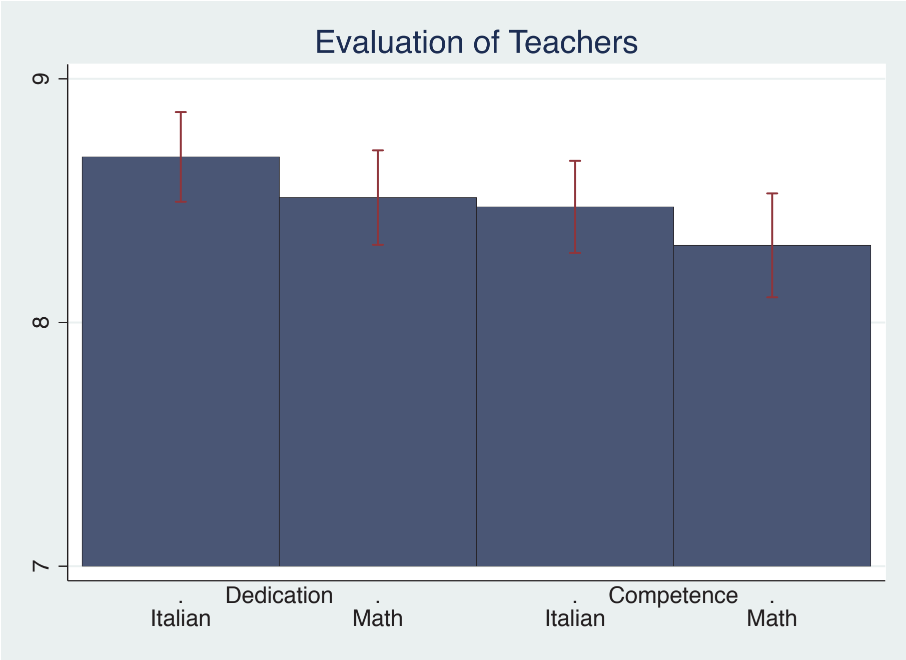


Figure A.6: Dedication and competence of teachers according with principals

A.3 Teacher Survey

A.3.1 Gender Implicit Association Test

We invite teachers to complete a seven-block IAT that was following the schematic overview presented in Figure A.7. Half of the teachers, randomly selected at individual level, completed the IAT, as presented in Figure A.7, while the other half completed the IAT with the blocks in the following order: 1, 5, 6, 7, 2, 3, 4 ("order incompatible" IAT). Furthermore, teachers were asked to complete a race IAT with male names and female names of Italian and immigrants. The order of race and gender IATs was randomized at individual level. In Table A.1, we check the influence of order of blocks on the IAT score. On average, there is a small difference in the IAT score between individuals that perform the order compatible and incompatible test. Hence, in all regression where there are no class (and therefore teacher) fixed effects we control for the order of IATs.

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

Blocks	Left Categories	Right Categories
1	Maschio (Male)	Femmina (Female)
2	Scientifico (Scientific)	Humanistic (Umanistico)
3	Maschio (Male) Scientifico (Scientific)	Femmina (Female) Humanistic (Umanistico)
4	Maschio (Male) Scientifico (Scientific)	Femmina (Female) Humanistic (Umanistico)
5	Humanistic (Umanistico)	Scientifico (Scientific)
6	Maschio (Male) Humanistic (Umanistico)	Femmina (Female) Scientifico (Scientific)
7	Maschio (Male) Humanistic (Umanistico)	Femmina (Female) Scientifico (Scientific)

Figure A.7: Schematic overview of the Gender Implicit Association Test

The stimuli presented within each category are summarized in Figure A.8. The improved IAT scoring procedure is used to calculate the bias from the reaction time of individuals to different associations and this method has been shown to be better account for method variance (Greenwald et al., 2003).

Categories	Stimuli
Maschio (Male)	Luca, Federico, Matteo, Alberto, Davide, Alessandro
Femmina (Female)	Anna, Martina, Laura, Giulia, Chiara, Alessia
Scientifico (Scientific)	Matematica (Math), Fisica (Physics), Scienze (Science), Chimica (Chemistry), Ingegneria (Engineering), Calcolo (Calculus)
Humanistic (Umanistico)	Lettere (Literature), Italiano (Italian), Filosofia (Philosophy), Letteratura (Literature), Storia (History), Lingue (Languages)

Figure A.8: Category Labels and Stimuli for the Implicit Association Tests

A.3.2 Teachers' Questionnaire

Factors influencing track choice

"Female students with the same math grade of males are less likely to attend a scientific track during high-school. According with your experience, how much can these factors influence the choice of females toward alternative tracks?" Answer in a scale from 1 to 5, where 1 means '*"Not at all"* and 5 means *"'A lot'*.

1. Low interest for scientific subjects
2. Low inclination for scientific subjects
3. Low self-esteem
4. Encouragement of the family toward alternative paths
5. Influence of gender predicament ("women are bad at math")

Factors influencing grading

"When you grade your students, which weight do you assign to the following components?" Answer in a scale from 1 to 5, where 1 means '*"A little"* and 5 means *"'A lot'*.

1. Performance in written exams
2. Performance in oral exams
3. Diligence in doing homeworks

Factors influencing track recommendation

"When you give the high-school track recommendation to your students, which weight do you assign to the following components?" Answer in a scale from 1 to 5, where 1 means '*"A little"* and 5 means *"'A lot'*.

1. Grades and performance at school
2. Predisposition and interests of the student
3. Parents' education
4. Economic resources of the family
5. Engagement of family in schooling

Explicit gender bias

Do you agree with the following sentences?

- *There are innate biological differences in math abilities of women and men*: '*Not at all*", "*A little*", "*A bit*","*A lot*", "*Absolutely*"
- *When jobs are scarce, men should have more right to a job than women*: '*Agree*", "*Neither Agree nor Disagree*", "*Disagree*"

A.4 Sample Selection

The purpose of this appendix is to investigate two issues related to sample selection: students who did not attend the standardized test score in grade 6 (or their test score was not correctly matched with data from the Ministry of Education) and teachers who did not complete the IAT test either because they were not teaching in the school anymore in 2016-17 or because they did not come to the meeting with enumerators.

In the school involved in this research project, 21,054 students attended the middle school exam in the years between 2013 and 2015. However, we do not have information on their initial test score for 12% of students. Table A.18 presents the differences between background characteristics and track choice of students, who attended the standardized test score in grade 6 and students who did not. The latter have an almost double probability of being immigrant and late in the school curricula in terms of grade achieved. They are in general more disadvantaged in terms of family background, achievements and they tend to choose a lower-level high -school track.

Table A.18: Summary Statistics of students by attendance to standardized test in grade 6

	Attendance	No Attendance	Diff.	se
Female	0.480	0.504	-0.024*	(0.011)
Late in school	0.367	0.093	0.274***	(0.007)
Immigrant	0.371	0.185	0.186***	(0.009)
Second Gen. Imm	0.081	0.079	0.002	(0.006)
High Edu Mother	0.331	0.428	-0.098***	(0.011)
High Occ Father	0.130	0.166	-0.036***	(0.008)
Med Occ Father	0.207	0.304	-0.097***	(0.010)
Std Math 8	-0.360	0.057	-0.417***	(0.021)
Std Reading 8	-0.471	0.074	-0.545***	(0.021)
High-school Track: Scientific	0.187	0.253	-0.066***	(0.010)
High-school Track: Classic	0.043	0.059	-0.016**	(0.006)
High-school Track: Other Acc	0.185	0.216	-0.031**	(0.010)
High-school Track: Technical Tech	0.196	0.185	0.011	(0.009)
High-school Track: Technical Eco	0.145	0.141	0.003	(0.008)
High-school Track: Vocational	0.245	0.146	0.099***	(0.009)
Track recommendation: Scientific	0.091	0.140	-0.049***	(0.009)
Track recommendation: Vocational	0.519	0.340	0.179***	(0.012)
Born South	0.046	0.029	0.017***	(0.004)
Observations	18586	2468		

Administrative data from MIUR and INVALSI.

The second issue regards data availability on teacher stereotypes. Indeed, enumerators were able to collect information only on around 80% of teachers currently working in the schools. Furthermore, teachers teaching to the three cohort graduating between 2013 and 2015 may have moved to a different school. Hence, we are able to obtain complete student-teacher data only on 9,309 students, 50% initial 18,586 students. Table A.19 provides evidence of whether student characteristics are correlated with the attendance of teachers of the IAT survey, within school. We find no significant difference among students of the school whose teacher was present and not present at IAT survey in terms of math ability, immigrant status and family background. There is only a small, but statistically significant, probability of male students being associated with a professor who completed IAT survey. Finally, in column 1, we also correlate the probability of teachers completing IAT with the type of class in terms of school hours without finding statistically significant effects.

Table A.19: Correlation between teacher attendance of the survey and student characteristics

Dependent Variable: Teacher who completed IAT								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Long School	0.055							0.054
	(0.043)							(0.043)
Female		-0.014**						-0.013**
		(0.006)						(0.006)
Std Math 6			0.004					0.001
			(0.005)					(0.008)
Std Math 8				0.006				0.005
				(0.005)				(0.008)
Immigrant					0.012			0.009
					(0.010)			(0.010)
HighEduMother						0.006		0.009
						(0.012)		(0.012)
HighOccFather							-0.015	-0.020
							(0.014)	(0.014)
MedOccFather							-0.000	-0.003
							(0.009)	(0.010)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	18530	18586	18586	18586	18586	18586	18586	18530
R^2	0.173	0.171	0.171	0.171	0.171	0.171	0.171	0.173

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports OLS estimate, where the dependent variable is a dummy variable which assumes value 1 if the teacher completed the IAT; 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 class level in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively. For few observations, we do not know the number of hours at school. This is the reason why the number of observation is slightly different in columns 1 and 8.

A.5 Implicit Bias of Italian Teachers

The main focus of the paper is on the role of math teacher gender bias and how it affects gender gap in math performance. However, I have also collected information through the same questionnaire administered to Italian teachers and the purpose of this appendix is to provide complementary evidence on how reading performance is affected by interaction with Italian teachers.

Table A.20 summarizes the main observable characteristics of Italian teachers used for this analysis: 91% of teachers are female, almost all have a full time contract, 26 percent are from the South of Italy and on average almost 24 years of experience as teachers. Female Italian teachers have a stronger gender bias, as measured by IAT score: they are faster in associating own gender with the subject they teach, as seen in Table A.21. This result is coherently with findings of Rudman et al. (2001) according to which individuals possess implicit gender stereotypes in self-favorable form because of the tendency to associate self with desirable traits.

Finally, in Table A.22, we report the correlation between student characteristics and teachers' implicit gender bias, measured by IAT score. The interaction between pupil gender and characteristics is never statistically significant, with few exception where the sign is in opposite direction with respect to the expectations that better females are assigned to less biased teachers.

The gender gap in reading is reversed compared to the one in math: females are 0.19 standard deviation better in Italian compared to males at the end of the middle school and the gap is increasing of 0.081 standard deviation from grade 6 to 8, as shown in the first two columns of Table A.23. This difference is similar to the one in most OECD countries (Fryer Jr and Levitt, 2010). However, teachers' gender stereotypes are not affecting this gap. Indeed, subsequent columns investigate the impact on the gender achievement gap in reading of Italian teacher gender bias. The results are not affected by the introduction of controls at student and teacher level as in Table 1.5. The effect of interest is close to and indistinguishable from zero. The bias of Italian teachers does have impact neither on males nor on females. Table A.24 examine the effect comparing students of the same gender enrolled in the same school in the same year (as in equation 1.2) and shows that there is no impact on both genders when controlling for teachers' characteristics that are crucial when there are only school by year fixed effects.

Teacher Bias

Table A.20: Summary Statistics from Italian Teachers' Questionnaire

Family and education					
Female	431	0.91	0.29	0.00	1.00
Born in the North	414	0.74	0.44	0.00	1.00
Age	412	51.37	7.44	33.00	66.00
Children	431	0.73	0.45	0.00	1.00
Number of children	308	1.79	0.82	0.00	5.00
Number of daughters	308	0.81	0.74	0.00	4.00
Low edu Mother	380	0.52	0.50	0.00	1.00
Middle edu Mother	380	0.38	0.49	0.00	1.00
High edu Mother	380	0.11	0.31	0.00	1.00
Advanced STEM	425	0.00	0.00	0.00	0.00
Degree Laude	369	0.29	0.46	0.00	1.00
Job characteristics					
Full time contract	417	0.99	0.11	0.00	1.00
Years of experience	417	23.70	9.59	2.00	43.00
Math Olympiad	425	0.00	0.00	0.00	0.00
Update Courses	425	0.93	0.25	0.00	1.00
Satisfy with teacher job	414	3.89	0.88	1.00	5.00
Implicit bias					
IAT Gender	431	0.39	0.39	-0.78	1.29
Self-reported explicit bias					
WVS Gender Equality	411	0.10	0.31	0.00	1.00
Gender Dif Innate Ability	402	1.37	0.65	1.00	3.00
Reason GenderGap: Interest for STEM	368	2.65	1.00	1.00	5.00
Reason GenderGap: Predisposition for STEM	342	2.16	1.05	1.00	5.00
Reason GenderGap: Low self-esteem	401	2.54	1.09	1.00	5.00
Reason GenderGap: Family support	400	2.94	1.07	1.00	5.00
Reason GenderGap: Cultural Stereotypes	398	2.17	1.17	1.00	5.00
Boys better in Invalsi	334	0.10	0.30	0.00	1.00
Girls better in Invalsi	334	0.53	0.50	0.00	1.00
Gender Equal in Invalsi	334	0.37	0.48	0.00	1.00
Observations	431				

Notes: First-hand data from teachers' questionnaire. We restrict the sample to teachers matched to students and therefore used in the main analysis of this paper. The balance table with the difference between teachers' matched and not matched with students' data is presented in Table A.2. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

Table A.21: Correlation between teachers' characteristics and Gender IAT Score

Dependent variable: raw IAT score of Italian teachers				
	Restricted sample		**All teachers**	
	(1)	(2)	(3)	(4)
Female	0.554***	0.555***	0.555***	0.555***
	(0.071)	(0.101)	(0.047)	(0.056)
Born in the North	-0.065*	-0.015	-0.048*	-0.008
	(0.033)	(0.051)	(0.025)	(0.032)
Age	-0.053*	-0.034	-0.001	0.003
	(0.029)	(0.041)	(0.014)	(0.017)
Age sq.	0.001*	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
High Edu Mother	0.013	-0.013	0.006	-0.001
	(0.040)	(0.053)	(0.027)	(0.032)
Children	-0.034	0.083	0.013	0.015
	(0.197)	(0.227)	(0.116)	(0.129)
Daughters	0.163***	0.198***	0.052	0.068*
	(0.038)	(0.046)	(0.032)	(0.035)
Degree Laude	-0.027	-0.008	0.030	0.053*
	(0.036)	(0.047)	(0.027)	(0.031)
Full time contract	0.072	-0.031	0.042	0.028
	(0.095)	(0.164)	(0.042)	(0.049)
Satisfy with teacher job	0.011	0.005	0.010	0.002
	(0.022)	(0.030)	(0.015)	(0.017)
Refresher courses	0.008	0.023	0.019	0.033*
	(0.024)	(0.030)	(0.017)	(0.020)
Explicit Bias	0.009	0.017	-0.018	-0.023
	(0.021)	(0.028)	(0.014)	(0.016)
Constant	1.112	0.602	-0.182	-0.290
	(0.755)	(1.105)	(0.343)	(0.436)
Obs.	431	431	844	844
R^2	0.244	0.403	0.244	0.325

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and own teacher characteristics; the unit of observation is teacher t in school s. Standard errors are robust and clustered at school level in parentheses. The description of all variables is the same as in Table 1.3. We include the order of IATs for math teachers (if the first one was the gender IAT and if the first associations were order compatible or not) and missing categories if the information is not available. The restrictd sample includes data expoits on teachers in the main regressions of this paper. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.22: Exogeneity of assignment of students to Italian teachers with different bias

Dependent Variable: Italian Teacher implicit gender bias (standardized)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.019	0.035	-0.020	0.021	0.016	-0.033	0.309
	(0.013)	(0.025)	(0.024)	(0.015)	(0.014)	(0.175)	(0.189)
Fem*HighEduMother		-0.029				-0.032	-0.002
		(0.035)				(0.031)	(0.050)
HighEduMother		0.047*				0.045**	0.010
		(0.027)				(0.022)	(0.028)
Fem*Med Occupation Father			0.049			0.056*	-0.024
			(0.035)			(0.032)	(0.048)
Med Occupation Father			-0.011			-0.025	0.053
			(0.029)			(0.027)	(0.038)
Fem*High Occupation Father			0.069*			0.068*	0.087
			(0.041)			(0.037)	(0.090)
High Occupation Father			0.044			0.038	-0.034
			(0.032)			(0.030)	(0.059)
Fem*Immigrant				-0.010		0.006	0.039
				(0.039)		(0.041)	(0.065)
Immigrant				-0.025		-0.033	-0.058
				(0.034)		(0.034)	(0.052)
Fem* Std Math grade 6					0.015	0.011	0.060**
					(0.016)	(0.014)	(0.023)
Std Math grade 6					-0.014	-0.026*	-0.086***
					(0.015)	(0.013)	(0.021)
Fem* Std Ita grade 5							-0.059*
							(0.033)
Std Ita grade 5							0.034
							(0.021)
Constant	0.121	0.091	0.122	0.126	0.125	-0.731	-2.072***
	(0.096)	(0.098)	(0.101)	(0.095)	(0.097)	(0.547)	(0.711)
School,year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Control	No	No	No	No	No	Yes	Yes
Obs.	8639	8639	8639	8639	8635	8635	1404
R^2	0.387	0.387	0.388	0.387	0.387	0.504	0.794

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and students' characteristics; 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 math teacher level in parentheses; the number of clusters is 353 in columns 1-6 and 137 in column 7. The variable "Fem" indicates the gender of the student, "HighEduMother" assumes value 1 if the mother has at least a 5 years diploma, "Medium Occupation Father" assumes value 1 if the father is a teacher or office worker, while "High Occupation Father" is 1 if the father is manager, university professor or an executive. "Immigrant" assumes value 1 is the student is not an Italian citizen, while "Std Mat grade 6" and "Std Ita grade 5" are the standardized test score in grade 6 in math and grade 5 in Italian respectively. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for the order of IAT in the questionnaire administered. For 55 students we do not observe the math test score in grade 6. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.23: Estimation of the effect of teachers' gender stereotypes on reading standardized test score in grade 8 - class FE regression

Dependent Variable: reading standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	0.194***	0.081***	0.080***	0.089***	0.564**
	(0.022)	(0.016)	(0.016)	(0.032)	(0.258)
Fem*Ita Teacher Bias			0.008	0.005	0.008
			(0.015)	(0.014)	(0.016)
Fem*Ita Teacher Fem					-0.011
					(0.052)
Fem*Ita Teacher Born North					-0.016
					(0.031)
Std Ita grade 6		0.730***	0.730***	0.693***	0.691***
		(0.016)	(0.016)	(0.018)	(0.018)
Constant	-0.003	-0.059***	-0.059***	-0.223***	-0.222***
	(0.011)	(0.008)	(0.008)	(0.024)	(0.024)
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	8639	8639	8639	8639	8639
R^2	0.185	0.591	0.591	0.603	0.604

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is Italian standardized test score in grade 8; the unit of observation is student i, in class c taught by Italian teacher t in grade 8 of school s. Standard errors are robust and clustered at Italian teacher level in parentheses; the number of clusters is 226. The number of fixed effects (classes) is 289. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between reading standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, degree with honour, update courses, age, type of contract, and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.24: Estimation of the effect of teachers' gender stereotypes on reading standardized test score in grade 8 - school FE regression

Dependent Variable: reading standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	0.197***	0.092***	0.091***	0.097***	0.586**
	(0.024)	(0.018)	(0.018)	(0.032)	(0.246)
Fem*Ita Teacher Bias			0.002	-0.001	0.005
			(0.014)	(0.014)	(0.015)
Ita Teacher Bias			0.042**	0.040**	0.019
			(0.019)	(0.019)	(0.017)
Fem*Female Ita Teacher					-0.023
					(0.056)
Female Teacher					0.142**
					(0.072)
Fem*North Ita Teacher					-0.005
					(0.030)
North Ita Teacher					-0.022
					(0.030)
Std Ita grade 6		0.722***	0.722***	0.679***	0.678***
		(0.016)	(0.016)	(0.017)	(0.017)
Constant	0.007	-0.068***	-0.064***	-0.229***	-0.249
	(0.021)	(0.017)	(0.024)	(0.032)	(0.168)
School, year FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	8639	8639	8639	8639	8639
R^2	0.114	0.543	0.544	0.559	0.561

Notes: This table reports OLS estimates of equation 1.2, where the dependent variable is reading standardized test score in grade 8; 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 Italian teacher level in parentheses; the number of clusters is 226. The number of fixed effects (school by cohort) is 146. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between reading standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, degree with honor, update courses, age, type of contract, education of the teacher' mother, and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

A.6 Bias in Grading

Previous literature has shown the importance of gender bias in grading (i.e the gender difference in standardized test score and unblind grades given by teachers) in affecting female improvements in math and university choice (Lavy and Megalokonomou, 2017; Lavy and Sand, 2015; Terrier, 2015). A natural question is whether IAT score affects bias in grading of teachers. I have information on grade given by teachers at the end of the semester.

As shown in Table A.25, girls on average get a higher grade compared to boys with the same standardized test score in math[1]. Females assigned to more biased teacher get a slightly lower grade, but the effect is small in magnitude and indistinguishable from zero. However, it should be consider that grades are categorical variable from 2 to 10, where 6 is the pass grade. As it can be clearly seen by Figure A.9, there is an extremely high bunching at the pass grade (6) and almost half of the students obtain this grade in math. Hence, it is not surprising that we don't detect an effect on this variable.

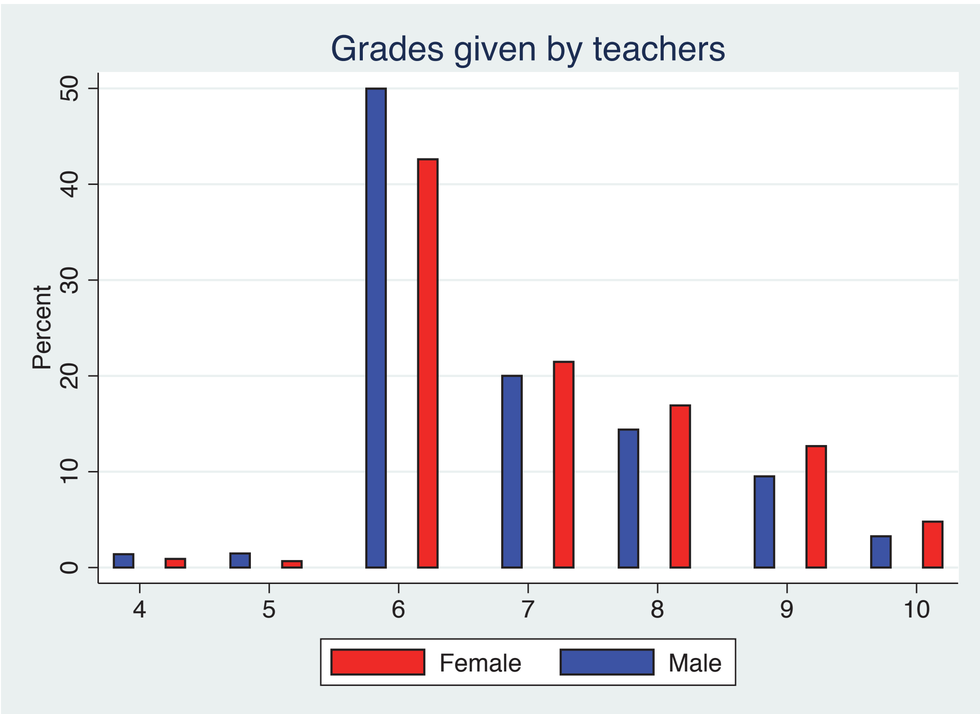


Figure A.9: Grades given by teachers

[1]Here, the results are shown controlling for the standardized test score in grade 6, but the magnitude and significance is very close when controlling for the standardized test score in grade 8.

Table A.25: Estimation of the effect of teachers' gender stereotypes on grading by teacher

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Math Grade end middle school**				**High Math Grade (>than6)**			
Fem	0.204***	0.354***	0.339***	0.260	0.070***	0.121***	0.116***	0.027
	(0.027)	(0.023)	(0.049)	(0.192)	(0.011)	(0.010)	(0.022)	(0.082)
Fem*Teacher Bias		-0.016	-0.018	-0.031		-0.004	-0.005	-0.013
		(0.025)	(0.025)	(0.025)		(0.010)	(0.010)	(0.011)
Fem* Teacher Fem				-0.013				-0.029
				(0.066)				(0.028)
Std Math grade 6		0.759***	0.716***	0.718***		0.262***	0.244***	0.244***
		(0.018)	(0.019)	(0.019)		(0.006)	(0.007)	(0.007)
Constant	6.980***	6.837***	6.645***	6.645***	0.509***	0.460***	0.388***	0.390***
	(0.013)	(0.012)	(0.037)	(0.037)	(0.005)	(0.005)	(0.016)	(0.017)
Obs.	9082	9082	9082	9082	9082	9082	9082	9082
R^2	0.118	0.433	0.446	0.447	0.104	0.331	0.345	0.346
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	No	Yes	Yes
Math Teacher Controls	No	No	No	Yes	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1.1, where the dependent variable is the grades given by teachers in columns 1-4 and a dummy for a grade higher than 7 in column 4-8; 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 math teacher level in parentheses; the number of clusters is 300. The number of fixed effects (classes) is 526. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

A.7 Conceptual Framework

I develop a simple conceptual framework, similar to Dee (2014) and based on the stereotype threat idea developed in social psychology that can help interpreting the results. In this model, student beliefs about own ability in math is a function of own true unobserved ability a_i and teachers' gender stereotypes s_t, defined as follows: $\alpha_i = f^i(a_i, s_t)$. The impact of the bias on own self-perception is an individual specific function: students with higher vulnerability to the stereotype that "girls are not good at math" will be more negatively impacted by teacher stereotype. This simple framework is flexible enough to capture heterogeneous perception of own ability among students with the same true unobserved ability a_i. For instance, a boy may perceive higher ability in math compared to a girl with the same unobserved talent for math a_i in the same class (i.e. exposed to the same teacher with stereotypes s_t). I assume that students' beliefs about own ability is a weakly decreasing function of teachers' stereotype, i.e. $\alpha_s \leq 0$. This is a testable assumption, which is supported by the empirical evidence available in Section 2.6 of the main paper.

In a simple framework, students choose effort and individual's utility can be represented as

$$u_i = \theta_i k(\alpha_i, e_i) - c(e_i) \tag{A.1}$$

where u is differentiable and the sufficient conditions for a local interior maximum hold, k is the benefits function, which depends on ability (α_i) and effort (e_i), and the cost c is paid according with the level of effort e exerted by individual i. The component θ_i introduces an exogenous heterogeneity and it captures observable difference across individuals in the returns to performance. In this simple framework, I do not introduce parametric assumptions on the utility function.

I am interested in how the optimal level of effort of students varies with stereotype of the teacher. The model implies that:

$$e_s^* = \frac{\theta k_{e\alpha} \alpha_s}{-(\theta k_{ee} - c_{ee})} \gtreqless 0 \tag{A.2}$$

The second order condition for a relative maximum implies that the second order derivative must be negative and therefore the denominator in equation (2) must be positive. Furthermore, I assumed that $\alpha_s \leq 0$, which implies that higher teacher stereotypes have a negative or null impact on self-perception of student i's ability, *ceteris paribus*. Hence, the optimal level of effort with respect to stereotype (e_s^*) depends on the complementarity or substitutability of effort and perceived ability ($k_{e\alpha}$).

Effort and perceived ability are often considered as complementary in the education production function (i.e. $k_{e\alpha} > 0$), so that a higher self-assessment of own capacities enhances the motivation to exert effort (Bénabou and Tirole, 2002). Hence, higher stereotypes will decrease the level of effort in equilibrium ($e_s^* \leq 0$). However, if the student increases effort as a reaction to a negative stance of the math teacher (i.e. $k_{e\alpha} < 0$), then the impact of stereotypes on effort is positive. As suggested also by Dee (2014), $e_s^* \geq 0$ is likely for instance if individuals of the stigmatized group consider the stereotype strongly improper and react with an *"I'll show you are wrong"* attitude. In the context of gender stereotypes, it would imply that talented female students may increase the level of effort when they interact with teachers with stronger bias in order to disprove the negative belief.

In the empirical counterpart of this model, I observe improvements in achievement test scores (P) and not directly effort[2], but I assume for simplicity that the derivative of performance with respect to effort is positive ($P_e > 0$) and I focus on the choice of the latter. Indeed, in the paper, I analyze whether improvements in achievement test scores are affected by teacher stereotypes. Assume two students, with the same gender, family background and math performance, are quasi-randomly assigned to two different teachers, respectively with stereotypes s_{t_i} and s_{t_j}, such that $s_{t_i} < s_{t_j}$. Then, if effort is complementary of students' perceived ability, the optimal level of effort (and therefore performance) of the student decreases with teachers' stereotypes. However, if $k_{e\alpha} < 0$, then $e_s^* > 0$. This theoretical results may explain why girls in the top of the initial ability distribution have slightly higher, but yet indistinguishable from zero, improvements in math when exposed to teachers with stronger gender stereotypes.

A.7.1 Extension of the Conceptual Framework

This simple framework can be extended to include teachers' behavior toward the pupils (β_{t_i}). This is an additional channel through which teacher bias (s_t) may impact on students' performance and choices. I define β_{t_i} as an individual specific function of (s_t): $\beta_{t_i} = h^i(s_t)$. Furthermore, I assume that teachers with higher stereotypes are less supportive toward member of the stigmatized group, i.e. $\beta_s \leq 0$. Unfortunately, I do not observe data on gender specific investment or interaction in the classroom between professors and students, but the social psychology literature described in Section 2.6 provides evidence in support of this assumption.

I extend the simple conceptual framework to analyze the impact of teachers' gender stereotypes on effort of students, as mediated by both student perception of own ability (α_i) and teacher investment toward pupils, in the form of either time or encouragement,

[2]Even ideally having information about the number of hours studied, it is not clear that this is necessarily a better measure of effort since the quality of time use is also essential in the learning process

(β_{t_i}). The individual chooses the level of effort in order to maximize:

$$u_i = \theta_i k(\alpha_i, e_i, \beta_{t_i}) - c(e_i) \tag{A.3}$$

where $\beta_{t_i} = h^i(s_t)$ and all other parameters and functions are defined as in equation (2).

The optimal level of effort with respect to teacher stereotypes is given by:

$$e_s^* = \frac{\theta(k_{e\alpha}\alpha_s + k_{e\beta}\beta_s)}{-(\theta k_{ee} - c_{ee})} \gtreqless 0 \tag{A.4}$$

In this extended framework, whether students increase or decrease their effort level when exposed to more biased teachers will depend both on the complementarity and substitutability of effort with both own perceived ability ($k_{e\alpha}$) and teacher behaviour $k_{e\beta}$. If both are complement, then the student will decrease the level of effort when exposed to a teacher with higher bias ($e_s^* < 0$). If both are substitute, we are in the case in which students work harder when exposed to more biased teachers to disprove the negative belief.

A.8 Math Performance: Talent vs. Effort

Research in social psychology shows that school teachers with higher gender stereotypes believe that males are more talented than females, even among students with comparable level of ability (Tiedemann, 2002). The performance of females is mainly attributed to effort.

In this Appendix, I provide evidence that teacher perception of gender differences in effort and ability mirrors (or is mirrored by) pupils own perception. I have information on a sample of around 18000 Italian students in grade 6 collected by the Italian National Evaluation Center (INVALSI) on the reasons why students believe they are performing well in both math and reading. In particular, students are asked: *"You solved all exercises in a math exam correctly. Why has it happened?"* and *"You need to tell your classmates about a text you read. You do it in a clear and precise way and everybody follows what you are saying. Why has it happened?"* Students can choose among different potential answers: "I was helped", "I was lucky", "It was easy", "I am good" and "I have exerted a lot of effort". Here I focus on the latter two answers. Female students are 7.6 and 2.3 percentage points less likely to say "I am good" in math and reading respectively than males (column 1 and 2, Table A.26). The difference in under-confidence between the two subjects is statistically significant for females, even controlling for individual level fixed effects (column 4). Finally, in column 5, I show that females are more likely than males to believe the most important reason of their success is effort. However, the difference between subjects is slightly lower, although statistically significant at 1 percent level. Unfortunately, in this dataset we do not have information on teacher gender beliefs to access whether they affect the pupil beliefs about the reason of successful performance in math and reading.

Table A.26: Correlation between subject, gender and own assessment

Dependent Variable: Reasons for performing well								
	Being good				**Exerting effort**			
	Math	Reading	Reading and Math		Math	Reading	Reading and Math	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.076***	-0.023***	-0.018***		0.139***	0.089***	0.088***	
	(0.005)	(0.006)	(0.006)		(0.007)	(0.007)	(0.007)	
Math			-0.025***	-0.026***			0.080***	0.080***
			(0.006)	(0.008)			(0.007)	(0.009)
Female*Math			-0.062***	-0.062***			0.032***	0.031**
			(0.008)	(0.011)			(0.009)	(0.013)
Math test	0.003***		0.001***		0.001***		-0.001***	
	(0.000)		(0.000)		(0.000)		(0.000)	
Reading test		0.000***	-0.000			0.005***	0.005***	
		(0.000)	(0.000)			(0.000)	(0.000)	
Individual FE	No	No	No	Yes	No	No	No	Yes
Obs.	16952	16979	33835	34028	16952	16979	33835	34028
R^2	0.042	0.001	0.018	0.531	0.025	0.038	0.048	0.605

Notes: Robust Standard Errors clustered at individual level in parentheses. This information is collected together with the standardized test score in grade 6. "Female" refers to the gender of students. "Math" is a dummy which assumes value 1 if the dependent variable is related to own assessment in math and 0 if it is related to own assessment in reading. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

2. Goals and Gaps: Educational Careers of Immigrant Children *(joint with E. La Ferrara and P. Pinotti)*

Abstract
We study the educational choices of children of immigrants in a tracked school system. We first show that immigrant boys in Italy enroll disproportionately into vocational high schools, as opposed to technical and academically-oriented high schools, compared to natives of similar ability. Immigrant girls, instead, choose similar schools as native ones. We then estimate the impact of a large-scale, randomized intervention providing tutoring and career counseling to high-ability immigrant students. Male treated students increase their probability of enrolling into the high track to the same level of natives, also closing the gap in terms of grade retention. There are no significant effects on immigrant females, who exhibit similar choices and performance as native ones in absence of the intervention. Increases in academic motivation and the resulting changes in teachers' recommendation regarding high school choice explain a sizable portion of the effect, while the effect of increases in cognitive skills is negligible. Finally, we find positive spillovers on immigrant classmates of treated students, while there is no effect on native classmates. *JEL: I24, J15* .

2.1 Introduction

Migrant flows have grown considerably over the past decades, increasingly involving families with children. In 2012, 12 percent of 15-year-old students in the average OECD country had an immigrant background (OECD, 2015b). The growing number of immigrant students has profoundly changed the challenges that schooling systems have to face in order to ensure skill development in a diverse student population and promote social cohesion. The ethnic gap in achievement test scores and socioemotional abilities increases substantially during childhood (Fryer Jr and Levitt, 2004; Heckman et al., 2006; Cunha and Heckman, 2007). The problem is exacerbated in schooling systems characterized by stratification in high school tracks, as early tracking may lead to the educational segregation of children from disadvantaged backgrounds in schools characterized by lower quality of education. This could ultimately have long term effects on the skills and occupational careers of children from immigrant families, reducing social mobility and creating unequal opportunities (Guyon et al., 2012; Brunello and Checchi, 2007).

This paper documents the extent of educational segregation of immigrant students and evaluates the effectiveness of an innovative program aimed at steering high-achieving immigrants towards high schools that fit their academic potential. We do so in the context of Italy, where the schooling system is characterized by tracking in the transition from middle to high school. While uncommon in the Anglosaxon world, this type of stratification is the norm among OECD countries, with the age of selection varying from 10 to 16 and an average of three high school tracks per country (OECD, 2013).

We start by showing that immigrants tend to choose vocational over technical or academic-oriented curricula, relative to native students with similar ability. We denote this phenomenon as ‘educational segregation’. Importantly, we are able to control for students’ ability using a standardized test administered at the beginning of middle school. The gap in high school choices is greater for male immigrants and it mirrors an analogous differential in failure rates and in the track recommendations received from teachers. The gap in track choice for boys persists along the entire distribution of ability, while for girls it is found at the low end of the distribution but not at the high end: medium and high-achieving immigrant girls choose the same tracks as native ones of comparable ability. This gender pattern is consistent with evidence from various countries that boys increasingly lag behind in educational attainment and that the female-male educational advantage is larger for low-socioeconomic status (SES) families (Figlio et al., 2016).

