Social Exclusion and Ethnic Segregation in Schools: The Role of Teacher's Ethnic Prejudice[*]

Sule Alan[†1], Enes Duysak[2], Elif Kubilay[2], and Ipek Mumcu[3]

[1]European University Institute, IT

[2]University of Essex, UK

[3]University of Exeter, UK

Abstract

Using data on primary school children and their teachers, we show that teachers who hold prejudicial attitudes towards an ethnic group create socially and spatially segregated classrooms. Leveraging a natural experiment where newly arrived refugee children are randomly assigned to teachers within schools, we find that teachers' ethnic prejudice, measured by an implicit association test, significantly lowers the prevalence of inter-ethnic social links, increases homophilic ties among host children, and puts refugee children at a higher risk of peer violence. Our results highlight the role of teachers in achieving integrated schools in a world of increasing ethnic diversity.

JEL Codes: I24, J15

Keywords: teacher effects; ethnic prejudice; social exclusion; ethnic segregation

[†]Corresponding Author: Department of Economics, European University Institute, 50014 Fiesole, Italy. Email: salancrossley@gmail.com.

1 Introduction

Ethnic prejudice and stereotypes are the root causes of ethnic tensions and conflicts world-wide (Arbatli et al., 2020). These harmful beliefs easily breed in ethnically diverse environments and tend to take hold in difficult sociopolitical contexts. In a world with growing anti-immigrant sentiments reinforced by massive population movements, understanding how ethnic biases shape our social interactions with the members of out-groups is crucial to building empowered and cohesive communities. Ethnically mixed schools represent an ideal setting to study the effect of ethnic biases and the formation of inter-ethnic relationships. While schools play an important role in building inter-ethnic cohesion, ethnically diverse schools, especially in socioeconomically disadvantaged settings, are susceptible to ethnicity-based conflicts, acts of social exclusion, and inter-group violence (Gradstein and Justman, 2002).[1] Students belonging to minority groups are generally at a higher risk of falling victim to these actions.

This paper examines how the school environment shapes young children's inter-ethnic relationships. Our focus is the role of teachers, in particular, teachers' implicit bias toward a minority group, in determining the nature of social ties formed among students. Teachers' implicit bias against a minority group may manifest itself as discriminatory behaviors, ranging from poor quality teacher-pupil interactions and neglect to outright unfair treatment of minority students, such as biased grading and disproportionate reprimanding (Burgess and Greaves, 2013, Alesina et al., 2018). Teachers' behaviors and attitudes toward ethnicity can implicitly guide students' socialization choices. Teachers who hold prejudicial attitudes

---

[1] Research on the effects of the ethnic composition of schools on students' inter-ethnic relations produces mixed results. As a prominent approach, inter-group contact theory predicts considerable social and private benefits of integrated schools (see, e.g. Paluck et al., 2019). See Alesina and Ferrara (2005) and Dahlberg et al. (2012) and references therein for the broader literature on the costs and benefits of ethnic diversity.

toward an ethnic group may consciously or unconsciously prevent students from forming inter-ethnic friendships, creating segregated classrooms where minority students tend to be socially and spatially excluded. Research shows that socially excluded minority children are more likely to engage in self-defeating behavior, reinforcing further social exclusion and weaker inter-ethnic interactions (Buhs et al., 2006).

We test whether there is a causal link between teachers' ethnic prejudice and students' inter-ethnic relationship by leveraging an ideal field setting. Our study site, Turkey, has received over 4 million refugees since the start of the Syrian Civil War in 2011. About 1 million of these refugees are school-aged children. As part of a multi-partnered EU initiative launched in 2016, the Turkish Ministry of Education has been placing refugee children in state schools based on their registered addresses. The study site provides us with a setting where i) the ethnic composition in schools changed due to a massive refugee influx and an active school placement program, ii) conditional on school, both host and refugee students were assigned to classrooms (teachers) randomly, and iii) we collect data from students and teachers in previously unavailable details. Our data set covers over 5000 2nd, 3rd, and 4th-grade primary school students and their teachers from 73 schools. About 13% of the students in our sample are refugees.

We measure teachers' ethnic bias using an Implicit Association Test (IAT), developed by Greenwald et al. (1998). The IAT aims to measure implicit attitudes toward a group using the difference in reaction times when individuals are asked to relate a group-specific concept to a negative or a positive attribute. Because it is hard to game, the test is considered free from social desirability bias inherent in explicit survey questions on socio-politically sensitive topics. In recent years, the IAT has been used extensively in economics research in the context of gender and ethnic discrimination in schools (see, e.g. Alesina et al., 2018, Carlana, 2019), ethnic discrimination in the workplace (Glover et al., 2017) and measuring ethnic bias in the lab (Berge et al., 2020). Even though the test captures the degree of implicit

ethnic bias in teachers, it is likely to be confounded with various other teacher characteristics. To the extent that these characteristics are also predictive of the outcomes we are interested in, our estimated effects would be biased. The differential impacts of the measured implicit bias on hosts and refugee children help mitigate these concerns. Nevertheless, to control as many possible confounds as possible to isolate the effect of teachers' ethnic bias, we collect detailed data from teachers. These include previously unavailable characteristics, such as teachers' fluid cognitive ability, measured by Raven's Progressive Matrices (Raven et al., 2004), and cognitive empathy, measured by Reading the Mind in the Eyes test (Baron-Cohen et al., 1997). In addition to standard demographic information and qualifications, we also collect detailed information to construct teaching styles, pedagogical practices, and motivation.

