Helsinki GSE

HELSINKI GSE DISCUSSION PAPERS 22 · 2024

The extent and consequences of teacher biases against immigrants

Ellen Sahlström
Mikko Silliman

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HANKEN

Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 22 · 2024

Ellen Sahlström and Mikko Silliman:
The extent and consequences of teacher biases against immigrants

ISBN 978-952-7543-21-4 (PDF)
ISSN 2954-1492

Helsinki GSE Discussion Papers:
https://www.helsinkigse.fi/discussion-papers

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, April 2024

The extent and consequences of teacher biases against immigrants

Ellen Sahlström, *Aalto University*
Mikko Silliman, *Norwegian School of Economics*[*]

March 11, 2024

Abstract

We study the extent and consequences of biases against immigrants exhibited by high school teachers in Finland. Compared to native students, immigrant students receive 0.06 standard deviation units lower scores from teachers than from blind graders. This effect is almost entirely driven by grading penalties incurred by high-performing immigrant students and is largest in subjects where teachers have more discretion in grading. While teacher-assigned grades on the matriculation exam are not used for tertiary enrollment decisions, we show that immigrant students who attend schools with biased teachers are less likely to continue to higher education.

Keywords: immigrants, discrimination, teachers, education policy
JEL Classification: I24, J15, J68

[*]Ellen Sahlström – ellen.sahlstrom@aalto.fi; and Mikko Silliman – mikko.silliman@gmail.com. Mikko is an affliate at CESifo and IZA. Ellen is grateful to funding from the Society of Swedish Literature in Finland (SLS) and Yrjö Jahnsson Foundation (YJS). We are thankful to the Matriculation Exam Board for access to the data and to Alex Hellsten, Outi Talvitie, and Tiina Tähkä for helping answer our questions about institutional details. We thank seminar participants at ETLA Economic Research, the Helsinki Graduate School of Economics, and Uppsala University, as well as Mark Chin, Christopher Cleveland, Peter Fredriksson, Ramin Izadi, Katrine Løken, Akshay Moorthy, Fritjof Sahlström, Matti Sarvimäki, Kirsten Slungaard Mumma, Erik Sørensen, Hanna Virtanen, Alexander Willen, and Björn Öckert for comments. All errors are our own.

Across nearly all European countries, immigrants receive lower grades and are less likely to attend higher education than their native peers (OECD, 2018a). There may be several reasons for these gaps, including differences in aspirations (Carlana et al., 2022), an unequal distribution of educational resources (Jackson and Mackevicius, 2021), or discrimination (Alan et al., 2023; Alesina et al., 2024). Nonetheless, as education often paves the way for economic well-being (Card, 1999; Zimmerman, 2014; Bhuller et al., 2017) as well as civic and political participation (Dee, 2004; Briole et al., 2022; Chetty et al., 2023), understanding and closing these gaps is an urgent political priority (European Commission, 2020).

We study the extent that teachers exhibit biases against immigrants and if such biases have long-term consequences. To measure teacher biases, we focus on an exceptionally clean natural experiment from Finland's digital high school exit exams. To provide students an initial sense of how they did on the exam, teachers grade student exams before the exams are randomly assigned blind graders who provide a final score determining the official grade. Both teachers and blind "censors" see exactly the same student responses, and teachers and censors are both asked to grade the exams based on a detailed scoring rubric. As in other contexts, our measure of bias is simply the difference between the teacher and censor provided scores. Importantly, while the test we study is high-stakes for students, the grades assigned by a teacher are in-and-of-themselves low-stakes: teacher grades have no direct bearing on high school graduation or enrollment in higher education.

Compared to native students, immigrant students receive 0.06 standard deviations lower scores from their teachers than from blind, external censors. The magnitude of this teacher bias is more than ten times larger than that by gender, and stable to the inclusion of socioeconomic covariates. These effects are driven almost entirely by teacher biases against high-achieving students from immigrant backgrounds. The bias in teacher-assigned scores is greatest in subjects like foreign languages and literature, where teachers have the most discretion in grading, while the mean bias is close to zero in mathematics. Since immigrant students are over-represented in foreign languages, the magnitude of the differences in teacher and censor scores falls when we condition on subject or teacher fixed effects, but exhibits otherwise similar patterns.

When we collapse measures of anti-immigrant bias to the teacher-level, we find that although the median teacher exhibits no bias in grading, there is a longer tail in negative bias, and the mean teacher bias is -0.02 SD. We then measure school-level anti-immigrant bias as the mean level of bias exhibited by teachers at each school. Again, we find that the median school shows little evidence of bias, but that there is a clear set of schools in which teachers consistently exhibit anti-immigrant bias (mean = -0.06 SD). Schools with anti-immigrant biases are primarily smaller schools located in rural areas, but no worse in terms of school quality for native students.

Next, we study whether attending schools with biased teachers shapes the later educational paths of students from immigrant backgrounds. Since we cannot yet observe graduation from secondary education or enrollment in tertiary education for students taking digital exams, we turn to prior cohorts of students. We then estimate the difference in school effectiveness for immigrants and natives, controlling for middle-school achievement and middle-school fixed-effects, as well as a rich array of family background characteristics. To facilitate comparison between these out-of-sample measures of differential school effectiveness and the measures of teacher biases against immigrants based on school-level mean teacher versus blind exam grades, we convert both measures to ranks and study their relationship. Despite measuring different objects, and being measured at different points in time, we see that schools where teachers show biases against immigrants in their grading practices are also the schools which are least effective in supporting immigrant students continue to higher education ($\beta = 0.15$, p-value = 0.027). This suggests that, conditional on ability, an immigrant student who attends a school in the quartile with the highest levels of bias against immigrants is roughly 16 percent less likely to enroll in higher education compared to an immigrant student enrolled in a school with the median level of bias against immigrants.

Our measures of teacher biases against immigrants are likely to be a lower-bound of the total extent of teacher biases students encounter for several reasons. First, our ability to detect biases in teacher grading behavior is likely attenuated in subjects where there is little scope for teacher judgement in assessments. Second, in the case that there is any bias against immigrants in blind grading – perhaps based on grammar – the teacher-censor gap in grading may be attenuated downward. Third, the measure of bias we focus on is from a context where there are almost no incentives for teachers to exhibit their biases.

Additionally, teacher biases are likely to influence students through other channels such as learning or aspirations rather than only grades (Diamond and Persson, 2016; Carlana et al., 2022). In this case, biased teachers may also harm student learning and thereby lower their scores (Bohren et al., 2022). To better understand these channels through which teacher biases affect students, we run a series of simulations where we study how the outcomes of immigrant students change if biases shape learning or aspirations, rather than only grades. We show that without changing application preferences for upper-secondary school tracks (aspirations), removing biased grading practices only marginally reduces the immigrant-native gap in admissions outcomes. Moreover, even a larger boost to student grades, reflecting increased student learning, does little to close the immigrant-native gap. Instead, both removing teacher biases and shifting the secondary school preferences of immigrant students closes immigrant-native gaps in secondary school tracks entirely.

Finally, we examine our estimates of teacher biases through the lens of existing models of discrimination. A simple model of taste-based discrimination (Becker, 1957) does not predict the pattern whereby the teacher-biases are concentrated amongst high-achieving immigrant students. Instead, at first glance, the results might appear to fit better with a model of statistical discrimination, whereby teachers have a prior that immigrant students are lower scoring (Arrow, 1972; Phelps, 1972; Fryer and Jackson, 2008). In this case, they might use immigrant background as a heuristic to help them assign grades faster. However, it turns out that immigrant students are nearly twice as likely as native students to achieve very top scores – suggesting that if teachers are employing statistically discrimination, they are doing so based on an inaccurate mental model (Bohren et al., 2023). That said, since teachers do not show signs of biases across other salient student characteristics, teachers would be employing statistical discrimination in a selective way, emphasizing immigrant background over other relevant dimensions of student background. More nuanced models of taste-based discrimination, whereby teachers have an aversion to seeing minority students in places of power or dislike the way that immigrant students achieve high scores fit our data better.

Our results extend the literature on teacher biases against immigrants by showing that exposure to biased teachers can have long-term consequences for students. In line with our results, existing research typically finds evidence of biases against immigrants in teacher grades. Comparing teacher-assigned grades to standardized test-scores, Burgess and Greaves (2013) find evidence of biases in grading towards students from different ethnic backgrounds amongst teachers in England. Likewise, in Sweden, evidence suggests that teachers assign lower scores to non-native students (Hinnerich et al., 2015). Further, Alan et al. (2023) document teacher biases against immigrants in Turkey, and show that these biases are linked to exclusionary classroom practices. Most recently, evidence from Italy links teacher biases against immigrants in grading to implicit biases, and shows that making teachers aware of their biases can change their grading practices (Alesina et al., 2024). However, not all studies find evidence of bias. A randomized experiment in the Netherlands where teachers were asked to assign grades on writing assignments where the name of the student was randomized to be from a native versus immigrant background showed no evidence of teacher biases against immigrants (Van Ewijk, 2011). Studying high-stakes math tests in Sweden, Diamond and Persson (2016) find no evidence of teacher biases by immigrant status or gender. In another paper from Sweden, Lindahl (2007) finds that teachers can even be more generous towards immigrant rather than native students. Using a nationally representative set of 12,000 high school teachers and the full set of high schools in Finland to study teacher biases against immigrants, we provide evidence that teachers exhibit anti-immigrant biases in their grading practices. Interestingly, we find that these biases are concentrated amongst high-achieving immigrant students – not those who might behave badly in class. Moreover, we show that teacher biases against immigrants are linked to reduced probabilities that students from immigrant backgrounds attend higher education.

We also contribute to the broader literature on teacher biases across other dimensions. One strand of this literature has focused on gender (Lavy, 2008; Hinnerich et al., 2011; Diamond and Persson, 2016; Lindahl, 2016; Lavy and Sand, 2018; Alne and Herstad, 2020; Terrier, 2020; Graetz and Karimi, 2022; Lavy and Megalokonomou, 2024). We find evidence of teacher biases against boys, but the magnitude of this bias is small, less than one tenth of the magnitude of bias against immigrants. Another important strand of this literature has examined race, primarily in the United States – but also in Brazil (Botelho et al., 2015; Chin et al., 2020; Rangel and Shi, 2020; Shi and Zhu, 2023).[1] Other papers study characteristics such as caste (Hanna and Linden, 2009) and idiosyncratic names (Figlio, 2005). Across these various dimensions of bias, only a handful of papers are able to link teacher bias in grading to longer-term outcomes, such as later course enrollment, track-choice, or enrollment in higher education (Lavy and Sand, 2018; Carlana, 2019; Lavy and Megalokonomou, 2024). Our results show evidence of grading biases even when teachers and blind-graders assess the exact same responses to the exact same questions, and that the magnitude of the bias we detect varies by the level of teacher discretion in grading. The pattern of results we find also challenges the notion that teacher biases can be explained by statistical discrimination. Moreover, we show that even when teacher-assigned grades themselves have no stakes attached to them, teacher biases can shape students' later educational trajectories, potentially by shaping student aspirations.

1 Institutions and data

Several institutional features make Finland a particularly interesting setting to study teachers biases against immigrants. First, while the Finnish education system has been internationally recognized for its lack of inequalities – with these studies often pointing to the role of teachers (Sahlberg, 2021) – there are large gaps between immigrants and natives in various life outcomes including education (e.g. Ansala et al., 2020). Second, Finland has experienced rapid increases in the rate of immigration, resulting in a system whereby the vast majority of teachers in the country had little or no exposure to immigrant peers during their own schooling. Third, the two-step grading process of the matriculation exam provides an exceptionally clean setting to identify teacher biases. Fourth and finally, linked administrative registries allow us to study the long-term effects of teacher bias, by linking together data spanning middle school, high school, and higher education.

1.1 Immigrants and their trajectories in Finnish schools

In 1990, less than one percent of people between the ages of 15 and 19 were considered as coming from an immigrant background (Figure 1).[2] That number has grown considerably, and today almost 10 percent of children aged 15-19 are from immigrant backgrounds. Since 2008, the number of children born to immigrants in Finland has also begun to grow. This migration is almost entirely from non-western countries (Figure A.1).

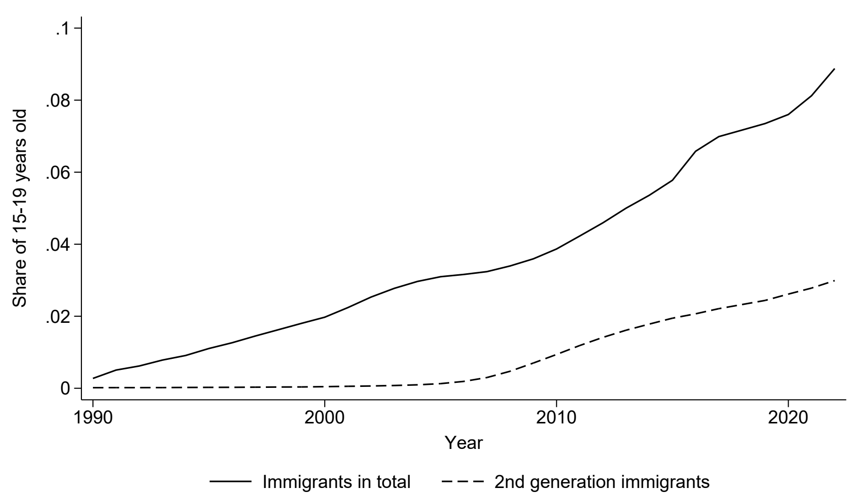
[1] A separate but related literature also documents positive effects of teachers and students sharing the same race (Dee, 2005; Lindsay and Hart, 2017; Scherer et al., 2021; Gershenson et al., 2022).