We then estimate the impact of an innovative program called “Equality of Opportunity for Immigrant Students” (EOP henceforth) that provided tutoring and career counseling to immigrant children displaying high academic potential. The curriculum of EOP included a number of meetings who helped students reflect on their aspirations and their potential through a series of psychological exercises based on Social Cognitive Career Theory (Lent et al., 1994). This paradigm views career development as a choice subject to contextual influences and constraints, so goals and self-efficacy are as important as cognitive skills in shaping individual careers. Importantly, treated students were never advised to choose one track (e.g., academic or vocational) over the others: they were only encouraged to weigh their talents and aspirations while making their decision.

We evaluate the effects of EOP leveraging random assignment of the program across schools as well as unique data on students' careers, cognitive and soft skills, and parental background. The program was offered in 70 schools randomly chosen from a sample of 145 in Northern Italy, with the remaining 75 serving as controls. We obtained restricted-use data on educational careers and standardized test scores for all students in treated and control schools from the Italian Ministry of Education. Based on this information, we selected the 10 immigrant students with the highest standardized test scores in grade 6: in treatment schools these students were invited to participate in EOP through grades 7 and 8; in control schools they were not offered such program.[1]

We find that EOP was remarkably successful in reducing educational segregation. Treated males have a 44 percent lower probability to be retained and a 12 percent higher probability of attending an academic or technical high school (as opposed to a vocational one) relative to males in the control group. Indeed, treated immigrant males chose more demanding schools in the same proportion as native males of comparable ability. In other words, EOP completely closed the immigrant-native track choice gap at the end of grade 8. The effects are in the same direction but smaller and not significant for girls (for whom no educational segregation had been identified in the first place). Therefore, EOP increases immigrants' enrollment into technical or academic schools only when counterfactual enrollment rates lie below those of comparable native students.

To shed light on the mechanisms underlying these effects, we collected data on academic performance and on psychological traits. We find that male treated students display an improvement in cognitive skills, as measured by their standardized test score at the end of grade 8. Importantly, their soft skills also improve: they have higher aspirations and more confidence in their own abilities, and they perceive that environmental barriers will play a smaller role in their future choices, relative to the control group. All these improvements seem to have been internalized by teachers, who recommend treated boys for more demanding high schools. No effect is found on teachers' recommendations for treated girls, whose academic performance and aspirations were unaffected by the program. Following the approach of Heckman et al. (2013), we decompose the impact of EOP on (males') ed-

[1]For the sake of exposition, in the rest of the paper we refer to these two groups as "treated" and "control", though we really mean "assigned to treatment" and "assigned to control". In other words, we present intention-to-treat estimates. We assess compliance with treatment assignment in Section 2.4.

ucational choices into experimentally induced changes in these observed mediating factors and changes in other (unmeasured) factors. We find that changes in aspirations and in teachers' recommendations induced by the treatment explain a sizable portion of the effect on track choice, while the effect of increases in test scores is negligible. Overall, these factors jointly explain 84 percent of the increase in the probability of choosing the high track, suggesting that other (unmeasured) factors account for a minor part of changes in educational choices.

We also distinguish between the effects of the two main components of EOP –career counseling and academic tutoring– exploiting discontinuous changes in the amount of academic tutoring around fixed thresholds of pre-program standardized test scores. We fail to detect significant changes in the outcomes of interest around such thresholds, suggesting that the effect of EOP can largely be attributed to career counseling. This last finding is consistent with the results of the decomposition analysis, i.e., that changes in motivation play a greater role than changes in test scores in explaining track choice.

Our administrative data allow us to follow all students in our sample through the first two years of high school. This is important in order to assess the longer term effects of EOP on students' educational careers, particularly the risks associated with enrolling into more demanding high school tracks. Reassuringly, treated students are no more likely to be held back in a grade, nor to have to take make-up exams after the summer (as opposed to being directly admitted to the next grade). They are also no more likely to drop out. Therefore, treated students did equally well as control ones, despite attending (on average) more demanding high schools.

Finally, we find evidence of positive spillovers of the intervention on immigrant classmates of treated students, while there is no effect on native classmates.

Our work is related to several strands of literature. The first comprises evaluations of interventions aimed at reducing inequality in educational achievement and opportunities. Several interventions have targeted low achieving students and provided a combination of information on school options and mentorship on soft skills. Some of these programs were successful in reducing grade retention and high school dropout rates (e.g., Goux et al., 2017; Martins, 2010); others had zero or negative effects (Rodriguez-Planas, 2012). Our program can be seen as complementary to the ones cited above, as it targets a different population –high achieving students– with the aim of aligning their potential to more ambitious goals.[2] In this respect, the framing of our intervention is comparable to programs that help low income students apply to better colleges (e.g., Hoxby and

[2]Different types of interventions have targeted students' inaccurate beliefs about their chances of succeeding in competitive academic environments. Bobba and Frisancho (2016) provided students with feedback on their performance in a (mock) exam for admission to selective high schools in Mexico, while Goodman (2016) estimates the effect of mandated exams on selective college admission in the US.

Turner, 2015; Bettinger et al., 2012), with the important difference that we work with a younger population that ends up being stratified within compulsory schooling. Indeed, in educational systems characterized by early tracking, high school choice constitutes the most critical juncture in students' careers.[3]

A second strand of literature looks at the role of soft skills. Heckman et al. (2006), Heckman and Kautz (2012), and Kautz et al. (2014) stress the importance of personality traits and motivations as factors that reduce social problems and are highly valued in the labor market. In addition, low aspirations and high perceived socioeconomic barriers can lead students to choose less demanding educational paths, perpetuating a negative cycle (Dalton et al., 2014; Genicot and Ray, 2014; Mookherjee et al., 2010). Recent contributions show that students from disadvantaged backgrounds often lack ambition and suffer from negative stereotypes, both at the high end (Hoxby and Avery, 2012) and at the low end (Guyon and Huillery, 2016) of the ability distribution. Our paper corroborates these findings for high ability students with an immigrant background and – importantly – shows that it is possible to modify aspirations and soft skills through a program that combines academically relevant information with psychological tools.[4]

Finally, our work speaks to the literature on tracking within education systems.[5] Some of the existing evidence points to the positive effects of tailored instruction (Duflo et al., 2011) and to a lack of negative consequences of school stratification in the long run (Dustmann et al., 2017). Other authors find that postponing school stratification or increasing the proportion of seats in academic tracks leads to improved educational outcomes (Malamud and Pop-Eleches, 2011; Guyon et al., 2012). Brunello and Checchi (2007) also show that parental background has stronger effects on labor market outcomes when tracking starts earlier. Our goal is not to assess what would be the effects of postponing or modifying high school tracking. Rather, we show that –within an existing tracking system– it is possible to reduce the mismatch between ability and high school choice for a population that may be particularly misinformed and disadvantaged: immigrant students.

The remainder of the paper is organized as follows. Section 2.2 describes the institutional setting and provides evidence of the educational segregation of immigrant children

[3]In contrast to this early stratification, Goldin and Katz (2009) refer to the US system as "open and forgiving".

[4]Alan and Ertac(forthcoming) show that non-cognitive skills such as patience and self-control are malleable. In particular, they show that a treatment targeted at improving intertemporal decision making in a classroom environment through training of teachers has significant impacts on experimental and real-life outcomes.

[5]For a review, see Betts (2011). Differently from systems where tracking takes the form of sorting higher ability students into specialized instruction (e.g., gifted programs, Card and Giuliano, 2016) or into magnet schools (e.g., Pop-Eleches and Urquiola, 2013) – within the same type of education – our context is one where tracking involves sorting into high schools with very different curricula and differential access to college.

in Italy. Section 2.3 illustrates the components of the intervention and our evaluation design. Section 2.4 describes the data and section 2.5 the results. Finally, Section 2.6 concludes discussing the policy implications of our findings.

2.2 Institutional background

2.2.1 Immigrants in Italian schools

Immigration is a relatively recent phenomenon in Italy. The number of (legal) foreign residents increased from 781,000 to 5 million between 1990 and 2015 – 1.4 and 8.3 percent of total residents, respectively. The majority of immigrants in Italy come from low and middle income countries, and are characterized on average by lower socioeconomic background than native households.[6] They are also younger and have more children than natives, so the share of foreigners among students is higher than their share in the total population.

At the beginning of the 2016/17 school year, immigrant children represented 10.8 percent of students in primary school, 9.7 percent in middle school, and 7.2 percent in high school (see Appendix Figure B.3). We observe a decline in immigrants' presence from middle to high school, reflecting higher dropout rates of immigrants in later grades compared to natives. Importantly, the share of immigrant students also differs between different types of high schools. Immigrants represent 12.5 percent of the student population in vocational schools, 8.5 percent in technical schools, and only 4.1 percent in academic schools. As we detail in the next section, these three types of high schools offer very different educational and employment opportunities.

2.2.2 Secondary education in Italy

Italian pupils normally enter formal schooling the year they turn 6 and the compulsory schooling age is 16. Pre-university education comprises five grades in elementary school, three grades in middle school, and five grades in high school. At the end of middle school, students must choose among three different types of high schools: vocational schools (*istituto professionale* and *formazione professionale*), technical schools (*istituto tecnico*), and academically-oriented schools (*liceo*). Students are free to enroll in whatever track they choose, and there is no tracking by ability.

The three tracks have the same duration, 5 years, but differ widely in terms of curricu-

[6]Figure B.1 in the Appendix shows the number of immigrants by nationality in 2015 (first 20 nationalities). Figure B.2 compares the income distribution across immigrant and native families, respectively.

lum, difficulty, and prestige.[7] Vocational schools focus on practical training in specialized manual, low-skilled jobs (e.g., plumber or hairdresser), while devoting a limited amount of time to general education. They are meant to prepare students for immediate employment at the end of high school. Technical and academic schools offer instead a comprehensive curriculum in math, humanities, and science. In principle, academic schools are primarily intended for students who want to pursue a university degree, whereas technical schools complement theory with practical training in specific non-manual jobs (e.g., accountant or graphic designer). Although enrollment in college is possible from all tracks, very few students who attended vocational education decide to obtain further education.[8] In practice, both academic and technical schools offer much better educational and employment prospects than vocational schools. Therefore, we define vocational schools as the "low-track", and we group technical and academic schools together into the "high-track".

Appendix Table B.1 compares average outcomes by track four years after graduation, separately by gender and for native vs. immigrant students.[9] Panel A shows that only 14.5 percent of Italians graduated from vocational schools, with no relevant differences by gender. They exhibit a much lower probability of pursuing tertiary education compared to high-track graduates: 20.5 percent, as opposed to 70.4 percent. College dropout rates in university also differ dramatically between the two groups, at 30.6 and 11.8 percent, respectively. In light of these figures, employment rates and salaries four years after graduation are not really informative about labor market prospects, as only a selected group of high-track graduates has already entered the labor market. However, we can compare the share of those "Not in Education, Employment or Training"(NEET), which reaches 29 percent among low-track graduates, 10 percentage points higher than among other graduates. This is particularly surprising, as in principle vocational schools should prepare students for immediate employment. Finally, graduates from vocational schools also have a higher probability – about one third – of regretting their choice. These figures are fairly similar by gender.

Panel B of Appendix Table A1 shows comparable statistics for immigrant students. Conditional on completing the same high school track, educational and occupational outcomes are remarkably similar to those experienced by natives. Also among immigrants, graduates from vocational schools exhibit lower enrollment into (and higher dropout rates

[7] Regarding duration, the only exception to the 5-year rule is a sub-track of the vocational track (*formazione professionale*) that lasts 3 years.

[8] Enrollment in college is not possible for students in the sub-track of the vocational track (*formazione professionale*) that lasts 3 years. In the 2015-16 school year, less than 4 percent of students enrolled in college had attended a vocational track. Furthermore, students who enroll in college after vocational school tend to take longer to complete college and to dropout at higher rates.

[9] The source of these data is the "Survey on Educational and Professional Paths of Upper Secondary School Graduates", conducted in 2015 by the Italian National Statistical Institute (ISTAT) on a representative sample of about 26,000 students graduating from high school in 2011.

from) tertiary education, as well as a higher prevalence of NEETs. A stark difference emerges, however, when we compare high school choice: 37 percent of immigrants (42 percent among males) graduate from the low track, compared to the aforementioned 14.5 (15.6 for males) of natives. In light of the (worse) outcomes experienced on average by low-track graduates, the over-representation of immigrants in this group raises concerns about immigrants' future career opportunities and, eventually, their prospects for successful integration and upward social mobility.

Of course, enrollment rates and outcomes across groups largely reflect endogenous sorting by ability and socioeconomic background. Below we provide a more informative comparison of transitions to high school between immigrants and natives, exploiting a unique dataset that matches administrative data on educational careers with standardized test scores and information on parental background.

2.2.3 Educational segregation

As we discuss in detail in Section 2.4, Italian students take a series of standardized tests of proficiency in reading and math at various points of their careers. These tests are known as INVALSI, from the name of the agency that administers them. The tests are identical for all students in a given grade and are blindly scored, so results are fully comparable across schools. Throughout the paper, we use the standardized test score obtained in grade 6 (INVALSI6) as a proxy of students' ability at the beginning of middle school.

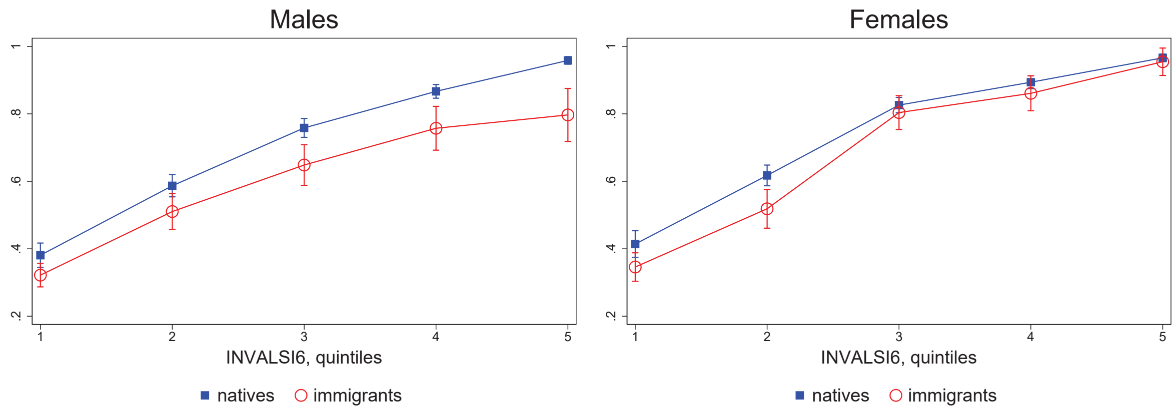
Using a unique dataset that matches the above scores with students' educational careers, we can compare the average probability of enrolling into the high track for native and immigrant students conditioning on their initial ability.[10] Figure 2.1 plots the probability of enrolling in a "high track" (academic or technical) by quintile of INVALSI6, separately for male and female students (left and right panel, respectively).

Squares (connected by a black line) refer to native students, while circles (connected by a grey line) to immigrants. Not surprisingly, the probability of choosing the high-track is increasing in INVALSI6 for all groups. However, such probability remains significantly lower for immigrant males than for native ones. The gap is larger in the upper part of the ability distribution, reaching 16 percentage points (or 17 percent of the mean) in the top quintile.[11] By contrast, the gap between immigrant and native females is much smaller and it is negligible in the upper part of the ability distribution.

[10] As we explain in Section 2.4, this dataset was constructed for the evaluation of our intervention, which took place in five large cities of Northern Italy. The data used in Figure 2.1 are those from schools in the control group, which were unaffected by our intervention.

[11] Appendix Table B.2 shows that the gap in educational choices between native and immigrant males persists when conditioning in addition on family background, as measured by parents' education, employment status, and occupation.

Figure 2.1: Probability of enrolling in the high track at the end of middle school, by quintile of standardized test score in grade 6 (INVALS6)



Notes: This figure compares the probability of enrolling in the high-track between immigrant and native students, by quintiles of performance in the standardized test in grade 6 (INVALSI6). The sample includes all students in the 75 control schools.

The literature has documented an educational gender gap in favor of girls during teenage years, possibly related to gender roles and norms imposing social control of daughters and more lax regulations for sons (Lopez, 2003). This gap is particularly pronounced among minorities.[12] The focus of our paper is not on the gender gap in high school enrollment, but on the gap between natives and immigrants: in particular, we investigate how psychological factors and academic performance contribute to differences in track choice, and test if the native-immigrant gap can be reduced through a specific policy. The fact that this gap differs between boys and girls, as shown in Figure 2.1, will help understand the differential impact of our intervention across genders.

2.3 The intervention

The intervention we evaluate was developed in collaboration with the Italian Ministry of Education and three bank foundations.[13] The program was called "Equality of Opportunity for Immigrant Students" (EOP) and aimed at aligning the goals and aspirations of

[12]Among others, Jackson and Moore III (2008) document that in the US black males lag behind black females on a range of key educational outcomes, e.g., high school graduation. Figlio et al. (2016) investigate the potential reasons for this and show evidence that family disadvantage disproportionately inhibits boys compared to girls (as opposed to explanations invoking a genetic or biological advantage of girls at birth or a pure neighborhood effect).

[13]The program was financed by three philanthropic institutions operating in Northern Italy, namely Fondazione CARIPLO, Compagnia di San Paolo, and Fondazione Cassa di Risparmio di Padova e Rovigo.

high-achieving immigrant students with their ability, in order to favor congruous educational choices at the end of middle school. The intervention took place during the last two years of middle school (grades 7 and 8) and was administered in a randomized fashion in five large cities of Northern Italy: Milan, Turin, Genoa, Brescia, and Padua.

The first dimension of targeting involved the definition of the school sample: schools were eligible to receive the program if they had at least 20 immigrant students, where 'immigrant' was defined as being a citizen of a country with lower GDP than Italy.[14] In the five cities there were 145 such schools: 70 were randomized into the treatment group and 75 in the control group.[15]

The second step was the definition of the target students. Because the goal of EOP was to reduce mismatch in track choice for high-achieving immigrants, within each school we selected the high-achievers as the 10 immigrant students with the highest standardized test score in grade 6 (INVALSI6). In the treatment schools these 10 students took part in the EOP program while in the control schools they did not. In both sets of schools, these top 10 immigrant students were surveyed and their academic performance and school choices were followed through administrative records. In our empirical analysis we will thus compare outcomes between the 10 immigrant students with the highest INVALSI6 scores in treated and control schools.

The EOP program consisted in a career choice consultancy that was developed based on Social Cognitive Career Theory (Lent et al., 1994). This paradigm views career development as a choice subject to contextual influences and constraints. Under this view, goals and self-efficacy are as important as cognitive skills in shaping individual careers.[16] Specifically, "persons with adequate skills but weak self-efficacy beliefs in a particular performance domain may prematurely rule out that domain from further occupational or academic choice consideration" (Brown, 2002). The goal of EOP was to help high-achieving immigrant students to identify educational and occupational goals congruous with their talents and to strengthen self-efficacy beliefs. It should be stressed that the approach was not to unconditionally push students towards high tracks, but to make them aware of existing opportunities and of their own skills and resources, so they could make more informed choices.

The protocol involved a total of 13 meetings and the program guidelines required participants to attend at least 75 percent of the meetings. All meetings were administered

[14]This excludes from the program immigrant children from high income European countries, for whom no educational segregation exists.

[15]Randomization at the school level lowers the risk of spillovers from the treated to the control group, compared to randomization at the individual-level. To enhance comparability between the two groups of schools we stratified randomization by province and school size.

[16]Self-efficacy can be defined as people's beliefs about their capabilities "to organize and execute courses of action required to attain designated types of performances" ; see Bandura (1986), p. 391.

by career counselors with graduate degrees in psychology and significant experience in career choice guidance for secondary school, especially with immigrant children. Some of the meetings were one-to-one, while others were in groups.

Students had five (one-to-one) meetings with a counselor and worked on tasks that prompted them to reflect on their goals, the personal resources needed to achieve such goals, and whether they already had or they needed to develop such resources. Examples of the tasks that students worked on include: (i) Think about your past life, indicate five study experiences and five other experiences that you have completed successfully. Consider now such experiences one by one and briefly indicate where and with whom it happened, what you did and which personal resources helped you doing well in that thing - your knowledge, skills, personality traits, motivations and everything you believe it was important to have; (ii) Choose a number of professions that interest you. For each of them, indicate which resources are needed (knowledge, skills, personality traits, motivations, ...) and divide them into "I have it" and "I need to develop it"; (iii) List the results you would like to achieve with your job, from the most to the least important.

Five other meetings were held in groups, where counselors provided treated students in each school with information about the Italian education system as well as peer guidance through videos displaying success stories of older immigrant students.

Two further meetings, respectively at the beginning and at the end of the intervention, were intended for parents. In the first meeting counselors described the content of the program, while in the second meeting they shared with parents aspirations and barriers perceived by students. Parents also received a brochure, translated into their mother tongue, summarizing the main options for secondary education in Italy.[17]

Finally, towards the end of the intervention career counselors met with teachers and discussed the educational path and high school track chosen by the students involved in the intervention.

Figure 2.2: Time Line

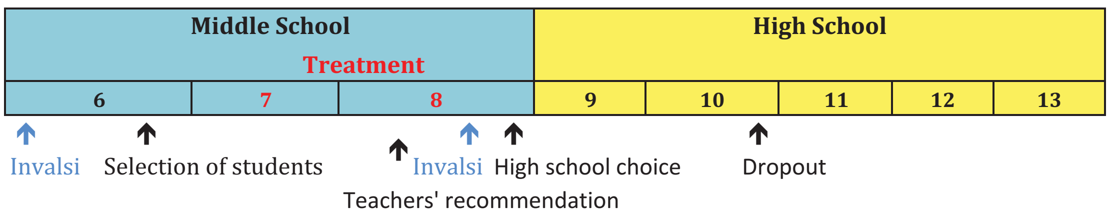


Figure 2.2 shows the timeline of the intervention and the realization of the main outcomes of interest. In grade 6 we selected the eligible students on the basis of the

[17]While it would be very interesting to disentangle the role that parents play in high school choice (e.g., Dustmann, 2004; Giustinelli, 2016), it was not possible to survey parents within our experiment.

standardized test score they got that year. EOP meetings started at the beginning of grade 7 and continued through grade 8, until the month of March. By the month of January of grade 8, students receive a formal "recommendation" by their teachers about the high school track that teachers deem most appropriate for them. This recommendation is not binding but it serves as a signal to the students and their families. In February all students have to pre-register for the high school they wish to attend through a web portal of the Ministry of Education. This choice can later be modified (though this is not very common), so we use the high school track in which students actually enroll at the end of grade 8 as our variable of interest.

A potential concern with an intervention like EOP is that participants who enroll in the high-track as a result of the program would subsequently experience difficulties in completing this (more demanding) high school. For this reason, the counseling and career choice module was accompanied by a module on Cognitive Academic Language Proficiency (CALP). CALP was not aimed at improving students' knowledge or their cognitive skills but, rather, at teaching them a method for studying several subjects – Italian grammar, geography, algebra, and geometry. Since the main motivation of the CALP module was to decrease the risk of subsequent failure for immigrant students enrolling in demanding tracks, students with a lower INVALSI6 were offered a higher number of CALP meetings. Specifically, students scoring below 65 out of 100 in INVALSI6 were invited to 29 meetings (55 hours tutoring). This group constituted 66 percent of all treated students. Students scoring between 65 and 80 (30 percent of the treatment group) were invited to 17 meetings (32 hours tutoring); finally, students scoring above 80 (4 percent of treatment group) were not invited to CALP sessions.

Due to ethical concerns we could not implement a fully factorial design; in particular we could not have a treatment arm that received career counseling without CALP. As explained above, it was considered that encouraging students to pursue ambitious goals without endowing them with the tools for succeeding in demanding high schools may have created a risk of harm. However, we can still assess the relative effectiveness of CALP by exploiting the different cutoffs for the number of CALP meetings described above. We do this in Section 2.5.2.

2.4 Data

2.4.1 School choice and academic performance

The first challenge in constructing the dataset for our analysis was to match information on school careers provided by the Ministry of Education (MIUR) with standardized test scores collected by the Institute for the Evaluation of the Italian Schooling System (IN-

VALSI), an independent public agency that monitors students' performance. This had never been done before and required a new protocol for collaboration among the two organizations. In fact, in order to preserve the anonymity of standardized tests scores, INVALSI and MIUR use different codes to identify each student and only the school has the crosswalk to match the two codes. The protocol of our study involved obtaining the match of INVALSI and MIUR records for all students completing grade 6 in 2012 (17,369 students in total) from the 145 middle schools in our sample.[18]

Administrative data from school registry

From the MIUR administrative registry we take the following variables that we use as outcomes for each student: (i) track choice at the end of grade 8; (ii) track recommended by teachers half-way through grade 8; (iii) grade retention for all grades between 6 and 9 (included); (iv) number of retakes taken after the summer of grade 9;[19] and (v) teachers' assessment of student's behavioral conduct during grades 8 and 9.

The registry also contains some information on students' background, in particular: citizenship, country of origin, date of birth and, of course, school and class attended throughout their careers.

INVALSI tests

Since 2010, INVALSI administers standardized reading and math proficiency tests to all students at the end of grades 2, 5, 6, 8, and 10. Such tests resemble those administered by the OECD Programme for International Student Assessment (PISA) to representative samples of 15-year old students. They consist of a series of questions including multiple choice as well as open ended questions, the exact structure of the test varying by grade.[20] Importantly, the test is identical for all students in a given grade, it is administered on the same day (at the end of the school year), and it is blindly scored, so results are fully comparable across schools in Italy. This is crucial for the purposes of our analysis, because it allows us to compare the educational choices of immigrant and native students holding constant their academic proficiency.

We use two test scores for the cohort of students who were in grade 6 in 2012. The first is the standardized test score for grade 6 (INVALSI6), which we include as a regressor in all specifications to control for students' initial ability. The second is the test score for

[18] As described in detail in the next section, our sample is representative of mid-to-large sized schools in urban areas of Northern Italy. Unfortunately, similar data are not publicly available for the entire student population.

[19] While through middle school pupils are either admitted to the next grade or retained, in high school they can also be admitted to the next grade conditional on re-taking (and passing) after the summer an examination in one or more subjects in which they were deficient during the year.

[20] In general, math questions are related to calculus, geometry, probability and algebra, while reading questions are related to text comprehension and grammar.

grade 8 (INVALSI8), which is one of our outcomes, as we want to test if EOP led to an improvement in academic performance.

2.4.2 Soft skills

We complement the above datasets with original survey data on soft skills collected at the end of grade 8. The goal of this survey is to allow us to better understand what mechanisms shape career-related interests and high school track choice. The survey was administered to all treated students and to a random 50 percent sample of control schools.[21] The questionnaire was developed by a team of psychologists based on Social Cognitive Career Theory and includes three main sections: (i) Goals. This comprises both educational (e.g., university degree, diploma or less) and occupational targets (e.g., blue collar, white collar, managerial or entrepreneurial jobs) that the student aims to achieve; (ii) Self-efficacy. This section includes a student's own assessment of the extent to which he or she possesses the skills and resources required to achieve the goals stated above, as well as broad notions of self esteem; (iii) Barriers. A series of questions elicit students' perceptions of environmental barriers, be they related to economic constraints, racial prejudice, or family preferences that differ from a student's own plans.

Following Thompson (2004), we summarize the individual variables described above into interpretable aggregates using factor analysis. This method extracts latent factors from subsets of psychological measures by maximizing (minimizing) the correlation across measures within (between) subsets. The measures associated with each factor and their respective loadings are reported in Appendix Table B.4. As discussed in Heckman et al. (2013), this approach is particularly suited for decomposing treatment effects between different mediating factors, as we do in Section 2.5.2.

2.4.3 Sample and randomization check

Our working sample at the inception of EOP comprises $1,217$ students: 597 in treated schools and 620 in control ones.[22]

[21] We chose not to administer the survey to half of the control schools because we wanted to be able to test if filling in a questionnaire on goals and perceived barriers may constitute a 'treatment' in itself. In Table B.3 we show that students in control schools involved and not involved in the soft skills questionnaire do not systematically differ in terms of high-school choice, grade retention, and test scores in grade 8.

[22] By construction, the sample should have comprised $1,451$ students, that is, $10 \times (70+75) = 1,450$ but in one school the 10th and 11th students obtained the same INVALSI6 score and were both eligible for the program. For 70 students it was impossible to match the MIUR and INVALSI identifiers, which reduced the sample to $1,381$ students. In addition, some students were retained in grade 6, moved to another school, or dropped out between the moment when they took the INVALSI6 test and the beginning of grade 7, leading to the above sample of $1,217$. Appendix Table B.5 shows that missingness is not selective across treatment and control schools.

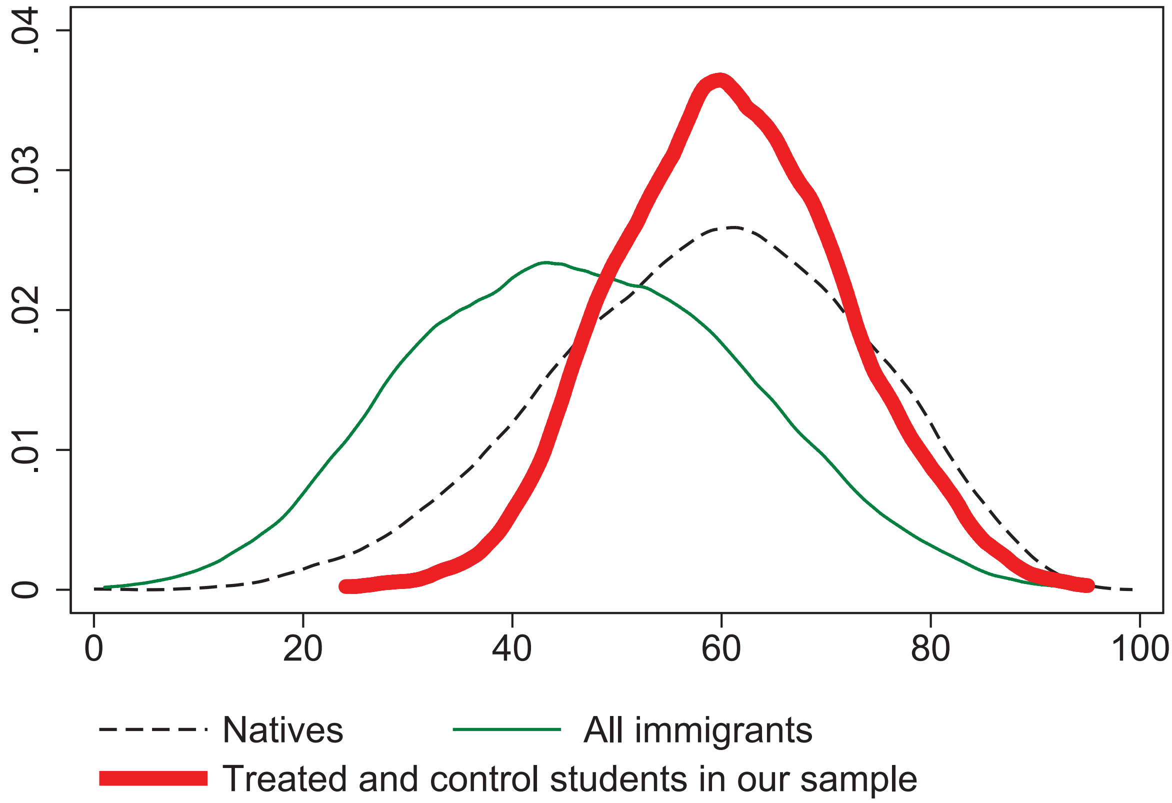
Table 2.1: Treated and control students, balance test

		Full Sample	Treated	Controls	Difference	P-value	Std. Difference
Panel A: Student characteristics							
Female		0.506	0.508	0.505	0.003	[0.93]	-0.006
Test score in grade 6 (INVALSI6)		60.82	60.93	60.71	0.224	[0.86]	0.005
First generation immigrant		0.555	0.547	0.561	-0.014	[0.73]	0.028
Born before 1999		0.257	0.242	0.273	-0.031	[0.25]	0.071
Brescia		0.179	0.165	0.194	-0.029	[0.67]	0.076
Genova		0.067	0.074	0.06	0.014	[0.75]	-0.056
Milan		0.496	0.476	0.516	-0.04	[0.65]	0.080
Padua		0.055	0.064	0.047	0.017	[0.68]	-0.074
Turin		0.203	0.223	0.184	0.039	[0.58]	-0.097
Panel B: Family characteristics							
Mother	Less than high school	0.350	0.348	0.352	-0.004	[0.93]	0.008
	High-school	0.464	0.481	0.447	0.034	[0.36]	-0.068
	Some post-secondary education	0.186	0.172	0.201	-0.029	[0.30]	0.075
	Blue collar	0.351	0.354	0.347	0.007	[0.86]	-0.015
	White collar	0.184	0.181	0.187	-0.006	[0.86]	0.015
	Unemployed	0.082	0.091	0.073	0.018	[0.48]	-0.066
	At home	0.384	0.374	0.393	-0.019	[0.65]	0.039
Father	Less than high school	0.332	0.330	0.335	-0.005	[0.91]	0.011
	High-school	0.499	0.505	0.495	0.01	[0.81]	-0.020
	Some post-secondary education	0.168	0.165	0.17	-0.005	[0.87]	0.013
	Blue collar	0.567	0.555	0.577	-0.022	[0.56]	0.044
	White collar	0.315	0.333	0.296	0.037	[0.29]	-0.080
	Unemployed	0.099	0.093	0.106	-0.013	[0.53]	0.043
	At home	0.020	0.019	0.021	-0.002	[0.85]	0.014

Notes: This table shows the number and characteristics of treated and control students in our sample. P-values for difference in means are reported in square brackets. The last column also reports the standardized difference between group averages.

Table 2.1 reports average characteristics of the treatment and control group at the start of our intervention. We distinguish between individual student characteristics (Panel A) and family background (Panel B). Half of the students in our sample are girls, 56 percent are first generation immigrants, and 26 percent were born before 1999 (the typical birth year of the cohort in our study). About 35 (33) percent of their mothers (fathers) have not completed high school, 46 (50) percent have a high school diploma, and 19 (17) percent have post-secondary education. Unemployment rates are about 8 percent for mothers and 10 percent for fathers. The share of mothers only working at home is 38 percent, while that of fathers is negligible. Among those working outside the home, 35 (57) percent of mothers (fathers) have a blue collar job, and 32 (18) percent a white collar job. Importantly, none of the student or family characteristics differ significantly between the treatment and control sample, indicating that our randomization was successful.

Figure 2.3: Distribution of standardized test score in grade 6 (INVALS6)



Notes: This figure compares the distribution of standardized test score in grade 6 (INVALSI6) across native students, immigrant students, and treated and control students in our sample.

Turning to the baseline academic performance of the students, Table 2.1 shows that the mean of the standardized test score INVALSI6 is 60.93 in the treatment group and 60.71 in the control one (not significantly different). To get a more complete picture, Figure 2.3 plots the distribution of INVALSI6 across three groups of students in our 145 schools: native students, all immigrant students, and the 10 immigrant students with the highest score (i.e., our treated and control groups). Although immigrants generally exhibit lower schooling performance than native students, the top 10 immigrant students in each school are comparable to natives in the medium-upper part of the distribution. As we previously showed in Figure 2.1, however, these immigrants choose high school tracks that are less prestigious (and less demanding) than those chosen by natives with comparable ability, particularly males.

2.5 Results

In this section we estimate the impact of EOP on educational choices and grade retention, separately for males and females, and we decompose treatment effects into several mediating factors. In addition, we estimate impacts on longer term outcomes and spillover effects on non-eligible students in treatment schools, comparing them to non-eligible students in control schools.

2.5.1 Educational choices and grade retention

In Table 2.2 we estimate the impact of EOP on high school track choice. The dependent variable is a dummy equal to 1 for students who choose the high track (which comprises academic and technical high schools) and 0 for those who choose the low track (vocational schools). The explanatory variable of interest is EOP, an indicator for whether a student attends a middle school that has (randomly) been selected to receive our intervention. The coefficient of this dummy should thus be interpreted as the intention-to-treat (ITT) effect of being assigned to treatment group. Odd-numbered columns condition on treatment only, while the specifications in even-numbered columns also include a squared polynomial in INVALSI6, a dummy for first generation immigrants, and province fixed effects. In all cases we cluster standard errors at the school level, the unit of randomization.