The literature on social networks offers powerful tools to study social interactions (Jackson, 2008). We utilize these tools to construct our outcome variables. Our primary outcomes are indicators of social exclusion of refugee students and ethnic segregation in classrooms. To construct these indicators, we elicit students' social networks within their classrooms. Specifically, students are asked to list the classmates i) they consider as their friends, ii) to whom they provide emotional and academic support, and iii) from whom they receive emotional and academic support, allowing the natural overlaps between these categories. Using the reported social ties, we construct student (node)-level measures of social exclusion and the degree of inter-ethnic interactions. We also construct classroom level measures (indices) of homophily. In our case, the host (refugee) homophily in a classroom concerns the above-expected numbers of social ties amongst the host students (refugee students) and gives us a measure of ethnic segregation in the classroom.

We first document that refugee students have significantly fewer social ties than host students. They have fewer friends and classmates who support them emotionally and academically. While host students receive, on average, 2.30 friendship nominations, refugee

4

students receive only 1.34 nominations. Considering that they arrived in their classrooms much later than their host classmates, this finding is not surprising. However, we find that teachers' ethnic bias significantly lowers the number of social ties enjoyed by refugee students. In classrooms where teachers have a stronger ethnic bias, refugee students are more likely to be socially excluded, i.e., they have fewer friends and fewer classmates who provide them emotional and academic support. For example, we find that a one standard deviation increase in teachers' ethnic bias leads to 0.24 (18%) fewer friendship ties enjoyed by refugee students.

We also estimate a significant effect of teacher bias on the degree of homophilic ties amongst host students. At the individual level, we find that as the teachers' ethnic bias increases, the number of friendship nominations extended from host students to refugee students decreases. In contrast, the number of friendship nominations extended from host students to host students increases significantly. This finding also emerges in our analysis of classroom-level homophily. We find that teachers who have a stronger ethnic bias create more ethnically segregated classrooms, measured as excessive numbers of homophilic social ties. Consistent with Currarini et al. (2009), we observe this excess homophily only among host students. A one standard deviation increase in teachers' ethnic bias increases friendship homophily among host students by about 13%.

We show that refugee students who are exposed to ethnically biased teachers face a higher risk of peer violence and bullying. We find that a one standard deviation increase in teachers' ethnic bias leads to 5 percentage point increase in bullying reported by refugee students, with a precise null effect on host students. Exposure to acts of social exclusion and victimization is likely to slow refugee students' progress in learning the host country's language. This, in turn, is likely to have a detrimental impact on their learning outcomes. Corroborating this conjecture, we find that teachers' ethnic bias has a significantly negative impact on refugee children's verbal ability in the host country's language. Specifically, a

one standard deviation increase in teachers' ethnic bias lowers refugee students' Turkish test scores by 0.15 standard deviations. The estimated effect on mathematics test scores is also negative and economically significant (0.07 standard deviation) but does not reach statistical significance. We estimate precise null effects on host students' achievement scores.

Our data allow us to rule out a reverse causality mechanism whereby refugee children's behavioral conduct influences their teacher's implicit bias as well as their relationships with their host classmates. To do this, we leverage the timing of our data collection and show that teachers' ethnic bias is insensitive to the duration of their exposure to refugee children in their classrooms. We then explore whether our results can be explained by the internalization of teachers' exclusionary classroom practices by students. Research shows that students who have good relationships with their teachers are more likely to be accepted by their classmates (see, e.g. Hughes et al., 2001). We find strong evidence that supports this mechanism. Students who are perceived as the teacher's favorites enjoy more friendship ties in our data, and refugee students are much less likely to enjoy such status in classrooms with biased teachers. In these classrooms, refugee students are likely to be spatially segregated and often seated at the back corners of the classroom, away from the teacher's attention. We also find some evidence that teachers transmit their ethnic attitudes to refugee children, possibly making them self-conscious about their own ethnicity.

Our contribution in this paper is twofold. First, to the best of our knowledge, this is the first paper that shows how teachers' implicit attitudes toward an ethnic minority group shape students' social networks, particularly their inter-ethnic relationships. Our setting allows us to identify this relationship causally by exploiting a natural experiment generated by a massive refugee crisis and the way a school placement policy is implemented by the host country. The second contribution pertains to the quality of our data. Because there is no secondary data available to answer the questions we pose in this paper, we collected detailed primary data from children and teachers using a diverse toolkit. Our data allow us to i)

6

construct multiple measures of social exclusion and ethnic segregation by utilizing the tools of social network theory, ii) control for previously unavailable teacher characteristics that may confound our ethnic bias measure, and iii) explore mechanisms through which teachers' ethnic bias might influence students' inter-ethnic relationships. Our paper also makes a broader contribution to the literature by highlighting the critical role of public education in building and protecting social capital. By identifying a particular factor that undermines this critical role, our study urges restorative and preventative policy actions targeting schools and communities that are at risk of ethnic tensions and conflict (Alan et al., 2021).

Our paper complements several strands of literature. A number of papers test the inter-group contact theory and show that prejudice toward out-groups may dissipate as intergroup relationships are allowed to take hold (e.g. Boisjoly et al., 2006, Carrell et al., 2019, Rao, 2019, Lowe, 2020, Mousa, 2020). Contrary to the predictions of intergroup contact theory, the literature on integration and conflict shows that negative sentiments against newly encountered out-group members may lead to inter-group conflict (Bazzi et al. (2019)). By showing the factors that mediate the effects of inter-ethnic contact, our paper complements these studies. Our paper also fits in the literature that studies the effects of negative attitudes toward minorities on various outcomes. In the context of education, Bergh et al. (2010) show that the achievement gap between ethnic groups in schools can be explained by the teachers' ethnic bias. Gershenson et al. (2018) and Papageorge et al. (2020) document that teachers' biased expectations for minorities may become self-fulfilling. Alesina et al. (2018) further show that making teachers aware of their bias decreases biased grading. Additionally, our paper contributes to the line of research that strives to understand the role of teachers in shaping children's socio-emotional skills (Gershenson, 2016, Jackson, 2018, Alan and Ertac, 2018, Alan et al., 2019, Liu and Loeb, 2021). Finally, by showing how social networks can be shaped and re-shaped via exogenous influences, we contribute to the empirical literature on social networks (Marmaros and Sacerdote, 2006).