[2] Statistics Finland defines a person as having an immigrant background if all known parents hold a foreign nationality, regardless of whether they themselves are born in Finland or abroad.

While immigrant-native gaps exist across the OECD, the magnitude of these gaps is large compared to otherwise similar countries (like Sweden, see Ansala et al., 2022). In fact, while more than nine in ten immigrant students in Sweden enroll in tertiary education, less than half of those in Finland do (Figure A.2). Moreover, the gap between immigrants and natives in Finland extends beyond education to a wide range of outcomes including labor market success as well as health (Sarvimäki, 2017; Ansala et al., 2020; Busk and Jauhiainen, 2022).

There may be several explanations for the magnitude of immigrant-native gaps in Finland. Immigrant students may attend lower-resourced schools (Jackson and Mackevicius, 2021), they may hold lower aspirations than native students (Alesina et al., 2024), and they may experience more bullying or discrimination from their peers (Strohmeier et al., 2011; Zacheus et al., 2019; Nshom and Croucher, 2017). An additional explanation is that teachers may hold prejudices or biases against immigrants or simply lack the cultural competencies necessarily for effectively teaching to a diverse student body – having had almost no immigrant classmates themselves given the recency of Finland's immigration history (Acquah et al., 2015).[3] Nonetheless, and while some larger cities in Finland have begun to provide extra resources to support schools serving large immigrant populations (Bernelius, 2013; Silliman, 2017), gaps between immigrant and native students in Finland persist.

Figure 1: The share of 15-19 year old's from immigrant backgrounds in Finland over time



Notes: Figure 1 shows the share of population in age group 15-19 with immigrant background, following the definition of Statistics Finland (i.e. all known parents born abroad). Second generation immigrants are born in Finland to immigrant parents, unlike first generation immigrants who are born abroad. Source: OECD (2018b).

[3]There is also some evidence of biases against immigrants in the labor market (Jaakkola, 2000; Ahmad, 2020).

1.2 The education system and Matriculation Exam

In Finland, students attend compulsory education through grade nine, after which most students apply to upper secondary school, here referred to as high school (see Figure A.3 for an overview of the education system).[4] High school programs are divided into two tracks of roughly equivalent size: the general (academic) and vocational tracks. The focus of our paper is on the general track, typically intended to prepare students for further studies in higher education. To graduate from the general track of high school, students are required to take and pass a number of national subject-specific exams. Grades on this high school exit exam (formally the Finnish Matriculation Exam), are also a central component of admission to most programs in tertiary education.

The Finnish Matriculation Exam consists of separate tests in different subjects, with graduating requiring passing grades in four tests. Literature is mandatory,[5] and three additional tests are chosen from at least three of the four groups (1) mathematics, (2) foreign language, (3) second national language, and (4) humanities and natural science (see Table A.1 for exact subjects). Students may take more than four tests, but have to define upon registration whether the test is one of the four mandatory tests in which a passing grade is required. Examinations are organized twice per year, in September and March, with a total of nine days of exams in different subjects spread out over three weeks. The timing of the test is nationally standardized, such that all students taking the exam in a specific subject take the test simultaneously in their own schools. Students can take exams that count towards their final matriculation exam certificate in a maximum of three subsequent examinations (ex. spring 2018, fall 2018, spring 2019), after which they must graduate. The first exams are often taken already during the second year of high school, followed by exams in the fall of the third year and the final exams in the spring before graduation the third year.

Each test is graded twice, first by the student's teacher and second by a randomly assigned external grader ("censor"). This dual grading of the exams will provide us the basis for estimating teacher biases. The teacher sees the full name of the student while grading, and will most likely know the student personally, particularly in small high schools.[6] Teachers assign preliminary scores based on scoring rubrics. This first step is required by law and provides the students with a preliminary evaluation of their performance shortly after the test. Tests are then randomly assigned to external graders who are not allowed to have any connection to students in any of the schools they are assigned to as graders. This censor-assigned score determines the final grade.[7]

[4]For the cohorts in our sample, mandatory schooling ended after 9th grade. Beginning from the cohort graduating 9th grade in 2021 (born 2005), an educational reform made schooling mandatory up to age 18 and removed completely the already small costs of attending secondary school by providing learning material for free. This does not directly affect this study.

[5]Literature is tested in Finnish or Swedish depending on school language, with the possibility to take a separate exam for non-native speakers.

[6]An average Finnish high school has 290 enrolled students (Statistics Finland, 2021).

[7]The Finnish matriculation exam is graded on a seven-step scale, where Improbatur (I, numeric grade 0) indicates fail,

Between 2016 and 2019, both the exam and its grading shifted to a digital environment. Since then, students' names are visible only to the teachers. Censors see the preliminary grade given by the teacher, but no names of student or school. An example of the grading system is shown in Figure A.4. Differences between teacher- and censor-assigned grades provide us with an exceptionally clean setting to study the extent of teacher biases.

Censors are chosen from highly qualified teachers at high schools and universities. Before the random assignment, each censor reports how many exams they are willing to grade. This is then taken into account in randomization, together with exam language (Finnish or Swedish), school, and subject, such that each censor grades a for them reasonable number of tests in their own field and in their native language. In most cases, all answers from one school and subject end up with the same censor. In small subjects, one censor might grade all or almost all tests taken. Teacher grades do not matter for the final score, but deviations from the teacher score with more than 25 points are confirmed by an additional second censor. Students see their teacher-assigned grades close after the exam, whereas the final grades are published mid-November and mid-May respectively for Fall and Spring examinations, a delay of roughly two months.

1.3 Data and sample

We merge data on the Matriculation Exam to data from administrative registries containing information on student background characteristics and educational outcomes.

Our main source of data is from the Finnish Matriculation Examination board (YTL, 2022). This data contains teacher and censor grades on all manually graded items from all digital exams. The raw exam data is at the item level. We aggregate this to the test level. To account for the fact that different exams are on different scales, and exams in different subjects contain different portions of manually graded questions, we standardize the raw scores from the manually graded portions of each exam to have a standard deviation of one and a mean of zero. This data-set contains student identifiers which we can link to Finnish administrative registries. Unfortunately, the identifiers for teachers and censors were not preserved in a format which we can link to other data. This data is available for students taking exams between 2016 and 2022.

Data on student background comes from an extended version of the population-wide FOLK module provided by Statistics Finland (2020d; 2020e). This data-set contains information on background characteristics necessary for characterizing students as immigrants and natives. We define immigrants as individuals whose all known parents are born abroad, regardless of child birth country.[8] These registers also contain additional demographic variables which we use to

followed by Approbatur (A=2), Lubenter approbatur (B=3), Cum laude approbatur (C=4), Magna cum laude approbatur (M=5), Eximia cum laude approbatur (E=6), and Laudatur (L=7).

[8]This definition of immigrants follows Statistics Finland's definition of what they call "foreign background".

construct variables for age, gender, municipality of residence, parental education and family income. Enrollment and degree are provided in the student and degree modules at Statistics Finland (2020a; 2020b).

Another source of data is the Secondary School Application Registry maintained by the Finnish National Board of Education (Statistics Finland, 2020c). This data-set contains information on middle school grades, high school application preferences, and admissions outcomes for students in their final year of compulsory education.

We conduct our analysis in two samples. First, to study the extent of teacher biases we are limited to the digital Matriculation Exam data, which is only available for cohorts born between 1999 and 2002. These students graduate middle school between 2015 and 2018. This sample contains 110,486 students and 388 schools. This is also the only sample in which we are able to identify teachers at all, as the only available teacher identifiers come from the Matriculation Exam data. We observe 12,044 teachers and 527 censors, grading a total of approximately 550,000 unique tests.

To study the longer-term consequences of teacher biases, we turn to prior cohorts of students, who we can follow through enrollment in higher education. This sample is restricted to all students graduating from middle school (9th grade) between the years 2011 and 2015. These individuals are born approximately during the years 1995 and 1999. This restriction leaves us with 146,306 students, close to half the cohort, in 440 schools. Digital exam data is not available for these cohorts, but we aggregate measures of teacher biases to the school level to study their longer-term consequences on educational outcomes.

Both samples include observations from all academic high schools in the population, with a decrease in number of schools due to school closures during the 2010's, when 50 schools were closed in the years between 2011 and 2016.

1.4 Descriptives

Table 1 shows the background characteristics of students in both of our estimation samples and for immigrants and natives separately. Immigrants come from families with lower income and less educated parents. Immigrants are also more likely to live in the in the capital region and other urban areas. Besides the small growth in the immigrant population over time, other differences between samples are small. While all Finnish high schools are included in the data, we can only estimate biases

Compared to natives, immigrant students also have lower middle school grades, and are less likely to continue to the academic track of high school, or enroll in tertiary education (Figure 2). Immigrant-native gaps in tertiary enrollment persist, even conditional on middle school grades (Figure 2c).

In the academic track of high school, immigrant and native students choose to take different subjects in the Matriculation Exam (Figure A.5). Immigrant students take fewer exams than native students. The only subjects where they are not under-represented are foreign languages – excluding English. The gap in exam participation in science subjects is also relatively small. In addition to taking fewer tests than natives, immigrant students also score considerably worse.

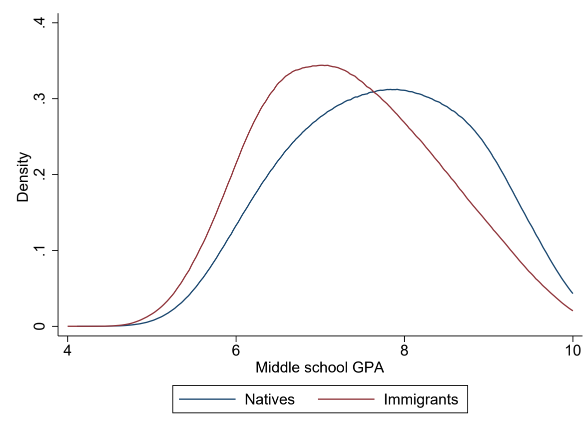
Table 1: Background characteristics of individuals in research sample

	Full cohorts 1995-1999		Academic track 1995-1999		Academic track, digital exam 1999-2002	
	Natives (SD)	Immigrants (SD)	Natives (SD)	Immigrants (SD)	Natives (SD)	Immigrants (SD)
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.509	0.510	0.425	0.437	0.415	0.404
	(0.500)	(0.500)	(0.494)	(0.496)	(0.493)	(0.491)
Family income (rank)	0.511	0.167	0.593	0.219	0.612	0.262
	(0.281)	(0.217)	(0.278)	(0.253)	(0.266)	(0.250)
College-educated mother	0.248	0.146	0.357	0.218	0.434	0.241
	(0.432)	(0.353)	(0.479)	(0.413)	(0.496)	(0.428)
College-educated father	0.214	0.109	0.324	0.173	0.363	0.180
	(0.410)	(0.311)	(0.468)	(0.378)	(0.481)	(0.384)
Native language Finnish	0.941	0.121	0.932	0.123	0.925	0.097
	(0.235)	(0.327)	(0.251)	(0.328)	(0.264)	(0.296)
Native language Swedish	0.054	0.013	0.061	0.015	0.067	0.022
	(0.226)	(0.115)	(0.239)	(0.122)	(0.251)	(0.147)
Capital region	0.247	0.518	0.293	0.557	0.297	0.563
	(0.431)	(0.500)	(0.455)	(0.497)	(0.457)	(0.496)
Urban municipality	0.667	0.893	0.705	0.904	0.719	0.896
	(0.471)	(0.309)	(0.456)	(0.294)	(0.449)	(0.305)
Semi-urban municipality	0.183	0.054	0.163	0.047	0.160	0.054
	(0.386)	(0.227)	(0.369)	(0.211)	(0.366)	(0.226)
Rural municipality	0.150	0.052	0.132	0.049	0.121	0.050
	(0.357)	(0.222)	(0.338)	(0.215)	(0.326)	(0.218)
Born abroad	0.007	0.621	0.008	0.564	0.012	0.541
	(0.086)	(0.485)	(0.090)	(0.496)	(0.107)	(0.498)
Students	279,246	13,294	147,440	6,372	105,450	5,036
High schools			440		388	

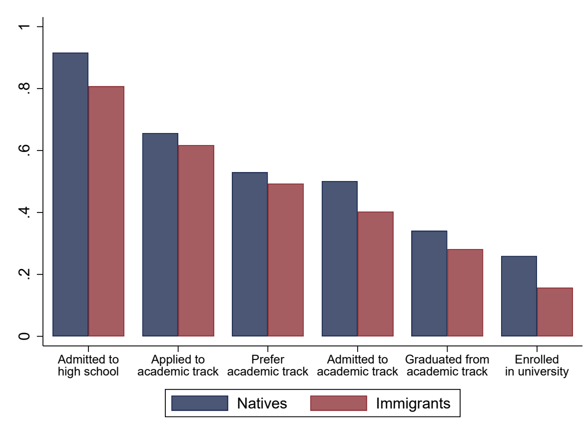
Notes: Mean background characteristics for the full relevant cohorts and the two samples used: the *long-term sample* consisting of academic track students from middle school cohorts 2011-2015 used in the estimation of effects of bias and the *digital-exam sample* used for defining teacher and school bias. Standard deviations shown in parentheses.