According to the univariate regression in column 1, assignment to EOP increases the probability of choosing the high track by 5 percentage points, on a baseline rate of 75 percent. As expected, given random assignment, such estimate is largely unaffected when controlling for student characteristics and province (column 2). However, the average effect masks important differences by gender. EOP increases males' enrollment into the high track by 8 to 9 percentage points, up from a baseline rate of 67.4 percent (columns 3 and 4). This is a 12 to 13 percent increase over the mean. By contrast, there is no effect

Table 2.2: The effect of EOP on educational choices

	Dependent variable: Choosing the high-track (= 1 if choose high track)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All immigrants		Male immigrants		Female immigrants		All immigrants	
EOP	0.051*	0.043*	0.091**	0.080**	0.011	0.009	0.091**	0.077**
	(0.027)	(0.024)	(0.037)	(0.035)	(0.034)	(0.031)	(0.037)	(0.034)
Female X EOP							-0.080*	-0.067
							(0.046)	(0.043)
Constant	0.750***	0.683***	0.674***	0.651***	0.824***	0.720***	0.674***	0.621***
	(0.019)	(0.038)	(0.028)	(0.048)	(0.022)	(0.049)	(0.028)	(0.041)
Mean dep. var. control	0.750	0.750	0.674	0.674	0.824	0.824	0.750	0.750
Observations	1,217	1,217	601	601	616	616	1,217	1,217
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.004	0.087	0.010	0.086	0.000	0.100	0.024	0.105

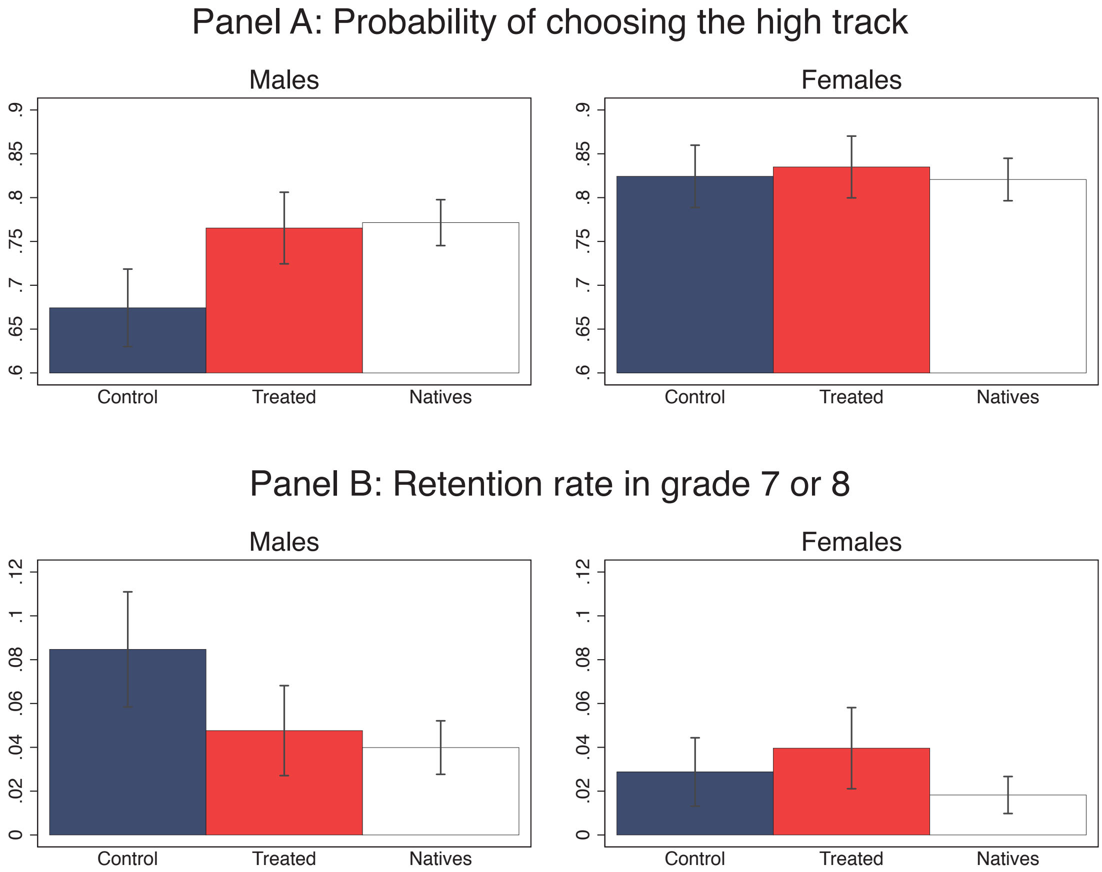
Notes: This table shows the effect of EOP on immigrant students' educational choices the end of middle school. The dependent variable is a dummy equal to 1 for students choosing the high-track (academic or technical schools) and equal to zero otherwise. EOP is a dummy equal to 1 for students in schools assigned to the treatment group and equal to zero for schools assigned to the control group. Specifications in columns (2), (4), (6), and (8) control in addition for a squared polynomial in INVALSI6, a dummy equal to 1 for first generation immigrants, and province fixed effects. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

on female students, who start, however, from a baseline enrollment rate of 82.4 percent (columns 5 and 6).

Panel A of Figure 2.4 compares enrollment in the high track for high-achieving immigrant students randomized into the control group (leftmost bar), treatment group (middle bar) and for a group of native students with comparable ability in the first year of middle school (rightmost bar). Specifically, we match each immigrant student in our sample with one native student who obtained an identical score in INVALSI6. By construction, these three groups of students had the same standardized test score in the first year of middle school. The figure shows that two years later the immigrant boys who received EOP make similar choices compared to natives who started off like them, while the untreated immigrants have a significantly lower probability of choosing the high track. EOP thus prevented the type of educational segregation that we documented in section 2.2. Interestingly, Figure 2.4 confirms that immigrant girls make similar choices as native girls even in the absence of intervention.

Taken together, the evidence in Figure 2.1, Table 2.2, and Figure 2.4 highlights a remarkable feature of the intervention: EOP influences educational choices only when counterfactual enrollment rates into the high track lie below those of comparable native students. Therefore, EOP seems to align immigrant students' goals and aspirations to those of native students when there is an initial misalignment, as opposed to just pushing

Figure 2.4: Track choice and grade retention of immigrants and comparable natives



Notes: These graphs shows the average probability (and associated confidence intervals) of choosing the high-track (top graphs) and being retained in grade 7 or 8 (bottom graphs) for treated students, control students, and a group of Italian students that are comparable in terms of schooling ability. Specifically, we match each immigrant student with a native student obtaining exactly the same score in INVALSI6.

all immigrant students towards the high track. This is a very desirable feature of EOP, as it lowers concerns that treated students may end up is schools that are too difficult for them. In Section 2.5.3 we provide direct evidence in this respect.

Table 2.3: The effect of EOP on grade retention

	Dependent variable: Grade retention (= 1 if repeat a grade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All immigrants		Male immigrants		Female immigrants		All immigrants	
EOP	-0.013	-0.013	-0.037*	-0.037*	0.011	0.009	-0.037*	-0.035*
	(0.016)	(0.016)	(0.021)	(0.021)	(0.019)	(0.019)	(0.021)	(0.020)
Female X EOP							0.048**	0.044*
							(0.024)	(0.023)
Constant	0.056***	0.073***	0.085***	0.097***	0.029***	0.048	0.085***	0.097***
	(0.010)	(0.028)	(0.016)	(0.028)	(0.010)	(0.043)	(0.016)	(0.030)
Mean dep. var. control	0.056	0.056	0.085	0.085	0.029	0.029	0.056	0.056
Observations	1,217	1,217	601	601	616	616	1,217	1,217
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.001	0.017	0.006	0.022	0.001	0.023	0.009	0.024

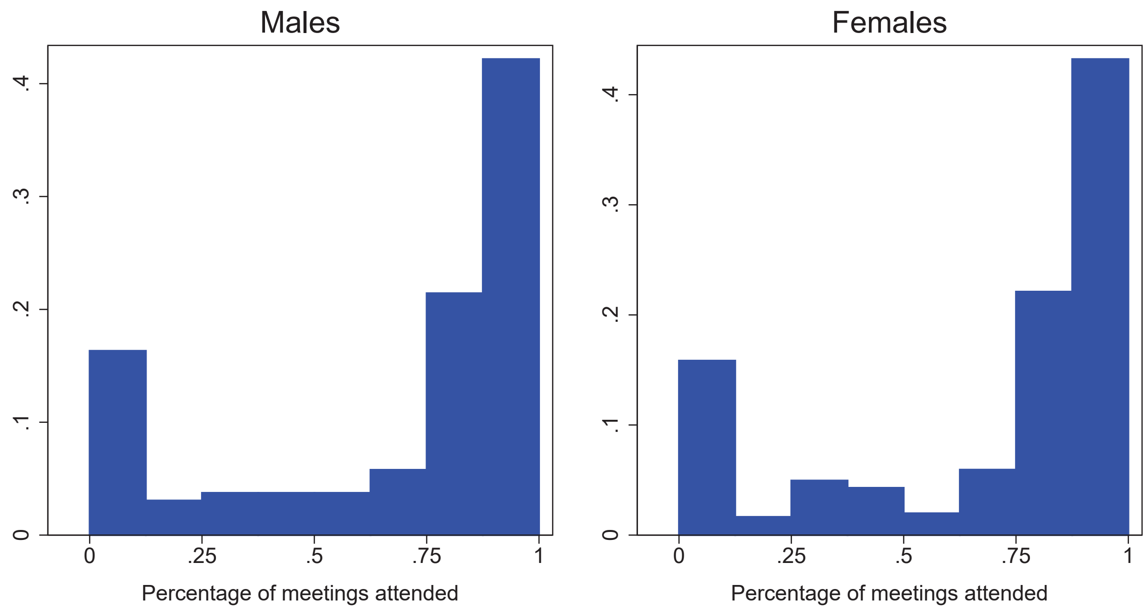
Notes: This table shows the effect of EOP on immigrant students' grade retention during middle school. The dependent variable is a dummy equal to 1 for students retained in grade 7 or 8, and equal to zero otherwise. EOP is a dummy equal to 1 for students in middle schools assigned to the treatment group and equal to zero for schools assigned to the control group. Specifications in columns (2), (4), (6), and (8) control in addition for a squared polynomial in INVALSI6, a dummy equal to 1 for first generation immigrants, and province fixed effects. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table 2.3 and Panel B of Figure 2.4 convey similar evidence for grade retention in grade 7 or 8. Although grade retention was not the primary outcome of interest of the intervention, it is arguably an important one. Indeed, grade retention is surprisingly high among male immigrant students in our sample: absent the intervention, it reaches 8.5 percent, as compared to only 4.2 percent for native males with a similar INVALSI6 score. This gap disappears in EOP schools, whereas there is neither a significant gap nor an effect for female students. In both respects, the effect on grade retention across genders is very similar to that on high school choice.

Overall, the results in Tables 2.2 and 2.3 point to sizeable and statistically significant effects induced by (random) assignment to the intervention for male immigrant students. Since there is one-sided non-compliance with treatment assignment, the average treatment-on-the-treated (ATT) effects on the subset of compliers are even larger. To assess the magnitude of the ATT, it is useful to start by describing the pattern of meetings attendance.

Figure 2.5 shows that the pattern is quite heterogeneous, with more than 40 percent of immigrant boys and girls attending at least 87.5 percent of the meetings, another

Figure 2.5: Meetings attendance of immigrant students assigned to EOP



20 percent attending between 75 and 87.5 percent of the meetings, and the remaining fractions attending less. Interestingly, about 15 percent of the students who were assigned to treatment ended up attending less than 12.5 percent of the meetings. Given this heterogeneity, there is no unambiguous way of defining treatment status. For this reason, in Table 2.4 we experiment with three alternative definitions.

In Panel A, we classify as treated all students attending at least one meeting (85 percent of the total sample). In Panel B we restrict the definition to students attending at least 75 percent of the meetings, in accordance with the program guidelines discussed in Section 2.3, which recommended attending at least this fraction. When adopting these definitions, the ATT effects on males range between a 9.4 to 12.5 percentage point increase in enrollment in the high track, and a 4.3 to 5.7 percentage point decrease in grade retention.

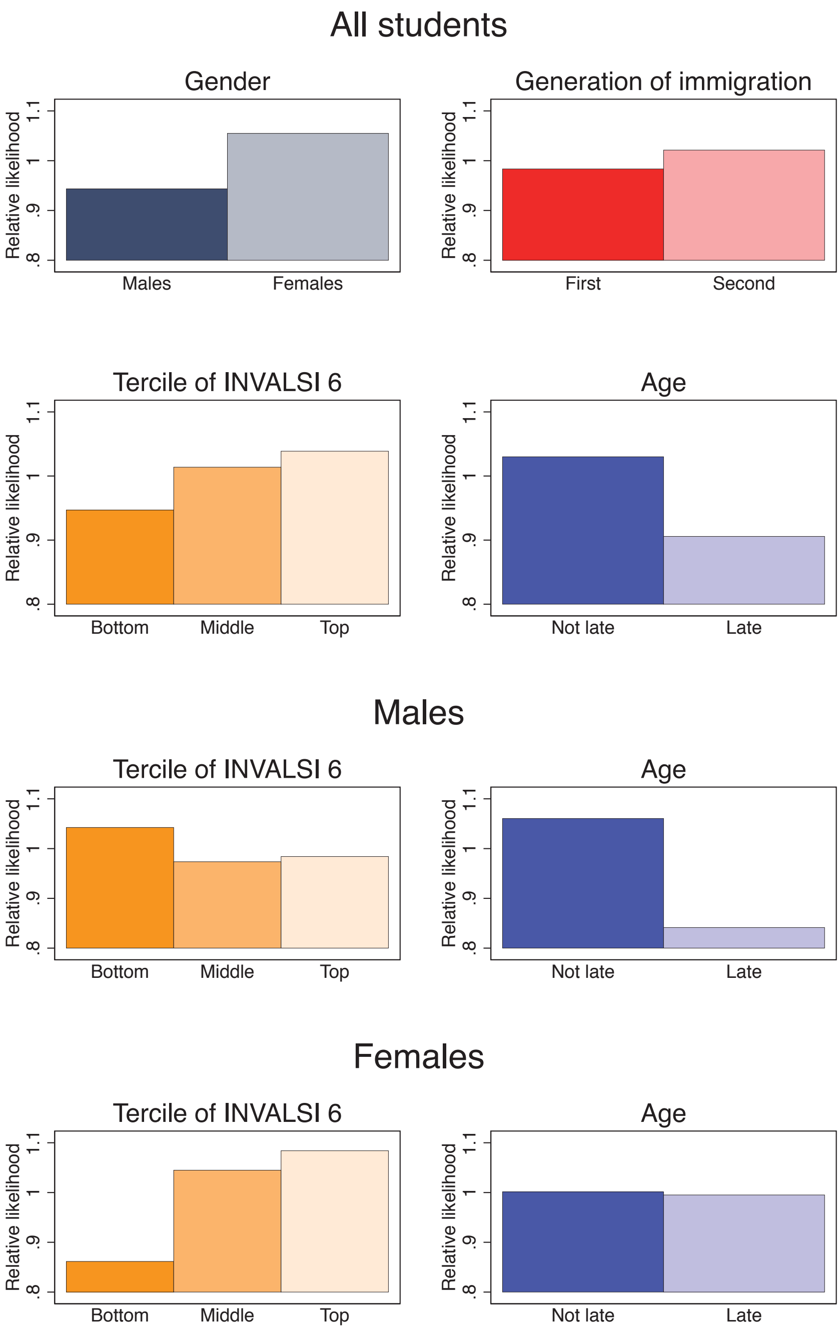
In Figure 2.6 we characterize compliers with treatment assignment, defining the treatment as in Panel B of Table 2.4, by the ratio of the first stage effect within specific sub-samples to the overall first stage (Angrist et al., 2016). Compliers are slightly more likely to be female, equally likely to be first and second generation immigrants, and more likely to be in the right grade ('Not late') given their age. The bottom panels of Figure 2.6 show that while female compliers are more likely to be from the top part of the initial ability distribution, male compliers are more likely to be from the bottom part, thus more in need for support.

Table 2.4: Effects of EOP, average treatment-on-the-treated (ATT)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.:	Choosing the high-track			Grade retention		
	All	Males	Females	All	Males	Females
Panel A: Treatment = 1 if attended at least one meeting						
ATT	0.051*	0.094**	0.010	-0.015	-0.043*	0.011
	(0.028)	(0.041)	(0.036)	(0.018)	(0.024)	(0.022)
Constant	0.682***	0.648***	0.720***	0.073***	0.098***	0.047
	(0.038)	(0.049)	(0.049)	(0.028)	(0.028)	(0.043)
Panel B: Treatment = 1 if attended at least 75% of meetings						
ATT	0.067*	0.125**	0.013	-0.020	-0.057*	0.014
	(0.037)	(0.054)	(0.047)	(0.024)	(0.032)	(0.029)
Constant	0.679***	0.641***	0.720***	0.074***	0.101***	0.047
	(0.038)	(0.050)	(0.049)	(0.027)	(0.028)	(0.043)
Panel C: Treatment = fraction of meetings attended						
ATT	0.064*	0.119**	0.013	-0.019	-0.055*	0.014
	(0.036)	(0.052)	(0.045)	(0.023)	(0.031)	(0.028)
Constant	0.681***	0.645***	0.720***	0.074***	0.100***	0.048
	(0.038)	(0.049)	(0.049)	(0.027)	(0.028)	(0.043)
Observations	1.217	601	616	1.217	601	616

Notes: This table shows the average-treatment-on-the-treated effect of EOP on immigrant students' educational choices (columns 1-3) and grade retention during middle school (columns 4-6). The ATT is computed as the ratio of the reduced form effect of EOP on such outcomes and the first stage effect on three alternative measures of compliance with treatment assignment: attending at least 1 meeting (Panel A), attending at least 75% of meetings (Panel B), and fraction of meetings attended (Panel C). All specifications control for a squared polynomial in INVALSI6, a dummy equal to 1 for first generation immigrants, and province fixed effects. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Figure 2.6: Compliers' characteristics



Notes: This figure shows compliers' characteristics ratios, i.e. the ratio of the first stage for student of a specific type (e.g., female/male) to the overall first stage. The instrument is the assignment to EOP and the endogenous variable is the probability of attending at least 75% of meetings. The figure illustrates the relative likelihood of compliers' gender, generation of immigration, tercile of INVALSI 6, and age.

In Panel C of Table 2.4 we measure treatment 'intensity' by the fraction of meetings attended. The corresponding ATT estimate suggests that one standard deviation increase in the number of meetings attended increases enrollment into the high track by 4.2 percentage points and reduces grade retention by 1.9 percentage point for males. Of course, this estimate rests upon the assumption that the effect increases linearly with the number of meetings.

More generally, all three approaches in Table 2.4 recover the ATT effect only under strong (and untestable) assumptions about the relationship between number of meetings attended and treatment intensity. For this reason, in the rest of the paper we focus on the intention-to-treat (ITT) effect of EOP.

Table 2.5: The effect of EOP on educational choices, heterogeneity

	Dependent variable: Choosing the high-track							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Males				Females			
EOP	0.100**	0.138*	0.077*	0.068*	0.022	0.145**	0.026	-0.008
	(0.042)	(0.074)	(0.046)	(0.040)	(0.040)	(0.070)	(0.044)	(0.036)
EOP*INVALSI6	0.038				0.117*			
	(0.057)				(0.064)			
EOP*INVALSI6, squared	-0.055				-0.095**			
	(0.036)				(0.048)			
EOP*Highly educated mother		-0.136				-0.196***		
		(0.089)				(0.074)		
EOP*First Gen. Immigrant			0.005				-0.031	
			(0.065)				(0.057)	
EOP*EU country				0.045				0.074
				(0.084)				(0.069)
Highly educated mother		0.206***				0.191***		
		(0.064)				(0.053)		
EU country				0.036				-0.085
				(0.066)				(0.055)
INVALSI6	0.180***	0.186***	0.199***	0.199***	0.142***	0.188***	0.197***	0.198***
	(0.042)	(0.029)	(0.029)	(0.030)	(0.045)	(0.040)	(0.039)	(0.039)
INVALSI6, squared	-0.024	-0.048***	-0.054***	-0.055***	-0.000	-0.039	-0.044	-0.043
	(0.027)	(0.018)	(0.019)	(0.019)	(0.037)	(0.031)	(0.030)	(0.029)
First Gen. Immigrant	-0.036	-0.038	-0.038	-0.052	-0.004	-0.009	0.010	0.010
	(0.034)	(0.035)	(0.050)	(0.038)	(0.029)	(0.028)	(0.041)	(0.030)
Constant	0.641***	0.580***	0.652***	0.659***	0.713***	0.622***	0.711***	0.726***
	(0.050)	(0.061)	(0.054)	(0.049)	(0.048)	(0.058)	(0.050)	(0.049)
Mean dep. var. control	0.674	0.674	0.674	0.674	0.824	0.824	0.824	0.824
Observations	601	601	601	601	616	616	616	616
R-squared	0.088	0.137	0.086	0.089	0.109	0.133	0.101	0.104

Notes: This table shows the heteroeneity of the effect of EOP on immigrant students' educational choices the end of middle school. The dependent variable is a dummy equal to 1 for students choosing the high-track (academic or technical schools) and equal to zero otherwise. EOP is a dummy equal to 1 for students in middle schools assigned to the treatment group and equal to zero for schools assigned to the control group. Highly educated mother is a dummy equal to 1 for students' whose mother has at least a high-school diploma. EU is a dummy equal to 1 for immigrants from EU-member countries. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

In Table 2.5 we examine heterogeneity of the ITT effect along additional dimensions beside gender; results for males and females are presented in columns 1-4 and 5-8, respectively. In columns 1 and 5 we interact assignment to EOP with a quadratic polynomial in the INVALSI6 score. The effect of EOP is the highest for students in the medium-upper part of the performance distribution (0.35 and 0.62 standard deviations above the mean for males and females, respectively). As for socioeconomic background, the negative coefficient on the interaction term between EOP and mother's education (columns 2 and 6) suggests that the effect is driven by students from disadvantaged families.[23] Instead, the effect does not differ between first vs. second generation immigrants (columns 3 and 7), nor between immigrants from EU vs. non-EU countries (columns 4 and 8). These results highlight the importance of addressing inequality of opportunity in educational choices that may be generated by differential access to information and support for low SES students.

2.5.2 Mechanisms

The results presented so far suggest that EOP had a strong effect on educational choices and grade retention of males, whereas the average effect was not significantly different from zero for females, although the effect was positive for females with low socioeconomic background. We next explore the mechanisms through which EOP impacted on such outcomes.

Personality and cognitive skills

In Table 2.6 we report the effects of EOP on the cognitive and personality skills described in Section 2.4.2.[24] Starting with personality skills, the intervention substantially increased students' aspirations, especially for males (+0.29 standard deviations) whereas the effect is weaker for females. EOP also reduced students' perceptions that their choices would be limited by barriers such as financial constraints, prejudice, or family plans. The effect is sizeable (a reduction in the index of barriers of 0.42 standard deviations) and virtually identical between males and females.[25]

Turning to cognitive skills, EOP increases the standardized test score in grade 8 (INVALSI8) for male students but not for females. The magnitude of the effect is +0.16 standard deviations, and is very similar for math and reading.

[23] We obtain analogous results when measuring parental background using father's education.

[24] The sample differs across columns because questionnaires measuring personality skills were administered only to a 50 percent random sample of control students, see Section 2.4.2.

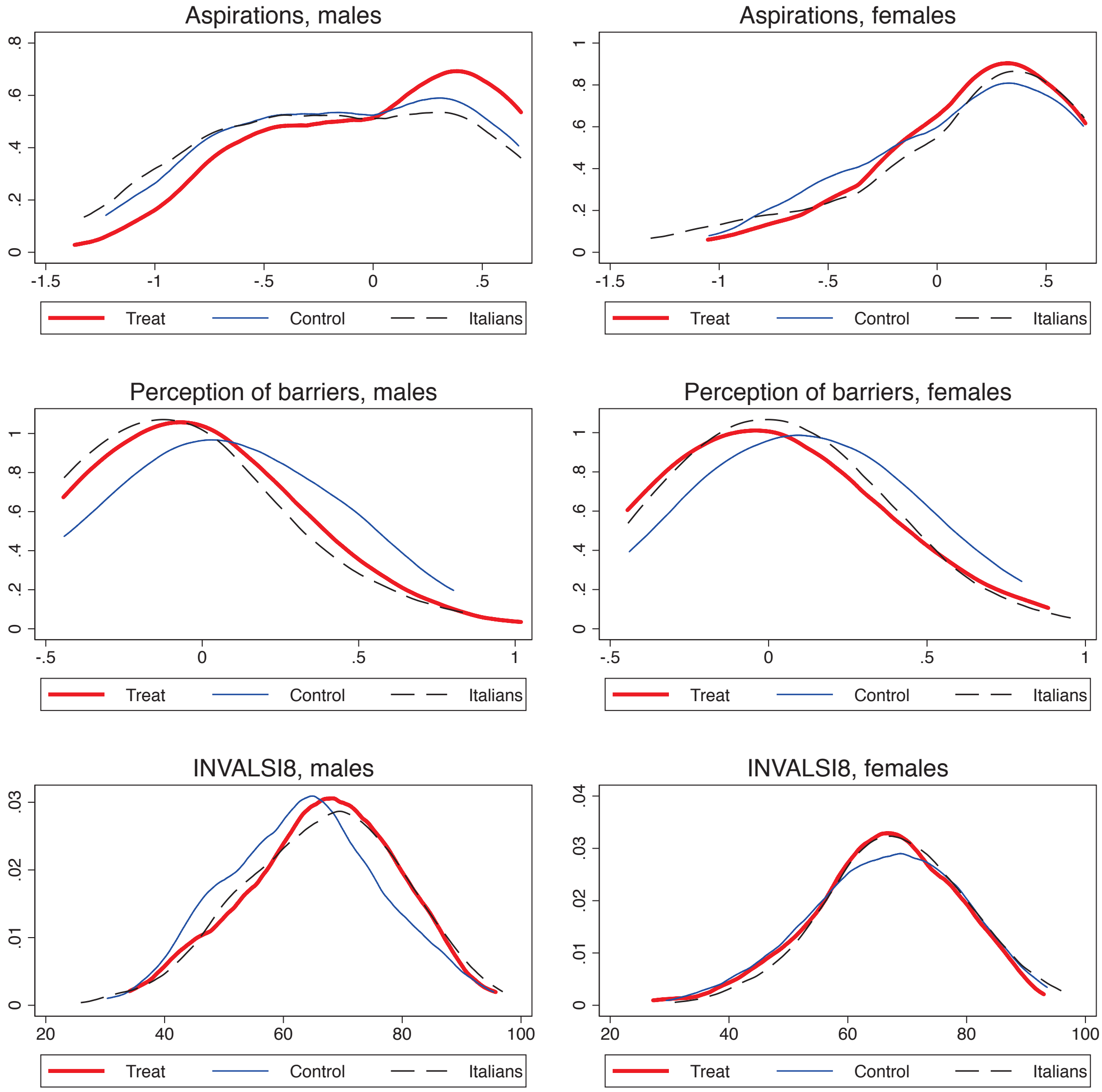
[25] Appendix Table B.6 shows that the effect is generally significant for the different psychological measures aggregated into the main principal components, even after we account for multiple hypothesis testing (last column).

Table 2.6: The effect of EOP on mediating factors

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.	Aspirations		Perception of barriers		INVALSI8	
EOP	0.171**	0.291***	-0.417***	-0.389***	0.076*	0.164***
	(0.076)	(0.104)	(0.081)	(0.122)	(0.040)	(0.052)
Female*EOP		-0.170		-0.032		-0.165**
		(0.125)		(0.163)		(0.064)
Constant	-0.049	-0.283**	0.240**	0.159	-0.318***	-0.404***
	(0.105)	(0.117)	(0.104)	(0.122)	(0.058)	(0.060)
Mean dep. var. control males	-0.209	-0.209	0.252	0.252	-0.129	-0.129
Mean dep. var. control females	0.248	0.248	0.420	0.420	0.057	0.057
Observations	687	687	687	687	1,094	1,094
R-squared	0.117	0.158	0.056	0.061	0.439	0.445
Dep. var.	INVALSI8, math		INVALSI8, reading		Teachers' Recomm.	
EOP	0.083	0.188***	0.069	0.139**	0.123***	0.172***
	(0.051)	(0.066)	(0.044)	(0.062)	(0.046)	(0.053)
Female*EOP		-0.204**		-0.126		-0.095
		(0.079)		(0.082)		(0.060)
Constant	-0.280***	-0.286***	-0.357***	-0.522***	0.393***	0.303***
	(0.075)	(0.073)	(0.067)	(0.076)	(0.041)	(0.042)
Mean dep. var. control males	0.008	0.008	-0.266	-0.266	0.371	0.371
Mean dep. var. control females	0.023	0.023	0.090	0.090	0.578	0.578
Observations	1,094	1,094	1,094	1,094	1,217	1,217
R-squared	0.383	0.388	0.293	0.320	0.125	0.150

Notes: This table shows the effect of EOP on several mediating factors. Aspirations and perception of barriers are the two principal components extracted from the psychological measures collected through students' questionnaires. The individual variables included in each index and their loading factors are reported in Appendix Table B.4. INVALSI8 is the score obtained in the standardized test at the end of middle school (grade 8). Teachers' recommendation is a dummy equal to 1 when the teacher recommends to enroll in the high-track and equal to zero otherwise. All specifications control for a squared polynomial in INVALSI6, a dummy equal to 1 for first generation immigrants, and province fixed effects. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Figure 2.7: Distribution of cognitive and personality skills across treated, controls, and comparable native students



Notes: These graphs shows the distribution of aspirations, perception of barriers, and INVALSI8 across treated students, control students, and a group of Italian students that are comparable in terms of schooling ability. Specifically, we match each immigrant student with a native student obtaining exactly the same score in INVALSI6.

Figure 2.7 compares the entire distribution of continuous intermediate outcomes – INVALSI8, aspirations, and perceived barriers – across treated, control, and native students who had the same standardized test score in grade 6. Overall, EOP produced sizeable effects on personality skills (reduced perception of barriers for both boys and girls, increased aspirations for boys), and it significantly improved cognitive skills of boys.

These changes in cognitive and personality skills induced teachers to revise their recommendations at the end of grade 8 (see the bottom-right panel of Table 2.6). On average, the probability that teachers recommend the high track is 12.3 percentage points higher for immigrant students in EOP schools, on a baseline of 46.4 percent in control schools. Interestingly, the change is substantially higher for male students: 17.2 percentage points, compared to 7.7 percentage points for female students. This is comforting, as teachers seem to update their beliefs based on *actual* changes in motivation and performance – greater for males, lower for females – as opposed to revising recommendations for all students involved in EOP.

Decomposing the treatment effect

Following Heckman et al. (2013), we decompose the treatment effect on educational choices into experimentally induced changes in the mediating factors in Table 2.6 and changes in other (unmeasured) factors. Assume the following linear model for the potential outcome when randomized into the treated ($d = 1$) and into the control group ($d = 0$):

$$Y_d = \tau_d + \sum_{j \in J} \alpha_d^j \theta_d^j + \beta_d \mathbf{X} + \epsilon_d, \quad d \in \{0, 1\}, \tag{2.1}$$

where Y is a dummy for choosing the high track, τ is the intercept, $\Theta = (\theta^j : j \in J)$ is the set of observed mediating factors (cognitive skills, personality traits, and teachers' recommendation), $\mathbf{X}$ is a vector of pre-program variables unaffected by the treatment (initial test score INVALSI6, generation of immigration, and province fixed effects), and ϵ_d is an error term. With the exception of $\mathbf{X}$, all variables and coefficients in equation (2.1) are allowed to depend on treatment assignment. In particular, τ_d captures the effect of experimentally induced changes in other (unobserved) determinants of Y, in addition to the observed mediating factors in Θ.

Separately identifying the components of the treatment effect attributable to τ_d and Θ, respectively, requires further assumptions as experimental variation allows us to consistently estimate the effects of EOP on measured factors and final educational decisions, but not the relationship between the former and the latter. Heckman et al. (2013) assume independence of observed and unobserved factors in the no-treatment state, conditional on the vector $\mathbf{X}$ of pre-treatment characteristics. Maintaining this assumption and imposing the additional testable restriction that coefficients do not vary with treatment assignment (respectively, $\alpha_d^j = \alpha$ for all j and $\beta_d = \beta$) allows us to decompose the effect of EOP as:

$$E\left(Y_1 - Y_0\right) = \sum_{j \in J} \alpha^j E\left(\theta_1^j - \theta_0^j\right) + \left(\tau_1 - \tau_0\right), \tag{2.2}$$

where $E\left(Y_1 - Y_0\right)$ is the average treatment effect; $E\left(\theta_1^j - \theta_0^j\right)$ is the average change induced in the j-th observed factor, and α^j is the associated effect on educational choices; finally, $(\tau_1 - \tau_0)$ is the effect due to other unmeasured factors. In Appendix Table B.7 we test and do not reject the structural invariance assumptions on α^j for all j and β.

Table 2.7: Decomposition of the effect of EOP on high-school choice, male students

	(1)		(2)		(3)	
	Explained	p-value	Explained	p-value	Explained	p-value
Aspirations	0.0413	[0.0005]	0.0359	[0.002]	0.0357	[0.001]
Barriers	-0.0095	[0.5756]	-0.006	[0.722]	-0.0024	[0.882]
INVALSI8			0.013	[0.103]	0.0101	[0.178]
Teachers' recommendation					0.0333	[0.046]
Total explained	0.0318		0.0429		0.0767	
EOP effect on choosing high-track	0.091		0.091		0.091	

Notes: This table decomposes the effect of EOP between changes in personality skills (aspirations and perception of barriers), increased schooling achievement (as measured by INVALSI8) and teachers' recommendations. As explained in Section 2.5.2, the decomposition follows the method devised by Heckman et al. (2013). Bootstrap standard errors clustered at school level generated from 1,000 iterations.

Table 2.7 shows decomposition (2.2) for the effect of EOP on males' educational choices.[26] Changes in personality skills explain about one third of the overall effect (column 1). However, this effect is entirely driven by aspirations, whereas perceptions of barriers do not seem an important mediating factor. This is consistent with the fact that males and females experience a similar decrease in perceived barriers, but educational choices change only for males. In column 2 we add schooling performance, as measured by the standardized test score in grade 8, as an additional mediating factor, and in column 3 we further add teachers' recommendation. Teachers' expectations on students' performance have been shown to play a crucial role in affecting academic performance and choices.[27]

[26]We only present results for males because treatment effects on females are not significantly different from zero.

[27]This phenomenon has been widely documented since the seminal work by Rosenthal and Jacobson (1968) as the Pygmalion and Golem effect, according to which respectively high expectations leads to an improvement in performance and low expectations lead to a decrease in performance. In particular, children of minority groups seem to have a greater advantage from positive expectations of teachers.

Indeed, we find that aspirations and teachers' recommendation are the most important factors, jointly explaining about two thirds of the treatment effect on educational choices. Improvements in schooling performance play a less crucial role, and the effect of perceived barriers remains not significantly different from zero. Overall, experimentally induced changes in measured skills and teachers' recommendation jointly explain 82 percent of the increase in the probability of choosing the high track, suggesting that other (unmeasured) factors account for a minor part of changes in educational choices. We obtain very similar results when we employ the alternative decomposition method by Gelbach (2016), see Appendix Table B.8.

Overall, these results confirm that EOP increased enrollment of high-achieving immigrants into the high track mainly by raising their educational and occupational aspirations. They suggest, in addition, that changes in aspirations and performance are incorporated into teachers recommendations, which may further amplify program effects.