The rest of the paper is organized as follows. Section 2 provides the background and the context for identification. Section 3 describes the data, the construction of our outcome measures, and the teachers' ethnic bias score. Our results are presented and discussed in Section 4. In Section 5 we explore potential mechanisms. We conclude in Section 6.

2   Background and Context for Identification

Our study site, Turkey, has received over 4 million refugees since the start of the Syrian Civil War in 2011. About 1 million of these refugees are school-aged children. As part of a multi-partnered EU initiative launched in 2016, the current Turkish Ministry of Education (MoE) policy is to place all school-aged refugee children in state schools based on their registered address. School administrators are mandated to admit refugee children located in their catchment areas and, upon admission, distribute them to classrooms based only on their age. Most refugee children lack essential Turkish language ability when they first arrive in their new schools. School administrators are mandated to distribute newcomers to classrooms as evenly as possible to avoid overwhelming teachers. School administrators tend to place refugee children in lower grades than what their age requires to facilitate better language learning. The objective of the school placement policy is to achieve faster integration through total immersion; therefore, at the time of our data collection, schools were not allowed to open separate classrooms for the refugee students. The placement policy was fully activated in the 2016-2017 academic year, with accelerated student placements in Winter 2017.

In addition to the refugee placement program, our identification strategy is powered by the way Turkish teachers are appointed to state schools and the way they are assigned to classrooms within schools. After completing their degree requirements, teachers are placed in a pool to be appointed to a public school in need. A new teacher typically has no say in which school he/she will be appointed to.[2] Currently, teachers may not ask to be re-

---

[2] In practice, they have no say in which district and even in which city they will eventually

appointed before completing at least four years (over six years in actual practice) of service in their current school. Requests to be re-appointed are honored if i) there is a school in need in the preferred city/district, and ii) the teacher has higher service points than her competitors who want the same location. The service points are accumulated based on years of service and the socioeconomic status (SES) of the catchment area, with low-SES areas yielding higher service points. As working in high-SES catchment areas is more desirable for teachers, there tends to be a high teacher turnover in low-SES district schools such as those in our sample. Even with a long tenure in the profession, it is extremely hard to be appointed to the generally desired (high-SES) districts.

Turkish primary schools provide an ideal setting to explore the effect of teachers' ethnic prejudice on students' outcomes because of two reasons: First, except for the relocations mentioned above, a primary school teacher teaches the students allocated to them from grade 1 to grade 4, after which students move to middle schools. School administrators are mandated to randomly allocate first graders to teachers through publicly held draws in the presence of parents. A primary school teacher spends considerable time with their pupils compared to a middle or high school branch teacher, so they are more likely to influence students' behaviors and attitudes. Second, the refugee school placement policy initially targeted primary schools. At the time of our data collection, very few refugee students were placed in middle or high schools.

3   Data and Outcomes

Our data set is a sub-sample of combined baseline data collected as part of two independently run randomized controlled trials (RCTs), implemented in the academic year of 2018-2019, in two Southeast provinces of Turkey, Sanliurfa and Mersin. Baseline data we use in the paper were collected in April-May 2018 in Sanliurfa at the end of the academic year, and October

---

end up. Teachers can refuse their placement, but this means giving up a secure public service job with relatively generous social security, which is rare.

2018 in Mersin, at the beginning of the academic year. These two provinces have received massive numbers of refugees since the start of the Syrian Civil War, and they have been part of the MoE's refugee placement program since its inception in 2016. Both RCTs include schools that are flagged as "socioeconomically disadvantaged" by their respective provincial authorities. The total number of schools for RCT 1 and RCT 2 is 87 and 77, respectively, with no overlap.[3]

To answer the research question we pose in this paper, we exclude several schools from the original samples. First of all, as our paper focuses on primary school students and teachers, we exclude all middle schools from our sample (27 middle schools in RCT 2). We also dropped 12 elementary schools that had not received refugee students yet.[4] From the remaining 125 elementary schools, we dropped 47 schools because they gave us singletons (single classroom), eliminating within-school variation in teachers' ethnic bias. Singletons arose due to (i) only one teacher signing up for the respective RCT in the school, (ii) dropping teachers who did not fill up the surveys or provide meaningful responses, and (iii) dropping classrooms with no refugees despite some refugee presence in the school.[5] There is generally

---

[3] The first RCT was eventually implemented using 80 out of 87 schools as teachers from 7 schools in Sanliurfa withdrew from the program after baseline.

[4] While all schools in both aforementioned RCTs were, and still are, part of the MoE's placement program, they received students at different intensity over time. In a matter of a single academic year after our data collection, all public schools in our sample received several waves of refugee influx.

[5] This sample selection has unlikely to have any significant external validity implications for us. First, a single teacher signing up for the RCT in a school is mainly due to high teacher turnover, which is a prominent characteristic of this region for all schools. The program was oversubscribed. We had to decline teachers who were unlikely to be in their school in the following academic year and some schools, by chance, ended up having only one teacher that fit our participation criterion. Second, given that all our schools and classrooms received refugees within a year, classrooms with no refugees at the time of our visit are unlikely to be different from our sample classrooms. In fact, we show in Section 5.1 that the distribution

no room for parents to choose their children's teachers in the state system, and parental involvement is rare in the socioeconomic segment our sample represents. Nevertheless, in addition to the statements from headteachers, we collected detailed testimonies from teachers about how, in practice, host and refugee children were assigned to teachers by their school administrators. Using these testimonies, we eliminated 5 more schools.