Figure 2: Immigrant-native gaps from middle school to tertiary enrollment

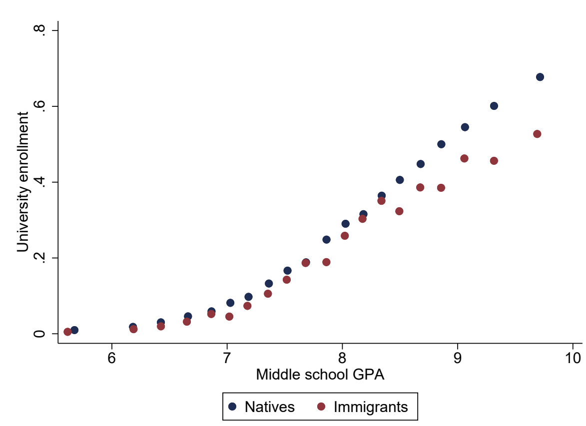
(a) Middle school GPA distribution



(b) Secondary school progression



(c) Tertiary enrollment



Notes: Middle school GPA in Figure (a) as raw average of grades in academic subjects, for cohorts included in the long-term sample. Figure (b) displays cohort share of natives and immigrants respectively at different stages of adolescent education. Academic track graduation is measured after 4 years of high school. University enrollment includes enrollment in universities of applied sciences and is measured 4 years after graduating middle school, one year after expected enrollment following the suggested duration of secondary education. Lastly, in Figure (c), we show that even conditional on GPA, immigrants are less likely to enroll in university.

2 Teacher biases against immigrants

We study whether there are differences between immigrants and natives in how grades assigned by teachers vary from those assigned by randomized external blind-graders on identical exam responses. Data from the Finnish high school exit exams allows us to estimate the magnitude of these gaps experienced by students, exhibited by teachers, and whether teachers with different degrees of bias cluster to particular schools. Our results show that compared to natives, the average immigrant receives 0.06 standard deviations lower grades from teachers than from blind-graders. We then show that the magnitude of variation in anti-immigrant bias in grading is similar to that in teacher effectiveness documented by other papers (Chetty et al., 2014), and that there is significant variation in teacher biases at the school level.

2.1 Estimating teacher biases

To estimate discrimination against immigrants through grading practices, we focus on a natural experiment stemming from the structure of the high school exit exams in Finland. High school exit exams in Finland are graded twice, first by teachers, and second by a randomly assigned set of centralized blind graders (termed a censor). In both cases, graders are given a rubric to base their marks on. However, when teachers grade assignments, they see the student's name. In contrast, the blind censors do not (as shown in Figure A.4). The teacher grade is intended to inform the students of how they are likely to fare on the exam before they receive their final grades. The official grade – and the only grade that counts – is the grade from the blind grader. The grades from these exams determine high school graduation, and are often part of the criteria for admission to university. That said, the teacher grades are no-stakes, having no bearing on high school graduation or admissions decisions.

Our measure of bias is based on the difference (γ_i) in teacher (T_i) versus blind (B_i) grades received by students (i).

$$\gamma_i = T_i - B_i \tag{1}$$

When this measure is negative, students receive lower grades from their teachers than from blind graders. In contrast, when this measure is positive, teachers assign higher grades than blind graders.

We then estimate the extent that immigrant students face bias from teachers through the following regression:

$$\gamma_{ijsc} = \alpha + \beta Immigrant_i + v'\mathbf{X}_i + \omega_c + \delta_s + \mu_j + e_{ijsc}. \tag{2}$$

We estimate whether there is a difference in the difference between teacher and blind grades received by immigrant and native students. We control for cohort (c) fixed-effects, and run these estimates with and without a varying set of individual controls ($v'\mathbf{X}_i$). In some specifications we also include fixed-effects for teachers (j) and/or schools (s). The coefficient of interest is β, which we interpret as the mean bias from teachers that immigrants experience. Negative values of β suggest that teachers unfairly penalize immigrant students in grading. In the case that blind graders exhibit biases against immigrants – perhaps based on syntax or grammar – our measures of bias against immigrants should be interpreted as lower-bounds of the true bias experienced by immigrants.

Since students select into different schools and take courses with different teachers, the mean bias experienced by immigrant students may differ from the mean bias teachers possess. Moreover, β from the above equation says little regarding the variation in teacher biases against immigrants: any bias the regression uncovers could be driven by a handful of very biased outliers, or a broader set of teachers with weaker biases.

To estimate the extent that each teacher exhibits anti-immigrant bias (γ_j), we calculate the mean difference in teacher and blind grades between immigrants received by immigrant (m) and native (n) students, for teacher (j) and student (i).

$$\gamma_j = \frac{\sum(T^m_{ij} - B^m_{ij})}{N^m_{ij}} - \frac{\sum(T^n_{ij} - B^n_{ij})}{N^n_{ij}} \tag{3}$$

Then, since γ_j will be estimated more reliably for some teachers compared to others depending on the number of immigrant and native students they teach, we adjust our estimates using empirical Bayes (see, for example, Kane and Staiger (2008) and Chetty et al. (2014)). To do this, we begin by estimating the reliability of each teacher's raw measures of bias.

$$\lambda_{jo} = \frac{\hat{\sigma}^2_{\gamma o}}{\hat{\sigma}^2_{\gamma o} + \frac{\hat{\sigma}^2_{\epsilon}}{EN_j}} \tag{4}$$

In the above equation, $\hat{\sigma}^2_{\gamma}$, is the variance of the immigrant bias, and $\hat{\sigma}^2_{\epsilon}$ is the overall variance in scoring, measured as the variance of scoring errors for natives. To account for the fact that the reliability of our bias measure may depend on the specific subject (o) in question, we sometimes also estimate and use subject-specific reliabilities. The effective sample size (EN) for each teacher is measured as:

$$EN_j = \sqrt{N^2_m + N^2_n} \tag{5}$$

Finally, we use these reliabilities (λ) to adjust our estimates of teacher biases by empirical Bayes by shrinking them towards the mean teacher bias. Since the mean teacher bias is attenuated in some subjects and not others – ex. based on how much leeway teachers might have in grading – we also sometimes shrink our estimates of teacher biases toward the mean bias in a particular subject (o) using subject specific reliabilities. Equation 6 defines our adjusted measure of teacher bias.

$$\tilde{\gamma}_{jo} = E[\gamma_j|o] + \lambda_{jo}(\gamma_{jo} - E[\gamma_j|o]) \tag{6}$$

We then define school-level bias (γ_s) against immigrants as the extent to which each school's teachers possess anti-immigrant biases, calculated as an average of teacher biases in the school. The idea behind this definition is that the extent that a student's educational trajectory is shaped by teacher biases may depend not just on the teachers they take classes with – but also on the teachers whose classes they explicitly avoid. While such an idea would be incorporated into our preferred measure of school bias in any case, we do not have teacher identifiers that go far back in time for when we estimate long-term effects – so we measure bias at the school level. We take care in how we operationalize this measure of school-level bias. First, because raw estimates of teacher bias (γ_j) do not account for measurement error in teacher biases, we adjust these using empirical Bayes ($\tilde{\gamma}_{jo}$), accounting for subject-specific reliabilities. Second, because the distributions of bias estimates of teachers in different subjects have different means – also likely due to the teacher leeway in grading practices – we de-mean the measures of reliability-adjusted estimates of teacher biases from their overall subject means. This adjusted measure gives us a subject-agnostic measure of the mean teacher biases possessed by adults in 322 high schools.

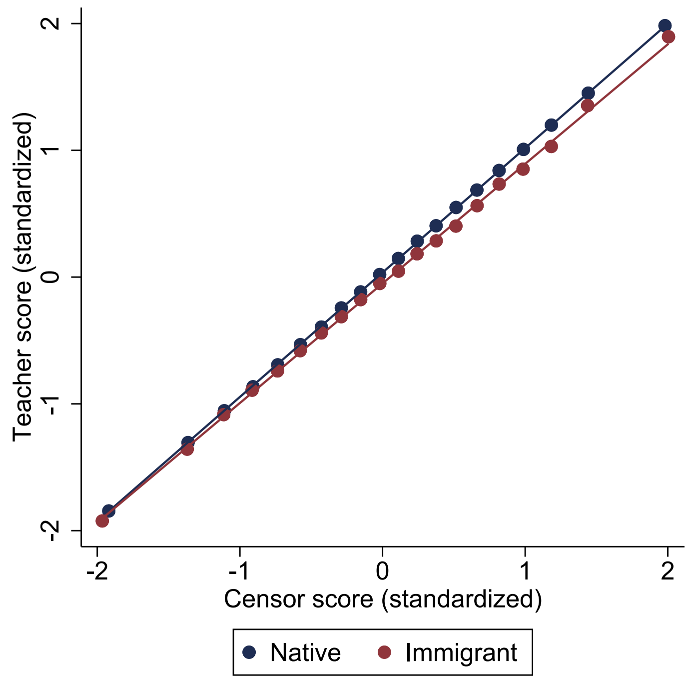
$$\gamma_s = \frac{\sum \tilde{\gamma}_{js}}{N_{js}} \tag{7}$$

The only high schools for which we are missing a measure of school-level bias are those where there were no students from immigrant backgrounds – and we could therefore not estimate teacher biases (see Table A.2 for schools with estimates of bias). Also, because of the rate of school consolidations in the past decades, we do not have measures of school-level biases for small high schools which were operating in our long-term sample, but not in the sample where we estimate biases.

2.2 Anti-immigrant bias experienced by students

As a first step to assess whether teachers exhibit biases against immigrants in their grading practices, we plot scores given by teachers against scores given by blind graders (censors) (Figure 3a). For natives, the teacher and censor scores almost exactly correspond to one another. Compared to native students with the same censor-assigned scores, immigrant students receive lower scores from their teachers. The gap between teacher and censor scores is particularly large amongst high scoring students from immigrant backgrounds.

Figure 3: Differences in teacher versus censor assessments for immigrants and natives separately



Notes: Figure 3 plots the relationship between teacher and censor (blind) assessments for immigrant and native students across the achievement distribution. Teacher scores are anchored in the standardized censor-scores.

We quantify the magnitude of these biases experienced by immigrant students using the regression specification in Equation 2 (Table 2). First, we report results from the simplest version of the regression, one without individual controls or fixed effects for schools or teachers. Compared to native students, students from immigrant backgrounds receive 0.06 standard deviations lower scores from teachers. Adding covariates measuring socioeconomic background suggests that the estimated anti-immigrant bias cannot be explained solely by socioeconomic background of the family. If anything, controlling for background increases the magnitude of the estimate of anti-immigrant bias. The estimate of bias from the simplest specification (Table 2, column (1)) is twelve times larger than our estimate of gender differences in teacher-censor grades – with girls receiving only 0.005 SD units higher scores than boys from their teachers compared to blind censors (Figure A.6) – and ten times larger the differences only by socioeconomic status (Table A.3).

Table 2: Bias against immigrant students in teacher assessments of matriculation exams

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	-0.061***	-0.061***	-0.070***	-0.069***	-0.068***	-0.070***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Male		-0.005***	-0.004***	-0.004***	-0.004***	-0.003***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Family income rank			-0.025***	-0.026***	-0.023***	-0.021***
			(0.002)	(0.002)	(0.002)	(0.002)
Parents unemployed				-0.013***	-0.012***	-0.011***
				(0.004)	(0.004)	(0.004)
Mother below college education					0.005***	0.003***
					(0.001)	(0.001)
Parents without own exam-experience						0.007***
						(0.001)
Student-by-test observations	554,361	554,361	554,361	554361	554,361	554,361
R^2	0.0018	0.0019	0.0023	0.0023	0.0023	0.0024

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

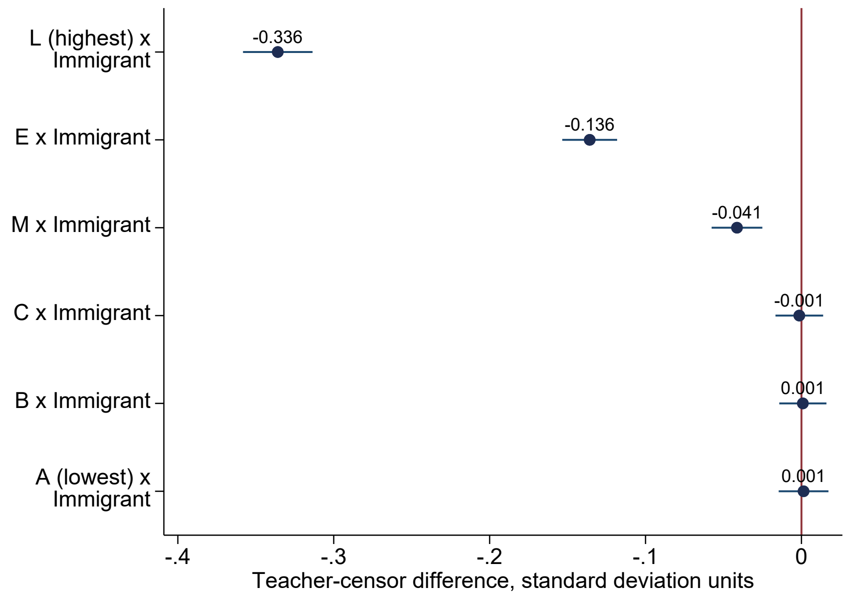
Notes: Results from Equation 2, using a specification including a varying vector of individual level controls. The regressions include cohort fixed effects. Standard errors in parentheses.

The bias against immigrant students is experienced equally by male and female students (Table A.4). Strikingly, however, anti-immigrant bias is concentrated almost exclusively amongst high-performing students from immigrant backgrounds (Figure 4).

Adding school fixed effects to the regression only reduces the immigrant-native gap in teacher-censor grades by about 10 percent (Table A.5). This suggests that the gap in teacher-censor grades is not driven by differences in the schools attended by native and immigrant students. However, limiting comparisons to students who take classes with the same teacher, the extent of bias is reduced by two-thirds, to 0.02 standard deviation units. Conditioning on subject fixed effects, the magnitude of this estimate is reduced still by about half to 0.01 standard deviation units.