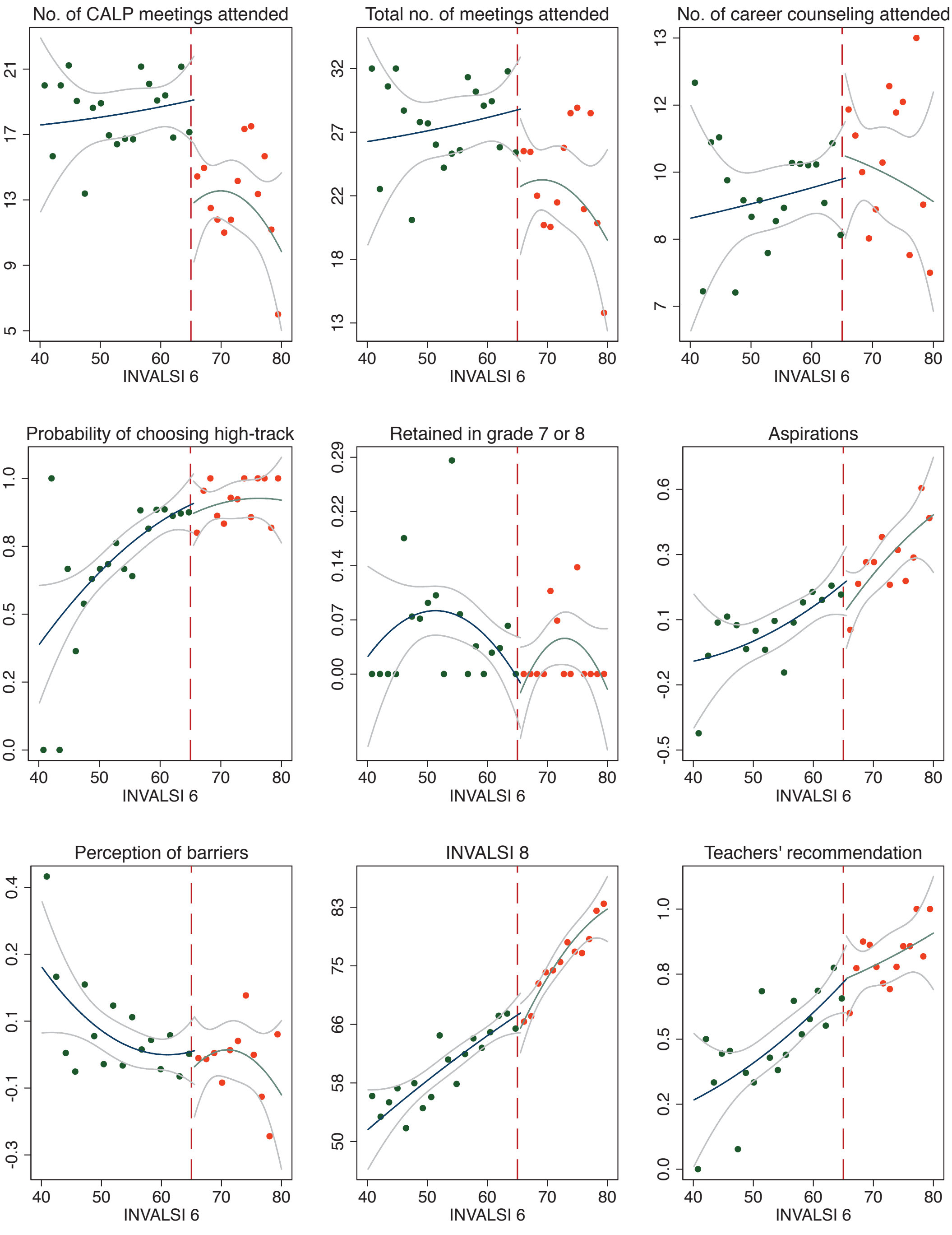
Career counseling and academic tutoring

A related question concerns the relative importance of the two components of the intervention: motivational and career counseling (the Social Cognitive Career Theory module) and specialized help for studying (the CALP module). Ideally, one would want to disentangle the contribution of each of these two components using a multi-arm treatment design. As explained in Section 2.3, however, ethical reasons prevented us from delivering the motivational treatment without also delivering help for studying.

As an alternative strategy, we compare outcomes between students in EOP schools that scored below and above the cutoff used to determine the number of meetings. In particular, while virtually every treated student was invited to 17 CALP meetings in grade 8, the students who had scored below 65 in the entry level test (INVALSI6) were invited to 12 additional CALP sessions in grade 7. This rule provides an ideal regression discontinuity design to isolate the effect of (additional) CALP meetings from the effect of the career orientation module. It does not allow to simulate a counterfactual in which CALP was not provided at all (extensive margin), but at least it sheds some insight on the intensive margin.

Indeed, the top three graphs in Figure 2.8 show that students with a score below 65 attended on average 5.5 more CALP meetings than students with a score above 65, whereas there are no differences in the number of other meetings. At the same time, the remaining graphs show that there is no significant discontinuity in the probability of choosing the high track or grade retention, nor are there significant differences in cognitive and personality skills between students on one side or the other of the cutoff. Therefore, neither final nor intermediate outcomes are significantly affected by an increase in the number of CALP meetings.

Figure 2.8: Effect of additional CALP meetings, regression discontinuity estimates



Notes: These graphs plot the number of meetings attended – distinguishing between career counselling and CALP modules – and treated students' outcomes against standardized test scores in grade 6 (INVALSI6). The vertical line indicates the cutoff score below which treated students are offered additional CALP meetings.

To conclude, EOP seems to operate mostly through motivational and career counseling, as opposed to specialized help for studying. These findings dovetail nicely with the evidence in Table 2.6 on aspirations being the main mediating factor, with a relatively smaller (though still significant) role of improvements in cognitive skills and school performance.

2.5.3 Longer term effects

The results presented so far confirm that EOP increased the probability of enrolling into academic and technical schools (high track) after grade 8. A potential concern is that such schools may prove too demanding for immigrant students from a disadvantaged background, even when considering students with relatively high ability. The reason is that immigrant students may face additional constraints compared to Italian students of similar ability. Examples of such constraints involve the lower degree of embeddedness in social networks that could help with studying challenging subjects, as well as financial constraints in paying for private tutoring (a practice sometimes used by Italian families when their children struggle in school).

Table 2.8: Effect of EOP on long-term outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Outcomes in grade 9				Outcomes in grade 10			
Dep. var.	Admitted to grade 10		Re-take courses		Dropout		Change school	
EOP	0.014	0.034	-0.099	-0.146	-0.034	-0.057	0.000	-0.005
	(0.032)	(0.042)	(0.072)	(0.115)	(0.027)	(0.039)	(0.022)	(0.034)
Female*EOP		-0.033		0.084		0.045		0.009
		(0.057)		(0.146)		(0.050)		(0.043)
Constant	0.380***	0.300***	0.346***	0.429***	0.343***	0.406***	0.161***	0.169***
	(0.044)	(0.048)	(0.126)	(0.150)	(0.038)	(0.041)	(0.043)	(0.047)
Mean dep. var. control males	0.411	0.411	0.251	0.251	0.365	0.365	0.119	0.119
Mean dep. var. control females	0.578	0.578	0.075	0.075	0.211	0.211	0.100	0.100
Observations	933	933	918	918	1,157	1,157	881	881
R-squared	0.054	0.077	0.025	0.028	0.063	0.083	0.011	0.011

Notes: This table shows the effect of EOP on immigrant students' outcomes in the first two years of high school, indicated above each column. EOP is a dummy equal to 1 for students in middle schools assigned to the treatment group and equal to zero for schools assigned to the control group. All regressions control in addition for a squared polynomial in INVALSI6, a dummy equal to 1 for first generation immigrants, and province fixed effects. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

In Table 2.8 we address this concern estimating the effect of (assignment to) EOP on performance during the first two years of high school. We consider four different outcomes. The first is the probability of being admitted to grade 10, the second year of high school (columns 1-2). The second is the number of make-up exams students need to take during the summer in order to avoid repeating the grade (typically no more than three exams).[28] The third outcome is the probability that a student drops out before completing grade 10 (columns 5-6), and the last outcome is the probability of changing school between grade 9 and 11 (columns 7-8).

For all these outcomes, we find that treated students are no more likely to experience difficulties compared to the control group. If anything, they are more likely to be admitted to grade 10, less likely to fail courses and less likely to drop out, although the estimated coefficients are not statistically significant.[29] It is worth stressing that the lack of a significant effect should not be seen as a shortcoming: given that our treated students were more likely to enroll in demanding high schools, the fact that they are doing as well as the control group (and if anything better, given the pattern of coefficients in Table 2.8) is actually a positive result.

2.5.4 Spillover effects

Our last piece of evidence relates to spillover effects. Any change in the achievement and educational choices of treated students may influence their peers – particularly those sharing the same immigrant background (Sacerdote et al., 2011). This may occur through several channels. Treated students could serve as role models to other students in their social network (Patacchini and Zenou, 2016). This may be particularly relevant in our context, because we showed that EOP had a strong impact on aspirations and barriers perceived by treated students. Another channel could be cognitive skills: given that EOP improved the academic performance of immigrant boys, this may have positive spillovers on their classmates. Another potential source of peer effects is time use. The fact that treated students spend more time studying or attending EOP sessions implies that they have less time to spend with their friends doing other (possibly less productive) activities.

We estimate the effect of EOP on treated students' classmates exploiting random assignment of the intervention across schools. Specifically, we include in the sample only classmates of treated and control students and estimate the following equation:

$$Y_{ics} = \alpha + \beta TreatedInClass_{cs} + \gamma \mathbf{X}_{ics} + \delta \mathbf{Z}_{cs} + u_{ics} \tag{2.3}$$

where i denotes the student, c the class, and s the school. $TreatedInClass_{cs}$ is a dummy

[28]The sample in columns 3-4 is smaller than in columns 1-2 because it only comprises students who did not fail outright and who did not dropout from high school. If a student is below the Pass level in many subjects, the decision of the school is typically to fail this student rather than give make-up exams after the summer.

[29]Ideally, we would also like to estimate the longer term effects on cognitive skills, as measured by the standardized test score in grade 10 (INVALSI10). Unfortunately, it is not possible to match MIUR registries for middle school students with their INVALSI tests in high school.

equal to 1 if the student belongs to a class with at least one treated student, and 0 if he/she belongs to a class with at least one control student (i.e., a class in a control school that contains at least one immigrant student whose INVALSI6 score was among the top 10 of the school). In this way, the sample used in the regression contains students from classes that are comparable in terms of having high-achieving immigrant students among them, some of which received EOP and others not. X_{ics} includes all the individual characteristics in our baseline specification (gender, dummy for first generation immigrants, and second degree polynomial in INVALSI6); Z_{cs} is a vector of controls at the school and class level (school size, class size, and percentage of immigrants in the class).

Table 2.9 shows the effect of being in the same class with a treated student on native (columns 1-3) and immigrant classmates (columns 4-6). Immigrant males who are in the same class with a treated student experience a decrease in grade retention (Panel B), whereas immigrant females in the same class with a treated student are more likely to enroll in the high track at the end of middle school (Panel A). Interestingly, the effects are comparable in size to those observed for the treated student themselves. This can be rationalized by observing that, by construction, non-treated immigrant classmates are selected to have lower academic performance at the beginning of middle schools (they were not among the top 10 immigrants of the school in terms of INVALSI6). It is thus possible that immigrant students down in the ability distribution are very responsive to (even indirect) treatment effects. This interpretation is reinforced when observing that immigrant females in the mid-to-lower part of the distribution make different choices from comparable native females in the absence of the intervention, whereas this is not the case for immigrant females in the upper part of the distribution (see Figure 2.1). Therefore it should not be surprising that, while EOP did not affect track choice of treated females (who were selected to be high performing and already made comparable choices to natives), it had an impact on non-treated females, for whom a 'choice gap' existed compared to natives of similar ability.

Why are male immigrant classmates not affected in terms of track choice (Panel A, column 5)? Our interpretation is that they are too far from the margin at which a demanding high school would become a preferable choice. This can be seen for example when comparing soft skills of immigrant boys and girls who did not qualify for the program. Female immigrant classmates of students eligible for EOP have aspiration levels that are closer to those of treated male students compared to other male immigrant classmates.[30]

[30]From the survey data on control schools, we caculated the average standardized level of aspirations of different subgroup of immigrant students. 'Control' boys and girls (that is, immigrants whose INVALSI6 was among the top 10) have an average of -0.11 and 0.13, respectively. Their immigrant classmates (not top 10) have an average of -0.26 if males, and -0.13 if females. Hence, in terms of average academic motivation, female immigrant classmates are closer than male classmates to students selected for the program. Elective affinity in soft skills may ease the diffusion of positive spillover effects of the program

Table 2.9: Peer effects in EOP schools

	(1)	(2)	(3)	(4)	(5)	(6)
	Native classmates			Immigrant classmates		
	All	Males	Females	All	Males	Females
Panel A - Dependent Variable: Choosing the high-track						
EOP class	0.001	0.008	-0.006	0.038	-0.009	0.091**
	(0.013)	(0.015)	(0.017)	(0.029)	(0.041)	(0.038)
Mean dep. var.	0.767	0.755	0.780	0.421	0.423	0.419
Observations	8,429	4,247	4,182	1,308	686	622
R-squared	0.197	0.205	0.193	0.097	0.096	0.125
Panel B - Dependent Variable: Retained in grade 7 or 8						
EOP class	-0.002	-0.005	0.002	-0.042*	-0.064**	-0.019
	(0.007)	(0.009)	(0.007)	(0.023)	(0.032)	(0.027)
Mean dep. var.	0.044	0.053	0.035	0.148	0.192	0.099
Observations	8,429	4,247	4182	1,308	686	622
R-squared	0.047	0.040	0.066	0.046	0.049	0.053
Panel C - Dependent Variable: INVALSI8						
EOP class	-0.017	-0.012	-0.021	-0.026	0.023	-0.069
	(0.030)	(0.034)	(0.032)	(0.062)	(0.077)	(0.081)
Mean dep. var.	0.129	0.139	0.119	-0.961	-0.960	-0.965
Observations	7,533	3,736	3,797	1,007	500	507
R-squared	0.591	0.603	0.580	0.330	0.349	0.319
Panel D - Dependent Variable: Teachers' raccomandation						
EOP class	-0.011	-0.012	-0.010	0.031	0.003	0.062
	(0.037)	(0.038)	(0.039)	(0.038)	(0.039)	(0.052)
Mean dep. var.	0.597	0.574	0.621	0.207	0.210	0.205
Observations	8,429	4,247	4,182	1,308	686	622
R-squared	0.223	0.235	0.218	0.115	0.130	0.112

Notes: This table shows the effect of being in the same class with an immigrant student randomized into the intervention on several outcomes of interest, indicated in the title of each panel. The sample includes only classmates of treated and control students. The main expanatory variable, EOP class, is a dummy equal to 1 for the classmates of treated students. All regressions control in addition for all the individual characteristics in our baseline specification (dummy for first generation immigrants, second degree polynomial of test score in grade 6, and province fixed effects) as well as for class size, percentage of immigrants in the same class, and school size. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Therefore, they may be closer to the threshold beyond which the choice of a demanding high school is affected. At the same time, the sizeable and significant reduction in the likelihood of grade retention for male immigrant classmates (Panel B, column 5) suggests that spillover effects did operate at that (lower) margin.

One last comment relates to native students. Columns 1-3 of Table 2.9 show that there are no significant differences in outcomes between the non-immigrant classmates of treated and control students. This finding is consistent with previous evidence that peer effects are particularly strong within groups with a similar background (see, e.g. Sacerdote et al., 2011).

Overall, the positive spillovers on immigrant classmates uncovered by our analysis have significant implications when assessing the success and cost-effectiveness of the intervention. We briefly discuss the latter in the concluding section.

2.6 Conclusions

Educational segregation is a significant risk in societies where school tracking occurs at an early age. This risk disproportionately affects students whose parents have less information about or are less integrated in the local education system, such as children of immigrants. We show that it is possible to reduce the mismatch created by early tracking through an innovative counseling program that provides a mix of soft skills training and academic tutoring. The program, known as EOP and implemented in a random sample of middle schools in Northern Italy, targeted high-achieving immigrant students selected on the basis of their test performance in grade 6. Two years later, immigrant boys assigned to treatment were 12 percent more likely than control ones to enroll in academic or technical high schools (as opposed to vocational ones). The program virtually closed the gap between native and immigrant boys in high school choice. No effect was found for girls, for which no mismatch was detected in the first place.

Thanks to detailed data on cognitive and non-cognitive outcomes, we are able to explore the channels through which the effect operates. Treated boys see improvements in grade 8 standardized test scores, as well as in motivation and aspirations, and perceive fewer barriers in their future. Also, teachers are more likely to formally recommend treated boys for demanding high school tracks. A variance decomposition exercise suggests that most of the impact on high school choice is mediated by aspirations and teachers' recommendations.

The significance and magnitude of our effects is noteworthy when compared to existing evidence from randomized field experiments that study partial derivatives of the human

through friendship network (Rapallini and Rustichini, 2016).

capital production function, recently summarized by Kautz et al. (2014) and Fryer Jr (2016). For instance, the latter shows that, among school interventions, only "high-dosage" tutoring (defined as "being tutored in group of 6 or fewer students for 4 or more days per week") has significant effects.[31] Our EOP treatment can be considered as "low dosage", given the above definition, yet its impact on treated boys is 0.188 standard deviations for math achievement and 0.139 standard deviations for reading.

Having established that EOP had sizeable and significant impacts, it is important to know if it is also cost-effective. A full fledged cost benefit analysis is not possible at this stage: on top of the challenge of quantifying non-pecuniary benefits and costs, students involved in our experiment have not yet completed secondary school and therefore lifetime earnings profiles are not observed (Heckman et al., 2017). The computation of the lifetime rate of return of EOP is therefore naturally based on assumptions about long-term outcomes such as college enrollment, earnings and unemployment. Although EOP may potentially have strong effects on health and criminal behavior, we present conservative estimates focusing our cost-benefit analysis only on social benefits coming from higher income taxes and public savings on unemployment insurance (Heckman et al., 2010; Eisenhauer et al., 2015). We examine the sensitivity of social rates of returns to a plausible range of assumptions. Appendix Table B.9 reports our calculations. Extrapolating the long-term benefits first only on those treated individuals who were directly affected through a reduction in grade retention or a change in their high-school choice toward a more demanding track, we estimate social rates of return between 3 and 5 percent. However, including the positive spillovers on the immigrant classmates of treated students (that we estimated in section 2.5.4), we estimate that the lifetime rate of return of EOP is between 7 and 9 percent, close to the historical return on equity.

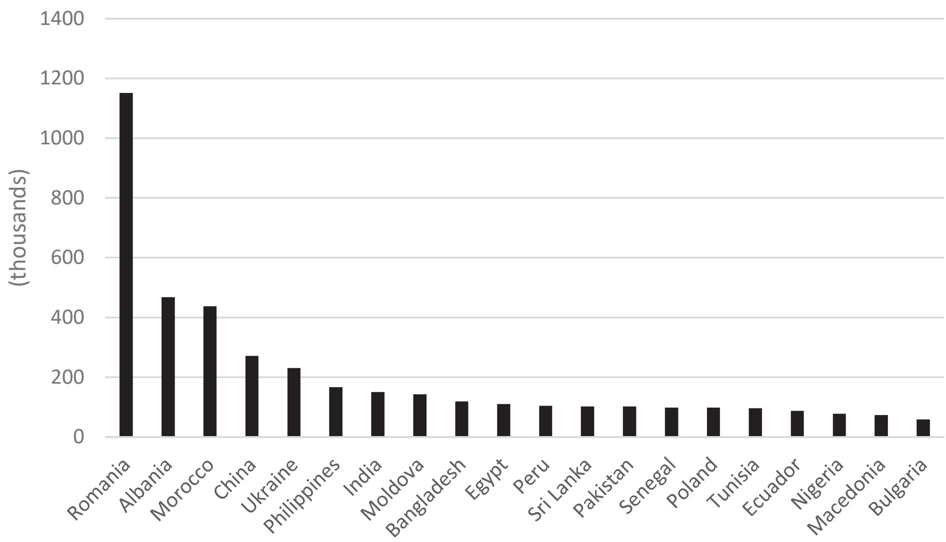
Our finding that soft skills played a more important role than improved test performance in determining high school choice suggests that scaled down versions of the program may be even more cost effective. For example, one could reduce the number of meetings with academic tutors and explore forms of delivering the information and motivational components of the program through teachers as opposed to dedicated counselors, to reduce costs. We leave this to future work.

[31] Fryer Jr (2016) finds that high-dosage tutoring leads to a meta-coefficient of 0.309 standard deviations for math achievement (with a standard error of 0.106) and 0.229 standard deviations for reading achievement (with a standard error of 0.033).

B. Appendix

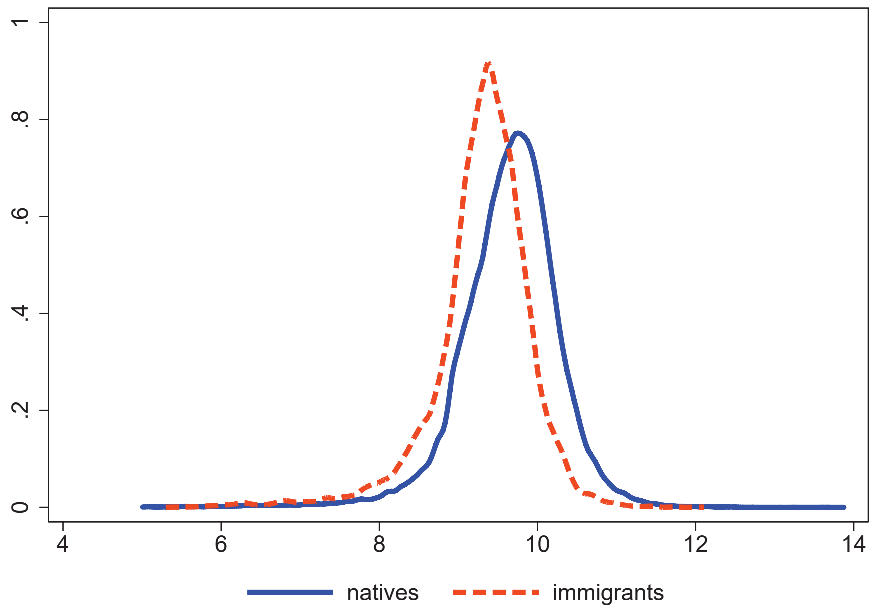
B.1 Additional Tables and Figures

Figure B.1: Immigrants in Italy by nationality, 2015



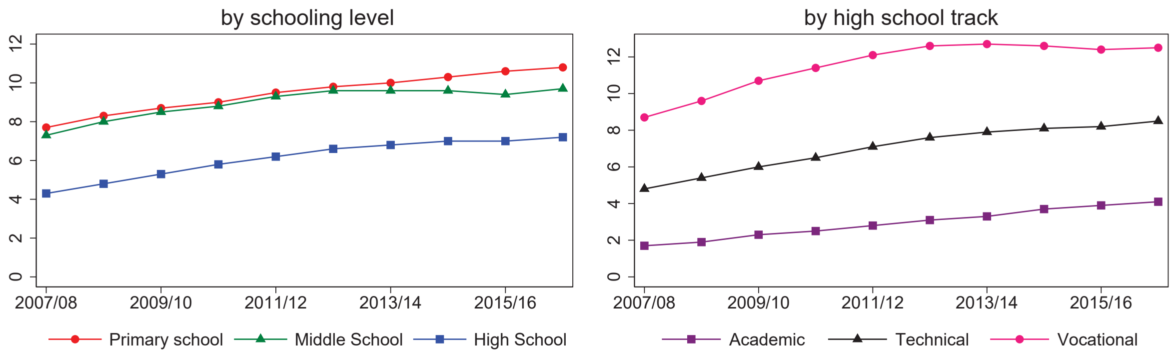
Source: ISTAT, "Demografia in Cifre", several years (www.demo.istat.it).

Figure B.2: Distribution of (log) income across native and immigrant families in Italy



Notes: This graph shows the distribution of (log) disposable income per equivalent adult at constant 2010 prices. *Source:* European Union Statistics on Income and Living Conditions (EU-SILC), 2007-2014.

Figure B.3: Percentage of immigrants over total students in Italy, by schooling level and high school track



Source: MIUR, "‘Portale dei dati sulla scuola"’ (dati.istruzione.it), several years.

Table B.1: Educational and occupational outcomes 4 years after graduation, by high-school track

	all students		males		females	
	high track	low track	high track	low track	high track	low track
Panel A: Native students						
Percentage of graduates by track	85.5	14.5	84.4	15.6	85.5	13.5
Ever enrolled into university	0.704	0.205	0.650	0.158	0.754	0.256
	(0.003)	(0.004)	(0.006)	(0.006)	(0.004)	(0.007)
Dropout rate in university	0.118	0.306	0.145	0.353	0.097	0.274
	(0.003)	(0.011)	(0.005)	(0.018)	(0.003)	(0.014)
Not in Employment, Education or Training (NEET)	0.199	0.291	0.189	0.264	0.208	0.320
	(0.003)	(0.005)	(0.005)	(0.007)	(0.004)	(0.007)
Regretting high school choice	0.267	0.318	0.266	0.304	0.269	0.333
	(0.003)	(0.005)	(0.005)	(0.007)	(0.004)	(0.008)
Panel B: Immigrant students						
Percentage of graduates by track	62.8	37.2	57.4	42.6	66.3	33.7
Ever enrolled into university	0.655	0.291	0.686	0.172	0.637	0.390
	(0.03)	(0.022)	(0.049)	(0.027)	(0.037)	(0.032)
Dropout rate in university	0.150	0.257	0.231	0.307	0.100	0.238
	(0.028)	(0.037)	(0.06)	(0.074)	(0.028)	(0.043)
Not in Employment, Education or Training (NEET)	0.264	0.294	0.237	0.238	0.280	0.340
	(0.027)	(0.022)	(0.045)	(0.03)	(0.035)	(0.031)
Regretting about high school choice	0.269	0.331	0.325	0.290	0.238	0.365
	(0.028)	(0.023)	(0.049)	(0.032)	(0.033)	(0.032)

Notes: This table shows average educational and occupational outcomes of students graduating from high school in year 2011 by gender and high school track; separate figures for native and immigrant students are presented in panel A and B, respectively. Standard errors are reported in parentheses.

Table B.2: Immigrant students' probability of choosing the high-track, controlling for socio-economic background

	Dependent variable: Choosing the high-track					
	(1)	(2)	(3)	(4)	(5)	(6)
	Males			Females		
Immigrant	-0.086***	-0.078***	-0.066***	-0.048***	-0.048***	-0.027
	(0.021)	(0.021)	(0.021)	(0.018)	(0.018)	(0.018)
Low-educated Mother		-0.164***	-0.134***		-0.173***	-0.136***
		(0.027)	(0.028)		(0.024)	(0.025)
Mid-educated Mother		-0.033*	-0.015		-0.064***	-0.046***
		(0.018)	(0.018)		(0.018)	(0.018)
Low-educated Father		-0.056**	-0.085***		-0.035	-0.038
		(0.023)	(0.024)		(0.024)	(0.024)
Mid-educated Father		0.031*	-0.005		0.041**	0.024
		(0.018)	(0.017)		(0.017)	(0.017)
Mother bluecollar			-0.044*			-0.057***
			(0.022)			(0.021)
Mother home/unemployed			-0.049			-0.118***
			(0.036)			(0.032)
Father bluecollar			-0.018			-0.032
			(0.026)			(0.024)
Father home/unemployed			0.000			-0.015
			(0.020)			(0.018)
Constant	0.711***	0.819***	0.855***	0.746***	0.845***	0.870***
	(0.009)	(0.017)	(0.017)	(0.008)	(0.016)	(0.017)
Observations	3,927	3,927	3,927	3,819	3,819	3,819
R-squared	0.213	0.244	0.252	0.204	0.227	0.234

Notes: This table shows how immigrant status influences the probability of choosing the high-track. The dependent variable is a dummy equal to 1 for students choosing the high-track. The main explanatory variable is a dummy equal to 1 for immigrant students. The sample includes all students in control schools. All regressions control in addition for a second degree polynomial of test score in grade 6 (INVALSI6), a dummy for first generation immigrants, and province fixed effects. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table B.3: The effect of completing the questionnaire on soft skills in control schools

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Demanding High-School		Grade Retention		Std Test Score grade 8	
School Questionnaire	-0.016	-0.004	0.008	0.007	-0.061	-0.085
	(0.039)	(0.057)	(0.021)	(0.033)	(0.100)	(0.125)
Fem*School Questionnaire		-0.029		0.004		0.035
		(0.069)		(0.036)		(0.140)
Fem		0.167***		-0.058**		0.168
		(0.050)		(0.025)		(0.111)
Constant	0.759***	0.676***	0.052***	0.081***	0.004	-0.081
	(0.027)	(0.042)	(0.014)	(0.023)	(0.069)	(0.087)
Observations	620	620	620	620	552	552
R-squared	0.000	0.031	0.000	0.015	0.001	0.015

Notes: This table tests whether control students in schools selected for the questionnaire differ in their high-school choice, grade retention and performance in the standardized test score from students in the control schools not selected for the questionnaire. The dependent variable is a dummy equal to 1 for students choosing the high-track in columns 1 and 2, a dummy equal to 1 if students are retained in grade 7 or 8 in columns 3 and 4 and the standardized value for the test score in columns 5 and 6. Standard errors clustered by school are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence level, respectively.

Table B.4: Principal component analysis, factor loadings

First principal component: Aspirations		
	Loadings	std. err.
Goal University	1	.
Self efficacy University	1.649	0.050
Self efficacy White collar	0.753	0.030
Self efficacy Manager	0.631	0.030
Second principal component: Perception of barriers		
	Loadings	std. err.
Barriers economic	1	.
Barriers family ideas	1.339	0.087
Barriers prejudice	0.837	0.063
Barriers family plans and marriage	1.001	0.074
Barriers self esteem	0.947	0.072

Notes: This table shows estimated factor loadings for the two principal components extracted from psychological measures; Satorra-Bentler robust standard errors are also presented. Measurements are categorical variables in a scale from 1 to 4. “Goal University" is the answer to the following question: *Thinking about your future, do you want to achieve an university degree?*. The measurements related to “Self efficacy" are the answers to the following questions: *Independently from your educational aim but thinking about your abilities, do you think you could get a university degree/ white collar job/ managerial job?*. The measurements related to “Barriers" are the answers to the following questions: *Do you think the following barriers could be an obstacle in the achievement of your educational aims? Economic resources/ The needs and ideas of your family/ Racial prejudice/ Family plans (children, marriage)/ Not feeling good enough.*

Table B.5: Initial vs. working sample

	Treated	Controls	Difference	P-value	Std. Difference
Initial sample	700	751			
Fraction missing match MIUR-INVALSI	0.043	0.053	- 0.010	[0.72]	-0.049
Number of students with available MIUR-INVALSI	670	711			
Fraction dropped between INVALSI6 and start of EOP	0.109	0.128	- 0.019	[0.51]	-0.059
Final sample	597	620			

Notes: This table shows the sample size of treated and control students in our sample. P-value for difference in means are reported in parenthesis. The last column also reports the standardized difference between group averages.

Table B.6: Treatment effect on soft skills (by survey question)

ITT on	Coefficient	p-value	p-value FWER
Group 1: aspirations			
Goal University	0.087	0.073	0.082
Self efficacy University	0.201	0.018	0.054
Self efficacy Whitecollar	0.215	0.001	0.007
Self efficacy Manager	0.150	0.027	0.062
Group 2: perception of environmental barriers			
Barriers economic	-0.182	0.008	0.019
Barriers family ideas	-0.203	0.002	0.007
Barriers prejudice	-0.280	0.000	0.001
Barriers family formation and marriage	-0.087	0.177	0.177
Barriers self esteem	-0.274	0.000	0.001

Notes: Robust standard errors clustered at school level. All regressions include generation of immigration, province and squared test score. P-values are adjusted for multiple hypothesis testing using the free step-down resampling method (Westfall and Young, 1993) to control the family-wise error rate (FWER). Measurements are categorical variables in a scale from 1 to 4. “Goal University" is the answer to the following question: *Thinking about your future, do you want to achieve an university degree?.* The measurements related to “Self efficacy" are the answers to the following questions: *Independently from your educational aim but thinking about your abilities, do you think you could get a university degree/ white collar job/ managerial job?.* The measurements related to “Barriers" are the answers to the following questions: *Do you think the following barriers could be an obstacle in the achievement of your educational aims? Economic resources/ The needs and ideas of your family/ Racial prejudice/ Family plans (children, marriage)/ Not feeling good enough.*

Table B.7: Specification test, Males

Outcome: Choosing the high-track	(1)		(2)		(3)	
Mediating factors: $H_0 : \alpha_1^j = \alpha_0^j$	Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value
Aspirations	1.50	[0.14]	1.34	[0.18]	1.68	[0.10]
Barriers	1.14	[0.26]	1.04	[0.30]	0.49	[0.63]
INVALSI8			0.14	[0.89]	0.60	[0.55]
Teachers' recommendation					0.60	[0.55]
Controls: $H_0 : \beta_1^i = \beta_0^i$	Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value
INVALSI6	0.72	[0.47]	0.55	[0.58]	0.88	[0.38]
INVALSI6 sq.	1.58	[0.12]	1.49	[0.14]	1.39	[0.17]
First generation immigrant	0.96	[0.34]	0.86	[0.39]	0.74	[0.46]
Prov BS	1.53	[0.13]	1.33	[0.19]	0.46	[0.65]
Prov GE	1.10	[0.28]	1.00	[0.32]	1.32	[0.19]
Prov MI	0.29	[0.77]	0.06	[0.96]	0.51	[0.61]
Prov PD	0.60	[0.55]	0.33	[0.75]	0.17	[0.86]
Prov TO	0.98	[0.33]	0.98	[0.33]	0.31	[0.76]
F-test	1.37	[0.21]	1.28	[0.26]	1.35	[0.23]

Notes: The first panel tests whether the treatment group regression coefficients in equation 2.2 are the same as the control group coefficients: $H_0 : \alpha_1^i = \alpha_0^i$, for each potential channel θ. The second panel tests whether the treatment group regression coefficients are the same as the control group coefficients: $H_0 : \beta_1^i = \beta_0^i$, for each potential control variable $\mathbf{X}$. In column (1), we consider only two mediating factors, i.e. aspirations and barriers, while in column (2) we include also the achievement test scores and in column (3) teachers' track recommendation.

Table B.8: Decomposition of the effect of EOP on high-school choice, male students (Gelbach, 2016)

	(1)		(2)		(3)	
	Explained	p-value	Explained	p-value	Explained	p-value
Aspiration	0.0354	[0.034]	0.0291	[0.042]	0.0264	[0.046]
Barriers	0.0046	[0.602]	0.0067	[0.446]	0.0042	[0.620]
Cognitive skills			0.0178	[0.085]	0.0105	[0.190]
Teachers' recommendation					0.0385	[0.011]
Total explained	0.0400	[0.032]	0.0537	[0.010]	0.0796	[0.001]
EOP effect on choosing high-track	0.091		0.091		0.091	

Notes: This table decomposes the effect of EOP between changes in personality skills (aspirations and perception of barriers), increased schooling achievement (as measured by INVALSI8) and teachers' recommendations. The decomposition follows the method devised by Gelbach (2016).

B.2 Cost Benefit Analysis

Table B.9: Cost Benefit Analysis

	Scenario 1	Scenario 2	Scenario 3
Parameters			
Discount rate	3%	3%	3%
Tax rate	28%	28%	28%
Higher salary per month (euros)	500	500	650
Lower unemployment probability	4%	4%	6%
Unemployment insurance benefit per month (euros)	1000	1000	1300
Number of beneficiaries	60	125	125
Costs and Benefits			
Total costs (thousand euros)	2,177	2,177	2,177
Higher taxes on wage (thousand euros)	3,344	7,006	9,108
Lower unemployment insurance (thousand euros)	955	2,002	3,904
Internal Rate of Return	2.8%	6.6%	8.8%

Notes: Although EOP program has potentially strong effects on health and on reduction of crime rates, we present conservative estimates focusing our cost-benefit analysis only on social benefits coming from higher income taxes and public savings on unemployment insurance. In the first scenario, we consider potential benefits only on 10% of students directly treated by EOP. In the second scenario, keeping all other assumptions constant, we consider also the additional spillovers on 5% of classmates of treated students (close to the share who did not fail the school year or decided to attend a more demanding track compared to classmates of control students). In the last scenario, we slightly reduce the unemployment probability, we slightly increase the expected average higher salary per month and the expected unemployment insurance benefit.