Our final sample covers over 5000 2nd, 3rd and 4th-grade students and 192 teachers from 73 primary schools. About 13% of our sample consists of refugee students. We collected our data by visiting all schools in person and spending over three lecture hours in every classroom. While we were collecting data from children, we asked teachers to fill up their own surveys in isolated rooms, so all student data collection took place in the absence of teachers. In what follows, we discuss our data collection toolkit and how we construct our outcome measures and variables of interest in detail.

3.1 Student Characteristics

We collected rich data on student characteristics using surveys and tests. To measure children's fluid cognitive ability, we implemented "Raven's Progressive Matrices" (Raven et al., 2004). To measure children's cognitive empathy (emotional intelligence), we implemented "Reading the Mind in the Eyes" test (Baron-Cohen et al., 1997). Because there are no centrally administered objective tests at the grade level we work with, we also implemented math and Turkish language tests in classrooms. We prepared these tests separately for each grade level, based on the national curricula. Our student survey includes standard demographic information and socioeconomic indicators, such as the availability of the internet at home and the father's employment status. We measured students' ethnic bias using survey questions as the IAT is not suitable for their age. We also elicited their bullying experiences in their schools. For this, students were asked to report the number of peers who regularly

---

of teacher ethnic bias in these classrooms is not different from the distribution in our sample classrooms.

hurt them (i) verbally, (ii) physically, and (iii) by ridicule. We provide our survey inventory in Online Appendix 3. Finally, we asked teachers to assign behavioral conduct grades to each student using a 1 to 5 scale with respect to 3 categories: (i) tendency to exhibit physical aggression, (ii) tendency to engage in verbal disputes, and (iii) temperament, with higher grades indicating worse behavior. We do not use these behaviour grades in our main analyses because they are available only at the end of the 2018-2019 academic year, not at baseline. However, we use them to provide support for our identification claims; see Section 3.5.

3.2 Social Networks in the Classroom

To elicit social networks within classrooms, we provided children with a user-friendly paper template. The template asked for nominations of classmates in three categories: friendship, emotional support, and academic support.[6] The template also asked students to write down the names of classmates whom they perceive as their teacher's favorite, excluding themselves. For each category, students were provided with three boxes to nominate (write down the names of) classmates. First, they were asked to nominate at most three best friends in the classroom and three of the teacher's favorites. Then they moved on to nominating at most three classmates from whom they get emotional support, then three classmates from whom they get academic support. Finally, they were asked to nominate at most three classmates whom they emotionally support and three classmates whom they academically support.

Before the elicitation began, children were instructed that they could also nominate friends who were absent that day. Moreover, they were told that the provided template allows them to nominate up to three classmates for each category, which means that they were allowed to nominate a minimum of zero and a maximum of three classmates. Finally, they were told that nominations across categories could overlap.[7] We piloted this elicitation

---

[6] For emotional and academic support, the exact wording reads as "helping a classmate who is sad" and "helping a classmate who needs help with homework", respectively.

[7] The template was designed as boxes to facilitate children to write one classmate in

method several times using templates that allow for 4 and 5 nominations. Our pilot sessions revealed that children have a hard time filling up larger templates; therefore, we decided to cap the number of nominations at three.

Using the elicited ties, we construct two sets of outcomes. The first set contains the standard in-degree centrality measures constructed using the individual (node) level information. These measures are simply the number of edges a node receives in a given category. A student's in-degree centrality for the friendship category gives the number of friendship nominations he/she receives from his/her classmates. The minimum number of nominations one can receive is zero, and the maximum is the number of students (minus the student himself/herself) in the classroom at the time of the elicitation. We construct the in-degree centrality for the other categories similarly. For example, in-degree centrality of a student for the "classmates who support me emotionally" category is the number of classmates who nominate him/her as a supportive classmate, i.e., someone who provides emotional support.

Our second set of measures focuses on the inter-ethnic content of network ties. The first of these measures contains several individual-level outcomes constructed using out-degree ties, i.e., nominations made by students. Here, we are interested in inter-ethnic nominations. We use the number of host nominations made by a refugee child and the number of refugee nominations made by a host child in all three categories as outcome variables. Note that none of these measures are independent of the class size or the proportion of refugee children in the classroom. Therefore, our analyses always control for these two classroom characteristics.

We also construct homophily indices for refugees and hosts separately as a measure of classroom-level ethnic segregation. Specifically, following Coleman (1958), we construct homophily indices for refugees and hosts separately as follows: Let R and H denote refugee and

each box neatly. We gave detailed examples before starting the elicitation to make sure they understood the procedures. Our examples were standard based on clearly written experimental instructions. The template sample is presented in Online Appendix 4.1.

host students, respectively. Denote the number of within-group ties of group i in classroom j as s_{ij}, and the total number of ties of group i in classroom j as t_{ij}, where $i \in \{H, R\}$. Then, s_{ij}/t_{ij} gives us the share of within-group (homophilic) ties for group i.

Denoting w_{ij} as the population share of group i in classroom j, the excess homophily of group i is defined as $(s_{ij}/t_{ij}) - w_{ij}$. The intuition behind this definition is simple. Say the share of refugees in a classroom is 10%. Then we would expect refugee students to form around 10% of their friendships with other refugee classmates if the ties are formed at random. If the observed share of within-group ties, s_{ij}/t_{ij}, is greater than the expected share, w_{ij}, we conclude that the group exhibits "excess homophily". In order to take into account groups with very large size w_{ij}, Coleman (1958) normalizes excess homophily by $1 - w_{ij}$, which is the maximum possible excess homophily that can be observed in group i.[8]

Then, Coleman's Homophily Index for group $i \in \{R, H\}$ in classroom j is:

$$C_{ij} = \frac{\frac{s_{ij}}{t_{ij}} - w_{ij}}{1 - w_{ij}}.$$

Note that since we have two groups (hosts and refugees), Coleman's measure gives us two normalized excess homophily scores, one for each group. We explore the effect of teachers' ethnic bias on both these scores. Figure A1 presents actual friendship ties from two classrooms in our data for illustration purposes. Both classrooms have a similar size (28 and 27 pupils) and a similar number of refugee students (6 and 8). It can be seen in visual clarity that classroom 1 is more integrated than classroom 2. Coleman Homophily Indices for the host (refugee) group are 0.937 (0.635) and -0.016 (0.17) and for classroom 1 and 2 respectively. Note that the refugee excess homophily index is not defined in classrooms with only one refugee student.