In contrast to school fixed-effects which had little effect on the magnitude of our estimates, the large decreases in the magnitude of the mean bias experienced by immigrants when we add teacher or subject fixed effects suggest that a substantial portion of teacher bias experienced by immigrant students is due to differences in the subjects selected by immigrant students. This may not be surprising, as we see large differences in the subject choices of immigrant and native students (Figure A.5). Moreover, we see that immigrants are over-represented only in foreign language courses – the courses where teachers are most likely to exhibit grading bias against immigrants (Figure 5).

Figure 4: Heterogeneity in bias against immigrants by final test grade

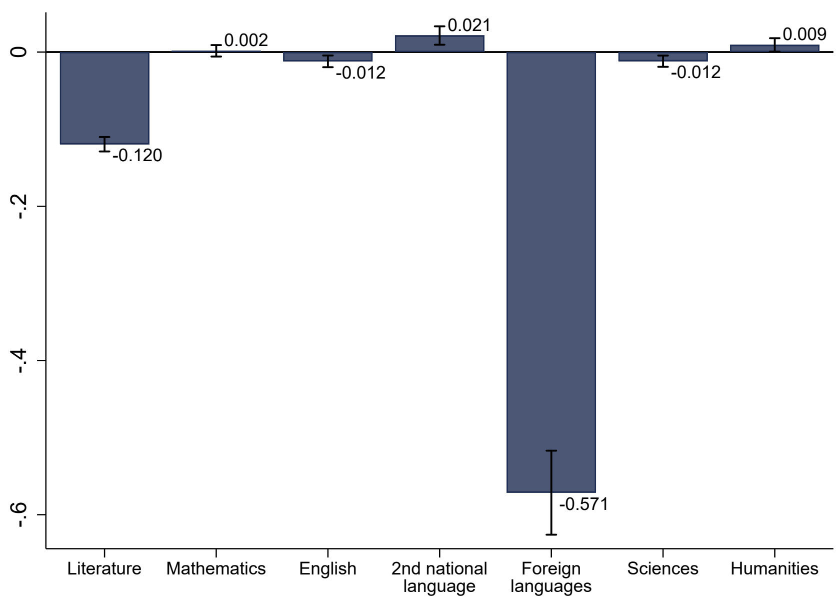


Notes: Figure 4 shows estimates of the teacher bias experienced by immigrant students at different points of the final grade distribution.

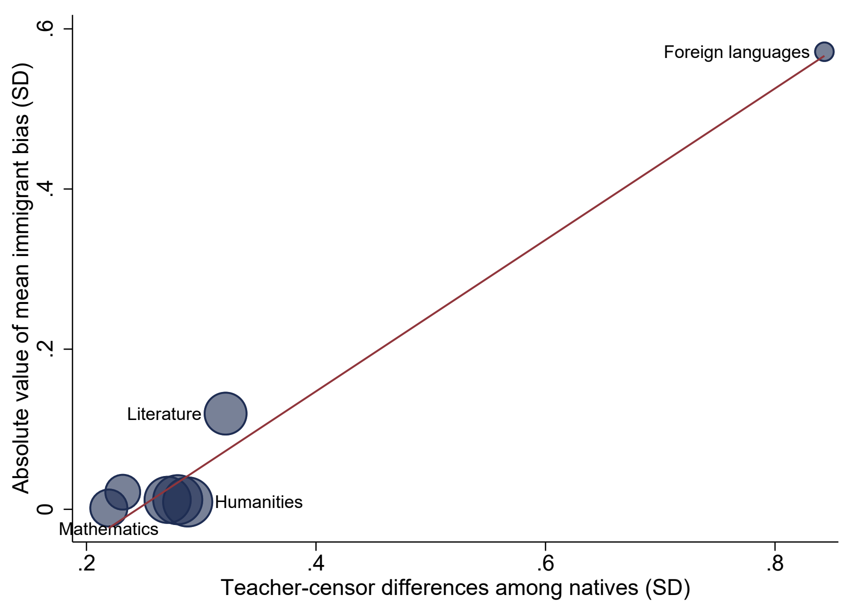
The differences in the estimates of teacher-biases by subject could result from real differences in bias held by teachers teaching different subjects. Alternatively, the differences could simply be due to the extent which certain types of items or tests allow for discrepancies in judgements regarding scoring. For language exams, for example, the teacher-graded component is largely based on a longer response to a prompt as well as a handful of open response questions. In contrast, the math questions require the student to show their steps in how they solve a math problem. We see this in the data, whereby the magnitude of our estimates of teacher biases correspond to the extent of teacher discretion in grading (Figure 5). In addition to the design-features of tests in between subjects, the types of teachers who grade different subjects could also matter. In cases where there is no expert in a particular foreign language in a school, a teacher with less expertise in that foreign language may be asked to grade the exam instead. In such cases, teachers with less domain expertise may assign additional weight to their perception of student ability. This may explain why we see far less evidence of bias in English exams compared to other foreign languages. Both sources of differences between subjects – test design and teacher expertise – may explain why we see stronger biases in some subjects rather than others. Nonetheless, if teachers hold similar prejudices regardless of the subject they teach (or grade), our aggregate estimates of teacher-biases based on grading discrepancies are likely to be significantly underestimated.

Figure 5: Differences in teacher-biases by subject

(a) Mean teacher bias against immigrants by subject group



(b) Teacher biases and teacher discretion in grading



Notes: Figure (a) presents the mean bias against immigrants exhibited by teachers across different subjects, estimated using the main specification in Equation 2. Figure (b) shows how these subject level estimates relate to teacher-censor differences when grading responses of native students. See subject groups in Table A.1.

The difference in the magnitude of the estimates of teacher-bias experienced by immigrant students between regressions with and without teacher effects tells us about the extent to which immigrant students are clustered amongst biased teachers. Perhaps counter-intuitively, the reduction in the magnitude of the estimate when comparing students who take classes with the same teacher suggests that teachers who exhibit the most anti-immigrant bias teach a disproportionate share of immigrant students. Paired with the finding that anti-immigrant bias is largest amongst high-performing students (Figure 4), this finding is consistent with an experience-based model of stereotypes, whereby a teacher's biases towards immigrants are shaped by their interactions with other immigrants (Burgess and Greaves, 2013). If other immigrants who they have interacted with perform poorly in their classes, it may be hard for a teacher to recognize high-performing immigrant students.

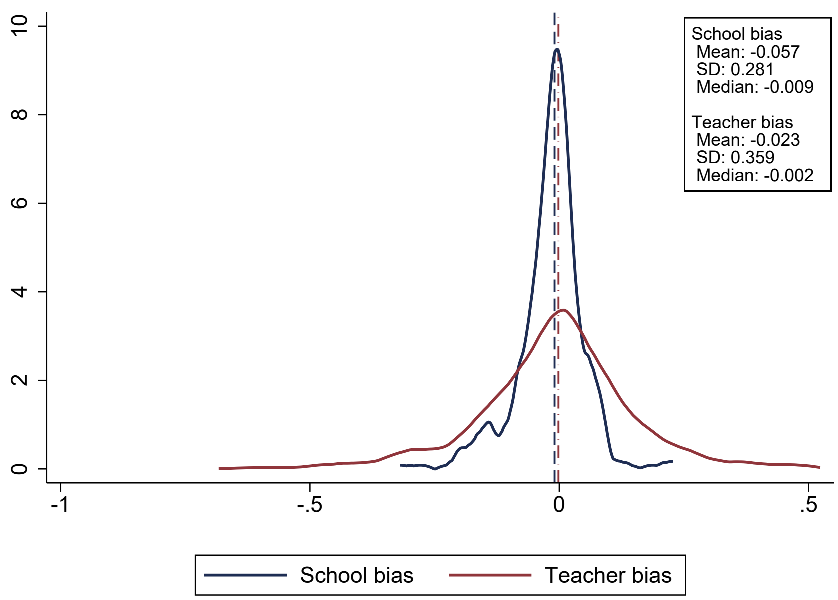
These estimates of teacher-bias experienced by students are based on comparing teacher-assigned grades to censor assigned-grades on identical responses. A potential concern for the validity of using this approach would be if particular teachers were systematically assigned censors of differing strictness. Since censors are randomly assigned schools at the subject level, the lack of a difference in estimates when we add subject versus censor fixed-effects supports the validity of our design (Table A.5).

The fact that teachers and blind-graders are assessing identical responses may also explain the smaller size of our estimates of teacher bias by gender compared to those in the existing literature (Lavy, 2008; Hinnerich et al., 2011; Diamond and Persson, 2016; Lindahl, 2016; Lavy and Sand, 2018; Alne and Herstad, 2020; Terrier, 2020; Graetz and Karimi, 2022; Lavy and Megalokonomou, 2024). As with our estimates of bias against immigrants, these estimates may be a lower-bound of the overall bias by gender. First, it might be that Finnish teachers are in fact less biased than other teachers. Second, since teacher scores are not based on course grades, there is less room for them to exhibit bias. Third, as mentioned earlier, there are few incentives for teachers to express bias. While these same factors would attenuate our estimates of bias against immigrants, the fact that the estimates of bias against immigrants are so much larger than those against boys further underscores the magnitude of teacher biases against immigrants.

2.3 Anti-immigrant bias across teachers and schools

Next, we study teacher-level variation in anti-immigrant biases more closely by plotting the distribution of our estimates of teacher bias ($\tilde{\gamma}_j$) in Figure 6. This figure presents several takeaways. First, the median teacher exhibits no bias against immigrants in their grading practices. Second, there is significant variation in how teachers grade the otherwise identical work of immigrant and native students. Third, there is a long left-tail in the extent of teacher-bias, suggesting that while most teacher exhibit relatively little bias towards immigrants, a handful of teachers drive the gaps between teacher and censor scores received by immigrant versus native students. The reliability of our measures of teacher biases across years is 0.52 – similar in magnitude to the reliability of teachers' gender biases, as estimated by Lavy and Megalokonomou (2024). A small note when interpreting these results is that we can only estimate teacher biases against immigrants for teachers who teach both immigrant and native students – given the limited extent that immigrant students are present in many rural areas in the country, we only have estimates of teacher biases for about five thousand of the twelve thousand teachers in the digital exam sample.

Figure 6: Distribution of school and teacher bias



Notes: Figure 6 presents the distribution empirical Bayes-corrected estimates of school and teacher biases, shrunken towards subject mean. One percent of each tail of the teacher distribution are excluded from the figure, altering the distributions slightly from the full statistics in the box. With the exclusion of these outliers, the mean and median measures of bias are largely unchanged, but the estimates of the standard deviations are more reasonable: 0.07 SD and 0.16 SD for schools and teachers respectively. See Figure A.7 for estimates which are shrunken toward the overall mean, not taking into account variation across subjects. $n_s = 322$, $n_j = 5170$

Last, we study the extent of school-level variation in teacher biases. We define bias at the school-level as the mean bias of teachers we observe working at that school. These estimates are plotted alongside the variation in teacher bias in Figure 6. While the median teacher exhibited almost no bias against immigrants, the median school does exhibit some bias (0.01 standard deviation units) against immigrants, suggesting some degree of clustering amongst teachers with similar biases. That said, there is significant variation in the extent of school-level bias, again with a longer left-tail, suggesting a relatively small group of schools which exhibit the most bias. In these most biased schools, conditional on blind assessments of their ability, immigrant students receive more than 0.1 standard deviations lower grades from their teachers than native students. By and large, these results suggest that the extent of anti-immigrant bias amongst teachers is greatest in small rural schools (Figure A.8). There is, however, no evidence of these schools being worse in terms of value added among natives, suggesting that the bias is not due to "bad teachers". Our measure of value added is described in detail in the following section.

3 Assessing the long-term consequences of teacher bias

In our setting, grades provided by teachers on the high school exit exams have no direct bearing on high school graduation or entry to tertiary education. As such, while they may be concerning in and of themselves, it is not obvious that our estimates of discrepancies in the grading practices of teachers towards immigrant and native students have any real-world consequences. On the other hand, however, biases exhibited in grading practices may capture underlying biases which may shape broader discriminatory behavior in teaching practices. For example, teachers may support the learning of immigrant and native students in differential ways (Gershenson et al., 2022), or teachers biases may influence student aspirations (Alesina et al., 2024).

A central goal of this paper is to understand whether teachers' biases in grading have real world economic costs. To provide evidence informing us of the extent that our measures of teacher biases map to real world consequences, we will compare our estimates of mean school-level teacher biases to differences in school effectiveness for immigrant and native students. Then, based on these estimates we perform back-of-the-envelope calculations to provide a sense of the potential costs of teacher biases on learning outcomes across students' entire educational careers.

3.1 Empirical strategy

If teacher biases extend beyond grading, exposure to biased teachers may harm immigrant students in terms of meaningful economic outcomes such as graduation or enrollment in higher education. A simple comparison of the extent a student is exposed to biased teachers and the student's later outcomes is, however, likely to be biased for several reasons. First, students have scope to choose which teachers they take classes with. Second, students attending a school with more biased teachers may differ from immigrant students attending a school with less biased teachers. Third, overall teacher or school effectiveness may be associated with the extent of teacher biases.

In our context, we are presented with additional data limitations. We do not yet observe long-term outcomes in the administrative data for the cohorts of students who enroll in the digital exams used to estimate teacher biases; and, for the students for whom we have long-term outcomes, we do not have measures of teacher biases.

With these challenges in mind, we study the potential consequences of teacher biases by relating our Matriculation exam-based measures of school-level biases to the extent that schools differentially improve immigrant versus native outcomes. We implement this strategy as follows.