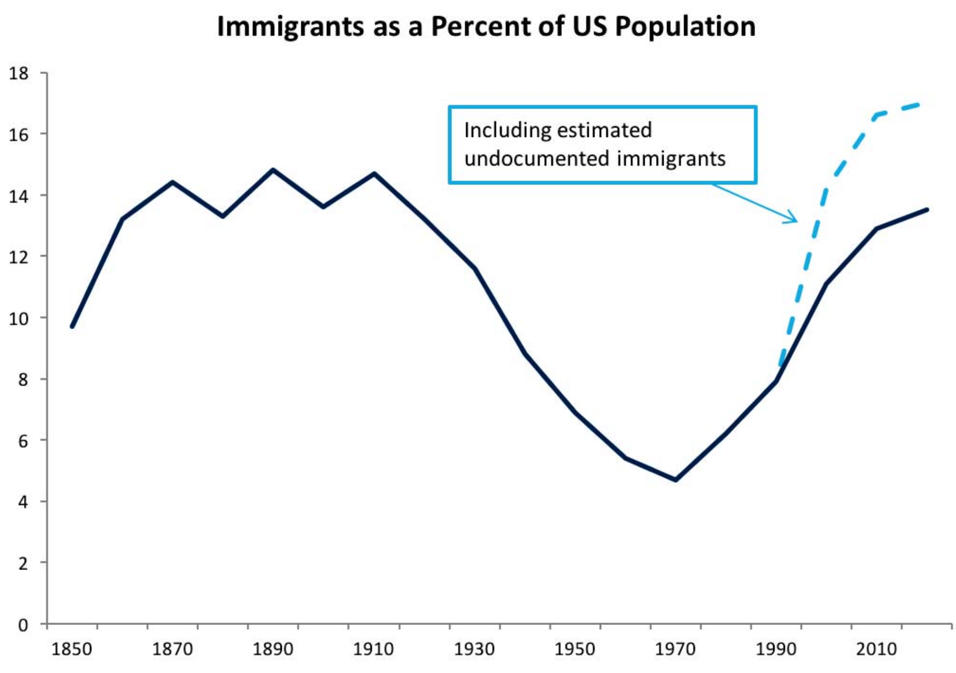
3. Happily Ever After: Immigration, Natives' Marriage, and Fertility *(joint with M. Tabellini)*

Abstract
In this paper, we study the effects of immigration on natives' marriage, fertility, and family formation across US cities between 1910 and 1930. Instrumenting immigrants' location decision by interacting pre-existing ethnic settlements with aggregate migration flows, we find that immigration raised marriage rates, fertility, and the propensity to leave the parental house for young native men and women. We show that these effects were driven by the large and positive impact of immigration on native men's employment and occupational standing, which increased the supply of "marriageable men". We also explore alternative mechanisms - changes in sex ratios, natives' cultural responses, and displacement effects of immigrants on female employment - and provide evidence that none of them can account for a quantitatively relevant fraction of our results. *JEL: J12, J13, J61, N32* .

3.1 Introduction

Between 1970 and 2010, the number of foreign born individuals living in the United States increased from 9 to roughly 40 million, with the share of immigrants over US population skyrocketing from 4.7 to 13 percent (Figure 3.1).[1] As for previous immigration waves in American history, alongside these trends, a heated debate on the economic, social, and political consequences of immigration has emerged (Porter, 2017). A large body of the literature has investigated the economic effects of immigration, testing in particular if immigrants lower natives' wages and employment (Card, 2001; Borjas, 2003; Borjas and Katz, 2007). A more recent set of papers has studied how immigration affects political outcomes and electoral results (Mayda et al., 2016; Halla et al., 2017). Somewhat surprisingly, however, much less is known about the impact of immigration on key social outcomes, such as marriage rates, fertility, and family formation among natives.

Figure 3.1: Immigrants as Percent of US Population



Notes: The solid line shows the number of legal immigrants as a percent of US population. The dashed line includes also the estimated number of illegal immigrants, available from 2000 onwards. Source: the number of legal immigrants comes from the Migration Policy Institute, while the number of illegal immigrants was taken from the Pew Research Center tabulations.

[1] US immigration statistics are underestimated because of the presence of large numbers of undocumented immigrants. According to some recent estimates (see Pew Center 2017), if undocumented immigrants were included, the share of foreign born over US population would be at least 4 percentage points higher (i.e. around 17 percent).

In this paper, we study how the inflow of immigrants to US cities between 1910 and 1930 affected marriage and fertility of young natives, as well as their decision to leave the parental house and set up an independent family unit. At the beginning of the twentieth century, 14 percent of the US population was foreign born, following the migration of more than 30 million Europeans between 1850 and 1915 (Figure 3.1). After 1915, however, World War I and the Immigration Acts (1921 and 1924) put an end to the Age of Mass Migration and drastically reduced immigration to the US (Abramitzky and Boustan, 2017). The key feature of these shocks is that they had heterogeneous effects across European countries. Since immigrants tend to cluster geographically along ethnic lines (Card, 2001), variation across sending regions mechanically translates into variation in the number as well as in the mix of immigrants received by US cities over time.

Exploiting this variation, we construct a “leave-out" version of the classic shift-share instrument often adopted in the immigration literature (Altonji and Card, 1991; Card, 2001). In particular, we predict the number of immigrants to US cities in a given year by interacting the geographic variation in historical settlements of different ethnic groups with the time-series variation in national flows from each sending region, net of the individuals that eventually settled in a given city's metropolitan statistical area (MSA).[2] The key identifying assumption behind the instrument is that the city-specific characteristics that attracted early movers from each ethnic group must not have a time-varying effect on local economic and social conditions in subsequent decades. For instance, this assumption would be violated if immigrants in 1900 settled in a given city anticipating subsequent economic growth. Below, we perform a number of checks - including testing for pre-trends and interacting year dummies with pre-migration city characteristics - to assess the validity of the instrument, and show that our results are robust to the use of alternative specifications.

Using this empirical strategy, we find that immigration increased marriage and fertility of native men and women. Our estimates are economically large, and suggest that a five percentage point (equivalent to a one standard deviation) increase in immigration raised natives' marriage rates and the children-to-women ratio by 2 and 3 percent respectively. When decomposing the increase in fertility between the intensive (i.e. more children per woman) and the extensive (i.e. more women having at least one child) margin, we document that the latter was quantitatively more important than the former. Specifically, our estimates imply that for every 10 new babies born from native women, 7 were due to the extensive margin, while only 3 were due to the intensive margin.

Exploiting the granularity of full count data, we explore which age groups were respon-

[2]We focus on European immigrants (see Table C.1 in the appendix for the complete list of sending countries), but results are robust to extending the analysis to all other non-European countries.

sible for the aggregate patterns just described. We show that the increase in both fertility and marriage was entirely driven by young couples, namely women (resp. men) aged 18-25 (resp. 20-27). Consistent with these findings, we also document that immigration induced young natives to leave their parental house earlier, and to set up an independent family unit.

In the second part of the paper, we investigate the mechanisms behind our main results. We provide evidence that they were driven by the large, positive impact of immigration on native men's employment, which increased the supply of "marriageable men" and made it easier for both men and women to marry, set up an independent household, and have kids. Specifically, our estimates suggest that for every ten new immigrants, one and a half more jobs were created for native men aged 20-35. Immigration also fostered natives' occupational standing, by inducing native workers to take up better jobs and move away from occupations more exposed to immigrants' competition, which tended to have lower skill requirements. Such large, positive effects on natives' employment were made possible by the fact that immigration increased firms' investment and productivity, in turn promoting industrialization and economic activity.

Next, we test a number of alternative mechanisms, and provide evidence that none of them can account for a quantitatively relevant fraction of our findings. First, we explore the possibility that immigration increased female marriage rates and fertility by altering sex ratios, i.e. the relative number of men and women, as more than 60% of immigrants entering the US at the time were young men. (Figure C.1) In contrast with this mechanism, however, immigration induced not only native women but also native men to marry more and to have more kids, suggesting that changes in sex ratios alone cannot be driving our main results. Furthermore, we provide evidence that the inflow of immigrants did not raise the probability that native women married and had kids with foreign born. In fact, consistent with the idea that marriage markets were highly segmented along ethnic lines, almost all native women were getting married with native husbands. We also unveil some interesting heterogeneous patterns, which depended on men's parentage. Specifically, even if immigration had on average a positive effect on marriage rates of native men, this did not happen for second generation men, who were probably more exposed to immigrants' competition in the marriage market (Angrist, 2002)
.

Second, we rule out the possibility that higher marriage rates among US born were the result of a cultural response by native couples aimed at "preserving" their own race (Bisin and Verdier, 2000; Spolaore and Wacziarg, 2016). In particular, building on the measure of linguistic distance from Chiswick and Miller (2005), we construct an index of cultural diversity, and show that the latter did not have any effect on natives' marriage rates. Lastly, we provide evidence that direct (negative) effects of immigration on female labor force participation, which might have induced women to first leave the labor force and then get married and have kids, cannot explain our key findings. Exploiting variation across age groups, we show that the decrease in labor force participation was limited to women whose marriage rates increased in response to immigration. Given the stigma attached to the work of wives outside the home at the beginning of the twentieth century, women were likely to quit their job as a consequence of marriage (Goldin, 2006).

Our results are related to several strands of the literature. First, we complement the recent paper by Autor et al. (2017) by showing that a positive (rather than a negative) shock to employment opportunities of men increases (instead of reducing) marriage, fertility, and financial independence of young couples. Despite the difference in the historical context - early twentieth century vs contemporaneous period - and in the source of the income shock - immigration vs trade - comparing results in this paper with those in Autor et al. (2017) suggests that some key policy-relevant parameters, such as the elasticity of marriage and fertility to income, can be stable over time. At the same time, however, while our estimates on fertility are in line with those in Kearney and Wilson (2017), differently from us, the latter paper does not find a positive effect of an employment boom on marriage rates. One possible interpretation for this difference is that the cultural environment might mediate the transmission of income shocks to social outcomes.

Second, our paper is related to the vast literature on the effects of sex ratios on marriage market outcomes of men and women. Focusing on the same historical context, Angrist (2002) exploits variation in sex ratios for second generation immigrants induced by the arrival of individuals from different countries. We complement this paper by showing that immigration can impact marriage rates and fertility in receiving countries not only by altering sex ratios for second generation immigrants, but also by affecting natives' employment.[3] Moreover, the differential effect of immigration on marriage rates of native men (positive for natives with native parents, but close to zero for second generation immigrants) is consistent with findings in Abramitzky et al. (2011), who show that in French regions where more men died during WWI, men (resp. women) were better (resp. worse) off in the marriage market.

Lastly, our paper contributes to the literature that explores the effect of immigration on female labor force participation and fertility. Findings in Furtado and Hock (2010) and Furtado (2016) suggest that the availability of lower cost childcare opportunities offered by the inflow of immigrants in recent decades allowed college educated women to both have more children and work longer hours, attenuating the negative correlation between

[3]We find a positive effect of immigration on natives' employment, which in turn induced natives to marry and have kids more often. However, it is possible that, when immigration decreases natives' employment, it also lowers fertility, marriage rates, and the propensity to set up independent households.

childbearing and labor force participation. This mechanism is unlikely to be at play in our context since, at the beginning of twentieth century, most women took care of their own children, and additional childbearing was assigned to black - and not immigrant - women. Goldin (1990, 2006) shows that in this historical period native women would quit their job upon getting married and having a child. Consistently with these findings, we show that the negative effect of immigration on female labor force participation was concentrated exclusively on women in the age group that experienced an increase in fertility and in marriage.

The paper is structured as follows. Section 2 describes the historical background. Section 3 presents the data. Section 4 lays out our empirical strategy, constructs the instrument for immigration, and reports first stage results. Section 5 investigates the effects of immigration on natives' marriage, fertility, and propensity to leave the parental house. Section 6, explores the mechanisms. Section 7 concludes.

3.2 Historical Background

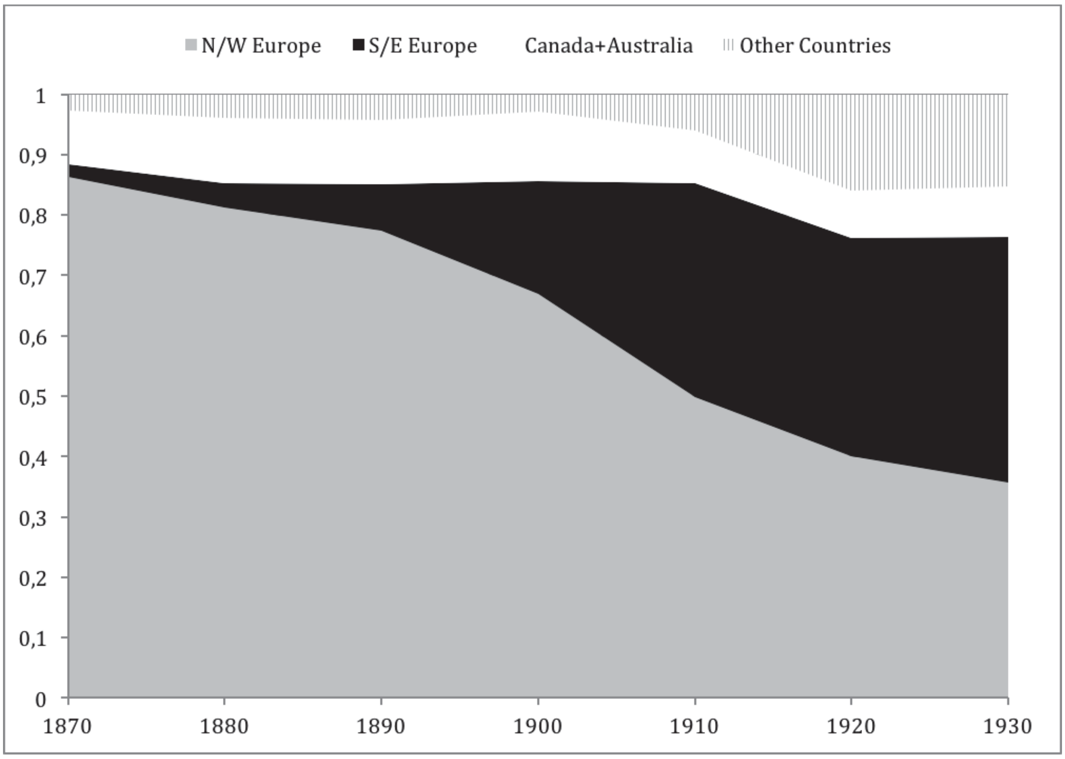
3.2.1 The Age of Mass Migration

Between 1850 and 1915, more than 30 million Europeans migrated to the United States. This massive migration episode took place in two waves: until 1890, most immigrants came from the British Isles, Germany, and Scandinavia; then, from the late 1880s, following the introduction of steam technology in shipping, which drastically reduced migration costs, immigration from Southern and Eastern Europe increased steadily (Keeling, 1999). In 1870, almost 90% of the foreign born came from Northern and Western European countries, whereas less than 5% of immigrants had arrived from Southern and Eastern Europe (Figure 3.2). By 1920, however, the situation had changed dramatically, and the share of immigrants born in new source countries was as high as 40%.

Europeans from new regions were culturally farther from natives and significantly less skilled than those from old sending regions (Hatton and Williamson, 1998, 2006). The shift in the composition of immigrants and concerns over their assimilation induced Congress to establish a commission that, between 1907 and 1911, studied the economic and social conditions of immigrants (Higham, 1955). In 1911, the Immigration Commission recommended the introduction of immigration restrictions, and in 1917, after decades of heated political debate, Congress passed a literacy test requiring that all immigrants entering the United States had to be able to read and write (Goldin, 1994).

Even before the adoption of the literacy test, in 1914, the Age of Mass Migration came to an abrupt end due to the onset of World War I, which drastically reduced European immigration between 1915 and 1919 (see Figure C.2). In 1920, despite the literacy test, migration flows increased again to their 1910 levels, fueling nativist movements and generating even stronger political pressure to adopt more effective measures to curb immigration. In response to the growing demand for immigration restrictions, in 1921 and 1924 Congress finally passed the Immigration Acts to limit the number of immigrants that could enter the United States in a given year by introducing country-specific quotas based on 1890 immigrants' population.[4]

Figure 3.2: Share of Foreign Born in the US



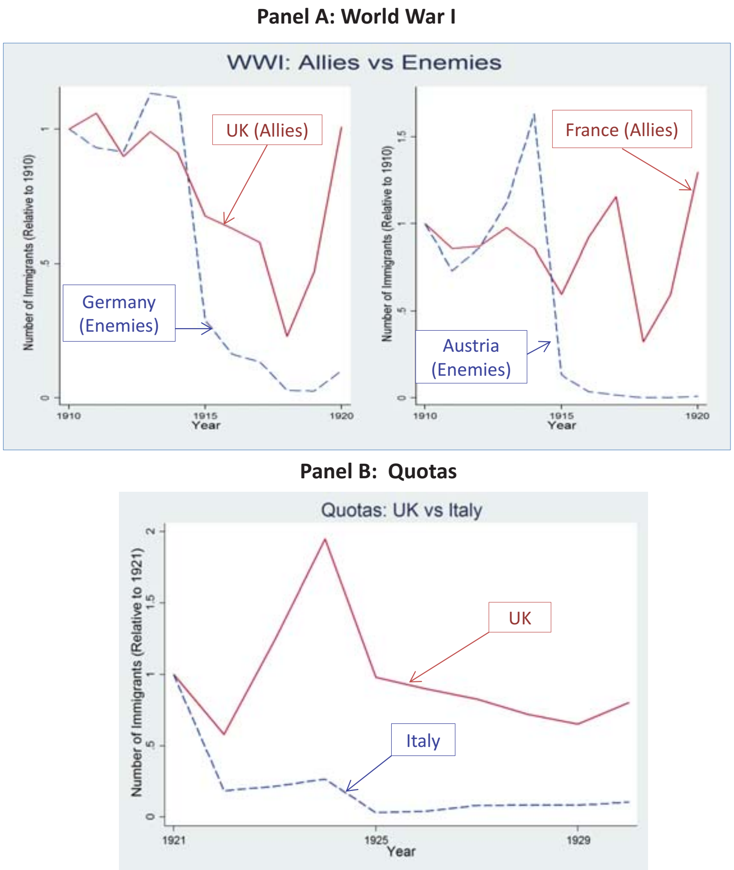
Notes: Share of immigrant stock living in the United States, by sending region and by decade. Source: Authors' calculations from IPUMS sample of US Census (Ruggles et al., 2015).

Both World War I and the Immigration Acts affected different sending countries in different ways. In particular, the reduction in immigration was more pronounced for European regions that were more directly involved in the War and which did not belong to the Allies (Figure 3.3, Panel A). Moreover, during the 1920s, quotas were set so as to limit the inflow of immigrants from new sending regions, while favoring that from old sources such as the UK, Germany, and Scandinavia (Figure 3.3, Panel B). Since immigrants tend

[4]With the 1924 National Origins Act, the total number of immigrants that could be admitted in a given year was capped at 150,000. In 1921, quotas were specified reflecting the 1910 composition of immigrants. However, they were rapidly changed to 1890 to limit immigration from new sending countries even further (Goldin, 1994).

to cluster along ethnic lines (Card, 2001), the post-1915 events generated substantial variation in the number as well as in the mix of immigrants received by US cities over time (Figure C.3): this is the variation that we exploit in our empirical analysis.

Figure 3.3: The impact of quotas and WWI on the share of immigrants in the USA.



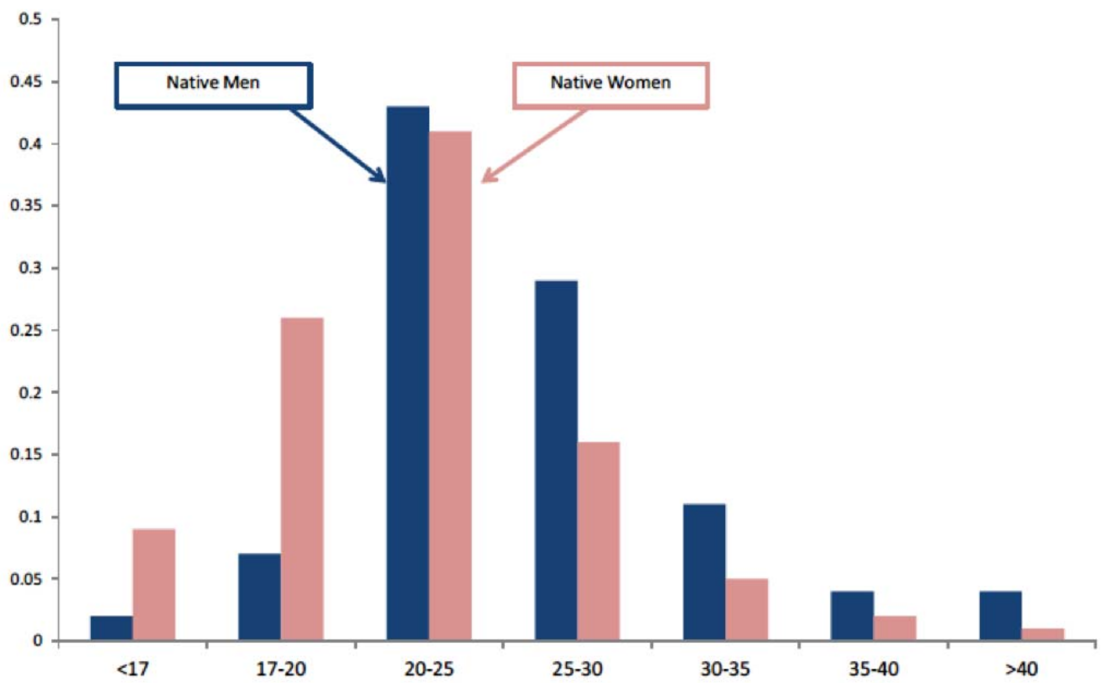
Notes: the figure plots the number immigrants from each European country that entered the United States in each year, scaled by the number of immigrants from that country in 1910 (Panel A) and 1921 (Panel B). Source: adapted from Tabellini (2017).

3.2.2 Immigration, Natives' Marriage, and Fertility

During the Age of Mass Migration, many prominent scholars expressed concerns over the effects of immigration on natives' fertility and marriage. As discussed in Leonard (2005), Edward Ross was among the first to propose the theory of “race suicide". According to this theory, not all immigrants were the same, and members of new, inferior races (i.e. immigrants from new sending regions) would eventually outbreed the “superior national stock" (i.e. natives and immigrants from old source countries) because industrial capitalism was conducive to the survival of the unfit (Leonard, 2005). More specifically, Francis A. Walker argued that “the native element failed to maintain its previous rate of increase because the foreigners came in such swarms", and natives were unwilling not only to engage in competition with “these new elements of the population", but also, they did not want “to brings sons and daughters into the world to enter that competition" (Walker, 1899, p. 424).

In contrast with these predictions, the inflow of immigrants might have increased marriage rates and fertility of native women by altering sex ratios (i.e. the relative number of men and women). At the time, more than 60% of immigrants entering the United States were young men between 20 and 35 (Figure C.1). Since in the early twentieth century the median age at first marriage was around 21 for women and 25 for men (Figure 3.4), even though marriage markets were highly segmented along ethnic lives, immigration might have made it easier for native women to find a mate and to have kids (Angrist, 2002).[5]

Figure 3.4: Marriage rates by age group and gender



Source: Authors’ calculations using IPUMS data.

Yet another possibility is that immigration affected natives’ marriage, fertility, and

[5]In particular, Figure 3.4 plots the distribution of the age at first marriage for native men and women in 1930 (the first year in which this question was asked in the US Census). See also estimates reported at https://www.thespruce.com/estimated-median-age-marriage-2303878.

trends in family formation by altering employment and occupational standing of native men. Historical accounts tend to view immigrants as one of the key determinants of American industrialization and economic development during the Age of Mass Migration. When describing the economic impact of European immigrants, historian Maldwyn Jones wrote that "The realization of America's vast economic potential has...been due in significant measure to the efforts of immigrants. They supplied much of the labor and technical skill needed to tap the underdeveloped resources of a virgin continent" (Jones, 1992, pp. 309-310). Similarly, John F. Kennedy argued that "every aspect of the American economy has profited from the contribution of immigrants" (Kennedy, 1964, p. 88).

During the Age of Mass Migration, the US economy had large potentials for growth. In this context, immigrants provided a cheap and unskilled supply of labor which could not only be absorbed, but that may have even allowed industries to expand (Foerster, 1924), in turn creating new job opportunities for native workers (Tabellini, 2017). It is thus possible that, by increasing the supply of "marriageable men", immigration raised fertility and marriage rates not only of native women, but also of native men. Moreover, if native men could find a stable job earlier in their working life, they might have been able to leave their parental house and set up their own household earlier. Somewhat ironically, then, immigration might have had exactly the opposite effect relative to what was argued by advocates of the theory of "race suicide".

3.3 Data

Our analysis is based on a balanced panel of the 180 US cities with at least 30,000 residents in each of the three census years from 1910 to 1930, and where at least some Europeans were living in 1900 (see Figure C.4 and Table C.2 for the complete list of cities). The dataset used in this paper was assembled using the decennial US Census of Population, made available by IPUMS (Ruggles et al., 2015).[6] From this source, we collected data on city population, on the number of immigrants by country of origin at the city and at the national level, and on most of the outcomes considered in our analysis, including marital status, relationship to the household head, and the number of children.[7] To investigate the mechanisms, we also collected data on employment, labor force participation, and occupation of native men of age 20 to 65 from the US Census, and on several measures of economic activity and industrialization from the 1904 to 1929 quinquennial Census of

[6]For 1900, we used the 5% sample, while for 1910, 1920, and 1930, we relied on the full count census datasets.

[7]See Table A1 for the list of European countries used in our work. To classify individuals based on their country of origin, we followed the classification made by IPUMS (Ruggles et al., 2015).

Manufactures (Tabellini, 2017).[8]

Table 3.1 reports the summary statistics for the main variables used in our analysis. City population ranges from more than 6.9 million (New York City in 1930) to as little as 30,200 (Pasadena in 1910). There is also wide variation in the fraction of immigrants across cities and over time, which was higher in the northeastern states of Connecticut, Massachusetts, New Jersey, and New York, and lower in the US South. As already discussed in Section 2, World War I and the Immigration Acts drastically reduced immigration: in 1910, the fraction of immigrants over city population was, on average, 0.18, but this number fell to 0.12 in 1930. The decline in the fraction of foreign born that entered the United States in the previous decade was even starker: for the average city, this number was 0.08 in 1910, but fell to 0.02 in 1930.

In Panel B of Table 3.1, we report the summary statistics of the main outcomes of this paper, i.e. marriage rates, fertility, and the propensity to leave the parental house for young native men and women. By the age of 33 for women and 35 for men, 65% of the native population was married. As shown in Table 3.2, among native women of native parentage, 74% were married to a native husband with both native parents, 20% to a husband with one or both foreign born parents, and only 6% to a foreign born husband. Instead, the probability of being married with a foreign born husband was as high as 25% for second generation women.[9]

Between 1910 and 1930, among women aged 18-33, the average children to women ratio was 0.65: 34% of native women had at least one child, while those who were mothers had on average almost 2 children each. Table 3.1 also suggests that the decision of leaving parents' home was strongly correlated with financial independence and with the choice of getting married: the proportion of men and women who were household head or spouse was close to marriage rates (45% and 43% for women and men respectively).

Finally, Panel C presents the summary statistics for the key labor market outcomes considered below. In 1910, the average employment to population ratio for native men aged 20-35 in our sample was 91%, and then fell to 84% in 1930, with the onset of the Great Depression. Average labor force participation for native women was 42%, with an increasing trend over time which was slowed down by the economic downturn in 1930. [10]

Immigration data are available for all the 540 city-year observations in our sample. However, for 1920, Sacramento (CA) and New Bedford (MA) had unreasonably low values for marriage, fertility, and the other demographic outcomes considered in our work, prob-

[8] In 1920, the US Census did not report employment status, but rather only an indicator for holding any gainful occupation. For this year, we imputed values from the latter to proxy for employment.

[9] In our sample, second generation women accounted for roughly one forth of all native women.

[10] Until 1930, the US Census classified individuals as participating in the labor force if they were holding any gainful occupation.

Table 3.1: Summary Statistics

	Count	Mean	SD	Min	Max
Panel A: City Demographics					
Fr. Immigrant	538	0.04	0.05	0.00	0.44
City Population (thousand)	538	190	511	30	6930
Panel B: Key Outcomes					
Marriage Rate of Women					
Aged 18-33	538	0.49	0.08	0.28	0.67
Aged 18-25	538	0.35	0.08	0.12	0.58
Aged 26-33	538	0.65	0.07	0.46	0.81
Marriage Rate of Men					
Aged 20-35	538	0.47	0.07	0.18	0.65
Aged 20-27	538	0.31	0.07	0.11	0.49
Aged 28-35	538	0.65	0.07	0.26	0.81
Fertility of Women 18-33					
Children to Women Ratio	538	0.65	0.12	0.40	1.00
Mothers to Women Ratio	538	0.34	0.05	0.21	0.49
Children to Mothers Ratio	538	1.90	0.11	1.59	2.27
Living with parents					
Women Aged 18-33	538	0.36	0.09	0.17	0.58
Men Aged 20-35	538	0.33	0.09	0.12	0.55
Living in own household					
Women Aged 18-33	538	0.45	0.08	0.25	0.67
Men Aged 20-35	538	0.43	0.06	0.18	0.60
Panel C: Labor Market					
Employment Men 20-35	538	0.90	0.05	0.71	0.98
Labor Force Participation Women 18-33	538	0.42	0.09	0.20	0.67

Notes: The Table shows the summary statistics of the main variables used in this paper for the 180 US cities with at least 30,000 residents in each Census year report. Source: Authors' calculations using IPUMS data.

Table 3.2: Characteristics of husbands of women aged 18-33

Husband	**Native**			**Immigrant**
	Native Parents	*Mix Parents*	*Foreign Parents*	
Native Wife				
Native Parents	0.74	0.08	0.12	0.06
Mix Parents	0.54	0.13	0.22	0.11
Foreign Parents	0.35	0.08	0.31	0.25
Immigrant Wife	0.12	0.03	0.09	0.76

Notes: The Table shows the probability of marriage with husband of different parentage for women aged 18-33 of the 180 US cities with at least 30,000 residents in each Census year report. Source: Authors' calculations using IPUMS data.

ably reflecting mis-reporting in the original documents. For this reason, in our baseline specification, we drop 1920 data for these two cities, but our results remain unchanged when all 540 city-year observations are included.[11]

3.4 Empirical Strategy

In this section, we present the baseline estimating equation (Section 3.4.1), construct the instrument for immigration (Section 3.4.2), and report first stage results (Section 3.4.3). To deal with the potential endogeneity of immigrants' location decision, we instrument the actual number of immigrants by interacting 1900 settlements of different ethnic groups with subsequent migration flows from each sending region, leaving out immigrants that eventually settled in the city's MSA.

3.4.1 Baseline Estimating Equation

To investigate the effects of immigration on natives' marriage, fertility, and family structure across US cities, we stack the data for the three Census years between 1910 and 1930, and estimate

$$y_{cst} = \gamma_c + \delta_{st} + \beta Imm_{cst} + u_{cst} \tag{3.1}$$

where y_{cst} is the outcome for city c in state s in Census year t, and Imm_{cst} refers to the fraction of immigrants received by city c in the previous decade, over city population. γ_c and δ_{st} are city and state by year fixed effects, implying that β is estimated from changes in the fraction of immigrants within the same city over time, compared to other

[11] Also, data from the Census of Manufactures were not available for Superior (WI), Washington DC in 1909 and 1919, and for Flint (MI), Galveston (TX), Huntington (WV), Lexington (KY), McKeesport (PA), Pueblo (CO), Quincy (IL), and Roanoke (VA) in 1929.

cities in the same state in a given year. Since city population could itself be an outcome of immigration, the number of immigrants is scaled by predicted (rather than actual) city population, constructed by multiplying 1900 population by average urban growth in the US, excluding that of the Census division where the city is located. Below, we also report results obtained when scaling immigration by 1900 population. Standard errors are clustered at the MSA level, and MSA boundaries are fixed to 1940 in order to keep geography constant.[12]

3.4.2 Instrument for Immigration

A priori, we may expect immigrants to be attracted to cities with better employment opportunities. Alternatively, immigrants might settle in otherwise declining cities, where house prices are lower. In either case, OLS estimates of equation (3.1) will likely be biased. To deal with this endogeneity problem, we construct a "leave-out" version of the shift-share instrument (Card, 2001). The instrument predicts the number of immigrants received by US cities over time by interacting 1900 settlements of different ethnic groups with subsequent migration flows from each sending region, excluding individuals that eventually settled in a given city's MSA. Formally, Imm_{cst} in (3.1) is instrumented with

$$Z_{cst} = \frac{1}{\hat{P}_{cst}} \sum_j \alpha_{jc} O_{jt}^{-M} \tag{3.2}$$

where $\hat{P}_{cst}$ is predicted city population; α_{jc} is the share of individuals of ethnic group j living in city c in 1900; and O_{jt}^{-M} is the number of immigrants from country j that entered the US between t and $t-1$, net of those that eventually settled in city c's MSA.[13]

The instrument constructed in equation (3.2) exploits two sources of variation: first, cross-sectional variation in the share of individuals from each ethnic group living in different US cities in 1900 (α_{jc}); second, time-series variation induced by changes in the total number of immigrants from any sending region entering the United States in a given decade (O_{jt}^{-M}). Appendix C.2 presents a simple example to illustrate graphically how the instrument combines them.

[12] In our baseline specification, we restrict attention to European immigrants that entered the United States during the previous decade, but results are robust to using immigrants' stock or considering immigrants from all sources.

[13] A similar "leave-out" strategy is used in Burchardi et al. (2016). Results are also robust to using a specification where the endogenous regressor, Imm_{cst}, is constructed by scaling the number of immigrants by actual (rather than predicted) city population, and is instrumented with Z_{cst} in (3.2), i.e. the predicted number of immigrants over predicted city population.

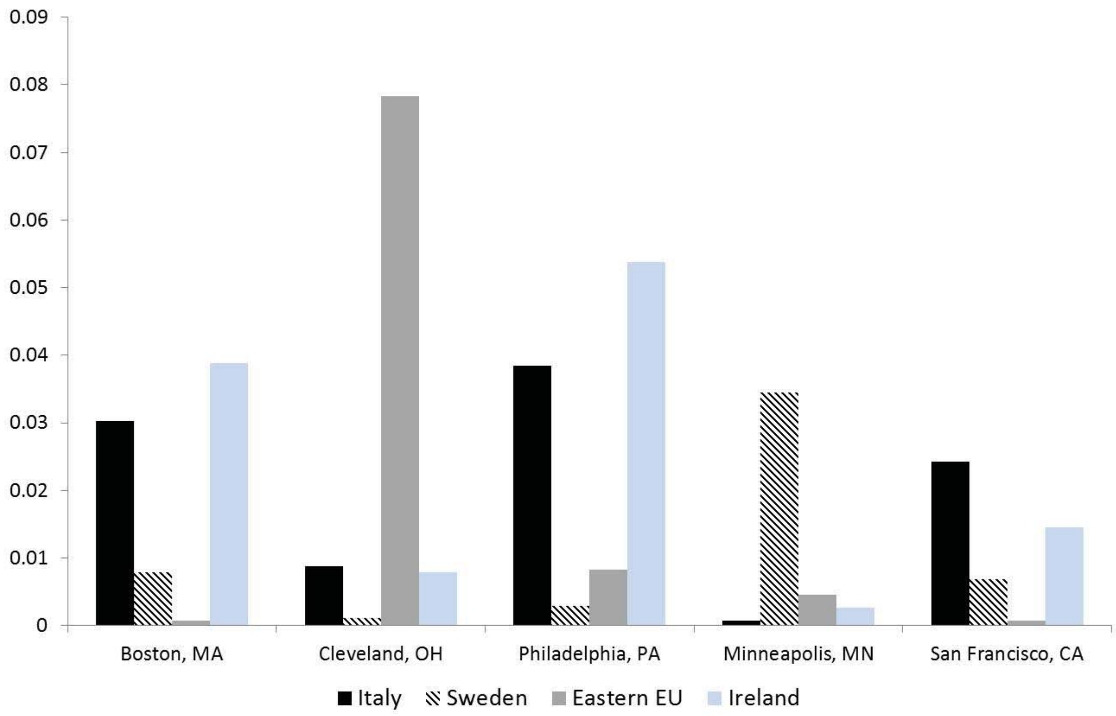
Geographic Variation in Immigrants' Settlements

The cross-sectional variation underlying the instrument in equation (3.2) is based on the idea that immigrants cluster geographically and when newcomers arrive, they tend to move where their ethnic community is larger because of social networks and family ties, and not because of local pull factors (Card, 2001; Stuart and Taylor, 2016). As documented in Sequeira et al. (2017), the gradual expansion of railroads during the nineteenth century is a strong predictor of the geographic distribution of immigrants in the US: places that gained access to the railroad just before an immigration boom received more immigrants in the following decade. Moreover, upon arrival, early settlers tended to locate in places that were relatively more attractive at that time. Since the timing of outmigration varied widely across European countries, depending on local political and economic conditions (Hatton and Williamson, 1998), different US regions were populated by different ethnic groups before 1900. Early settlers then acted as a catalyst for subsequent migrants from the same ethnic group (Lafortune and Tessada, 2014).

To visually display the degree of geographic concentration of different ethnic groups, Figure 3.5 plots the share of individuals from different European regions living in selected US cities in 1900.[14] For example, while Italian communities were present in Boston, Philadelphia, and San Francisco, they were practically non-existent in Minneapolis. On the other hand, while almost 4% of Swedes living in the US in 1900 were settled in Minneapolis, less than 1% of them were located in north-eastern cities like Philadelphia or Boston. Finally, in 1900, more than 8% of Eastern Europeans were living in Cleveland, but their share in the other cities displayed in Figure 3.5 was well below 1%. Presenting a similar example for Ohio, Figure C.5 shows that differences in immigrants' settlements existed not only across, but also within states. This is important since our empirical strategy exploits only within state variation in immigration.