---

[8] In cases where there is excess heterophily, i.e. $s_{ij}/t_{ij} < w_{ij}$, the measure is normalized by w_{ij} instead of $1 - w_{ij}$. This adjustment ensures that the Coleman Homophily Index is between -1 and 1.

3.3 Teacher Characteristics

We collected detailed information from teachers, including demographics, education, experience, tenure, cognitive abilities, teaching styles, and motivation. We explain these after explaining the way we implemented our Implicit Association Test (IAT) and construct our ethnic bias measure.

3.3.1 Teacher's Ethnic Bias: An Implicit Association Test

Developed by Greenwald et al. (1998), the Implicit Association Test aims to elicit implicit attitudes toward a group. Because it is hard to game, the test is considered free from social desirability bias inherent in surveys on socio-politically sensitive topics.[9] Since mass refugee settlement and the governments integration efforts are politically sensitive topics in Turkey, rather than eliciting ethnic bias via explicit survey questions, we chose to implement an implicit association test. Despite its benefits, the literature documents multiple reliability and validity issues regarding the IAT. In particular, its test-retest reliability has been shown to be quite disappointing relative to well-known cognitive tests (Lane et al., 2007). Evidence on the validity of the IAT is mixed at best. In a recent meta-analysis, Forscher et al. (2019) provide evidence that while IAT captures deep-seated biases and correlates reasonably well with behavior, it is also sensitive to short-term stimuli. These issues imply measurement error of an unknown structure and certainly have implications for our results. For one thing, if the measurement error is classical (although there is no guarantee that it is), it works against finding any meaningful relationship between ethnic bias (the latent trait our IAT measure is meant to capture) and the outcomes of interest (socialization choices of children). We will revisit this issue when we present our results.

---

[9] For example, Avitzour et al. (2020) show that teachers' implicit gender bias, but not explicit bias, explains boy-favoring grading in math.

Traditionally, the IAT is implemented using computers or tablets. However, using the same protocol, the IAT can be administered using paper and pen in settings where using computers are impractical. Lemm et al. (2008) developed a data collection protocol for a paper-and-pen IAT as an alternative to the computer-based collection. In a paper-and-pen IAT, the respondent observes a word or an image (stimulus), presented in the middle of a line. The two attributes to associate the stimulus are located on the left and the right of the stimulus. The respondent is instructed to mark the attribute he/she associates with the stimulus. The implicit association is quantified by the number of correct responses within a given period. Note that what paper-and-pen IAT measure is an approximation of a respondent's reaction time.[10]

We designed our paper-and-pen IAT using four attribute categories; good, bad, Turkey, and Syria. We first allowed the teachers to familiarize themselves with the words and images (stimuli) listed under each attribute category. The first part, "good vs. bad," and the second part, "Turkey vs. Syria," are both single attribute comparisons, and each involves 14 lines (14 stimuli). Teachers were given 14 seconds to complete each of these parts. For example, in the first part (good vs. bad), the respondent sees the word "sad" in the middle of the line and is expected to attribute it to the category "bad," which is located on the right of the stimulus. After completing the attribution of 14 stimuli in 14 seconds, the respondent moves to the second part. Here, the same exercise is done for categories Turkey and Syria. For example, the respondent sees the word "Aleppo" and expected to attribute it to the category "Syria," which is located on the right of the stimulus. After completing parts 1 and 2, the respondent moves to the last two parts of the test, which contain two attribute comparisons. In part 3, the categories are "Good/Turkey vs. Bad/Syria" and in part 4, "Good/Syria vs. Bad/Turkey." These parts have 28 lines each (28 stimuli), and teachers were given 30 seconds to complete each part. For example, in part 4, the respondent observes the word

---

[10] For the applications of paper-and-pen IAT, see Vargas et al. (2007) and Lemm et al. (2008).

"enemy" and expected to attribute it to "Bad/Turkey," as it belongs to the "bad" attribute category. Each part was administered with a timer. We provide our IAT sheets in Online Appendix 5. The first two parts of the task can be considered as warm-up sessions. We use the last two parts to construct our measure of ethnic bias. Following the protocol by Lemm et al. (2008), we construct our implicit bias score as follows:

Let A and B denote the number of correct answers in part 3 (Good/Turkey vs. Bad/Syria) and part 4 (Good/Syria vs. Bad/Turkey), respectively. Then, the ethnic bias score η is:

$$\eta = \frac{X}{Y} * \sqrt{X - Y},$$

where $X = max\{A, B\}$ and $Y = min\{A, B\}$. If $B > A$, then the resulting values are multiplied by -1 to retain the direction of the effect.[11]

Higher values of score η indicate a stronger implicit bias toward Syria and Syrians. It is important to emphasize that the purpose of the test is not to label individuals as racist or tolerant. Rather, it aims to show how variation across individuals is related to variations in actual behaviors and outcomes (Greenwald et al., 2015). After constructing the score in the way described above, we standardize it to mean zero and standard deviation of 1. This transformation allows us to utilize only the variation across teachers and facilitates a more intuitive interpretation of our coefficient estimates. Online Appendix Figure A2 presents the distribution of the standardized ethnic bias score in our data, showing substantial variation across teachers.