First, we turn to earlier cohorts of students (starting high school between 2011 and 2016) who we can follow through high school graduation and tertiary enrollment. While high school is expected to last three years, we study whether students enroll in tertiary in the fifth year after applying to high school to allow for the fact that most men complete their up to one year of mandatory military service directly after high school.

We then estimate the extent that different high schools systematically affect the outcomes (L) of students from the same middle school and who are similar in terms of their prior academic performance and family background. In the absence of admissions lotteries (Deming et al., 2014; Angrist et al., 2024), we follow the existing literature, and estimate school effectiveness using a value-added model (Jackson et al., 2020). Since we are not making inferences based directly on these regression results themselves, it is simpler to run the regressions separately for immigrants and natives, rather than through a regression where each term is interacted with immigrant status.

$$L_{ics}^{immigrants} = \alpha_s^{immigrants} + \rho_m + \sum_{m=1}^{M} \beta_m \Big(\mathbf{1}[m_i = m] \times GPA_{im} \Big) + \omega_c + v'\mathbf{X}_i + e_{ics} \tag{8}$$

$$L_{ics}^{natives} = \alpha_s^{natives} + \rho_m + \sum_{m=1}^{M} \beta_m \Big(\mathbf{1}[m_i = m] \times GPA_{im} \Big) + \omega_c + v'\mathbf{X}_i + e_{ics} \tag{9}$$

The objects of interest in these regressions are the estimates of high school fixed effects (α_s), from both the immigrant and native sample. Importantly, these estimates of high school fixed effects are based on comparisons of students (i) who attend the same middle-schools (ρ_m) and received similar grades in those schools (GPA_{im}), but who attended different high schools. We also include a detailed vector of controls for family background – including family-income rank and parental education ($v'\mathbf{X}_i$) – as well as cohort fixed-effects (ω_c).

To measure whether schools are more effective for immigrant versus native students, we simply take the difference of the value-added estimates by school for these two groups:

$$\hat{\alpha}_s^{immigrants} - \hat{\alpha}_s^{natives} = \Delta Eff\widehat{ectiv}eness_{VAM} \tag{10}$$

There may be several reasons that schools are more effective for one set of students than another. This measure should not be interpreted as a direct measure of school-level teacher-biases. For example, schools may vary in their effectiveness for immigrant and native students based on the extent of discrimination immigrant students face from other students or differences in the availability of targeted programming focused on supporting immigrant students. However, one of the reasons that schools may be more effective for natives or immigrants is if teachers in that school hold biases against or in favor of one of the two groups.

To assess whether teacher biases might have long-term consequences, we relate our measures of teacher biases based on grading ($\hat{\gamma}_s$) to differences in school effectiveness for immigrants and natives ($\Delta \widehat{Effectiveness}_{VAM}$). Since these two measures are based on different scales, we convert both to ranks before relating them to one another.

$$\Delta \widehat{Effectiveness}_{VAM}^{rank} = \alpha + \beta \hat{\gamma}_s^{rank} + \varepsilon_i \tag{11}$$

Even in the case that teacher biases are the exclusive driver of immigrant-native differences in school effectiveness, both types of biases are estimated with error, and we should expect that their relationship is attenuated downward due to measurement error. While the fact that these two estimates come from entirely different cohorts of students ensures that there is no mechanical bias stemming from having the same set of students in the exam data and in the higher education enrollment data, it is also likely to attenuate the relationship we observe since the temporal distance between the two samples makes the samples distinct in other ways. For example, in that period, the teachers in secondary schools are shuffled as older teachers retire, new teachers are hired, and teachers move between schools. Still, our estimates suggest that teacher biases are relatively stable over time and estimated average-to-year correlations within teachers are in line with previous literature (Staiger and Kane, 2014; Lavy and Megalokonomou, 2024). And, if teacher biases are correlated within schools, school-level biases may be relatively unchanged even with some teacher turnover. Additionally, the number and type of immigrants attending schools has changed somewhat between these two samples.

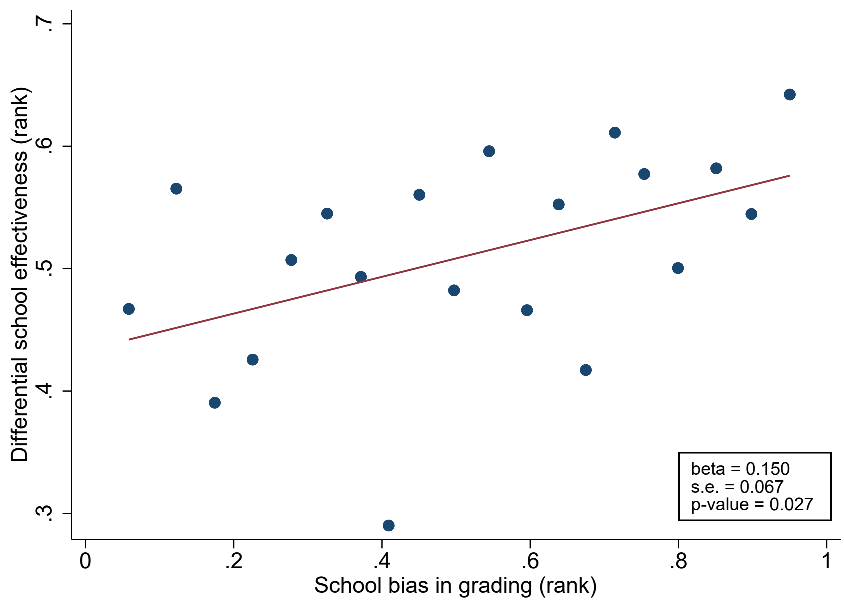
3.2 Are teacher-biases in grading associated with gaps in tertiary enrollment?

We find that the immigrant-native differences in school effectiveness are closely tied to our measures of teacher-biases (Figure 7). Schools that exhibit anti-immigrant bias – as measured through teacher versus blind grades on exams – are also the schools which are least effective in supporting immigrant students enroll in tertiary education ($\beta = 0.15, p = 0.027$). Put another way, an immigrant student is 4 percentage points (27 percent from a baseline of 15 percent) less likely to enroll in tertiary education within five years if they attend a high school within the top 25th percentile of anti-immigrant bias compared to the median high school.

In contrast, we do not detect any relationship between the extent of teacher biases in schools and immigrant graduation rates (Table A.6). This is in line with the observation that teacher biases against immigrant students are experienced almost exclusively by high-performing students – almost all of whom graduate from secondary school on time.

In line with our prior results, these estimates suggest that teacher-biases against immigrants may play an important role in explaining immigrant-native gaps in tertiary enrollment rates amongst otherwise similar high-performing students – but are unlikely to explain differences in secondary school graduation rates for students at the lower end of the academic performance distribution.

Figure 7: Differences in teacher grading bias reflect immigrant-native gaps in long-term value added



Notes: This figure reports the relationship between the school-level measures of bias in teacher end of high school exam grades and differences in school value added in terms of tertiary enrollment for immigrants and natives. The regression line is estimated using the following equation: $\Delta \widehat{Effectiveness}_{VAM} = \alpha + \beta\hat{\gamma}_s + \varepsilon_i$, where each school is weighted by its effective sample size.

3.3 How do teacher-biases shape educational careers?

While the dual grading of digital end of high school exams provides us with an exceptionally clean snapshot of teacher biases in the country – it is unlikely that biases towards immigrants are limited only to high school teachers. We might expect that middle and perhaps also elementary school teachers posses similar biases in grading. Additionally, biases are likely to exhibit themselves through other avenues – such as inhibiting student learning, or limiting student aspirations. The relationship between the teacher biases in low-stakes assessments and tertiary enrollment outcomes supports the idea that teacher bias may manifest itself through other mechanisms than just grading bias (Figure 7). In this sense, the direct measures of bias we estimate are likely to underestimate total bias experienced by immigrant students (Bohren et al., 2022).

To study how teacher biases shape educational trajectories, we perform simulations where we vary the extent that teacher biases affect student grades and student aspirations. We base our simulations around upper-secondary school admissions – a context where we directly observe student grades and preferences for secondary school tracks – and a context where teacher assigned grades are typically the sole factor determining admission to secondary schools (see, for example, Silliman and Virtanen (2022) or Huttunen et al. (2023)). If middle school teachers exhibit anti-immigrant grading biases, the middle school grades of immigrant students are likely to be lower than under objective grading.

We simulate admissions outcomes under scenarios where the grades of immigrant students are increased to better reflect their true ability (Table A.7). In the first scenario, we increase immigrant GPA's by 0.05 units, in the second by 0.10 units, and in the third 0.15 units. The first of these scenarios reflects teacher grading biases against immigrants comparable to those we estimate in secondary schools. However, to capture the idea that teacher grading biases might shift student learning, we include two additional scenarios – boosting the grades of immigrant students slightly more.

The data show that while only 1 percent of native students are not qualified for admission to any of the secondary school programs they apply to, 4 percent of immigrant students have grades which do not qualify for them for any secondary school program. More than a third of this gap goes away when we adjust middle school grades to account for teacher grading biases (Row 1, columns 2 and 3). While only 8 percent of native students have middle school grades which do not qualify them for admission to their first ranked secondary school program, 17 percent of immigrant students have middle school grades below the admissions requirements of their first ranked program. Between 11 and 44 percent of this gap goes away when we adjust for teacher grading biases (Row 2). Still, the third row of the table shows that adjusting middle school grades to account for teacher grading biases does almost nothing to close the gap in the probability that immigrant students would be admitted to the academic track.

If teacher biases against immigrants shift the aspirations of immigrant students, the application preferences of students might be different in the absence of teacher biases. In Panel B of Table A.7, we study whether each student's grades are above the minimum admissions cutoff experienced by students in their middle school applying to academic track upper secondary programs. This simulation suggests that 97 percent of native students would have middle school grades that qualify them for some academic program – had they applied. In contrast, only 92 percent of immigrant students would have been qualified for some academic program. This gap is only half as large as the true gap in admissions to academic track secondary programs. Moreover, combined with even small

GPA boosts for immigrant students, the entire gap in qualification for the academic track disappears.

Together, the results from this simulation suggest that while biases in teacher grading may explain a portion in native-immigrant gaps in post-compulsory educational trajectories, these gaps are largely driven by differences in application preferences rather than only differences in grades.

4 Explaining teacher bias against immigrants

Teachers may exhibit biases against a specific group of students for several reasons. A central area of research in the recent literature on discrimination has been to understand the reasons behind bias and discrimination across contexts (Guryan and Charles, 2013; Bertrand and Duflo, 2017).

Under a model of *taste-based discrimination*, biased teachers treat immigrants differently from natives due to prejudice or dislike (Becker, 1957). At its simplest, a model of taste-based discrimination predicts that teachers who hold biases against immigrants assign lower grades to immigrants than natives who have otherwise similar responses. The pattern of bias we see in our data shows that teachers only exhibit biases in their assessments of high-performing immigrants (Figure 4). This pattern is inconsistent with a model where teachers assign lower grades to members of one group across the board. Interestingly, if low-performing immigrants are also the most disruptive or poorly behaving, then our results are also inconsistent with a model where teachers would under-assess immigrant students who show bad behavior.

Alternatively, under *statistical discrimination*, teachers treat groups differently due to beliefs about group ability based on prior exposure to members of that group (Arrow, 1972; Phelps, 1972). For example, if a teacher is typically exposed to low-scoring immigrant students, they may develop beliefs whereby they associate immigrant background with lower-performance. When scoring student responses, they might use this mental model as a shortcut in their assessments, find it hard to believe that an immigrant student may score high, and assign a lower score to these students than they might deserve (Fryer and Jackson, 2008; Burgess and Greaves, 2013). Providing some evidence in support of this model, Figure A.5b shows that immigrant students are over-represented amongst low-scoring students, and that the average native student scores above the average immigrant student.

Still, where models of statistical discrimination typically assume that there may be efficiency gains from discrimination (Arrow, 1972), teachers do not have any incentives to exhibit bias in grading to shape student outcomes. One possibility is that teachers are using information about immigrant background as a heuristic to speed up the grading process. Moreover, immigrant students are about twice as likely as native students to achieve the very highest scores in the distribution: immigrant students are actually over-represented as high performers (Figure A.5b). This points to the possibility of *inaccurate statistical discrimination*, whereby teachers may make grading decisions based on biased-beliefs regarding immigrant and native abilities – but that these beliefs are not grounded in the true distributions of immigrant and native performance (Bohren et al., 2023). One way to square these two points could be that teacher beliefs regarding immigrant and native performance are accurate – but only within the context of their schools. In line with this, Figure A.8 suggests that teachers with anti-immigrant biases are most concentrated in rural schools serving low-income immigrants.

While these results may point to the role of statistical discrimination in explaining at least some of the teacher-biases we document, there remains a puzzle. Figure A.9 shows that, on average, boys score worse than girls. The magnitude of this difference is 140% of the difference between immigrant and native students. If teachers used group-differences in performance to inform their grading practices across the board, we should expect to see more evidence of teacher biases against boys (Figure A.10). The absence of significant gender biases suggests that either teachers are not discriminating against boys due to inaccurate beliefs about their mean performance – or, that teachers are selectively employing some form of statistical discrimination – by immigrant status, but not gender. The selective application of statistical discrimination by only some salient group characteristics brings back the possibility of taste-based discrimination, perhaps in combination with statistical. When we examine whether teachers exhibit more bias toward specific groups of immigrant students, we see that teacher biases are greatest towards eastern European and Russian immigrants, and if anything slightly smaller towards students from Muslim majority countries or Africa (Table A.8). This pattern of results suggests that teacher biases against immigrants are not, at least in their entirety, explained by skin color or religion. Despite having adequate statistical power to detect these differences by immigrant group, however, these results should be interpreted with caution, since students from different immigrant backgrounds are in different parts of the test-score distribution, attend different schools, and have different teachers.