[14] See also Abramitzky and Boustan (2017) for a discussion of the geographic concentration of Europeans in the United States during the Age of Mass Migration.

Figure 3.5: Share of Immigrants from Selected Regions in US Cities, 1900 (Alternative)



Notes: This graph shows the share of individuals of European ancestry living in US cities in 1900, for selected ethnic groups. Source: Authors' calculations using IPUMS data.

Identifying Assumptions and Instrument Validity

The key identifying assumption behind the instrument is that cities receiving more immigrants (from each sending area) before 1900 must not be on different trajectories for the evolution of economic and social conditions in subsequent decades. Said differently, outmigration from European regions must be independent of cross-city pull factors systematically related to 1900 settlers' country of origin. For example, between 1910 and 1920, immigration to the US was higher from Poland than from Norway. The exclusion restriction would be violated if this happened because cities that in 1900 had attracted more Poles were growing more than cities where more Norwegians had moved to in 1900.

Another threat to the validity of the identifying assumption is that the characteristics of cities that attracted early immigrants might have time-varying, confounding effects on migration patterns as well as on changes in the outcomes of interest. It is possible, for instance, that larger urban centers attracted more immigrants in the nineteenth century, and that these cities kept growing more also in subsequent decades. In turn, more sustained economic growth may have increased marriage and fertility of natives, invalidating the instrument constructed in equation (3.2). To deal with these and similar issues, we perform several robustness checks, which we describe below when presenting our main results.

3.4.3 First Stage Results

First stage results for the relationship between actual and predicted immigration are reported in Table 3.3, after partialling out city and state by year fixed effects. In column 1, the dependent variable is the fraction of immigrants over actual city population, and the regressor of interest is the baseline instrument constructed in equation (3.2). Columns 2 and 3 replicate column 1 by scaling both the actual and the predicted number of immigrants by, respectively, 1900 and predicted population. In all cases, the F-stat is very high, and there is a strong and significant relationship between the endogenous regressor and the instrument.

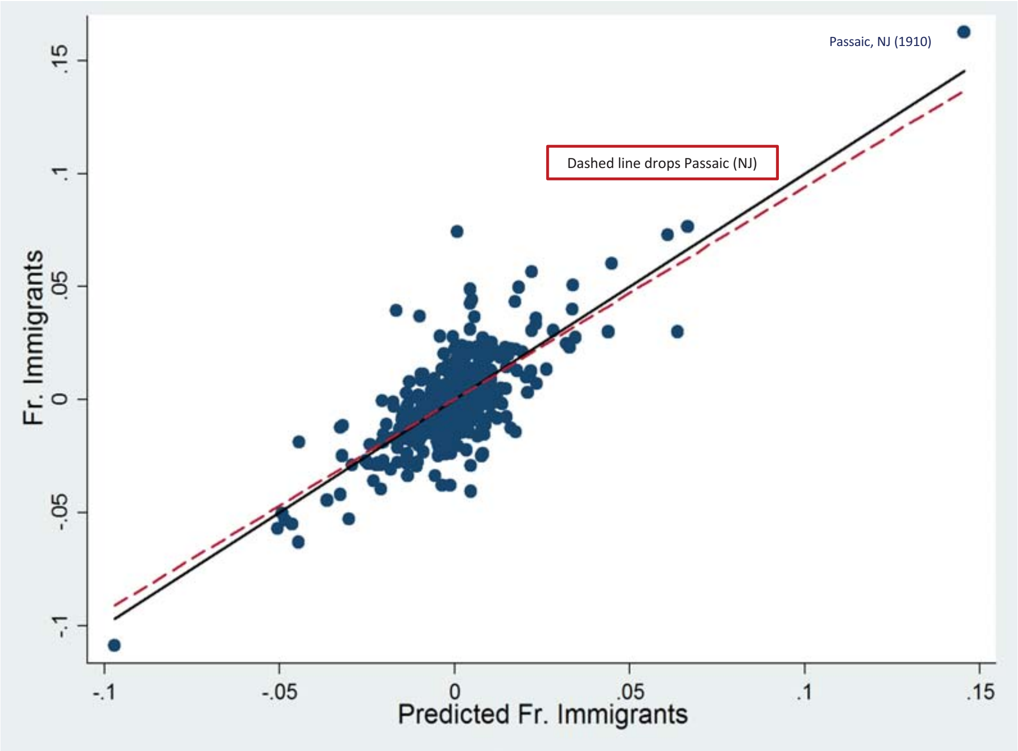
Table 3.3: First Stage

	Dep. Variable: Fraction of Immigrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Z	0.830***	0.944***	0.990***	0.905***	0.889***	0.986***
	(0.053)	(0.071)	(0.063)	(0.090)	(0.086)	(0.066)
Immigrants over	*Actual pop.*	*1900 pop.*	*Predicted pop.*	*Predicted pop.*	*Predicted pop.*	*Predicted pop.*
Year interacted with 1900				*Immigrants and city population*	*Value added by manufacture*	*Fr. native women married*
F-stat	249.3	175.3	251.3	100.2	107.5	224.5
Obs.	538	538	538	538	526	538

Notes: the sample includes a balanced panel of the 180 US cities with at least 30,000 residents in each Census year 1910, 1920, and 1930. In Col 1 the actual number of immigrants is scaled by actual population, and the instrument is the leave-out version of the shift-share IV in equation (2) (Section 3.4.2). Cols 2 and 3 replicate Col 1 by scaling the actual and predicted number of immigrants by, respectively, 1900 and predicted population. From Col 3 onwards, Table 3.3 presents results from specifications where both the predicted and the actual number of immigrants are scaled by predicted population. Cols 4 to 6 include the interaction between year dummies and, respectively: the (log of) 1900 city and immigrants population; the (log of) 1904 value added by manufacture per establishment; and the marriage rate of native women in 1900. F-stat refers to the K-P F-stat for weak instrument. All regressions partial out city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.6 plots the residual scatterplot of the regression reported in column 3. As it appears, the city of Passaic (NJ) experienced a large drop in immigration between 1910 and 1930, and one may be concerned that, for this reason, it influences the strength of the first stage. However, omitting this city barely affects the slope of the regression line (see red dashed line in Figure 3.6 and additional results in Tabellini, 2017). From column 3 onwards, Table 3.3 presents estimates for specifications where both the actual and the predicted number of immigrants are scaled by predicted city population, and explores the stability of the baseline specification to the inclusion of interactions between year dummies and 1900 city characteristics.

Figure 3.6: First Stage



Notes: The y-axis (resp. x-axis) reports the actual (resp. predicted) number of immigrants over predicted city population in each of the three Census years, 1910, 1920, and 1930. Each point in the scatter diagram represents the residual change in a city's actual and predicted fraction of immigrants after partialling out city and year by state fixed effects. The predicted number of immigrants is constructed as discussed in Section 3.4.2 in the text. Predicted city population is obtained by multiplying 1900 city population with average urban growth, excluding that of the Census division where a city is located. The solid line shows the regression coefficient for the full sample (coefficient=0.990, standard error=0.063). The dotted (red) line shows the regression coefficient obtained when dropping the city of Passaic, NJ (coefficient=0.940, standard error=0.068).

First, we augment the specification reported in column 3 by interacting the 1900 (log of) city and immigrants' population (column 4).[15] Next, in columns 5 and 6, we include interactions between year dummies and, respectively, the 1904 (log of) value added by manufacture and men in 1900. Even though the F-stat falls relative to column 1, it remains well above conventional levels. Also, and importantly, neither the economic nor the statistical significance of coefficients is affected.

Overall, Table 3.3 suggests that there is a strong relationship between actual and predicted immigration, which is robust to the use of different specifications and alternative ways of constructing the instrument.

3.5 Main Results

In this Section, we present three sets of results. First, immigration had a positive and large effect on marriage rates of both native women and native men (Section 3.5.1). Second, the inflow of immigrants raised natives' fertility by increasing the share of young women with at least one child (Section 3.5.2). Third, immigration induced native young men and women to anticipate the age at which they chose to leave their parental house (Section 3.5.3).

3.5.1 Immigration and Marriage Rates of Natives

In Table 3.4, we study the impact of immigration on natives' marriage focusing on the age groups with the highest marriage rates, i.e. women aged 18-33 and men aged 20-35.[16] In Panel A (resp. Panel B), the dependent variable is the fraction of native women (resp. men) who were married. OLS results of equation (3.1) are presented in column 1, while column 2 reports 2SLS estimates for the baseline specification, where we instrument the fraction of immigrants over predicted population using the "leave-out" shift-share instrument described in Section 3.4.2. Throughout the paper, we always report the mean of the dependent variable at baseline and the F-stat associated with first stage results shown in Table 3.3.

Starting from Panel A, both OLS and 2SLS estimates suggest that immigration increased marriage rates for native women aged 18-33.[17] These effects are not only sta-

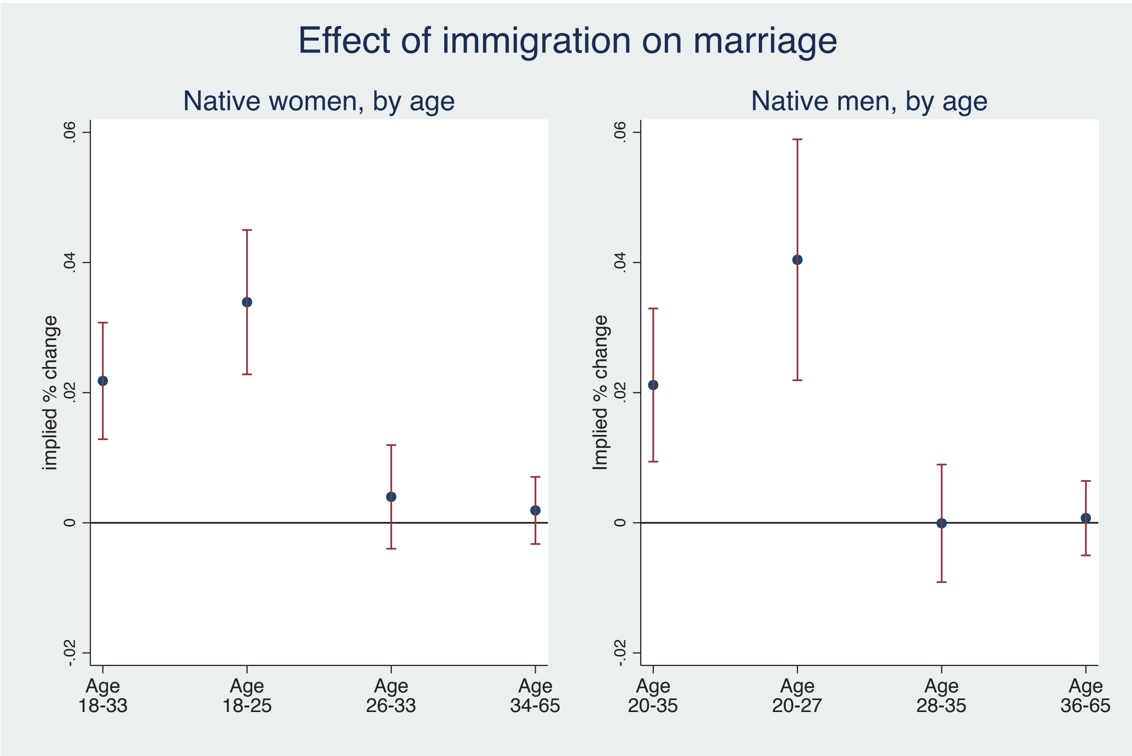
[15]This check is important since the instrument mechanically predicts higher immigration to cities that had a larger 1900 fraction of immigrants, and, at the same time, larger ethnic enclaves might have direct and time-varying effects on economic and social conditions.

[16]As discussed in Section 3.2.2, the median age at first marriage was around 21 for women and 25 for men (Figure 3.4).

[17]Both the OLS and 2SLS coefficients reported in columns 1 and 2 respectively are positive and significant, with the latter being only slightly smaller than the former.

tistically significant but also economically relevant: the coefficient in column 2 implies that one standard deviation increase in the fraction of immigrants raised marriage rates of native women aged 18-33 by 2.2% relative to the 1910 mean (see Figure 3.7). Panel B documents a similar pattern (both qualitative and quantitative) for native men aged 20-35: a five percentage point increase in immigration (equivalent to a one standard deviation) raised men's marriage rates by 2.1% relative to their baseline mean.[18] How do these estimates compare to the existing literature? Our findings are quantitatively close to those obtained in Autor et al. (2017), who document that, over the last thirty years, a one percentage point increase in import competition from China lowered female marriage rates by 1.8%

Figure 3.7: The impact of immigration on marriage rates by gender and age groups



Notes: This graph shows the impact of one standard deviation increase of the fraction of immigrants on the increase in marriage rates with respect to the mean value in 1910. We report the standardized coefficients by age group and for men and women separately.

[18]OLS estimates are sensitive to the inclusion of three cities (Duluth, Superior, and Tacoma) for which in 1910 marriage rates were very low. The mean value of marriage rates of men aged 20-35 in 1910 is 24 percentage points lower compared to the mean value of the same cities in 1920 and 23 percentage points lower compared to other US cities in our sample in 1910. The latter effect corresponds to 4.6 lower standard deviations in the marriage rates of men in these cities compared to the rest of our sample. In the Appendix Table C.3, we present estimates of OLS and 2SLS results with and without these three cities. Once we restrict the sample, OLS and 2SLS are closer in magnitude.

Subsequent columns of Table 3.4 explore the robustness of our baseline results. First, in column 3, we test for pre-trends by regressing the 1900 to 1910 change in marriage rates against the 1910 to 1920 instrumented change in immigration. Reassuringly, in both Panel A and Panel B, the coefficient on immigration is statistically indistinguishable from zero and different from that reported in column 2. Next, in columns 4 and 5, we augment our baseline specification by interacting year dummies with the (log of) 1900 city and immigrants' population and the 1900 marriage rate, respectively. This exercise is performed to check that results in column 2 are not due to city-specific characteristics that may have simultaneously attracted more immigrants before 1900 and affected the evolution of natives' marriage rates in subsequent decades. In all cases, the point estimate remains statistically significant and quantitatively close to that estimated in the baseline specification. Finally, in column 6 we provide evidence that results are robust to scaling both the actual and the predicted number of immigrants by 1900, rather than predicted, population.

Up to now, we reported results for the "marriage-relevant" age groups by gender. In Figure 3.7, we separately document the effect of immigration on marriage rates of native men and women for different age groups. All of the effect estimated in Table 3.4 comes from the youngest cohorts: one standard deviation increase in the fraction of immigrants raised marriage rates of native women aged 18-25 and men aged 20-27 by 3.4% and 4.0%, respectively, relative to their baseline means. Instead, the effect of immigration is not statistically significant for older cohorts. The point estimates and standard errors related to this figure are reported in Appendix Table C.4, where we also show the probability of being never married for the oldest cohorts. While immigration had no effect on the probability of being never married for men, it lowered the likelihood that women aged 34-65 remained unmarried.

3.5.2 Immigration and Natives' Fertility

In Table 3.5, we study how exposure to immigration affected fertility of native women in our sample of 180 cities. The first two columns focus on the overall fertility rate, defined as the children to women ratio, while in subsequent columns we separately analyze the effect of immigration on the extensive and the intensive margin. We define the former as the share of women with at least one child, and the latter as the children to mothers ratio. In odd (resp. even) columns, the dependent variable is the total number of children in the household (resp. children below the age of 5). Since full-count data allow to match mothers with children only if they are living in the same household, we restrict the sample to women aged 18-33, whose children are likely to live with their parents.

Both OLS and 2SLS results, reported in Panels A and B respectively, document a

Table 3.4: Immigration and Marriage of Natives

	(1) OLS	(2) 2SLS	(3) Pre-trends	(4) 2SLS	(5) 2SLS	(6) 2SLS
Panel A: Dep. Var. Marriage Rate of Women (Age 18-33)						
Fr. Immigrant	0.238***	0.209***	0.128	0.329***	0.197***	0.154***
	(0.057)	(0.044)	(0.204)	(0.058)	(0.053)	(0.027)
F-stat		251.3	318.4	100.2	107.5	175.3
Mean dep. var. in 1910	0.45	0.45	0.45	0.45	0.45	0.45
Obs.	538	538	178	538	538	538
Panel B: Dep. Var. Marriage Rate of Men (Age 20-35)						
Fr. Immigrant	-0.006	0.190***	0.078	0.181***	0.217***	0.121***
	(0.135)	(0.054)	(0.092)	(0.059)	(0.061)	(0.038)
F-stat		251.3	318.4	100.2	107.5	175.3
Mean dep. var. in 1910	0.42	0.42	0.42	0.42	0.42	0.42
Obs.	538	538	178	538	538	538
Pre-period			*Yes*			
Year by 1900 city and imm. pop				*Yes*		
Year by 1900 fr married					*Yes*	
Imm over 1900 pop						*Yes*

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is the fraction of women married in the age range 18-33 in Panel A and the fraction of men married in the age range 20-35 in Panel B. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

positive and significant relationship between immigration and fertility of native women. The point estimate in column 1 of Panel B implies that a one standard deviation increase in immigration raised the children to women ratio by 3.3% relative to its 1910 mean. When decomposing this effect along the extensive and the intensive margin, we note that immigration increased the number of women with children by 2.4%, and raised the average number of children per woman by 1%. Said differently, for every ten new babies born from native women, seven were due to the extensive margin, while three to the intensive margin. The magnitude of the effect is similar when we restrict our attention to children below the age of 5.[19]

Between the late nineteenth century and the 1930s, the US went through a demographic transition, with a reversal of the positive relationship between income and economic growth (Galor and Weil, 2000). The fertility rate of the total white population declined substantially, with the birth rate moving from almost 50 per thousand population in 1850 to 20 per thousand in 1930 (Zelnik, 1959). However, in our analysis, the inclusion of state by year fixed effects takes care of these national trends, since the effect of immigration is estimated from changes in the fraction of immigrants within the same city over time, as compared to other cities in the same state in a given year. Moreover, as noted by Easterlin (1961), the decline in fertility was driven by rural areas; instead, fertility of the urban native (white) population remained stable in this time period.[20]

In Appendix Table C.5, we separately report the effect of immigration on fertility of native women by age groups. As for marriage, the effect is driven mainly by native women aged 18-25, especially on the extensive margin: one standard deviation increase in immigration raised the number of women in the younger age cohort with at least one child by 3.1%.

[19]As before, our estimates are quantitatively in line with those from Autor et al. (2017), who, for the more recent period, find that a 1 percentage point increase in import competition from China reduced fertility by 2.8%.

[20]Guinnane et al. (2006) find that fertility of immigrants in the late nineteenth and early twentieth century was higher than that of natives, but converged to US standards for second generation immigrants.

Table 3.5: Immigration and Fertility of Native Women

	Dep. Variable: Fertility of Native Women (aged 18-33)					
	Children to Women Ratio		Mothers to Women Ratio		Children to Mothers Ratio	
	All Children	Children<5	All Children	Children<5	All Children	Children<5
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Fr. Immigrant	0.291**	0.200***	0.132**	0.123***	0.149	0.209**
	(0.131)	(0.056)	(0.052)	(0.036)	(0.159)	(0.085)
Panel B: 2SLS						
Fr. Immigrant	0.431***	0.194***	0.165***	0.105***	0.342***	0.111**
	(0.072)	(0.037)	(0.027)	(0.026)	(0.103)	(0.052)
F-stat	251.3	251.3	251.3	251.3	251.3	251.3
Mean Dep.Var.	0.650	0.340	0.340	0.250	1.900	1.010
Obs.	538	538	538	538	538	538

Notes: this Table presents results of 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is: in column 1 (column 2), the total number of children (toddlers) with native mother in the age range 18-33 over the total number of women in the age range 18-33, in column 3 (column 4) the fraction of women in the age range 18-33 who have children (toddlers) and in column 5 (column 6) the average number of children (toddlers) per mother in the age range 18-33. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.5.3 Household Formation

In Table 3.6, we provide evidence that immigration anticipated the choice of natives to leave their parental house, and set up their own independent family unit. In the first two columns, we focus on women aged 18-33, while in subsequent columns we report the effects of immigration on men aged 20-35.[21]

Specifically, the coefficients in Table 3.6 imply that one standard deviation increase in immigration raised the probability of living in an independent family unit by 2.4% for women and 2.2% for men, relative to the mean in 1910. This effect is quantitatively close to that estimate for marriage rates, suggesting that the decisions of getting married and of leaving the parental house were both part of a unique lifetime plan. Interestingly, focusing on the contemporaneous period, Autor et al. (2017) find that one percentage point increase in import competition from China, not only decreased marriage rates and fertility, but also lowered the probability of living with the spouse by 1.6%.

In Figure C.6, we provide evidence that the effect of immigration on the probability of leaving the parental house was driven by women aged 18-25 and men aged 20-27: for these age groups, a five percentage point increase in immigration raised the probability of setting up their own household by more than 3%. Incidentally, these cohorts also experienced the largest increase in marriage and fertility because of immigration.

Stitching together the three sets of results presented in this section, our estimates paint a coherent picture of how immigration affected family formation, marriage rates, and fertility of native men and women in the urban early twentieth century US. The inflow of immigrants induced natives to get married more (and, possibly, earlier); this decision was accompanied by the choice of leaving the parental house and set up an independent family unit. In a period in which oral contraception was not yet available (Bailey, 2006), higher fertility was probably mechanically related to marriage and family formation decisions.

3.6 Mechanisms

In this section, we explore the mechanisms behind the results presented above. In Section 6.1, we start by documenting that immigration raised employment and occupational standing of native men, and then argue that such higher supply of "marriageable men" was the key driver of the positive effect of immigration on natives' marriage and fertility

[21] Both OLS and 2SLS results, reported in Panels A and B respectively, are statistically significant and close in magnitude for women. As for marriage rates, for men, OLS estimates are instead sensible to the inclusion of three cities (Duluth, Superior, and Tacoma). In these cities, only 20% of men were household head in 1910, as compared to 42% in 1920 or 39% in the other cities of our sample in 1910. 2SLS estimates are instead unaffected by the inclusion of these three cities.

Table 3.6: Immigration and Living Choices of Natives

	(1)	(2)	(3)	(4)
Dep. Var.	**Living with Parents**	**Living in Own House**	**Living with Parents**	**Living in Own House**
	Women 18-33		**Men 20-35**	
Panel A: OLS				
Fr. Immigrant	-0.383***	0.231***	-0.493***	-0.004
	(0.086)	(0.057)	(0.131)	(0.137)
Panel B: 2SLS				
Fr. Immigrant	-0.285***	0.204***	-0.316***	0.171***
	(0.043)	(0.040)	(0.045)	(0.056)
F-stat	251.3	251.3	251.3	251.3
Mean dep. var.	0.370	0.418	0.317	0.387
Obs.	538	538	538	538

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable are described on the top part of the Table. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

shown in Section 5. Next, we provide evidence that changes in sex ratios (Section 6.2), natives' cultural responses (Section 6.3), and direct effects of immigration on native female labor force participation (Section 6.4) cannot account for a quantitatively relevant fraction of our main findings.

3.6.1 Natives' Employment and the Supply of Marriageable Men

In two important contributions, Wilson (1987) and Wilson (1996) argues that the decline in marriage and the rise in the share of single-mother households in the US during the last forty years have been, at least in part, due to deteriorating employment opportunities in manufacturing. Along these lines, exploiting exogenous variation in exposure to import competition from China across US local labor markets, Autor et al. (2017) find that job losses in manufacturing caused a steep decline in marriage rates and a significant increase in the proportion of single-mother households. In this section, we investigate the possibility that a similar mechanism, with the opposite sign, was at play in our context. Specifically, we advance and empirically test the hypothesis that immigration had a positive effect on natives' marriage, fertility, and patterns of family formation by increasing employment and occupational standing of native men, in turn raising the supply of "marriageable men".

In Table 3.7, we study the effects of immigration on natives' employment to population ratio, focusing on men in the "marriageable relevant" age range, i.e. 20-35 (see Section 5.1). As for Table 3.4, columns 1 and 2 estimate the baseline specification (see equation (3.1)) with OLS and 2SLS respectively. In both cases, there is a strong and positive relationship between immigration and natives' employment. The coefficient in column 2, which is quantitatively very close to OLS results reported in column 1, implies that a five percentage points increase in immigration (equivalent to one standard deviation) raised natives' employment to population ratio by 0.9% relative to its 1910 mean. Said differently, for every ten new immigrants, one and a half more jobs were created for native men aged 20 to 35.

Table 3.7: Immigration and Employment of Native Men

Dep. Var. : Natives' Employment to Population Ratio (Men, Age 20-35)						
	(1) OLS	(2) 2SLS	(3) Pre-trends	(4) 2SLS	(5) 2SLS	(6) 2SLS
Fr. Immigrant	0.151***	0.152***	-0.071	0.094**	0.130**	0.113***
	(0.043)	(0.044)	(0.124)	(0.041)	(0.053)	(0.033)
F-stat		251.3	318.4	100.2	107.5	175.3
Mean dep. var. in 1910	0.911	0.911	0.911	0.911	0.911	0.911
Obs.	538	538	180	538	538	538
Pre-period			*Yes*			
Year by 1900 city and imm. pop				*Yes*		
Year by 1900 value added manuf.					*Yes*	
Imm over 1900 pop						*Yes*

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is the natives' employment to population ratio in the age range 20-35 for men. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As documented in Figure C.7, the effect of immigration is slightly larger for men in the age range 20-27, but remains positive and statistically significant also for those aged 28-35. The point estimate is positive and quantitatively very similar, albeit not statistically significant, also for older natives, i.e. those in the age range 36-65.[22] As we did in Table 3.4, we next test the robustness of our baseline specification in subsequent columns of Table 3.7. First, as for marriage rates, there is no evidence of pre-trends (column 3). Second, results are robust to interacting year dummies with the 1900 log of city and immigrants population (column 4) and log of value added by manufacture (column 5). Third, our estimates are unchanged when scaling both the actual and the predicted number of immigrants by 1900, rather than predicted, city population (column 6).[23]

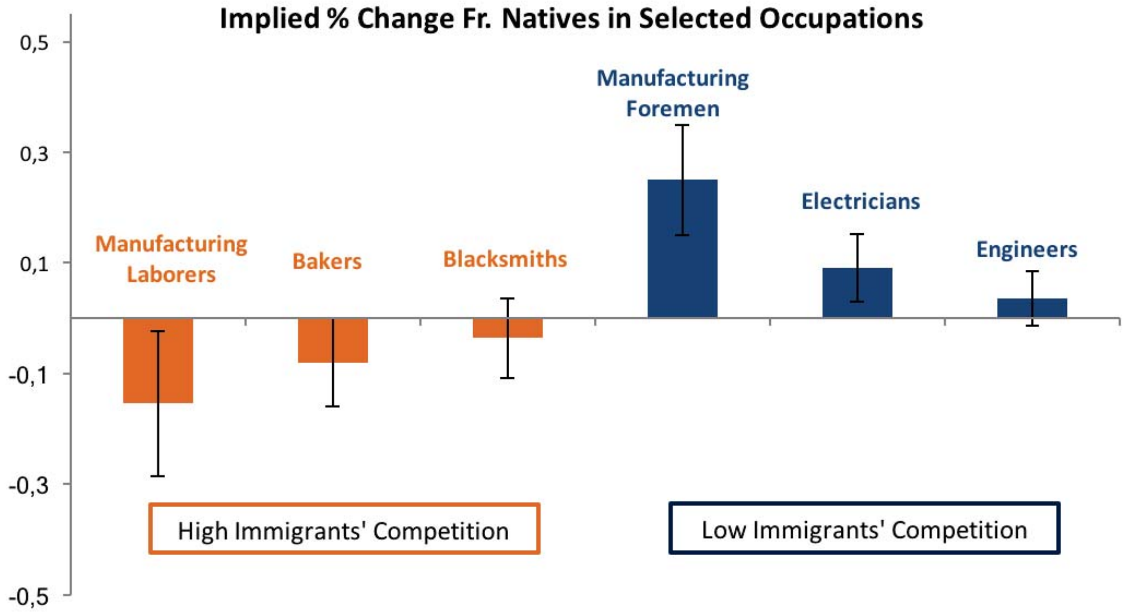
The positive effects of immigration on natives' employment estimated in Table 3.7 are in contrast with some of the results from the contemporary immigration literature such as Borjas (2003), Borjas and Katz (2007), and Dustmann et al. (2017) who find a negative effect of immigration on natives' labor market outcomes. Our findings are also somewhat different from those of some contemporaneous cross-city design studies that estimate a zero effect of immigration on natives' wages (Card, 2001,0). However, they are consistent with a recent body of the literature which documents a positive impact of immigrants on natives' occupational mobility (Foged and Peri, 2016), and more specifically for this historical period with Tabellini (2017).

In line with the latter works, in Figure 3.8, we show that immigration induced natives to leave occupations that were more exposed to immigrants' competition and to take up jobs where immigrants were prevented from entering, because of skill and language mismatch or because of discrimination. Specifically, Figure 3.8 plots the implied effect (expressed in percent change) of a one standard deviation increase in immigration on the fraction of native men aged 20-35 employed in specific occupations (see also Table C.6 in the appendix).

[22]Very similar results are obtained in Tabellini (2017), who studies the effects of immigration on natives' employment for natives in the age range 15-65 in the same sample of cities. Tabellini (2017) also shows that immigration had a positive and large effect on natives' occupational standing, measured as the log of occupational scores.

[23]For many other robustness checks, see Tabellini (2017).

Figure 3.8: Natives' Occupation Mobility (20-35)



Notes: the figure plots the percent change in the fraction of natives aged 20-35 in each occupation (relative to the 1910 mean) implied by a one standard deviation increase in immigration, according to 2SLS estimates (with corresponding 95% confidence intervals).

The first three (orange) bars starting from the left refer to occupations that were highly exposed to immigrants' competition:[24] as it appears, immigration had a negative effect on the share of young natives working in these occupations. This effect is statistically significant and economically large especially for manufacturing laborers - one of the jobs with the highest exposure to immigrants' competition and with the lowest skill requirements. Moving rightward along the graph, the three (blue) bars on the right show that immigration increased the fraction of natives employed in more skilled and less exposed occupations such as manufacturing foremen, electricians, and engineers. The pattern displayed in Figure 3.8 can be effectively summarized using the words of the economist and statistician Isaac Hourwich who, in 1912, noted that "the effect of immigration upon the occupational distribution of industrial wage earners has been the elevation of the English-speaking workmen to the status of an aristocracy of labor, while the immigrants have been employed to perform the rough work of all industries" (Meyer, 1981).

For natives' employment to increase, immigration must have also stimulated economic activity, inducing firms to create new jobs. Otherwise, absent changes in labor demand, it would be hard to reconcile the labor supply shock induced by immigration with the

[24]We proxy for the degree of immigrants' competition using the ratio of the probability that natives and immigrants held a given occupation in 1910 (see also Table C.6 and Tabellini (2017)).

positive employment effects estimated above. Consistent with this idea, in Table C.7, we show that immigration had a positive and large effect on value added and the value of products per establishment (columns 1 and 2), establishment size (column 3), capital accumulation (column 4), and total factor productivity (column 5).[25] Specifically, the coefficients in Panel B of Table C.7 imply that a one standard deviation increase in immigration increased industrial production by roughly 10% relative to its 1910 level. Such sizeable effects are not only consistent with the historical literature reviewed in Section 2, but can also explain the positive employment effects estimated in Table 3.7. On the one hand, industrial expansion allowed the economy to absorb the large supply shock by creating new jobs for both high and low skilled workers. On the other, it provided natives with opportunities for skill upgrading (see also Tabellini, 2017 for a more extensive discussion).

Overall, this section documents that immigration boosted natives' employment and induced men to take up better jobs. We argue that, in turn, the larger pool of "marriageable men" was responsible for the positive effects of immigration on natives' marriage rates, fertility, and propensity to leave the parental household earlier. In Table C.8, we provide an additional piece of evidence consistent with this interpretation by showing that immigration lowered the share of children below the age of 10 born from native parents living in a household where the father was unskilled (column 1).[26] Similarly, even if the coefficient is not statistically significant at conventional levels, there is a positive relationship between immigration and the share of children of native parentage whose father was employed. These results suggest that, because of immigration, children of native parentage were likely to grow up in a better environment at home.

Consistent with the latter observation, as it appears from Table C.9, immigration increased the fraction of sons of native parentage aged 6-14 who were enrolled in school (column 1).[27] Somewhat interestingly, though, we do not find a similar effect for daughters (column 4), even if the 1910 average enrollment was very similar for boys and girls. One possible explanation for this pattern is that families were credit constrained and, as more resources became available, parents chose to first invest them in sons rather than in daughters. Especially in an urban context, higher employment opportunities brought about by immigration might have increased the opportunity cost of schooling, in turn

[25] Panel A and Panel B report, respectively, OLS and 2SLS results. We proxy for capital utilitazion using the log of horsepower (column 4), and estimate the effects of immigration on productivity (column 5), assuming a Cobb-Douglas production function with two factors of production, capital and (homogeneous) labor.

[26] OLS and 2SLS results are reported respectively in Panel A and Panel B. Very similar results are obtained when focusing on the share of families rather than on the share of children (see columns 3 and 4).

[27] As before, Panel A and Panel B report OLS and 2SLS results respectively.

inducing some boys to opt out of high school. Indeed, column 3 of Table C.9 shows that immigration had a negative and significant effect on enrollment of sons of native parentage aged 15-18.

3.6.2 Changes in Sex Ratios

The literature has documented that sex ratios, i.e. the relative number of men and women, can be an important determinant of marriage and family formation decisions (Angrist, 2002; Abramitzky et al., 2011). Since more than 60% of immigrants entering the United States at the beginning of the twentieth century were young men (Figure C.1), immigration likely altered sex ratios, possibly increasing the availability of potential mates for native women. However, in this section, we argue that this channel cannot explain a relevant fraction of our main results.

First, while changes in the relative number of men and women might have contributed to the increase in marriage rates and fertility of native women documented above, they cannot explain why immigration also raised native males' marriage rates.[28] Second, as we show in Table 3.8, where we explore the characteristics of partners of native women aged 18-33, only 4% of native women had a foreign born husband as of 1910. Also, the increase in marriage rates for men and women was quantitatively similar (see Table 3.4), suggesting that natives, in most cases, were marrying with each other.

Focusing on results reported in Table 3.8, in Panel A, we find that one standard deviation increase in immigration raised the probability of getting married with a husband of native parentage by around 6% for all native women, irrespective of their parentage (columns 2 and 5). Instead, while the effect of immigration on the probability of having a foreign born spouse for native women was indistinguishable from zero (column 3), it was positive and significant for second generation women (column 6). Yet, focusing on the relevant age group (i.e. 18-33), since second generation women who had a foreign born husband represented less than 2.5% of all native women, the implied effect of immigration on the overall marriage rate of native women was negligible.[29] Finally, Panel B documents that these effects were mirrored by a corresponding increase in fertility precisely for couples with higher marriage rates, in turn supporting the idea that immigration raised natives' fertility by fostering marriage in an era when oral contraception was not yet available (Bailey, 2006).

[28]Indirectly, higher competition in the marriage market may have induced men to increase their investment in education and on-the-job training and their earnings, as suggested by Becker (1981) in his notion of male "efficiency" (see also (Angrist, 2002)). However, even in this case, changes in sex ratios should have had a stronger impact on women as compared to men.

[29]In the age group 18-33, second generation women were 25% of native females, and their probability of marrying with a foreign born was 10% at baseline (see the last column of Table 3.8).