---

[11] Note that the formula does not allow for $Y = 0$. We do not have a case where either A or B is zero, so we set a minimum $B = 1$. Then as part 3 and part 4 contain 28 questions each, we have $-145.5 < \eta < 145.5$.

3.3.2 Teacher Ability, Teaching Styles and Motivation

We are aware that our ethnic bias measure may be confounded by some teacher characteristics that may be relevant for predicting the student outcomes we consider. To address this, in addition to standard information on demographics and qualifications, we collected information from teachers in detail that is, to the best of our knowledge, previously unavailable in any data set. We measured the teachers' cognitive ability and cognitive empathy using Raven's Progressive Matrices and the Reading the Mind in the Eyes test, respectively. We also elicited teaching styles (pedagogy) through survey questions.[12] Using these questions, we construct five main teaching styles: Growth vs. Fixed mindset, Inquiry-based vs. Lecture-based teaching, Modern vs. Traditional teaching, Extrinsic vs. Intrinsic motivator, and Warmth vs. Distant.

It is also argued that conditional on ability, the teacher's effort and motivation are essential indicators of his/her quality (Mbiti et al., 2019). However, it is hard to observe and measure these characteristics. A commonly applied method to capture motivation is to measure professional satisfaction (Pool, 1997). For this, we asked several questions regarding job satisfaction and dedication to the teaching profession to construct a motivation measure, which we refer to as "professional satisfaction". Finally, we asked teachers their assessment of their own competency in teaching. Example questions for each of the categories mentioned above are presented in Online Appendix 3.

3.4 Descriptive Statistics

Table 1 presents a summary of student characteristics separately for host and refugee children. The last column gives the p-value from the test of equality in means across the two groups. It is clear that host and refugee student characteristics are significantly different.

---

[12] Pedagogy has been shown to be important in determining student outcomes; see Bietenbeck (2014), Hidalgo-Cabrillana and Lopez-Mayan (2018).

On average, refugee children are one year older than host children, and their fathers are more likely to be unemployed. Apart from gender composition, only two characteristics are not different across host and refugee students. These are fluid cognitive ability (IQ, measured by Raven's Progressive Matrices) and behavioral conduct grades assigned by teachers, with some evidence of better conduct with respect to verbally disruptive behavior (significant at the 10% level). As expected, refugee students perform significantly worse than host children in standardized math and verbal (Turkish) ability tests. In particular, refugee children scored 0.36 and 0.50 standard deviations lower in math and Turkish, respectively.

Under the title "Network Variables" in Table 1, we summarize our individual (node)-level social exclusion measures. The first two provide friendship nominations received (in-degree ties) and friendship nominations made (out-degree ties) followed by nominations received for the teacher's favorite student status. The next four present emotional and academic support received and provided. Finally, the last three provide average out-degree ties that involve refugee students. These measures clearly show that refugee students have significantly fewer friends, and they provide and receive less support from their classmates than host children. Significant differences between host and refugee students in these measures indicate severe ethnic segregation in classrooms. Note also that bullying is prevalent in our sample for both hosts and refugees. Nevertheless, refugee students are significantly more likely to experience bullying from their peers compared to host students, and they report higher numbers of peers who bully them.

Table 2 presents teacher demographics and classroom characteristics. About 38% of our teachers are male, with an average age of 35. The average teaching experience stands around 12 years with considerable variation, ranging from 1 year to 29. Panel 1 and 2 in Figure A3 present the distribution of teacher's (standardized) cognitive ability (Raven test score) and cognitive empathy scores (Reading the Mind in the Eyes Test score), respectively. In both cognitive measures, we observe significant variation across teachers. The proportion

19

of refugee students in classrooms varies greatly, ranging from 2% to 47%, with the average standing at 15%. The last six rows give the summary statistics of Coleman's Homophily Indices for host and refugee groups in the classroom. Recall that this index varies between -1 and 1, with higher values indicating more homophilic ties. We observe high average host homophily, with substantial variation across classrooms. Variation in refugee homophily is large as well, but it has a much lower average.

3.5   Internal Validity

As mentioned in Section 2, despite the MoE's mandate of randomly assigning pupils across classrooms, anecdotal evidence suggests that the rule may not always be adhered and school principals sometimes give in to pressure from some parents. This type of selection may lead to ability sorting and masks the actual teacher effect on student outcomes. Given the way we choose our sample, this selection is highly unlikely in our sample. Nevertheless, we perform several balance checks to make sure that our results are internally valid. Table 3 presents the coefficient estimates obtained by regressing the IAT score on pre-determined characteristics of students for the full sample and separately for hosts and refugees, controlling for school fixed effects and grade dummies (grade 2-3-4). We added fluid IQ and cognitive empathy scores to these regressions as these are considered innate and known to develop early in childhood (Schneider et al., 2014). We observe that none of the pre-determined student characteristics bear any relation to teachers' ethnic bias. The estimated coefficients are small and statistically insignificant. Furthermore, we cannot reject the hypothesis that, conditional on school, all coefficients are jointly zero; see p-values from the F-test.

We also check evidence of ability sorting by exploring whether teachers' characteristics, including their ethnic bias, predict students' cognitive ability. Online Appendix Table A1 presents the predictive power of teacher characteristics on pupil's fluid cognitive ability scores (Raven scores) for hosts and refugees separately. We observe no evidence of any association between teachers' characteristics, including their ethnic bias, and students' cognitive ability.