It is also possible that teachers exhibit biases in their assessments in the absence of statistical discrimination. For example, if teachers are averse to students from minority groups reaching positions of power, their biases in assessments may only extend to high-performing immigrant groups. Alternatively, teacher biases may reflect additional information teachers gain through their interactions with the students their teach. For example, there have been recurring controversies surrounding admissions to elite universities in the United States, where minority groups have been discriminated against for being too bookish or for lacking extracurriculars (Karabel, 2005; Arcidiacono et al., 2022). In our setting, it could be possible that teachers see minority students working too hard, or overdoing it, to achieve top marks. To the extent that this is true, we might expect high-achieving girls to receive a similar penalty from teachers – but this view is not supported by our data (Figure A.6 and Figure A.10).

5 Conclusion

The dual grading of digital end of high school exams, once by teachers and another time by blind centralized graders, provides a clean setting in which to measure teacher biases. We show that compared to native students, immigrant students receive 0.06 standard deviation units lower grades from teachers than from blind graders. High-performing immigrant students experience the largest bias in teacher assessments. We find that biases against immigrants are driven by a minority of teachers – and the median teacher shows no sign of bias. Our estimate of biases against immigrants are more than ten times larger than estimates of teacher biases by gender, and largest in subjects where teachers have the most discretion in grading.

In our setting, the teacher-assigned grades are not used for decisions regarding high school graduation or enrollment in higher education. Still, schools which exhibit bias according to teacher grades are the schools from which immigrant students are least likely to continue their educations in higher education. We show that this is likely to be because teachers' biases in grading are underlied by biases in other teaching practices – such as in how teachers shape students' academic ambitions. If teacher biases in grading reflect only a portion of teacher biases, then any grading-based measure of teacher biases is an underestimate of the entirety of biases experienced by students. As such, while blind-grading is a simple solution to reduce the direct impacts of biases in grading, it may have limited effects on how teacher biases shape student trajectories.

Together, our results highlight the role of teacher biases in shaping the educational trajectories of immigrant students. These results underscore the significance of using blind graders in high-stakes measures of academic performance. Nonetheless, we show that grading biases are likely to extend beyond grading in and of itself (see also Alan et al. (2023)). As teachers are incredibly important in shaping students' educational pathways (Jackson et al., 2014), it is crucial to find ways to ensure that teachers are effective for all students in today's increasingly diverse classrooms. Recent papers show that the cultural competencies required to teach to diverse students are malleable and can be learned (Edmonds, 2022; Gershenson et al., 2022). Existing research also suggests that biases in teacher grading practices can be reduced by making teachers aware of their own biases (Alesina et al., 2024).

This paper has documented the role of teachers in contributing to immigrant-native gaps in educational outcomes. Still, inequalities in immigrant-native outcomes extend beyond education (Ansala et al., 2022), and reducing these inequalities requires a deeper understanding of the various forces underlying these gaps.

References

Acquah, E. O. et al. (2015). Responding to changing student demographics in finland: A study of teachers' developing cultural competence.

Ahmad, A. (2020). Do equal qualifications yield equal rewards for immigrants in the labour market? *Work, Employment and Society*, 34(5):826–843.

Alan, S., Duysak, E., Kubilay, E., and Mumcu, I. (2023). Social exclusion and ethnic segregation in schools: The role of teachers' ethnic prejudice. *Review of Economics and Statistics*, 105(5):1039–1054.

Alesina, A., Carlana, M., La Ferrara, E., and Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, Forthcoming.

Alne, R. and Herstad, E. I. (2020). Heterogeneity in educational gender biases and the effect on labor market outcomes. *Available at SSRN 3479209*.

Angrist, J., Hull, P., Pathak, P. A., and Walters, C. (2024). Credible school value-added with undersubscribed school lotteries. *Review of Economics and Statistics*, 106(1):1–19.

Ansala, L., Åslund, O., and Sarvimäki, M. (2022). Immigration history, entry jobs and the labor market integration of immigrants. *Journal of Economic Geography*, 22(3):581–604.

Ansala, L., Hämäläinen, U., and Sarvimäki, M. (2020). Age at arrival, parents and neighborhoods: understanding the educational attainment of immigrants' children. *Journal of Economic Geography*, 20(2):459–480.

Arcidiacono, P., Kinsler, J., and Ransom, T. (2022). Asian american discrimination in harvard admissions. *European Economic Review*, 144:104079.

Arrow, K. J. (1972). Some mathematical models of race discrimination in the labor market. *Racial discrimination in economic life*, pages 187–204.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago press.

Bernelius, V. (2013). Eriytyvät kaupunkikoulut: Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. *Handbook of economic field experiments*, 1:309–393.

Bhuller, M., Mogstad, M., and Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4):993–1030.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, pages 1–45.

Bohren, J. A., Hull, P., and Imas, A. (2022). Systemic discrimination: Theory and measurement. Technical report, National Bureau of Economic Research.

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

Briole, S., Gurgand, M., Maurin, E., McNally, S., Ruiz-Valenzuela, J., and Santín, D. (2022). The making of civic virtues: a school-based experiment in three countries. IZA Discussion Paper No. 15141, Institute of Labor Economics.

Burgess, S. and Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3):535–576.

Busk, H. and Jauhiainen, S. (2022). The careers of immigrants in finland: Empirical evidence for genders and year of immigration. *Journal of International Migration and Integration*, pages 1–22.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3:1801–1863.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

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

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges. Working Paper 31492, National Bureau of Economic Research.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chin, M. J., Quinn, D. M., Dhaliwal, T. K., and Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, 49(8):566–578.

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*, 88(9-10):1697–1720.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165.

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School choice, school quality, and postsecondary attainment. *American Economic Review*, 104(3):991–1013.

Diamond, R. and Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Technical report, National Bureau of Economic Research.

Edmonds, L. (2022). Role models revisited: HBCUs, same-race teacher effects, and Black student achievement.

European Commission (2020). Action plan on integration and inclusion 2021–2027.

Figlio, D. N. (2005). Names, Expectations and the Black-White Test Score Gap. Working Paper 11195, National Bureau of Economic Research.

Fryer, R. and Jackson, M. O. (2008). A categorical model of cognition and biased decision making. *The BE Journal of Theoretical Economics*, 8(1):0000102202193517041357.

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C. A., and Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, 14(4):300–342.

Graetz, G. and Karimi, A. (2022). Gender gap variation across assessment types: Explanations and implications. *Economics of Education Review*, 91:102313.

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

Hanna, R. and Linden, L. (2009). Measuring discrimination in education. Working Paper 15057, National Bureau of Economic Research.

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

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2015). Discrimination against students with foreign backgrounds: Evidence from grading in swedish public high schools. *Education Economics*, 23(6):660–676.

Huttunen, K., Pekkarinen, T., Uusitalo, R., and Virtanen, H. (2023). Lost boys? Secondary education and crime. *Journal of Public Economics*, 218:104804.

Jaakkola, M. (2000). Finnish attitudes towards immigrants in 1987-1999. *Finnish Yearbook of Population Research*, pages 129–161.

Jackson, C. K. and Mackevicius, C. (2021). The distribution of school spending impacts. Working Paper 28517, National Bureau of Economic Research.

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., and Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, 2(4):491–508.

Jackson, C. K., Rockoff, J. E., and Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annu. Rev. Econ.*, 6(1):801–825.

Kane, T. J. and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.

Karabel, J. (2005). *The chosen: The hidden history of admission and exclusion at Harvard, Yale, and Princeton*. Houghton Mifflin Harcourt.

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

Lavy, V. and Megalokonomou, R. (2024). Persistency in Teachers' Grading Biases and Effect on Longer Term Outcomes: University Admission Exams and Choice of Field of Study. *American Economic Journal: Applied Economics*, Forthcoming.

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

Lindahl, E. (2007). Comparing teachers' assessments and national test results-evidence from sweden. Technical report, IFAU-Institute for Evaluation of Labour Market and Education Policy.

Lindahl, E. (2016). Are teacher assessments biased?–evidence from sweden. *Education economics*, 24(2):224–238.

Lindsay, C. A. and Hart, C. M. (2017). Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina. *Educational Evaluation and Policy Analysis*, 39(3):485–510.

Nshom, E. and Croucher, S. (2017). Perceived threat and prejudice towards immigrants in finland: A study among early, middle, and late finnish adolescents. *Journal of International and Intercultural Communication*, 10(4):309–323.

OECD (2018a). *Education at a Glance 2018*.

OECD (2018b). *Indicator B7 How equitable are entry and graduation in tertiary education?*

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review*, 62(4):659–661.

Rangel, M. A. and Shi, Y. (2020). First impressions: The case of teacher racial bias. IZA Discussion Paper No. 13347, Institute of Labor Economics.

Sahlberg, P. (2021). *Finnish lessons 3.0: What can the world learn from educational change in Finland?* Teachers College Press.

Sarvimäki, M. (2017). Labor market integration of refugees in finland. *VATT research reports*, 185.

Scherer, E., Cleveland, C., and Ivester, R. (2021). The effects of teacher-student demographic matching on social-emotional learning. In *SREE 2021 Conference*. SREE.

Shi, Y. and Zhu, M. (2023). "model minorities" in the classroom? positive evaluation bias towards asian students and its consequences. *Journal of Public Economics*, 220:104838.

Silliman, M. (2017). Targeted funding, immigrant background, and educational outcomes: Evidence from helsinki's "positive discrimination" policy. Technical report, VATT Institute for Economic Research.

Silliman, M. and Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics*, 14(1):197–224.

Staiger, D. O. and Kane, T. J. (2014). Making decisions with imprecise performance measures: The relationship between annual student achievement gains and a teacher's career valueadded. *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*, pages 144–169.

Statistics Finland (2020a). EDUC OPISK Students. Helsinki: Statistics Finland Research Services (accessed December 12, 2023).

Statistics Finland (2020b). EDUC TREK Degrees and qualifications module. Helsinki: Statistics Finland Research Services (accessed December 12, 2023).

Statistics Finland (2020c). EDUC TYHR Entrance to secondary education module. Helsinki: Statistics Finland Research Services (accessed December 12, 2023).

Statistics Finland (2020d). FOLK Basic module. Helsinki: Statistics Finland Research Services (accessed December 12, 2023).

Statistics Finland (2020e). FOLK Child-Parent module. Helsinki: Statistics Finland Research Services (accessed December 12, 2023).

Statistics Finland (2021). Suomen virallinen tilasto (SVT): Opiskelijat ja tutkinnot [verkkojulkaisu]. [Online; accessed 8-December-2023].

Statistics Finland (2023). Suomen virallinen tilasto (SVT): Väestörakenne [verkkojulkaisu]. [Online; accessed 4-October-2023].

Strohmeier, D., Kärnä, A., and Salmivalli, C. (2011). Intrapersonal and interpersonal risk factors for peer victimization in immigrant youth in finland. *Developmental psychology*, 47(1):248.

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

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

YTL (2022). Finnish Matriculation Exam registry [Ylioppilastutkintorekisteri]. Helsinki: Finnish Matriculation Examination Board.

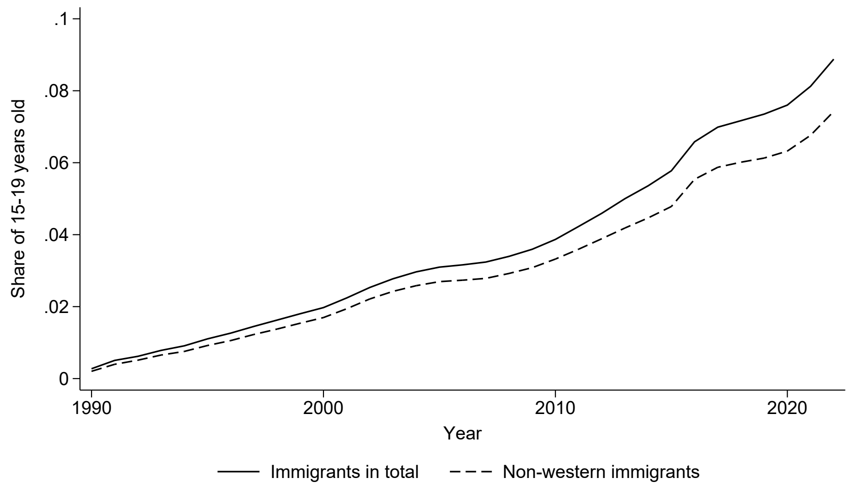
YTL (2023). Arvostelupalvelun käyttöohje. https://www.ylioppilastutkinto.fi/fi/asiointipalvelut/ilmoittautumiset-tutkintoon-ja-koesuoritusten-arvostelu/arvostelupalvelun. Helsinki: Finnish Matriculation Examination Board (accessed November 23, 2023).