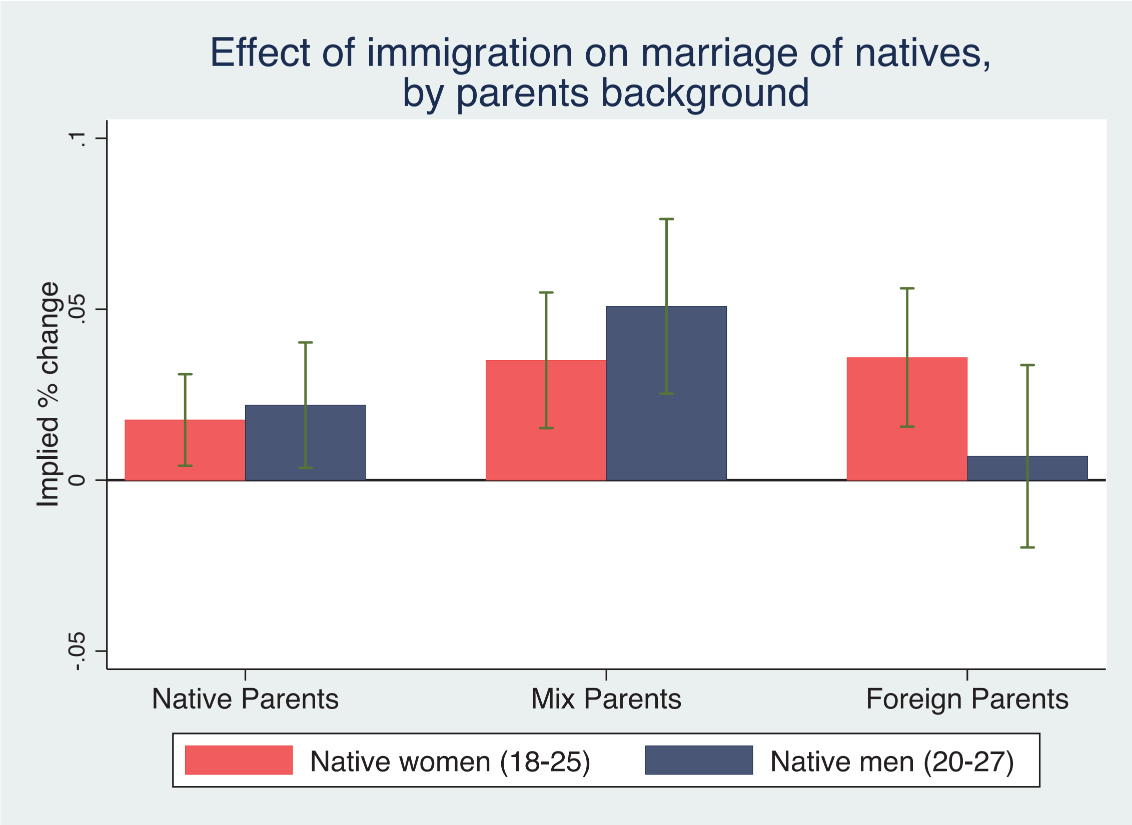
Table 3.8: Immigration, Marriage Rate, and Fertility of Native Women aged 18-33 (2SLS results)

	(1)	(2)	(3)	(4)	(5)	(6)
	All Native Women			**Second Generation Women**		
Husband	**All**	**Native Parentage**	**Immigrant**	**All**	**Native Parentage**	**Immigrant**
Panel A: Marriage rate						
Fr. Immigrant	0.209***	0.309***	0.001	0.193***	0.169**	0.178***
	(0.044)	(0.046)	(0.020)	(0.071)	(0.066)	(0.046)
Mean dep. var.	0.47	0.27	0.04	0.45	0.14	0.10
Panel B: Fertility (Children to Women Ratio)						
Fr. Immigrant	0.431***	0.443***	-0.005	0.359**	0.177*	0.259**
	(0.072)	(0.087)	(0.053)	(0.162)	(0.103)	(0.127)
Mean dep. var.	0.65	0.35	0.07	0.58	0.19	0.17
F-stat	251.3	251.3	251.3	251.3	251.3	251.3
Obs.	538	538	538	538	538	538

Notes: this Table presents results of 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. In panel A, the dependent variable is the marriage rate of women aged 18-33 by husband parentage. In panel B, the dependent variable is the children to women ratio by father parentage. We consider only children of women aged 18-33. For example, in column 2 of Panel B, the dependent variable is the number of children with native mother aged 18-33 and father with a native parentage over the number of native women aged 18-33. Columns 4-6 focus on women who are second generation immigrants. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Having established that most of the effects of immigration were not driven by native women marrying foreign born husbands, in the last part of this section, we study how the inflow of immigrants affected marriage prospects of second generation men and women, via changes in sex ratios. Sex ratios can have important implications for the marriage market of second generation immigrants, both directly and indirectly through the allocation of bargaining power within the couple. For example, in the same historical context of our paper, Angrist (2002) finds that a higher relative number of men in their own ethnic group improved marriage prospects of second generation females. Figure 3.9 documents a pattern in line with this idea: because of immigration, marriage rates of second generation women aged 18-25 increased twice as much as those of women of native parentage. Similarly, while immigration had a positive and large effect on marriage rates for men of native parentage, it did not have any significant impact for second generation men. This finding is consistent with the idea that immigrants increased competition in the marriage market for second generation men. In Table C.10, we separately report the effect of immigration on marriage rates of native men and women for different age groups and parentage, and document that all of the effect comes from the youngest cohorts represented in Figure 3.9 (that is, women aged 18-25 and men aged 20-27).

Figure 3.9: The impact of immigration on marriage rates by parentage



Notes: This graph shows the impact of one standard deviation increase of the fraction of immigrants on the marriage rates of men and women by parentage with respect to the mean value in 1910. We report the standardized coefficients.

To sum up, even though sex ratios were affected by immigration, they can hardly explain the increase in marriage rates of natives with native parentage, a group for which the relative number of men and women in the reference population was not significantly affected. Since natives of native parentage were by far the largest group among US born individuals, their decisions disproportionately affected natives' overall marriage and fertility.

3.6.3 Preservation of "Natives"

Opposition to immigration was widespread during the Age of Mass Migration, with a heated aversion towards individuals coming from non Anglo-Saxon and non English-speaking countries (Abramitzky and Boustan, 2017; Leonard, 2016). Since immigrants from Southern and Eastern Europe were linguistically and culturally far from natives (Hatton and Williamson, 2006), it is possible that natives reacted to immigration by marrying more and having more kids, in order to preserve their own race and culture (see Section3.2.2).

The role of culture in affecting marriage and fertility decisions has been stressed, among others, by Bisin and Verdier (2000) and Fernández and Fogli (2006), who study the transmission of cultural norms among second generation immigrants in the US. More broadly, social interactions can influence the diffusion of cultural norms and have historically contributed to the convergence of fertility rates, both within and across countries (Spolaore and Wacziarg, 2016). For instance, Daudin et al. (2016) find that the demographic transition at the end of the nineteenth century in France was affected by the diffusion of low-fertility norms through internal migration.

To test if native men and women changed their family formation decisions to preserve their own culture, we analyze whether the effect on marriage rates and fertility was stronger when natives were exposed to linguistically farther individuals (which we take as a proxy for cultural distance). Specifically, we construct an index of immigrants' linguistic distance from English, $LD_{ct} = \sum_j (sh^j_{ct} L^j)$, where sh^j_{ct} is the share of ethnic group j among the foreign born population of city c in year t, and L^j is the linguistic distance from English of country j, computed in Chiswick and Miller (2005).[30] To ease the interpretation of results, which are reported in Table 3.9, we standardize our measure of linguistic distance by subtracting its mean and dividing it through its standard deviation. Differently from what we would have expected if this mechanism was driving our results, marriage rates were not differentially affected by immigrants with different linguistic distance from English. These results thus suggest that cultural considerations were unlikely to explain our key findings.

[30]We instrument the actual ethnic shares sh_{ct} using the same logic of the instrument constructed in equation (2).

Table 3.9: Immigration, Linguistic Distance, Employment and Marriage Rate of Natives (2SLS results)

Dep Var.:	**Employment Men 20-35**	**Marriage Rate Women 18-33**			
Own Parents		**All**	**Native**	**Mix**	**Immigrants**
	(1)	(2)	(3)	(4)	(5)
Fr. Immigrant	0.136**	0.207***	0.126**	0.282***	0.274***
	(0.066)	(0.060)	(0.064)	(0.091)	(0.102)
Ling. Distance	0.001	0.000	0.001	-0.004	-0.007
	(0.004)	(0.003)	(0.004)	(0.005)	(0.006)
Mean Dep. Var.	0.340	0.257	0.277	0.642	0.603
Obs.	538	538	538	538	538

Notes: this Table presents results of 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is the marriage rate of the groups described in each panel. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6.4 Increased Labor Market Competition for Women

From the end of the nineteenth century to the 1920s, female workers were mainly young, unmarried, and from low-income households (Goldin, 2006). Most women were employed as piece workers in manufacturing, as private household workers or laundresses, or in clerical jobs. Upon getting married, women typically quitted their jobs because of the stigma attached to wives working outside their home (Goldin, 1990; Cherlin, 2014). As shown in Table C.11, in our sample of cities, the 1910 average labor force participation of native women aged 18-25 was 0.49, but was substantially lower for older women (0.33 and 0.25 for women aged 26-33 and 34-65, respectively).[31]

Studying the link between immigration, female labor force participation, and fertility, Furtado (2016) shows that the availability of lower cost childcare opportunities brought about by immigration induced native women to have more kids and work longer hours. In contrast with these results, at the beginning of the twentieth century, immigration may have increased competition in the labor market for women, in turn inducing them to first leave their job and then, as a consequence, to get married and have more children (Angrist and Evans, 1998). While possible, this interpretation seems to be inconsistent with the historical context studied in our paper: at that time, as already discussed above, women most frequently took care of their own children, and used to quit their job upon marriage. Moreover, even though immigrants provided a cheap and unskilled supply of labor, which in principle might have displaced women, during the Age of Mass Migration, the US economy had large potential for economic expansion (Higgs, 1971). Thus, the displacement of female workers due to immigration seems unlikely, even more so as immigrants were more closely substitutes for men than for women, and we showed above that immigration increased natives' employment (see Section 6.1).

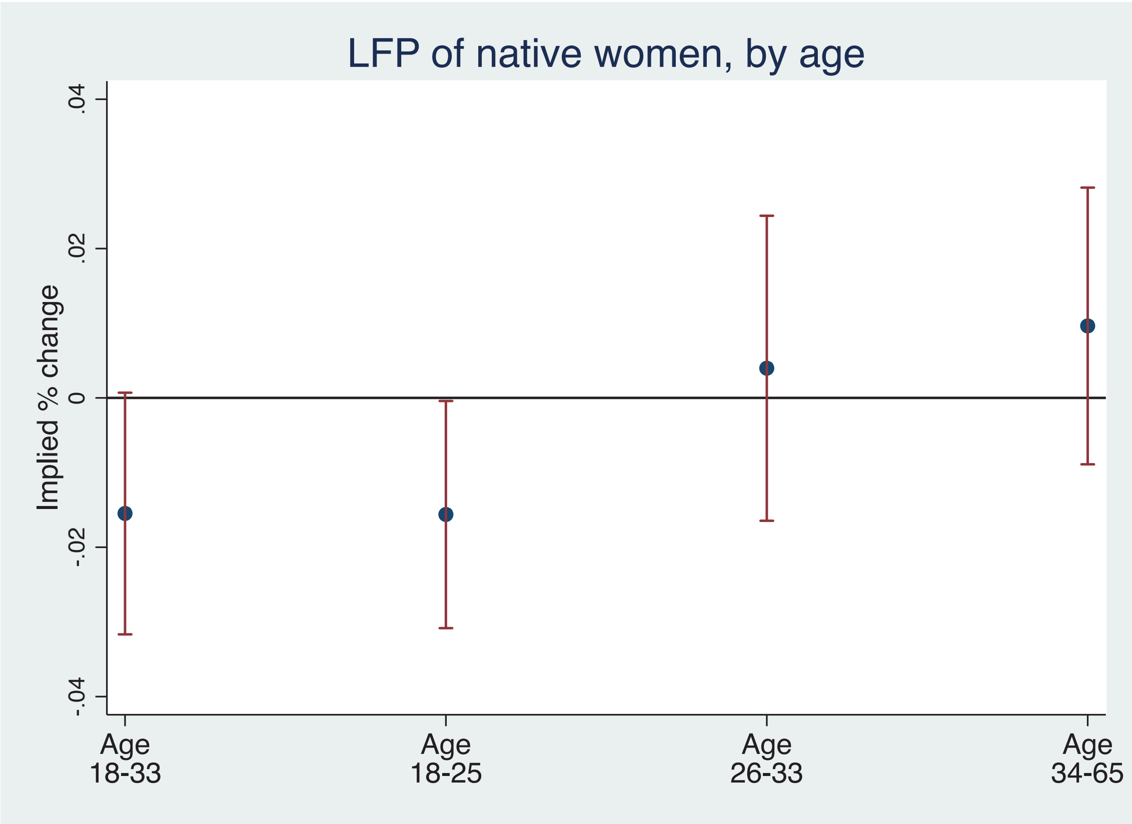
In line with this discussion, in Table C.11, we document that immigration decreased labor force participation only for native women in the age group that experienced a significant increase in marriage rates (i.e., women aged 18-25). The impact is instead indistinguishable from zero for all older age cohorts, including women between 26 and 33 years old, among which one third was in the labor force.[32] In Figure 3.10, we report the implied coefficients for the effect of a one standard deviation increase in immigration, and show that female labor force participation in the age group 18-25 fell by 1.6% relative to its 1910 mean. Incidentally, this effect is only slightly smaller (in absolute value) than the increase in marriage induced by immigration for women in the same age group (see Figure 3.7). Our interpretation of these results is that immigration first induced native

[31] Goldin (2006) notes that labor force participation of married women may be underestimated before 1940 because they were often reluctant to report that they had a job.

[32] Furthermore, women aged 26-33 were likely to work in the same sectors and occupations as women aged 18-25.

women to marry and have children, and then, as a consequence of the latter two decisions, to leave the labor force.

Figure 3.10: The impact of immigration on labor force participation by age groups of women



Notes: This graph shows the impact of one standard deviation increase of the fraction of immigrants on the decrease in labor force participation of women with respect to the mean value in 1910. We report the standardized coefficients separately by age group.

3.7 Conclusions

Today, immigration is at the forefront of the political debate, and there are increasing concerns over its economic and social consequences. If we look at American history, however, this is not the first time that immigration is such a relevant and controversial issue. In fact, at the beginning of the twentieth century, following the inflow of more than 30 million Europeans, the share of foreign born in the US population was even higher than it is today, and opposition towards immigration was widespread.

In this paper, we exploit plausibly exogenous variation in the number of European immigrants to US cities between 1910 and 1930 induced by WWI and the Immigration

Acts to study the impact of immigration on marriage rates, fertility, and the propensity to leave the parental house for young native men and women. We find that, by promoting industrial expansion and economic activity, immigration increased the supply of native “marriageable" men who, because of their better employment prospects and occupational standing, became more attractive spouses. This, in turn, fostered natives’ marriage rates for both men and women, and induced young adults to leave their parents’ house earlier in their life. Higher marriage rates, in a period when oral contraception was not yet available, raised natives’ fertility, mainly by increasing the number of women with at least one child (extensive margin).

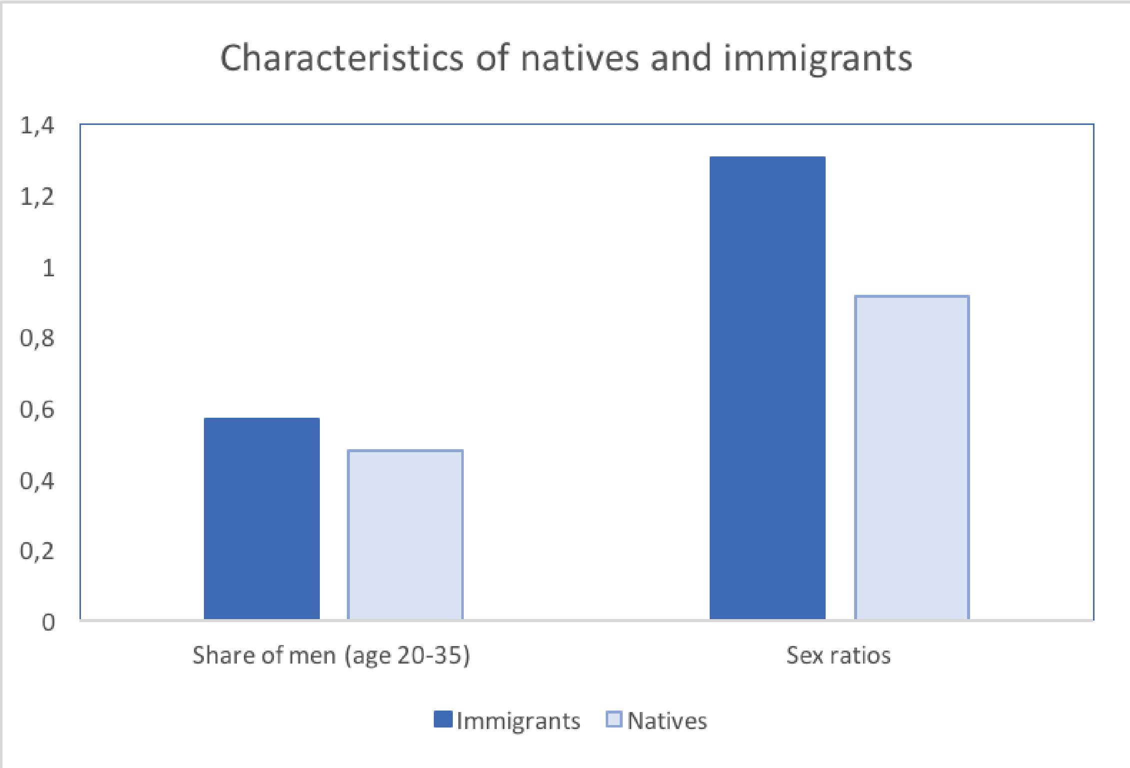
In our context, the inflow of immigrants was largely beneficial to natives’ economic and social outcomes. However, this does not imply that immigration will always promote fertility and marriage among young natives. In fact, if immigrants increase labor market competition, they may deteriorate, rather than promote, family stability as well as the environment where children grow up. Moreover, while we showed that in the early twentieth century, immigration to US cities affected marriage rates and fertility of natives mostly through (positive) income shocks, other channels may be at play in other settings. These observations suggest that one needs to be careful when extrapolating our results to other contexts.

Findings in this paper provide motivation for future work in at least two directions. First, in this study, we have not explored how changes in the supply of “marriageable men" affected the quality of the match between husbands and wives. If higher marriage rates were associated with worse matching between partners, this might have increased divorce rates and family instability, in turn lowering children’s well-being (Stevenson and Wolfers, 2007; Lundberg et al., 2016). Second, this setting seems ideal to study the dynamics of cultural assimilation - between immigrants and natives as well as between different ethnic groups - using intermarriage as a proxy for the latter. These are ideas that we plan to investigate in our future research.

C. Appendix

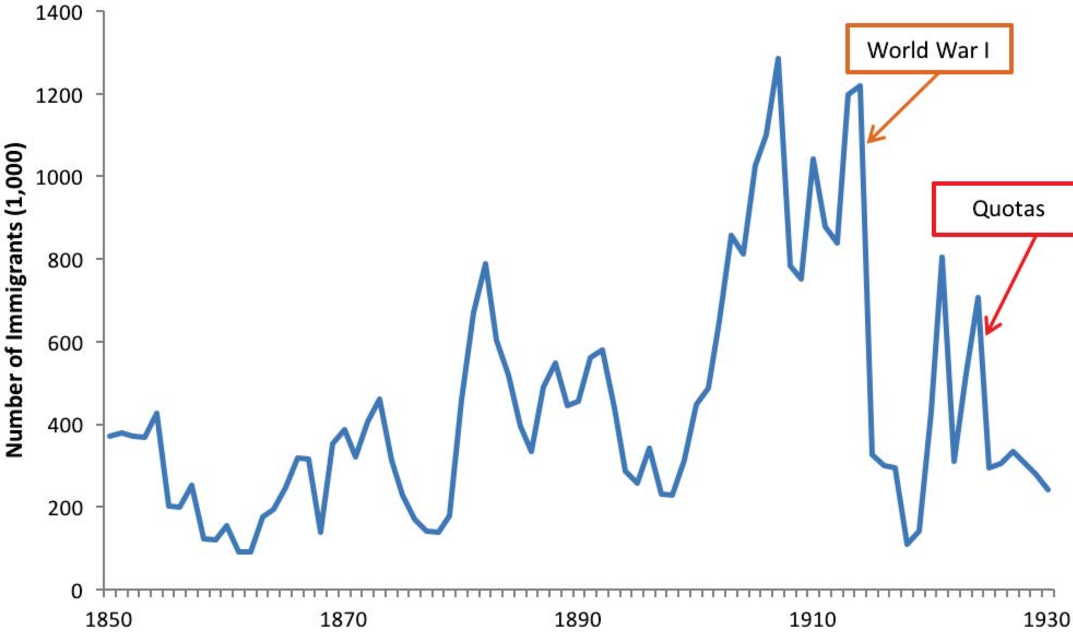
C.1 Additional Tables and Figures

Figure C.1: Summary Statistics on share of men and sex ratios for natives and immigrants (in 1910)



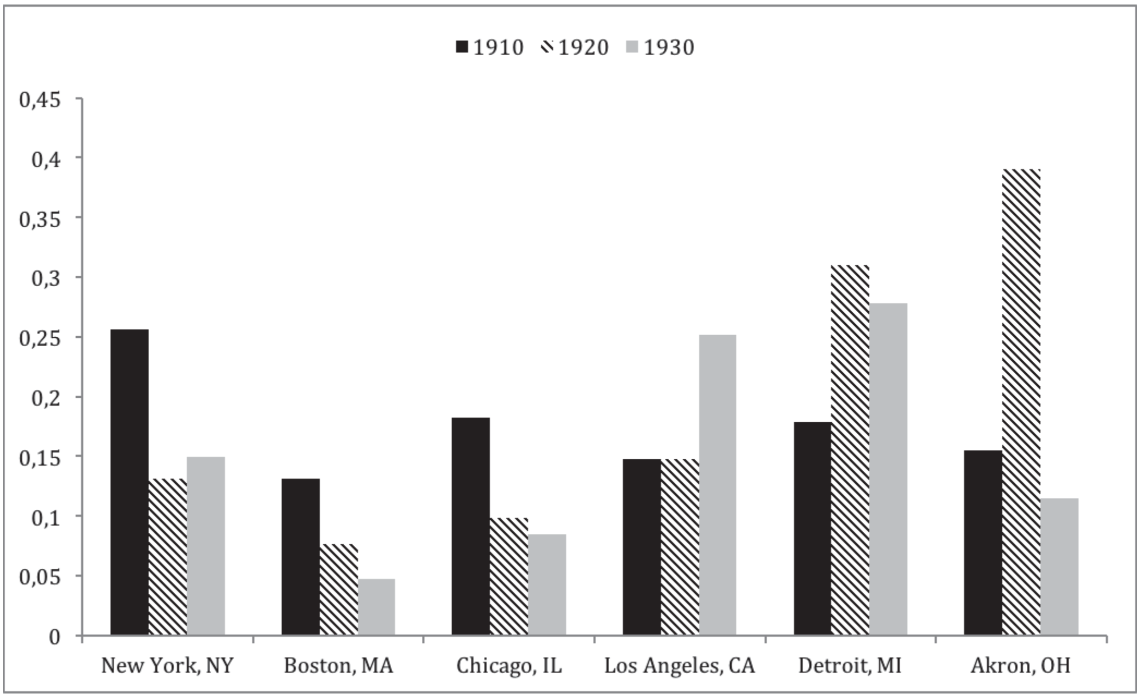
Notes: The sex ratios is defined as the number of native men (resp. immigrant men) in the age group 20-35 over the number of native women in the age group 18-33 (resp. immigrant women). Source: Authors' calculations from IPUMS sample of 1910 US Census (Ruggles et al., 2015).

Figure C.2: Total Number of Immigrants (in Thousands)



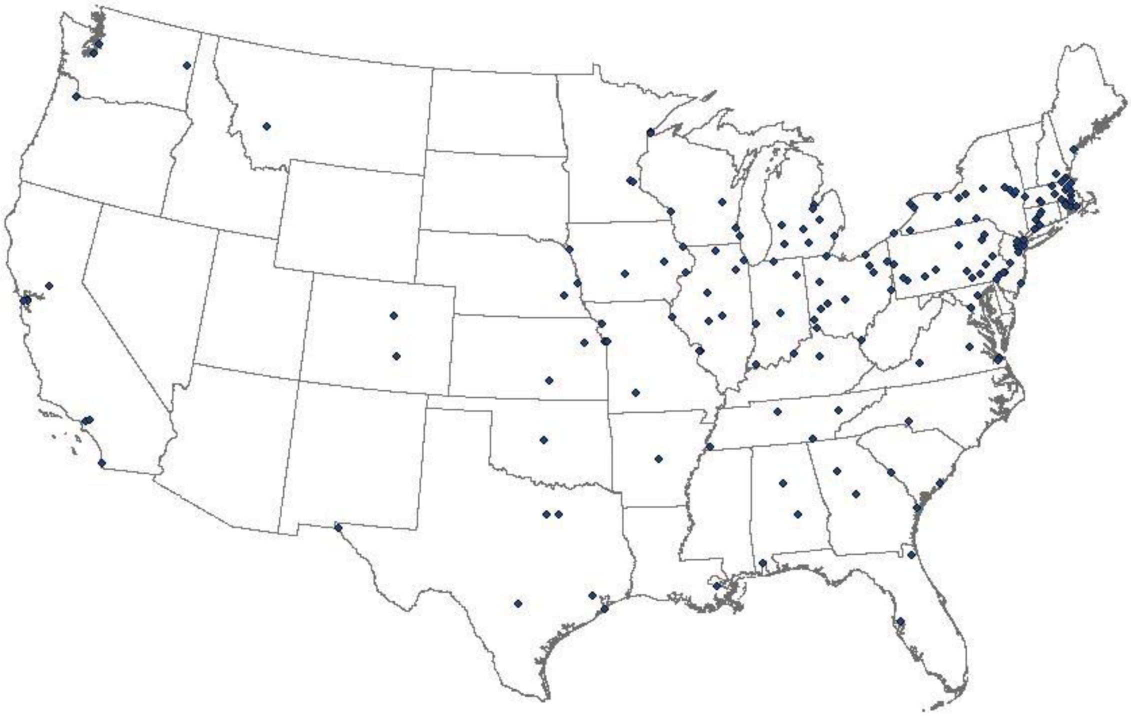
Notes: Annual inflow of immigrants to the United States (1850-1930). Source: Migration Policy Institute.

Figure C.3: Recent Immigrants Over 1900 City Population, by Decade



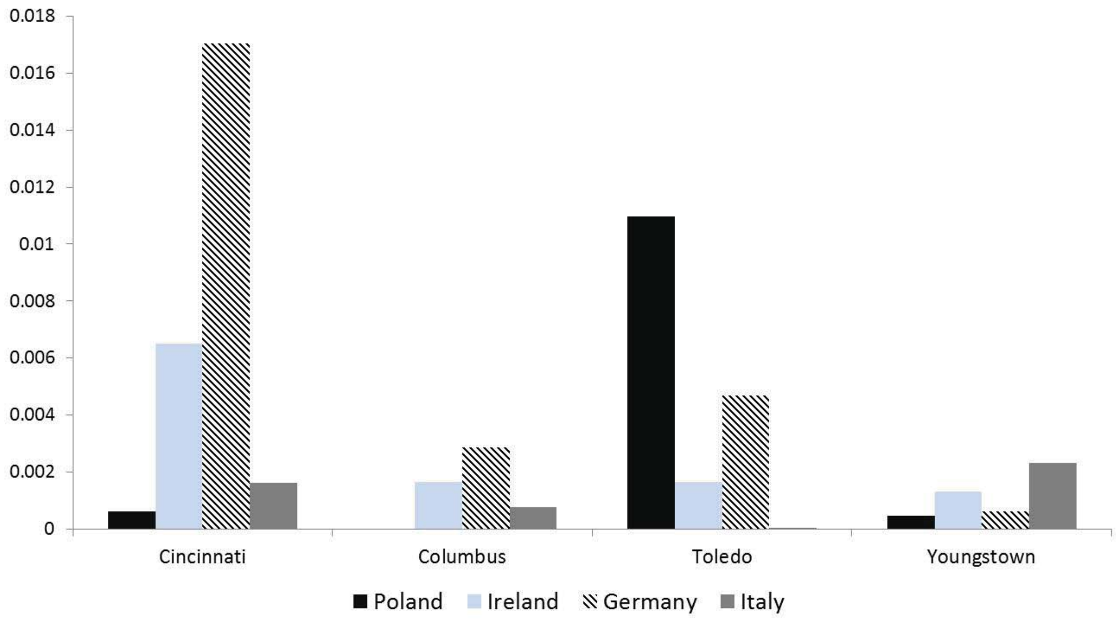
Notes: Number of European immigrants that arrived in the United States in the last decade over 1900 city population, for selected cities and by decade. Source: Authors' calculations from IPUMS sample of US Census (Ruggles et al., 2015).

Figure C.4: 180 Cities in the Balanced Panel.



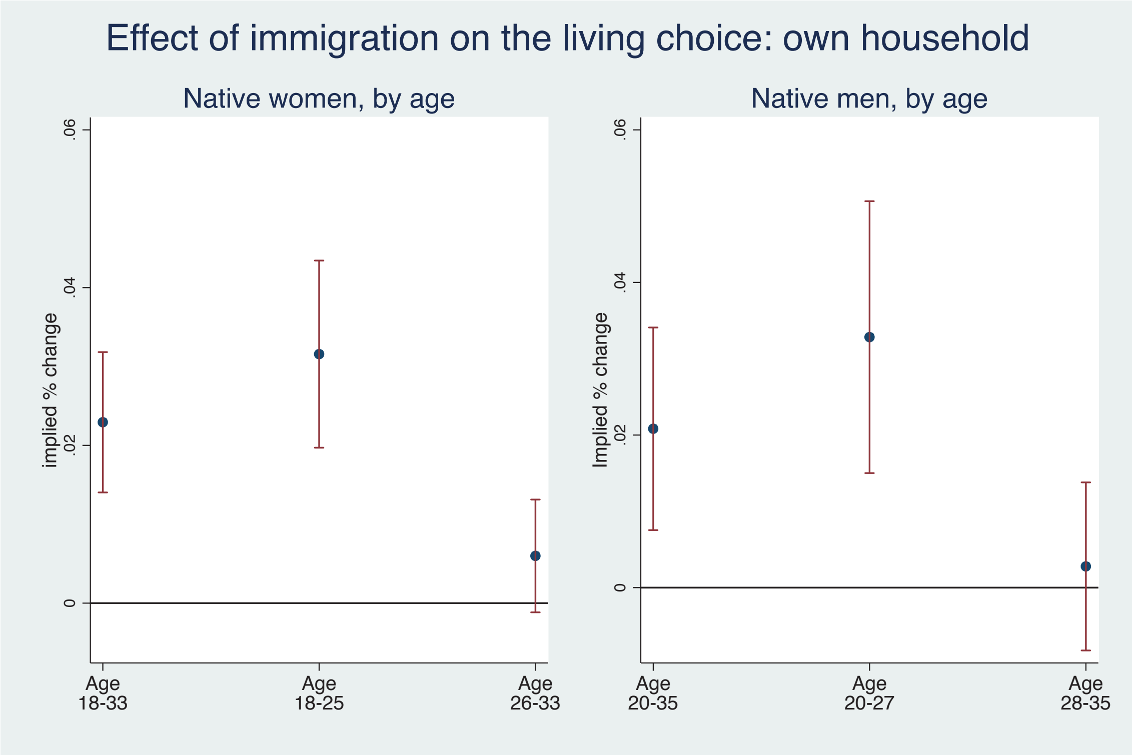
Notes: The map plots the 180 cities with at least 30,000 residents in each of the three Census years 1910, 1920, and 1930.

Figure C.5: Share of Immigrants from Selected Regions in Ohio, 1900



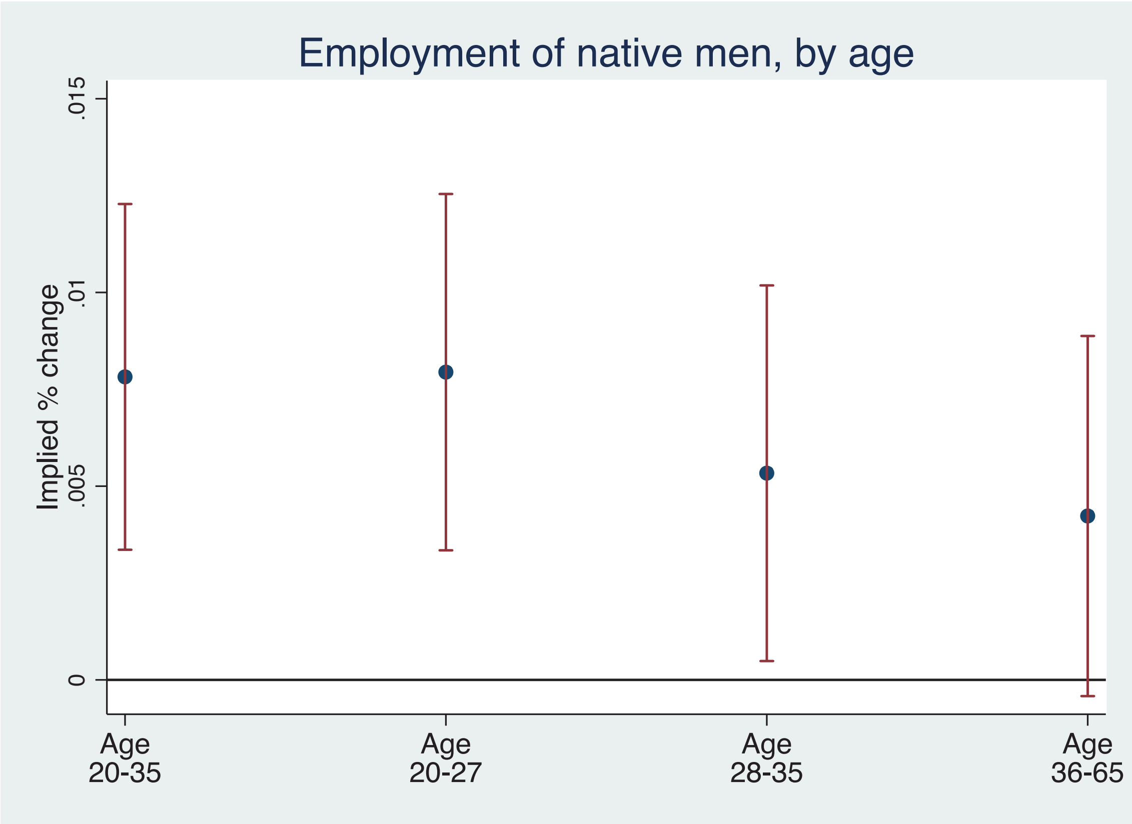
Notes: This graph shows share of individuals of European ancestry living in selected cities of Ohio in 1900, for selected ethnic groups. Source: Authors' calculations using IPUMS data. .

Figure C.6: The impact of immigration on the choice of creating an independent family unit, by gender and age group.



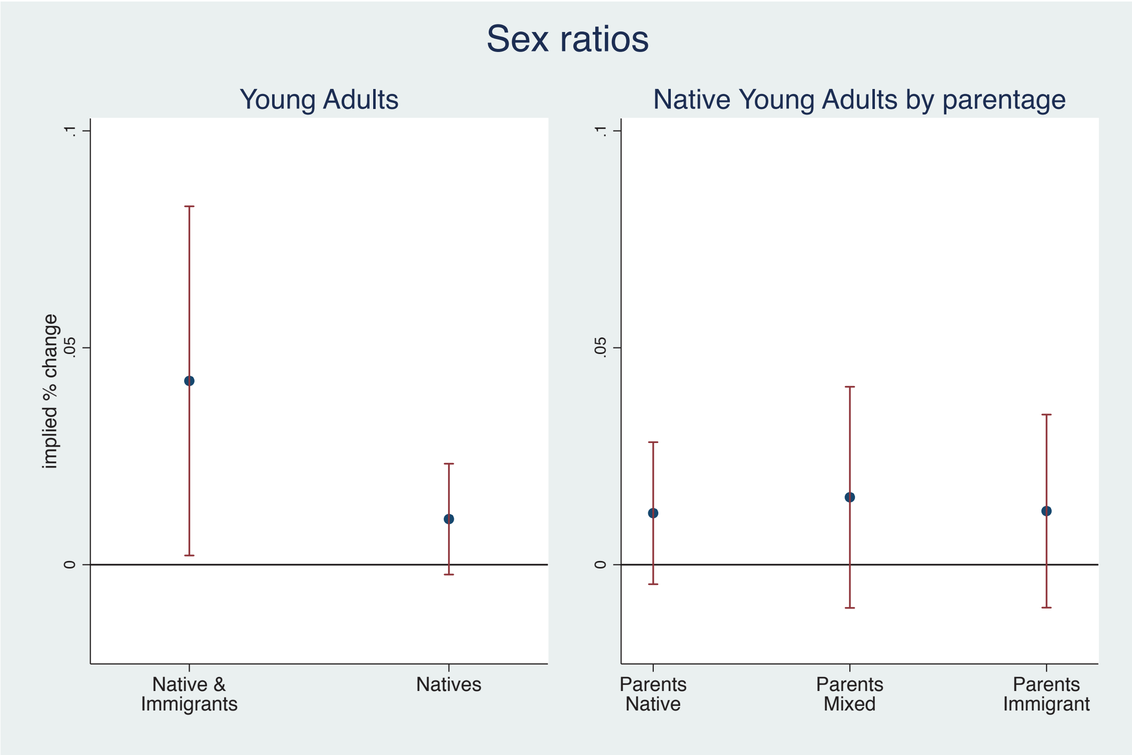
Notes: This graph shows the impact of one standard deviation increase of the fraction of immigrants on the probability of being the household head or spouse by age and gender.