We conduct a similar internal validity check to see whether school administrators did follow the rule of random allocation of refugee students across classrooms. Online Appendix Table A2 presents the predictive power of teacher characteristics on the proportion of refugee students in classrooms. We find that school fixed effects, class size, and grade level together explain 81% of the variation in the proportion of refugee students across classrooms. None of the teacher characteristics, including the teacher's IAT score, have any predictive power on the ratio of refugees in the classroom. Almost all coefficient estimates are virtually zero. In summary, we find no evidence of ability sorting or non-random allocation of students, either refugees or hosts, across teachers.

4 Results

Before estimating the effect of teachers' ethnic bias on our outcomes of interest, we analyze what our ethnic bias score captures in terms of classroom and teacher characteristics. This analysis aims to inform us about the extent to which our ethnic bias measure may be confounded by some teacher characteristics. Table 4 presents the results from regressing teacher's ethnic bias (the standardized IAT score) on classroom and teacher characteristics. For the former, we use class size and the proportion of refugee students in the classroom. For the latter, we use teacher's demographic characteristics, cognitive abilities, teaching styles, and motivation. Teacher's cognitive ability and the refugee concentration in the province of their desired work location emerge as prominent predictors of his/her ethnic bias score. A one standard deviation increase in teacher's Raven score is associated with a 0.37 standard deviation decline in ethnic bias score, and this association is significant at the 1% level. The proportion of refugees in provinces where teachers would rather work is also predictive of their ethnic bias at the 5% level. These preferred provinces are almost always where teachers grew up in our context. Note again that the proportion of refugees in the class does not predict the teacher's ethnic bias. Nevertheless, we will control for all characteristics presented in this table when estimating the effect of teachers' ethnic bias on our outcomes of interest.

4.1 Empirical Specification

Our empirical specification for individual-level outcomes is as follows:

$$y_{ics} = \alpha_0 + \alpha_1 \text{TBIAS}_{cs} + \text{SC}'_{ics}\beta + \text{TC}'_{cs}\gamma + \delta_s + \varepsilon_{ics}, \tag{1}$$

where y_{ics} is the outcome of interest for child i in classroom c in school s. TBIAS_{cs} is the standardized measure of teacher's ethnic bias (the IAT score), the variable of interest. Vector SC_{ics} contains observables for student i in classroom c school s that are likely predictive of the outcome y. These include gender, age, standardized cognitive ability, cognitive empathy scores, a dummy variable for students with learning difficulties and the number of semesters the student exposed to refugees in the classroom. Vector TC_{cs} contains teacher and classroom characteristics, including class size, grade dummies, and the proportion of refugees, and δ_s are school fixed effects. The coefficient of interest is $\hat{\alpha}_1$. Throughout the paper, we estimate the above empirical model for host and refugee students separately, and test the equality of coefficient estimates. We cluster standard errors at the classroom level. While we use a broad set of controls in all our analyses, for the sake of space, we present only the estimates of the coefficient of interest, that is, the coefficient on teacher's ethnic bias score, $\hat{\alpha}_1$. For classroom level outcomes (ethnic segregation/homophily measures) we also perform a non-parametric estimation and provide visually the non-parametric relationship between teacher's ethnic bias TBIAS_{cs} and segregation in the classroom (homophily), controlling (linearly) for the aforementioned classroom level characteristics and school fixed effects.

4.2 Teachers' Ethnic Bias and Social Exclusion of Refugee Children

Table 5 presents the effect of teachers' ethnic bias on friendship and support ties for host and refugee students separately. The first two columns give the estimated impact of teachers' ethnic bias on the number of friendship nominations received (in-degree centrality). Columns 3 and 4 give the estimated effects of teachers' bias on the total number of classmates a student reports to receive emotional support from, plus the number of such nominations she receives

from her classmates. Columns 5 and 6 present the same estimates for the academic support category. Columns 7 and 8 give the effects on the total number of classmates a student reports to provide emotional support to, plus the number of such nominations she obtains from her classmates. Columns 9 and 10 present the same estimates for the academic support category. A simple illustration (for a hypothetical class size of four) of how we construct these outcomes is given in Online Appendix 4.2.

The estimated impact on the number of friendship nominations a refugee child receives (in-degree centrality) is negative and statistically different from that of a host child; see the last row for the test for the equality of impact estimates. A one standard deviation increase in teachers' ethnic bias score leads to, on average, 0.24 fewer friendship nominations that refugee children receive. To put this estimate into perspective, the average number of friendship nominations received by refugee children is 1.34 (see Table 1), which is about 42% less than the ties enjoyed by the host children. 0.24 fewer ties imply an 18% decline in the friendship ties that refugee children enjoy due to a one standard deviation increase in teachers' ethnic bias. Teachers' bias has an opposite, albeit small, effect on host students, which is statistically significant at 5%. Similarly, for the other social ties categories, we observe that as teachers' ethnic bias increases, the number of social ties a refugee student enjoys declines significantly. Take emotional support received for an example (Column 4). A one standard deviation increase in teachers' ethnic bias score leads to an average of 0.21 fewer emotional support ties a refugee student enjoys. The results for academic links are similar. We estimate no significant impacts on host students' emotional and academic support ties.