Zacheus, T., Kalalahti, M., Varjo, J., Saarinen, M., Jahnukainen, M., Mäkelä, M.-L., and Kivirauma, J. (2019). Discrimination, harassment and racism in finnish lower secondary schools. *Nordic Journal of Migration Research*, 9(1):81–98.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

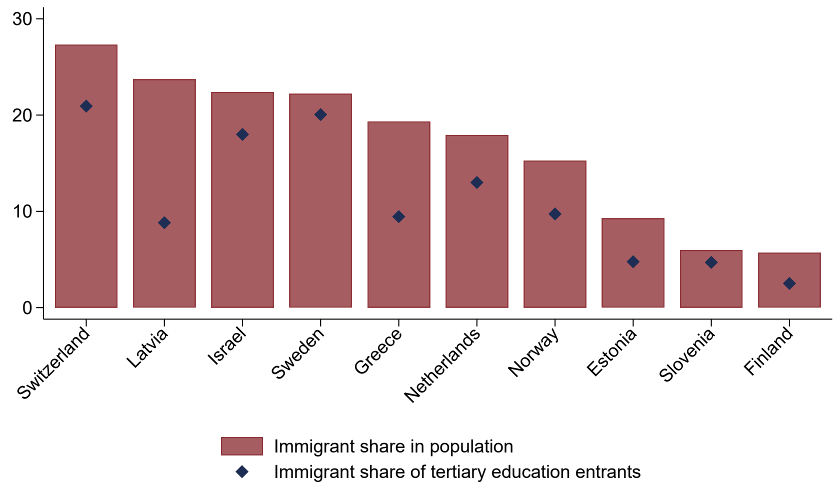
A Appendix

Figure A.1: Upper-secondary school aged western and non-western immigrants



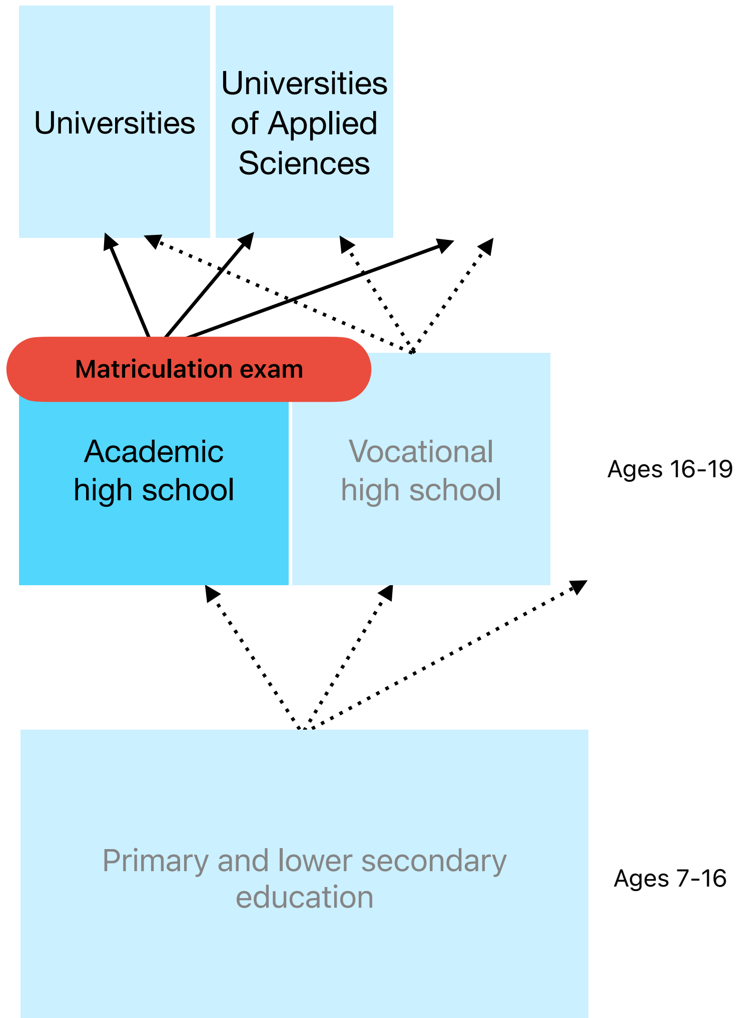
Notes: This figure shows the immigrant share broken down by country of origin. Non-western immigrants excludes the individuals with a background in western countries, defined as Northern or Western Europe, the United States, Canada, Australia, or New Zealand. Source: Statistics Finland (2023).

Figure A.2: Tertiary enrollment of immigrants across several OECD countries



Notes: Figure A.2 shows immigrant share of the population in 10 European countries, compared to immigrant share of entrants to tertiary education. Sources: OECD (2018b).

Figure A.3: The education system in Finland

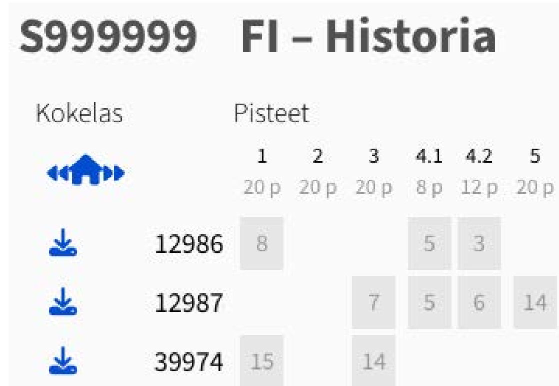


Notes: An illustration of the Finnish education system relevant for the cohorts of this study. Academic high school can also be referred to as general track of upper secondary school. Lower secondary school is also referred to as middle school.

Figure A.4: Digital grading system

(a) Teacher view, showing student name

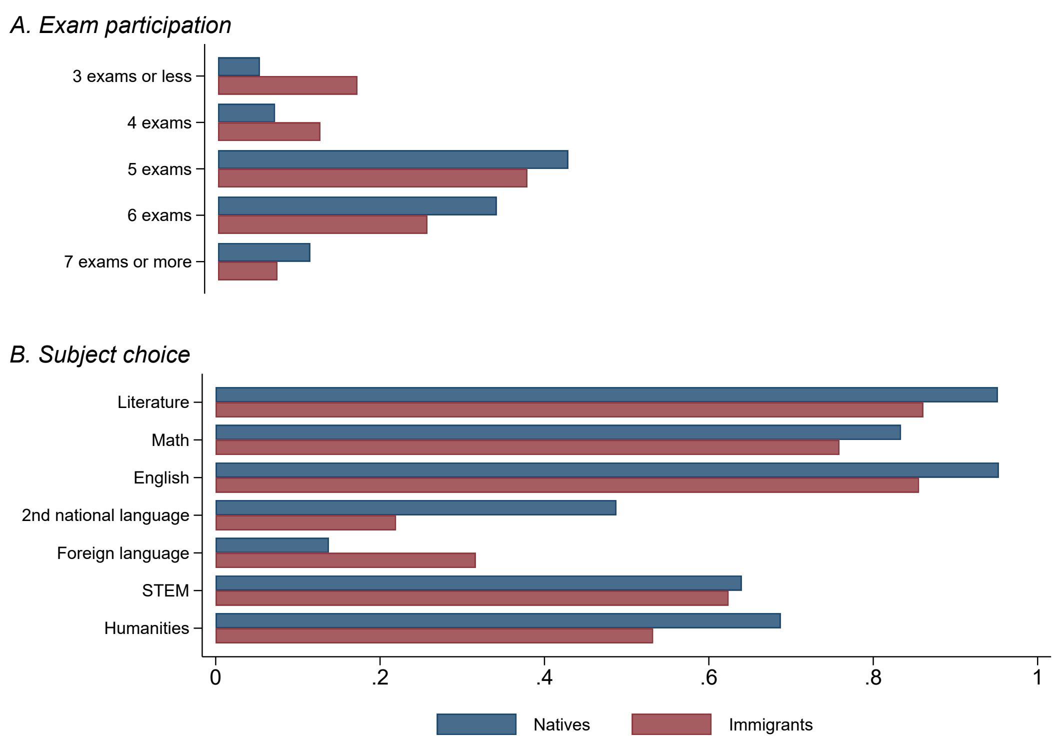
(b) Censor view, student name and school undisclosed



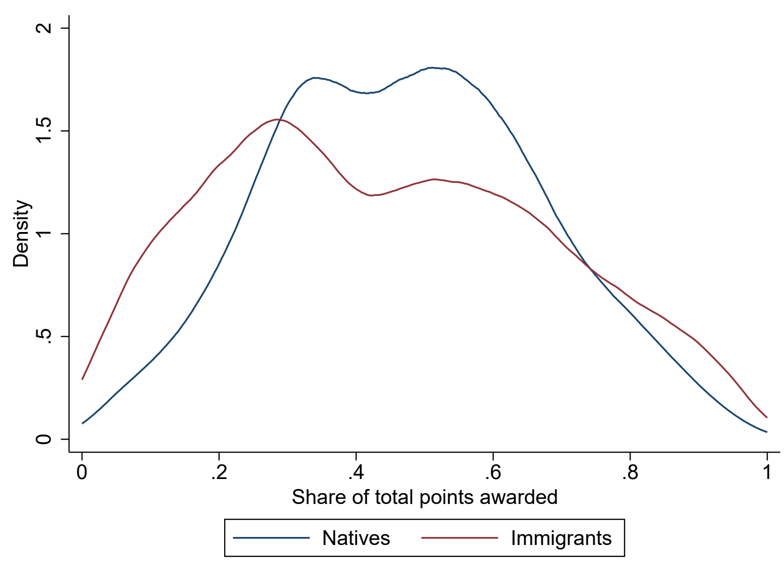
Notes: Examples of the digital grading system for teachers and censors respectively. The teacher sees the name of the student and the answer by clicking the white box. Teachers assign a score to each student and question. Censors see these teacher-assigned scores, but no names of the student, teacher, or school. Similarly, censors view the submitted answer by clicking the, now graded, squares, and they then assign students blind scores. (YTL, 2023)
Translations of Finnish text in figure:
Historia = History
Aloita arviointi klikkaamalla jotain solua. = Begin grading by clicking a cell.
Kokelas = Test taker
Pisteet = Points
Sukunimi, Etunimi = Last name, First name

Figure A.5: End of high school exam participation and results for immigrants and natives

(a) Exam participation



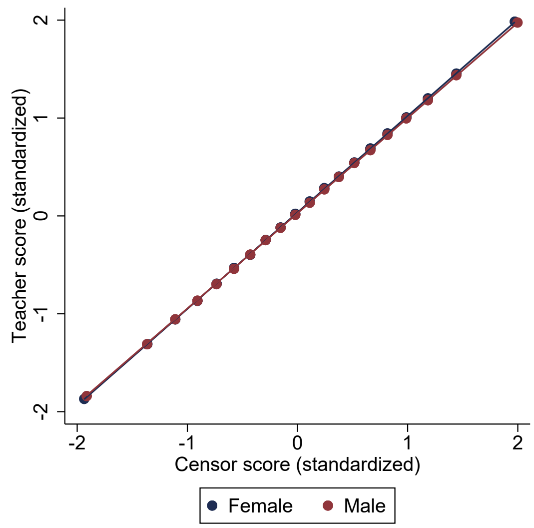
(b) Exam performance



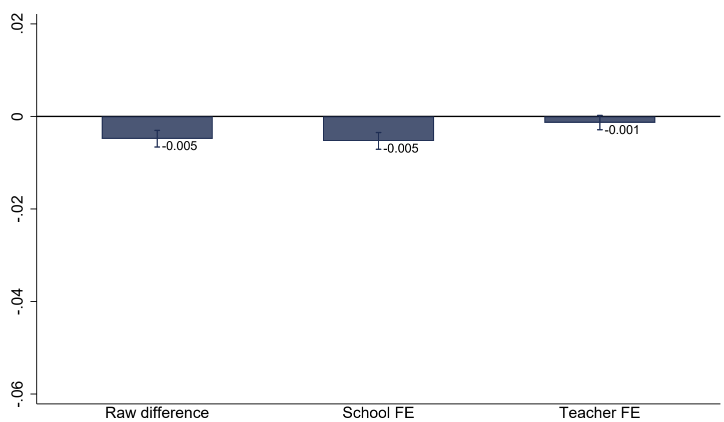
Notes: Panel A in the above figure shows number of exams taken by natives and immigrants respectively. Note that 4 passed exams are required for graduating. Panel B show the share of students taking at least one exam from the subject group. To eliminate possible bias stemming from taking exams during years for which we do not have data, the sample includes only high school students born in 2001 (n = 28,472). Matriculation exam test scores in the lower figure are shown as a share of maximum test score, based on points given by censor (final points), including all available test scores.

Figure A.6: Estimates of gender bias

(a) Correlation between teacher and censor scores

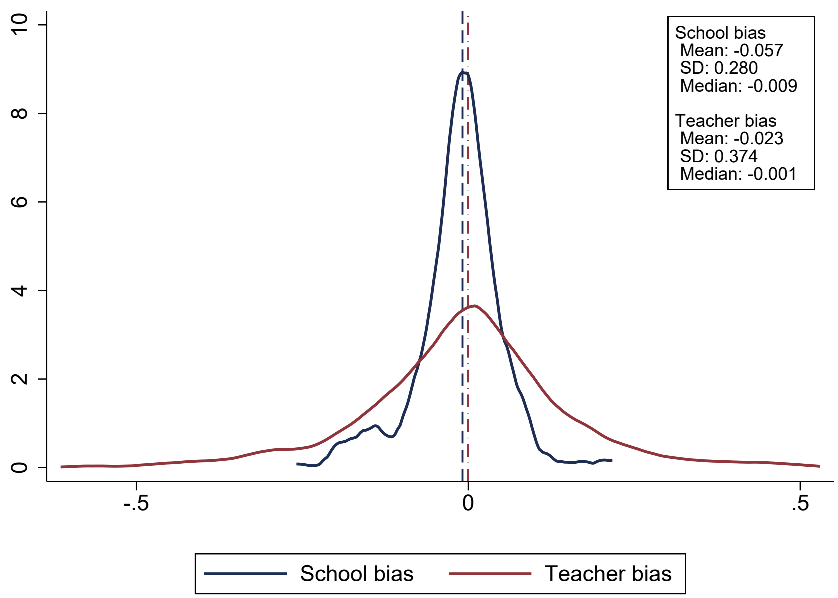


(b) Average teacher bias faced by male students



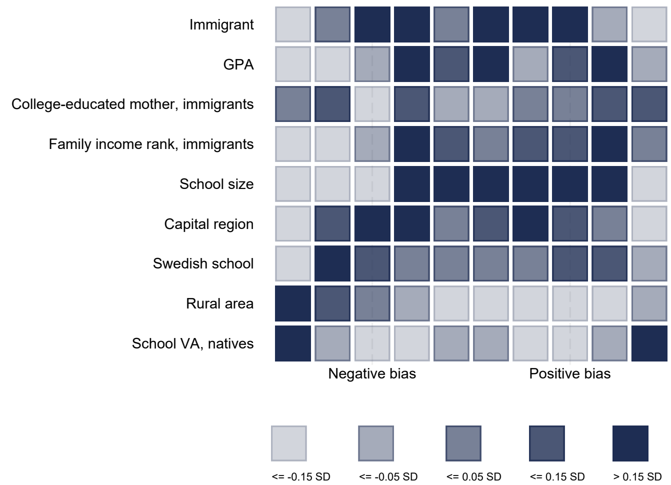
Notes: These figures replicate Figures 3, but this time by gender rather than immigrant background. Figure (a) plots teacher- and censor-assigned scores for female and male students separately. Figure (b) follows Equation 2, with coefficients reported for gender (Male = 1) instead of immigrant.