Figure C.7: The impact of immigration on the choice of living with parents by age groups and gender



Notes: This graph shows the impact of one standard deviation increase of the fraction of immigrants on the decrease in the probability of living with parents with respect to the mean value in 1910. We report the standardized coefficients separately by age group and gender.

Figure C.8: The impact of immigration on the choice of creating an independent family unit, by gender and age group.



Notes: This graph shows the impact of one standard deviation increase of the fraction of immigrants on the sex ratios or young adults, i.e. the number of men in the age group 20-35 over the number of women in the age group 18-33. The first bar shows the impact for the whole population (natives+ immigrants) living in the 180 US cities with at least 30,000 residents in each Census year. The following bars present the sex ratios for natives, divided by parentage.

Table C.1: Sending Regions

UK	Russia
Ireland	Eastern Europe (Yugoslavia, Czechoslovakia, etc.)
Denmark	Austria-Hungary
Finland	Switzerland
Norway	France
Sweden	Belgium-Netherlands
Germany	Greece-Portugal-Spain
Poland	Italy

Table C.2: The impact of immigration on marriage rate by parentage

Akron, OH	Elizabeth, NJ	McKeesport, PA	Saint Joseph, MO
Albany, NY	Elmira, NY	Memphis, TN	Saint Louis, MO
Allentown, PA	Erie, PA	Milwaukee, WI	Saint Paul, MN
Altoona, PA	Evansville, IN	Minneapolis, MN	Salem, MA
Amsterdam, NY	Everett, MA	Mobile, AL	San Antonio, TX
Atlanta, GA	Fall River, MA	Montgomery, AL	San Diego, CA
Atlantic City, NJ	Fitchburg, MA	Mount Vernon, NY	San Francisco, CA
Auburn, NY	Flint, MI	Nashville, TN	Savannah, GA
Augusta, GA	Fort Wayne, IN	New Bedford, MA	Schenectedy, NY
Baltimore, MD	Fort Worth, TX	New Britain, CT	Scranton, PA
Bay City, MI	Galveston, TX	New Castle, PA	Seattle, WA
Bayonne, NJ	Grand Rapids, MI	New Haven, CT	Sioux City, IA
Berkeley, CA	Hamilton, OH	New Orleans, LA	Somerville, MA
Binghamton, NY	Harrisburg, PA	New York, NY	South Bend, IN
Birmingham, AL	Hartford, CT	Newark, NJ	Spokane, WA
Boston, MA	Haverhill, MA	Newton, MA	Springfield, IL
Bridgeport, CT	Hoboken, NJ	Niagara Falls, NY	Springfield, MA
Brockton, MA	Holyoke, MA	Norfolk, VA	Springfield, MO
Buffalo, NY	Houston, TX	Oakland, CA	Springfield, OH
Butte, MT	Huntington, WV	Oklahoma City, OK	Superior, WI
Cambridge, MA	Indianapolis, IN	Omaha, NE	Syracuse, NY
Camden, NJ	Jackson, MI	Oshkosh, WI	Tacoma, WA
Canton, OH	Jacksonville, FL	Pasadena, CA	Tampa, FL
Cedar Rapids, IA	Jamestown , NY	Passaic, NJ	Taunton, MA
Charleston, SC	Jersey City, NJ	Paterson, NJ	Terre Haute, IN
Charlotte, NC	Johnstown, PA	Pawtucket, RI	Toledo, OH
Chattanooga, TN	Joliet, IL	Peoria, IL	Topeka, KS
Chelsea, MA	Kalamazoo, MI	Perth Amboy, NJ	Trenton, NJ
Chester, PA	Kansas City, KS	Philadelphia, PA	Troy, NY
Chicago, IL	Kansas City, MO	Pittsburgh, PA	Utica, NY
Cincinnati, OH	Knoxville, TN	Pittsfield, MA	Washington, DC
Cleveland, OH	La Crosse, WI	Portland, ME	Waterbury, CT
Columbus, OH	Lancaster, PA	Portland, OR	Wheeling, WV
Covington, KY	Lansing, MI	Portsmouth, VA	Wichita, KS
Dallas, TX	Lawrence, MA	Providence, RI	Wilkes-Barre, PA
Davenport, IA	Lexington, KY	Pueblo, CO	Williamsport, PA
Dayton, OH	Lima, OH	Quincy, IL	Wilmington, DE
Decatur, IL	Lincoln, NE	Quincy, MA	Woonsocket, RI
Denver, CO	Little Rock, AR	Racine, WI	Worcester, MA
Des Moines, IA	Los Angeles, CA	Reading, PA	Yonkers, NY
Detroit, MI	Louisville, KY	Richmond, VA	York, PA
Dubuque, IA	Lowell, MA	Roanoke, VA	Youngstown, OH
Duluth, MN	Lynn, MA	Rochester, NY	
East Orange, NJ	Macon, GA	Rockford, IL	
East St. Louis, IL	Malden, MA	Sacramento, CA	
El Paso, TX	Manchester, NH	Saginaw, MI	

Notes: This table lists the 180 US cities with at least 30,000 residents in each Census year.

Table C.3: Immigration and Marriage of Native Men aged 20-35

	(1)	(2)	(3)	(4)
	All Sample		**Restricted Sample**	
	OLS	2SLS	OLS	2SLS
Fr. Immigrant	-0.006	0.190***	0.077	0.147**
	(0.135)	(0.054)	(0.082)	(0.063)
F-stat		251.3		251.3
Mean dep. var. in 1910	0.42	0.42	0.43	0.43
Obs.	538	538	529	529

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report in columns 1 and 2. In columns 3 and 4, we exclude three cities (Duluth, Superior, and Tacoma) with an extraordinary low level of marriage rate of men aged 20-35 in 1910. The dependent variable is the fraction of men married in the age range 20-35. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Immigration and Marriage of Natives - 2SLS results

	(1)	(2)	(3)	(4)	(5)
Panel A: Women					
		Marriage rate			**Never Married**
Age Groups	**18-33**	**18-25**	**26-33**	**34-65**	**34-65**
Fr. Immigrant	0.209***	0.229***	0.053	0.025	-0.082***
	(0.044)	(0.038)	(0.054)	(0.035)	(0.019)
Mean dep. var.	0.47	0.34	0.65	0.63	0.15
Obs.	538	538	538	538	538
Panel B: Men					
		Marriage rate			**Never Married**
Age Groups	**20-35**	**20-27**	**28-35**	**36-65**	**36-65**
Fr. Immigrant	0.190***	0.236***	-0.001	0.011	-0.026
	(0.054)	(0.055)	(0.059)	(0.045)	(0.035)
Mean dep. var.	0.45	0.30	0.65	0.73	0.14
Obs.	538	538	538	538	538

Notes: this Table presents results of 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is the fraction of women married in the different age range in Panel A and the fraction of men married in the different age range in Panel B. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: Immigration and Fertility of Native Women - 2SLS results

	Dep. Variable: Fertility of Native Women					
	Children to Women Ratio		Mothers to Women Ratio		Children to Mothers Ratio	
	Age 18-25	Age 26-33	Age 18-25	Age 26-33	Age 18-25	Age 26-33
	(1)	(2)	(3)	(4)	(5)	(6)
Fr. Immigrant	0.234***	0.357***	0.131***	0.071**	0.160**	0.391***
	(0.049)	(0.122)	(0.025)	(0.034)	(0.080)	(0.137)
F-stat	251.3	251.3	251.3	251.3	251.3	251.3
Mean Dep.Var.	0.330	1.050	0.210	0.500	1.530	2.090
Obs.	538	538	538	538	538	538

Notes: this Table presents results of 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is: in column 1 and 2, the total number of children with native mother over the total number of women in the age range, in column 3 and 4 the fraction of women who have children and in column 5 and 6 the average number of children per mother. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: Native Men (20-35) in Selected Occupations

	High Immigrants' Competition			**Low Immigrants' Competition**		
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction Natives	**Manuf. Laborers**	**Bakers**	**Blacksmiths**	**Manuf. Foremen**	**Engineers**	**Electricians**
Panel A: OLS						
Fr. Immigrant	-0.079	-0.008*	-0.009	0.025***	0.017	0.007
	(0.058)	(0.004)	(0.005)	(0.004)	(0.014)	(0.006)
Panel B: 2SLS						
Fr. Immigrant	-0.117**	-0.008**	-0.005	0.030***	0.033***	0.010
	(0.050)	(0.004)	(0.005)	(0.006)	(0.011)	(0.007)
F-stats	251.3	251.3	251.3	251.3	251.3	251.3
Mean dep. var.	0.038	0.005	0.007	0.006	0.018	0.014
Natives/Immigrants Ratio (1910)	0.220	0.231	0.750	3.500	3.667	4.200
Obs.	538	538	538	538	538	538

Notes: this table presents results for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year 1910, 1920, and 1930 (see Table C.2 in the appendix). The dependent variable is the fraction of native males in age range (20-35) working in the occupation reported at the top of each column. Panels A and B report, respectively, OLS and 2SLS results. *Fr. Immigrants* is the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2 (see (2) in the main text). F-stat refers to the K-P F-stat for weak instrument. All regressions include city and state by year fixed effects. The mean of each dependent variable at baseline is shown at the bottom of the Table. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Additional Outcomes on Economic Activity (2SLS)

Dep. Variables	(1) **Log Value Added per Establishment**	(2) **Log Value of products per Establishment**	(3) **Log Establishment Size**	(4) **Log horsepower**	(5) **TFP**
Panel A: OLS					
Fr. Immigrant	2.057***	2.264***	2.195***	1.267***	0.295
	(0.703)	(0.704)	(0.614)	(0.475)	(0.358)
Panel B: 2SLS					
Fr. Immigrant	2.889***	3.549***	2.532***	1.906***	1.013*
	(0.954)	(1.214)	(0.815)	(0.705)	(0.540)
F-stat	270.5	270.5	270.5	270.5	270.5
Obs.	525	525	525	525	525

Notes: this Table presents results for a balanced panel of the 178 US cities with at least 30,000 residents in each Census year 1910, 1920, and 1930, and for which data were reported in the Census of Manufacture between 1909 and 1929. Panels A and B report, respectively, OLS and 2SLS results. The dependent variable is: the log of value added per establishment in Col 1; the log of value of products per establishment in Col 2, the log establishment size in Col 3; the log of horsepower in Col 4; and total factor productivity (TFP) in Col 5. *Fr. Immigrants* is the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2 (see (2) in the main text). F-stat refers to the K-P F-stat for weak instrument. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Immigration and Living Choices of Natives

	(1)	(2)	(3)	(4)
Dep. Var.	**Share of children < 10 (native parents)**		**Share of families (children < 10, native parents)**	
	Father employed	**Father unskilled**	**Father employed**	**Father unskilled**
Panel A: OLS				
Fr. Immigrant	0.052	-0.027	0.032	-0.035
	(0.049)	(0.075)	(0.045)	(0.076)
Panel B: 2SLS				
Fr. Immigrant	0.049	-0.138^{**}	0.024	-0.171^{***}
	(0.037)	(0.061)	(0.034)	(0.063)
F-stat	251.3	251.3	251.3	251.3
Mean dep. var.	0.908	0.332	0.901	0.318
Obs.	538	538	538	538

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable are described on the top part of the Table. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: Immigration and Education of Native Children

	Dep. Var.: Fraction attending school					
	Sons of natives			**Daughters of natives**		
	(1)	(2)	(3)	(4)	(5)	(6)
Age group:	**Age 6-14**	**Age 15-18**	**Age 19-24**	**Age 6-14**	**Age 15-18**	**Age 19-24**
Panel A: OLS						
Fr. Immigrant	0.007	-0.081	0.010	-0.025	0.059	-0.006
	(0.040)	(0.080)	(0.019)	(0.041)	(0.076)	(0.027)
Panel B: 2SLS						
Fr. Immigrant	0.067***	-0.100**	0.011	0.017	0.044	-0.042*
	(0.025)	(0.049)	(0.017)	(0.027)	(0.059)	(0.023)
Dep. var:	.933	.241	.015	.936	.22	.013
Obs.	538	538	538	538	538	538

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable are described on the top part of the Table. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.10: Immigration and Marriage Rate of Natives by parentage (2SLS results)

Dep. Variable: Marriage rate						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:	**Women Age 18-25**			**Women Age 26-33**		
Own Parents	**Native**	**Mixed**	**Foreign**	**Native**	**Mixed**	**Foreign**
Fr. Immigrant	0.127***	0.192***	0.211***	0.068	0.163	0.117
	(0.049)	(0.055)	(0.061)	(0.042)	(0.113)	(0.112)
Mean Dep. Var.	0.340	0.257	0.277	0.642	0.603	0.587
Obs.	538	538	538	538	538	538
Panel B:	**Men Age 20-27**			**Men Age 28-35**		
Own Parents	**Native**	**Mixed**	**Foreign**	**Native**	**Mixed**	**Foreign**
Fr. Immigrant	0.139**	0.227***	0.035	-0.029	0.013	0.047
	(0.059)	(0.058)	(0.067)	(0.075)	(0.091)	(0.078)
Mean Dep. Var.	0.297	0.210	0.233	0.623	0.575	0.561
Obs.	538	538	538	538	538	538

Notes: this Table presents results of 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is the marriage rate of the groups describled in each panel. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.11: Immigration and LFP of Native Women - 2SLS results

	LFP of Native Women				
	(1)	(2)	(3)	(4)	(5)
	Age 18-33	Age 18-33	Age 18-25	Age 26-33	Age 34-65
	OLS	2SLS	2SLS	2SLS	2SLS
Fr. Immigrant	-0.084	-0.115*	-0.135**	0.023	0.042
	(0.083)	(0.061)	(0.067)	(0.061)	(0.041)
F-stat		251.3	251.3	251.3	251.3
Mean dep. var.	.42	.42	.49	.33	.25
Obs.	538	538	538	538	538

Notes: this Table presents results of OLS and 2SLS for a balanced panel of the 180 US cities with at least 30,000 residents in each Census year report. The dependent variable is the labor force participation of women in the different age range. *Fr. Immigrants* refers to the fraction of immigrants arrived in the previous decade over predicted city population, and is instrumented using the baseline version of the instrument constructed in Section 3.4.2. The mean of dependent variables is shown at the bottom of the Table. KP F-stat is the Kleibergen-Paap F stat for joint significance of instruments. All regressions include city and state by year fixed effects. Robust standard errors, clustered at the MSA level, in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.2 Graphical Example

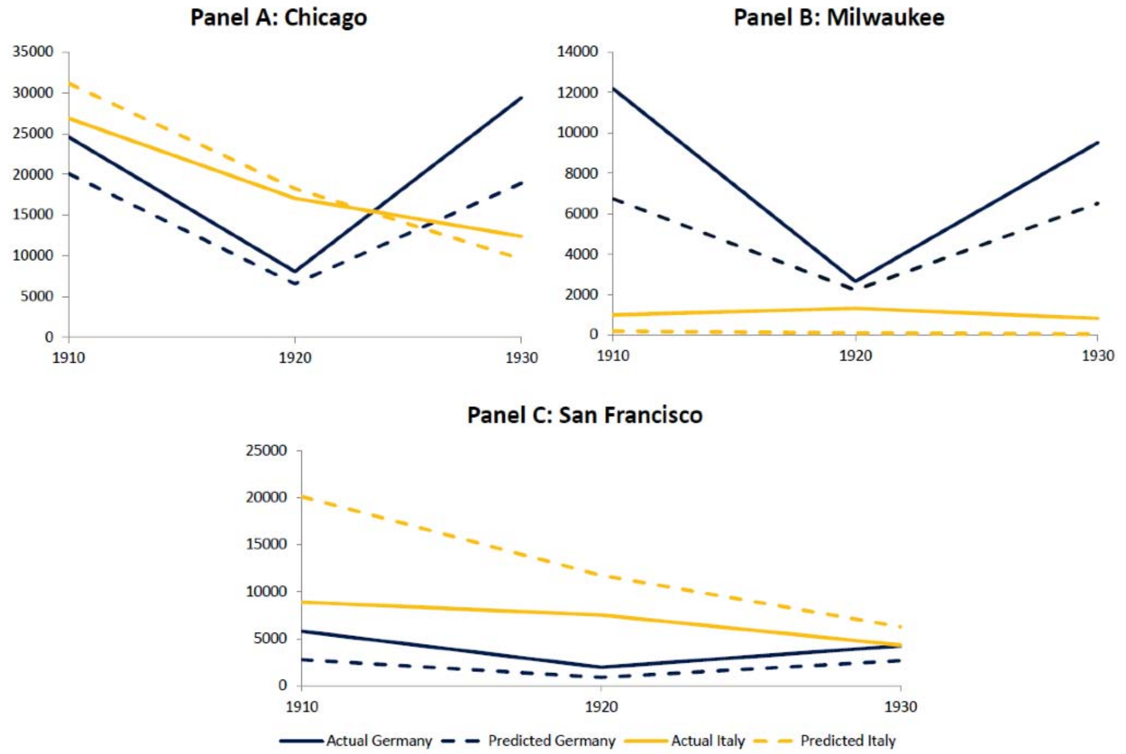
As discussed in the main text, the shift-share instrument exploits two sources of variation: first, cross-sectional variation in the share of individuals from each ethnic group living in different US cities in 1900; second, time-series variation due to changes in the total number of immigrants from any sending region entering the United States in a given decade. To illustrate how the two sources of variation are combined by our instrument, Figure C.9 presents a simple example for two ethnic groups (Germans and Italians) and three cities (Chicago, Milwaukee, and San Francisco).

German immigration fell between 1910 and 1920 due to WWI, but rebounded after 1920, as the quotas were quite generous with respect to Germany. Instead, between 1910 and 1930, Italian immigration declined monotonically. Starting from Panel A, Chicago had large Italian and German communities in 1900. In line with the aggregate flows, both the actual (straight lines) and the predicted (dotted lines) number of Italians (yellow lines) and Germans (blue lines) arriving in Chicago fell between 1910 and 1920. However, after 1920, while Italian immigration continued its decline, Chicago experienced a positive immigration shock from Germany.

Next, Panel B presents the example of Milwaukee, a city with a relatively large German community, but with almost no Italians in 1900. As a result, for this city, variation in immigration was driven by changes in German, and not Italian, immigration. Finally, while very few Germans were living in San Francisco in 1900, Italian settlements were fairly large in this city. As documented in Panel C, the actual and predicted immigration shock for San Francisco was due to the decline in Italian immigration, and only marginally to the inflow of Germans after 1920.

The instrument in equation (3.2) extends this example to many cities and many ethnic groups, but the logic behind it can be grasped by looking at the patterns in Panels A to C of Figure C.9

Figure C.9: Actual and Predicted Immigration



Notes: This Figure reports the actual and predicted number of Italians and Germans arrived during the previous decade to Chicago (Panel A), Milwaukee (Panel B), and San Francisco (Panel C), in 1910, 1920, and 1930. Predicted immigration is obtained from the instrument constructed in equation (2) in the main text. Source: from IPUMS sample of US Census (Ruggles et al., 2015).

References

Abramitzky, R. and L. P. Boustan (2017). Immigration in american economic history. *Journal of Economic Literature*.

Abramitzky, R., A. Delavande, and L. Vasconcelos (2011). Marrying up: the role of sex ratio in assortative matching. *American Economic Journal: Applied Economics 3*(3), 124–157.

Alan, S. and S. Ertac (forthcoming). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*.

Alan, S., S. Ertac, and I. Mumcu (2017). Gender stereotypes in the classroom and effects of achievement. *Working Paper*.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018). Revealing stereotypes: Teacher bias and immigrants performance. *Mimeo Bocconi Univeristy*.

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

Altonji, J. G. and D. Card (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market*, pp. 201–234. University of Chicago Press.

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? evidence from america's second generation. *The Quarterly Journal of Economics 117*(3), 997–1038.

Angrist, J. and W. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016). Stand and deliver: Effects of boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics 34*(1), 275–318.

Autor, D., D. Dorn, G. Hanson, et al. (2017). When work disappears: manufacturing decline and the falling marriage-market value of men. *NBER Working Paper 23173*.

Bailey, M. J. (2006). More power to the pill: the impact of contraceptive freedom on women's life cycle labor supply. *The Quarterly Journal of Economics 121*(1), 289–320.

Banaji, M. R., B. A. Nosek, and A. G. Greenwald (2004). No place for nostalgia in science: A response to arkes and tetlock. *Psychological Inquiry 15*(4), 279–310.

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs, NJ Prentice Hall.

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

Baron-Cohen, S. (2003). *The Essential Difference: Men, Women, and the Extreme Male Brain.* Allen Lane, London.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: does exposure reduce bias? *The Quarterly Journal of Economics 124*(4), 1497–1540.

Becker, G. (1981). *A Treatise on the Family.* Harvard university press.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics 117*(3), 871–915.

Bertrand, M., D. Chugh, and S. Mullainathan (2005). Implicit discrimination. *American Economic Review*, 94–98.

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

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012). The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment. *The Quarterly Journal of Economics 127*(3), 1205–1242.

Betts, J. (2011). The economics of tracking in education. *Handbook of the Economics of Education 3*(341-381), 4.

Bharadwaj, P., G. De Giorgi, D. Hansen, and C. Neilson (2016). The gender gap in mathematics: Evidence from low-and middle-income countries. *Economic Development and Cultural Change*.

Bisin, A. and T. Verdier (2000). Beyond the melting pot: Cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics 115*(3), 955–988.

Bobba, M. and V. Frisancho (2016). Learning about Oneself: The effects of signalling academic ability on school choice. *TSE Working Paper No. 660, Toulouse School of Economics*.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2017). Stereotypes. *The Quarterly Journal of Economics*.

Bordalo, P., K. B. Coffman, N. Gennaioli, and A. Shleifer (2016). Beliefs about gender. *NBER Working Paper*.

Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics 118*(4), 1335–1374.

Borjas, G. J. and L. F. Katz (2007). The evolution of the mexican-born workforce in the united states. In *Mexican immigration to the United States*, pp. 13–56. University of Chicago Press.

Brown, D. (2002). *Career choice and development*. John Wiley & Sons.

Brunello, G. and D. Checchi (2007). Does school tracking affect equality of opportunity? new international evidence. *Economic Policy 22*(52), 781–861.

Burchardi, K. B., T. Chaney, and T. A. Hassan (2016). Migrants, ancestors, and investments. *National Bureau of Economic Research*.

Burns, J., L. Corno, and E. La Ferrara (2016). Interaction, stereotypes and performance. evidence from south africa. *Working Paper*.

Campa, P., A. Casarico, and P. Profeta (2010). *Gender culture and gender gap in employment*, Volume 57. Oxford University Press.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Card, D. (2005). Is the new immigration really so bad? *The Economic Journal 115*(507).

Card, D. and L. Giuliano (2016). Can tracking raise the test scores of high-ability minority students? *The American Economic Review 106*(10), 2783–2816.

Card, D. and A. A. Payne (2017). High school choices and the gender gap in stem. Technical report, National Bureau of Economic Research.

Carlana, M., E. La Ferrara, and P. Pinotti (2017). Goals and gaps: Educational careers of immigrant children. *Mimeo Bocconi Univeristy*.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Cherlin, A. J. (2014). *Labor's love lost: the rise and fall of the working-class family in America*. Russell Sage Foundation.

Chiswick, B. R. and P. W. Miller (2005). Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of Multilingual and Multicultural Development 26*(1), 1–11.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*.

Cooper, H. M. and T. L. Good (1983). *Pygmalion grows up: Studies in the expectation communication process*. Longman Publishing Group.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *The American Economic Review 97*(2), 31.

Cvencek, D., A. N. Meltzoff, and A. G. Greenwald (2011). Math–gender stereotypes in elementary school children. *Child development 82*(3), 766–779.

Dalton, P. S., S. Ghosal, and A. Mani (2014). Poverty and aspirations failure. *The Economic Journal*.

Dasgupta, N. and A. G. Greenwald (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology 81*(5), 800.

Daudin, G., R. Franck, and H. Rapoport (2016). The cultural diffusion of the fertility transition: Evidence from internal migration in 19th century france. *CESifo Working Paper*.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review 95*(2), 158–165.

Dee, T. S. (2014). Stereotype threat and the student-athlete. *Economic Inquiry 52*(1), 173–182.

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

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review 101*(5), 1739–74.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers 56*(2), 209–230.

Dustmann, C., P. A. Puhani, and U. Schönberg (2017). The long-term effects of early track choice. *The Economic Journal 127*(603), 1348–1380.

Dustmann, C., U. Schonberg, and J. Stuhler (2017). Labor supply shocks, native wages, and the adjustment of local employment. *The Quarterly Journal of Economics 132*(1), 435.

Easterlin, R. (1961). The american baby boom in historical perspective. *The American Economic Review 51*(5), 869–911.

Eisenhauer, P., J. J. Heckman, and E. Vytlacil (2015). The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy 123*(2), 413–443.

Ertl, B., S. Luttenberger, and M. Paechter (2017). The impact of gender stereotypes on the self-concept of female students in stem subjects with an under-representation of females. *Frontiers in Psychology 8*.

Fernández, R. and A. Fogli (2006). Fertility: The role of culture and family experience. *Journal of the European Economic Association 4*(2-3), 552–561.

Ferrer-Esteban, G. (2011). Beyond the traditional territorial divide in the italian education system. effects of system management factors on performance in lower secondary school. Technical report.

Figlio, D., K. Karbownik, J. Roth, M. Wasserman, et al. (2016). Family disadvantage and the gender gap in behavioral and educational outcomes. Technical report, National Bureau of Economic Research.

Foerster, R. F. (1924). *The Italian emigration of our times*, Volume 20. Harvard University Press.

Foged, M. and G. Peri (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics 8*(2), 1–34.

Fryer, R. G., S. D. Levitt, and J. A. List (2008). Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study. *The American Economic Review 98*(2), 370–375.

Fryer Jr, R. G. (2016). The production of human capital in developed countries: Evidence from 196 randomized field experiments. Technical report, National Bureau of Economic Research.

Fryer Jr, R. G. and S. D. Levitt (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics 86*(2), 447–464.

Fryer Jr, R. G. and S. D. Levitt (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 210–240.

Furtado, D. (2016). Fertility responses of high-skilled native women to immigrant inflows. *Demography 53*(1), 27–53.

Furtado, D. and H. Hock (2010). Low skilled immigration and work-fertility tradeoffs among high skilled us natives. *The American Economic Review 100*(2), 224–228.

Galor, O. and D. N. Weil (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *The American Economic Review*, 806–828.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics 34*(2), 509–543.

Genicot, G. and D. Ray (2014). Aspirations and inequality. *NBER Working Paper*.

Giustinelli, P. (2016). Group decision making with uncertain outcomes: Unpacking child–parent choice of the high school track. *International Economic Review 57*(2), 573–602.

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

Goldin, C. (1990). *The gender gap: An economic history of American women*. New York: Cambridge University Press.

Goldin, C. (1994). The political economy of immigration restriction in the united states, 1890 to 1921. In *The regulated economy: A historical approach to political economy*, pp. 223–258. University of Chicago Press.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *The American Economic Review*.

Goldin, C., L. F. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives 20*(4), 133–156.

Goldin, C. D. and L. F. Katz (2009). *The race between education and technology*. Harvard University Press.

Goodman, S. (2016). Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics 98*(4), 671–684.

Goux, D., M. Gurgand, and E. Maurin (2017). Adjusting your dreams? the effect of school and peers on dropout behaviour. *Economic Journal 127*(602), 1025–1046.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology 74*(6), 1464.

Greenwald, A. G., B. A. Nosek, and M. R. Banaji (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., T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji (2009). Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of personality and social psychology 97*(1), 17.

Große, N. D. and G. Riener (2010). Explaining gender differences in competitiveness: gender-task stereotypes. *Jena economic research papers*.

Guinnane, T. W., C. M. Moehling, and C. Ó. Gráda (2006). The fertility of the irish in the united states in 1910. *Explorations in Economic History 43*(3), 465–485.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Culture, gender, and math. *Science 320*(5880), 1164–1165.

Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *The Journal of Economic Perspectives 20*(2), 23–48.

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

Guyon, N. and E. Huillery (2016). Biased aspirations and social inequality at school: Evidence from french teenagers. *LIEPP Working Paper*.

Guyon, N., E. Maurin, and S. McNally (2012). The effect of tracking students by ability into different schools a natural experiment. *Journal of Human Resources 47*(3), 684–721.

Halla, M., A. F. Wagner, and J. Zweimüller (2017). Immigration and voting for the far right. *Journal of the European Economic Association*, jvw003.

Hatton, T. J. and J. G. Williamson (1998). *The age of mass migration: Causes and economic impact*. Oxford University Press on Demand.

Hatton, T. J. and J. G. Williamson (2006). International migration in the long run: Positive selection, negative selection, and policy. In *Labor Mobility and the World Economy*, pp. 1–31. Springer.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–86.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2017). Returns to education: The causal effects of education on earnings, health and smoking. *Journal of Political Economy*.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour economics 19*(4), 451–464.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010). The rate of return to the highscope perry preschool program. *Journal of public Economics 94*(1), 114–128.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Higgs, R. (1971). *The transformation of the American economy, 1865-1914: An essay in interpretation.* Wiley New York.

Higham, J. (1955). *Strangers in the land: Patterns of American nativism, 1860-1925.* Rutgers University Press.

Hoxby, C. M. and C. Avery (2012). The missing" one-offs": The hidden supply of high-achieving, low income students. *NBER Working Paper*.

Hoxby, C. M. and S. Turner (2015). What high-achieving low-income students know about college. *The American Economic Review 105*(5), 514–517.

Hyde, J. S., E. Fennema, and S. J. Lamon (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin 107*(2), 139.

Jackson, J. F. and J. L. Moore III (2008). The african american male crisis in education: A popular media infatuation or needed public policy response. *American Behavioral Scientist 51*(7), 847–853.

Jones, M. A. (1992). *American immigration.* University of Chicago Press.

Kautz, T., J. J. Heckman, R. Diris, B. Ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *Report prepared for the Organisation of Economic Co-operation and Development, Paris, http://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non- Cognitive-Skills-to-Promote-Lifetime-Success.pdf*.

Kearney, M. S. and R. Wilson (2017). Male earnings, marriageable men, and nonmarital fertility: Evidence from the fracking boom. *National Bureau of Economic Research*.

Keeling, D. (1999). The transportation revolution and transatlantic migration, 1850-1914. *Research in Economic History 19*, 39–74.

Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology 141*(2), 165–173.

Kennedy, P. J. F. (1964). *A nation of immigrants.* New York: Harper & Row Publishers.

Kiefer, A. K. and D. Sekaquaptewa (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology 43*(5), 825–832.

Kugler, A. D. et al. (2017). Choice of majors: Are women really different from men? Technical report, CEPR Discussion Papers.

Lafortune, J. and J. Tessada (2014). Smooth (er) landing? the role of networks in the location and occupational choice of immigrants. *Mimeo, Instituto de Economía, Pontificia Universidad Católica de Chile September 2013*.

Lane, K. A., M. R. Banaji, B. A. Nosek, and A. G. Greenwald (2007). Understanding and using the implicit association test: Iv. *Implicit measures of attitudes*, 59–102.

Lavy, V. and R. Megalokonomou (2017). Persistency in teachers' grading biases and effect on longer term outcomes: University admission exams and choice of field of study. *Working Paper*.

Lavy, V. and E. Sand (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases. *NBER Working Paper* (w20909).

Lent, R. W., S. D. Brown, and G. Hackett (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of vocational behavior 45*(1), 79–122.

Leonard, T. C. (2005). Retrospectives: Eugenics and economics in the progressive era. *The Journal of Economic Perspectives 19*(4), 207–224.

Leonard, T. C. (2016). *Illiberal reformers: race, eugenics, and American economics in the Progressive Era.* Princeton University Press.

Lopez, N. (2003). *Hopeful girls, troubled boys: Race and gender disparity in urban education.* Psychology Press.

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel (2015). Understanding ethnic identity in africa: Evidence from the implicit association test (iat). *American Economic Review 105*(5), 340–45.

Lundberg, S., R. A. Pollak, and J. Stearns (2016). Family inequality: Diverging patterns in marriage, cohabitation, and childbearing. *The Journal of Economic Perspectives 30*(2), 79–101.

Malamud, O. and C. Pop-Eleches (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics 95*(11), 1538–1549.

Martins, P. S. (2010). Can targeted, non-cognitive skills programs improve achievement? evidence from epis. *IZA Working Paper*.

Mayda, A. M., G. Peri, and W. Steingress (2016). Immigration to the us: A problem for the republicans or the democrats? *National Bureau of Economic Research Working Papers*.

McConnell, A. R. and J. M. Leibold (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of experimental Social psychology 37*(5), 435–442.

Meyer, S. (1981). *The five dollar day: Labor management and social control in the Ford Motor Company, 1908-1921.* State University of New York Press.

Mookherjee, D., D. Ray, and S. Napel (2010). Aspirations, segregation, and occupational choice. *Journal of the European Economic Association 8*(1), 139–168.

Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in math test scores: The role of competition. *The Journal of Economic Perspectives 24*(2), 129–144.

Nosek, B. A., M. R. Banaji, and A. G. Greenwald (2002). Math= male, me= female, therefore math≠ me. *Journal of personality and social psychology 83*(1), 44.

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

Nosek, B. A., F. L. Smyth, N. Sriram, N. M. Lindner, T. Devos, A. Ayala, Y. Bar-Anan, R. Bergh, H. Cai, K. Gonsalkorale, et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences 106*(26), 10593–10597.

OECD (2013). PISA 2012 results: What Makes Schools Successful? Resources, Policies and Practices, (Volume IV). *OECD Publishing http://dx.doi.org/10.1787/9789264201156-en*.

OECD (2014). Are boys and girls equally prepared for life?

OECD (2015a). The abc of gender equality in education: Aptitude, behaviour, confidence.

OECD (2015b). Immigrant students at school- easing the journey towards integration. *OECD Publishing http://dx.doi.org/10.1787/9789264249509-en*.

Patacchini, E. and Y. Zenou (2016). Racial identity and education in social networks. *Social Networks 44*, 85–94.

Pop-Eleches, C. and M. Urquiola (2013). Going to a better school: Effects and behavioral responses. *The American Economic Review 103*(4), 1289–1324.

Porter, E. (2017). Can immigration hurt the economy? an old prejudice returns. The New York Times, February 14, 2017.

Rapallini, C. and A. Rustichini (2016). Elective affinities matter as much as ethnicity in multi-ethnic schools. *Journal of Economic Behavior & Organization 131*, 243–262.

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

Reuben, E., M. Wiswall, and B. Zafar (2015). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*.

Riegle-Crumb, C. and M. Humphries (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society 26*(2), 290–322.

Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the united states. *American Economic Journal: Applied Economics 4*(4), 121–139.

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

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

Rudman, L. A., A. G. Greenwald, and D. E. McGhee (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin 27*(9), 1164–1178.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015). Integrated public use microdata series: Version 6.0 [dataset]. *Minneapolis: University of Minnesota*.

Sacerdote, B. et al. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education 3*(3), 249–277.

Sadker, M. and D. Sadker (2010). *Failing at fairness: How America's schools cheat girls*. Simon and Schuster.

Sequeira, S., N. Nunn, and N. Qian (2017). Migrants and the making of america: The short-and long-run effects of immigration during the age of mass migration. *National Bureau of Economic Research*.

Spencer, S. J., C. M. Steele, and D. M. Quinn (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology 35*(1), 4–28.

Spolaore, E. and R. Wacziarg (2016). Fertility and modernity. *UCLA CCPR Population Working Papers*.

Steele, C. M. and J. Aronson (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology 69*(5), 797.

Stevenson, B. and J. Wolfers (2007). Marriage and divorce: Changes and their driving forces. *The Journal of Economic Perspectives 21*(2), 27–52.

Stuart, B. A. and E. J. Taylor (2016). Social interactions and location decisions: Evidence from us mass migration. *Unpublished working paper*.

Tabellini, M. (2017). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *Working Paper*.

Terrier, C. (2015). Giving a little help to girls? evidence on grade discrimination and its effect on students' achievement. *Working Paper*.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association.

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in mathematics 50*(1), 49–62.

Walker, F. A. (1899). *Discussions in economics and statistics*, Volume 2. H. Holt.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, Volume 279. John Wiley & Sons.

Wilson, J. W. (1996). When work disappears: The world of the new urban poor. *New York: Alfred A Knopf*.

Wilson, W. (1987). The truly disadvantaged: The inner city, the underclass, and public policy. chicago: Univ. *Press, Chicago*.

Zelnik, M. (1959). *Estimates of Annual Births and Birth Rates for the White Population of the United States from 1855 to 1934.* Ph. D. thesis, Princeton University.