Figure A.7: Distribution of school and teacher bias (without subject de-meaning)



Notes: Distribution of empirical Bayes-corrected school and teacher biases. One percent of each tail of the teacher distribution are excluded from the figure, altering the distributions slightly from the full statistics in the box. This Figure differs from Figure 6 in that estimates are not corrected for different subject means. ($n_s = 322$, $n_j = 5170$)

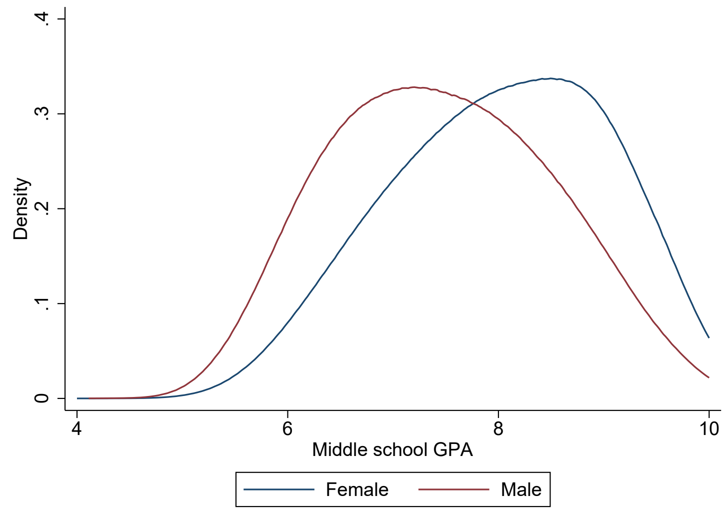
Figure A.8: Schools differ in other aspects than only bias



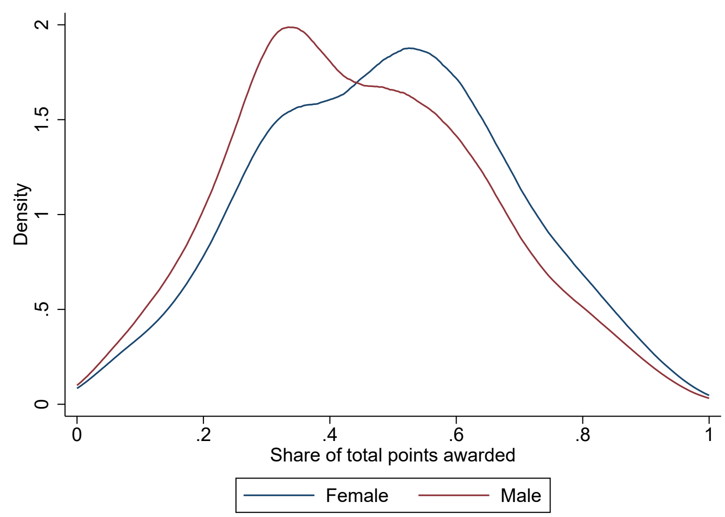
Notes: This figure shows size of deviation (measured in standard deviations) from mean value of several characteristics among schools in the digital-exam sample, by level of school bias. Estimated school bias is on the x-axis, divided into deciles.

Figure A.9: Academic performance by sex

(a) Middle school GPA

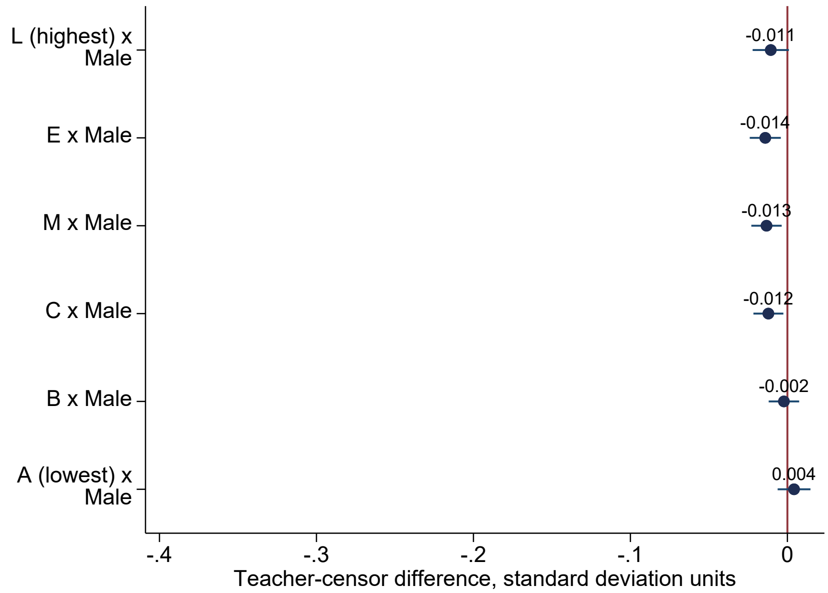


(b) Exam performance



Notes: Panel A of this figure shows the distribution of GPA at the end of middle school for female and male respectively. Panel B shows matriculation exam test scores as a share of maximum test score, based on points given by censor (final points), including all available test scores. To eliminate possible bias stemming from individual students taking part of the exams during years for which we do not have data, the sample includes only high school students born in 2001 (n = 28,472).

Figure A.10: Heterogeneity in teacher-censor difference by student performance and sex



Notes: The figures show estimates for the difference in scoring between female and male students by final matriculation exam grade ranging from A, the lowest grade, to L, the highest.

Table A.1: Tests within each subject group

Subject group	**Included tests (syllabus level)**
Literature	Literature and writing in Finnish, Swedish or Sapmi (mother tongue); Listening, reading and writing in Finnish or Swedish (for students with foreign mother-tongue)
Math	Mathematics (basic or advanced)
English	English (basic or advanced)
2nd national language	Finnish, Swedish (intermediate or advanced)
Foreign language	French, German, Russian and Spanish (basic or advanced); Italian, Inari Sami, North Sami, Skolt Sami, Latin and Portuguese (basic)
STEM	Biology, Chemistry, Physics, Psychology
Humanities	Ethics, Philosophy, Geography, History, Health studies, Evangelical Lutheran religion, Orthodox religion, Social studies

Notes: Division of tests available in the Finnish matriculation exam into subject groups. Literature (one of the options) is mandatory, tests from the other groups are chosen according to set criteria, such that the total number of tests is at least 4.

Table A.2: Characteristics of schools for which we can estimate bias

	All schools (SD)	Schools where bias can be estimated (SD)
Immigrant	0.038	0.046
	(0.062)	(0.066)
Male	0.406	0.407
	(0.105)	(0.102)
Final grade (avg)	4.090	4.101
	(0.509)	(0.526)
Exam takers	81.915	92.790
	(70.052)	(71.173)
Capital region	0.198	0.226
	(0.399)	(0.419)
Urban municipality	0.521	0.582
	(0.500)	(0.494)
Semi-urban municipality	0.196	0.198
	(0.397)	(0.399)
Rural municipality	0.284	0.220
	(0.451)	(0.415)
Swedish language school	0.090	0.099
	(0.287)	(0.299)
Schools	388	322

Notes: Average characteristics of schools in the digital-exam sample. The first column includes all schools, in column 2, only schools for which we have estimates of bias are included. In practice, this excludes schools with no immigrant students, as that prevents us to estimate teacher bias.

Table A.3: Average teacher bias faced by students with different backgrounds

	(1)	(2)	(3)	(4)
Immigrant	-0.061***			
	(0.002)			
Male		-0.005***		
		(0.001)		
Low-income			0.006***	
			(0.001)	
College-educated mother				-0.007***
				(0.001)
Obs.	554,361	554,361	554,361	554,361

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Results from Equation 2, altering the explaining variable. Low-income is measured as having a family income below median. The regressions include cohort fixed effects. Standard errors in parentheses. One observation is one student-test-observation.

Table A.4: Interactions between immigrant background, gender, and socioeconomic status

	(1)	(2)	(3)	(4)
Immigrant	-0.061***	-0.059***	-0.042***	-0.060***
	(0.002)	(0.003)	(0.005)	(0.002)
Male		-0.005***		
		(0.001)		
Immigrant x Male		-0.007		
		(0.004)		
Low-income			0.013***	
			(0.001)	
Immigrant x Low-income			-0.032***	
			(0.005)	
Mother education				-0.009***
				(0.001)
Immigrant x Mother education				-0.013***
				(0.005)
Obs.	554,361	554,361	554,361	554,361

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Results from Equation 2, controlling for different background characteristics interacted with immigrant background. The regressions include cohort fixed effects. Standard errors in parentheses. One observation is one student-test-observation.

Table A.5: Average teacher bias faced by students with immigrant backgrounds

	(1)	(2)	(3)	(4)	(5)
Immigrant	-0.061***	-0.055***	-0.020***	-0.008***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Cohort FE	Yes	Yes	Yes	Yes	Yes
School FE		Yes			
Teacher FE			Yes		
Subject FE				Yes	
Censor FE					Yes
Obs.	554,361	554,361	552,058	554,361	554,360

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Results from Equation 2, with additional censor fixed effects. Standard errors in parentheses.

Table A.6: Relationship between differences in long-term value added and differences in grading

	(1)	(2)	(3)	(4)	(5)
			A. University enrollment		
Grading bias (rank)	0.086	0.130*	0.104	0.109	0.148**
	(0.066)	(0.067)	(0.066)	(0.066)	(0.068)
			B. Secondary graduation in 3 years		
Grading bias (rank)	-0.026	-0.027	0.008	0.026	0.012
	(0.066)	(0.067)	(0.066)	(0.066)	(0.068)
			C. Secondary graduation in 4 years		
Grading bias (rank)	-0.041	-0.039	-0.007	0.018	-0.000
	(0.066)	(0.067)	(0.066)	(0.066)	(0.068)
Subject specific		Yes			Yes
Emp. Bayes corrected			Yes	Yes	Yes
Emp. Bayes by subject				Yes	Yes
Obs.	256	255	256	256	254

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Like Figure 7, this table reports the relationship between the school-level measures of bias in teacher grading of the matriculation exam and differences in school value added in terms of tertiary enrollment for immigrants and natives. The estimate stem from the equation: $\widehat{BIAS}_{VAM} = \alpha + \beta\hat{\gamma}_s + \varepsilon_i$, where each school is weighted by its effective sample size ($\sqrt{N_{imm}^2 + N_{nat}^2}$). In addition to school value added in terms of tertiary enrollment, we include estimates of school value added in terms of secondary graduation. Both sides of the equation are measured in ranks.

Table A.7: Simulations of how teacher grading biases could affect post-compulsory educational trajectories

	Natives	Immigrants			
		Original GPA	0.05 GPA boost	0.10 GPA boost	0.15 GPA boost
		Panel A: Holding preferences fixed			
Admitted anywhere	0.99	0.96	0.97	0.97	0.97
Admitted first choice	0.92	0.83	0.84	0.86	0.87
Admitted academic	0.52	0.42	0.42	0.43	0.43
		Panel B: Changing preferences			
Admitted academic	0.97	0.92	0.96	0.97	0.97

Notes: This table studies how teacher grading biases against immigrants could affect post-compulsory educational trajectories. The first column shows the probability that the middle school grades of native students place them above the minimum admissions cutoff for various outcomes. For immigrants these are shown in the following columns. In column two, immigrant grades are unchanged from those which are observed in the data. The grades of immigrants are boosted in columns 3-5 to account for various levels of grading biases – or teacher biases which result in less learning, and thereby lower grades. In Panel A, we maintain all application preferences as observed in the actual data. In Panel B, we modify application preferences to see whether each student's grades would have qualified them for the academic track with the minimum GPA threshold, for students from their middle school.

Table A.8: Heterogeneity in teacher bias against immigrants by immigrant country of origin

	(1)	(2)	(3)	(4)	(5)
Immigrant	-0.061***	-0.040***	-0.072***	-0.073***	-0.039***
	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)
Immigrant x Non-Western		-0.027***			
		(0.005)			
Immigrant x Africa & West Asia			0.042***		
			(0.005)		
Immigrant x Muslim-majority				0.039***	
				(0.004)	
Immigrant x East Europe & Russia					-0.059***
					(0.004)
Obs.	554,361	554,361	554,361	554,361	554,361

Notes: Results from Equation 2, including a dummy for background country of origin. The regressions include cohort fixed effects. Standard errors in parentheses. One observation is one student-test-observation.