4.3 Teachers' Ethnic Bias and Ethnic Segregation in the Classroom

The above analyses explore the effect of teachers' bias on the number of social ties for host and refugee students separately. Our next question pertains to the impact of teachers' bias on the formation of inter-ethnic ties. Table 6 presents the estimated effects on students' inter-ethnic nominations. The first two columns present nominations made by host children,

and column 3 and 4 present nominations made by refugee children. Panel 1 presents the results based on friendship ties, Panel 2, and Panel 3 emotional and academic support ties. Panel 1 confirms our earlier findings that teachers' ethnic bias lowers the number of friendship nominations refugees receive. A one standard deviation increase in teachers' ethnic bias score leads to 0.02 fewer refugee classmate nominations and 0.05 more host classmate nominations made by host children. What is particularly striking in this panel is that we estimate a sizable negative impact on the number of refugee nominations made by refugee students and a positive impact on the number of host nominations made by refugee students. For the former, we find that a one standard deviation increase in teachers' ethnic bias leads to 0.06 fewer friendship nominations extended by refugee students to refugee students. While this estimate does not reach significance, the qualitatively similar estimates we obtain for emotional and academic support categories do. These findings are consistent with Currarini et al. (2009), which shows that preferences toward socializing with the same ethnicity imply more popularity for the members of the dominant ethnic group than those of minority groups. Our results show how teachers' negative attitudes toward a minority can play a role in reinforcing this result.[13]

The above results imply that ethnically biased teachers are likely to create ethnically segregated classrooms. Such classrooms would be characterized by strong in-group (homophilic) ties with minimal inter-group socialization. To test this, we estimate the effect of teachers' ethnic bias on the level of excess homophily within host and refugee students separately. Column 5 and 6 in Table 6 present the results. Consistent with our individual-level analyses, we estimate significant effects of teachers' ethnic bias on host homophily. A one standard deviation increase in teachers' ethnic bias leads to 0.09 units (about 13%) increase in host ho-

---

[13] Because the outcome variables used in Table 6 are constructed using only out-degree ties and therefore capped at 3 by design, we also experimented with a truncated regression model. The results are similar in size and precision and presented in Online Appendix Table A3.

mophily index concerning friendship ties. Both absolute and relative effect sizes concerning emotional and academic support ties are similar (about 10% and 13% increase, respectively). We find no evidence of an impact on excess homophily within refugee students, which is also consistent with our individual-level results. Our non-parametric regressions exhibit these results in visual clarity; see Online Appendix Figure A4.

4.4 Teachers' Ethnic Bias, Peer Violence and Achievement

Ethnic segregation likely leads to ethnic tensions and conflict in the school ground. In ethnically segregated schools, socially excluded minority students may be at a higher risk of falling victim to verbal and physical bullying. Recall that refugee students report higher exposure to bullying than host students do in our sample. Specifically, 93% (89%) of refugee (host) students report being bullied regularly at school. Against this background, we estimate the effect of teachers' ethnic bias on the probability of reporting bullying and the number of reported bullies. Columns 1 to 4 in Table 7 present the results for hosts and refugees separately. Refugee students who are exposed to teachers with stronger ethnic bias face a higher risk of falling victim to peer bullying. A one standard deviation increase in teachers' ethnic bias leads to a 5 percentage points increase in the probability of being bullied by peers and 0.46 more reported bullies. We estimate no effect on host students.

Our results so far strongly suggest that teachers have a significant role in the social exclusion of refugee students in their new schools.[14] This exclusion is likely to slow refugee students' progress in learning the host country's language and negatively impact their achievement outcomes. Columns 5 to 8 in Table 7 present results that corroborate this prediction. We estimate that teachers' ethnic bias has a significantly detrimental effect on refugee stu-

---

[14] We also considered absenteeism as an outcome and estimated the effect of teachers' ethnic bias on the "share of absent children on the day of our visit controlling for classroom and teacher characteristics. We find no effect on the share of absent, either for hosts or refugees.

dents' verbal ability in the host (Turkish) language. A one standard deviation increase in teachers' ethnic bias score lowers refugee students' Turkish test scores by 0.15 standard deviation. The estimated effect on mathematics scores is negative and economically significant (0.07 standard deviation) but does not reach statistical significance.[15] Note that the teachers' ethnic bias has no impact on host students' achievement scores. These precise null estimates confirm that our ethnic bias measure does not proxy an unobserved teacher quality after controlling for available teacher characteristics.[16]

4.5 Robustness Checks

Overall, our results are strikingly consistent across different outcomes. This consistency indicates that while measured with some error and possibly being confounded, the variation in IAT scores provides a valuable signal about the underlying implicit ethnic bias in teachers. In fact, when we re-estimate our models without other teacher characteristics, we obtain surprisingly similar results, suggesting that the IAT, despite all its shortcomings, seems to capture a distinct trait in our setting (see Online Appendix Table A4).

Another way to check whether our estimated effects reflect the effect of teachers' ethnic bias rather than some unobserved teacher characteristics is to absorb all time-invariant

---

[15] When we estimate the model for achievement scores using the full sample and interacting teachers' ethnic bias with the refugee dummy, we find that teachers' ethnic bias has a statistically significant negative effect on refugees' math scores without any effect on host students. A one standard deviation increase in teachers' ethnic bias lowers the math test scores of refugee students by about 0.10 standard deviation, significant at the 5% level.

[16] Figlio and Özek (2019) and Figlio and Özek (2020) analyze the educational outcomes of immigrants, refugees, and native students. We also check how refugee exposure affects host students' academic outcomes and find that a one percentage point increase in classroom refugee share reduces host Turkish and Math scores by 0.01 standard deviations (p-values; 0.047 and 0.062, respectively). We do not find any significant effects on refugee students (p-values; 0.303 and 0.332).

teacher characteristics using class fixed effects. Of course, this specification prevents us from identifying the effects on hosts and refugees separately. However, it allows us to identify the differential effect of teachers' ethnic bias using the interaction of teachers' bias variable with the refugee dummy. This specification helps us assess the impact of unobserved confounds by comparing the estimated differential effects with those implied by our main specification. Online Appendix Table A5 presents the coefficient estimates for socialization, achievement, and bullying outcomes, also presenting the differential impacts implied by our main results using school fixed effects ("Main Specification (refugee-host)"). We find that both specifications yield remarkably similar results for most of our outcomes, suggesting that our main specifications successfully isolate the variation in the IAT score that reflects the variation in ethnic bias.

Finally, we conducted another robustness check by leveraging the fact that we have some classrooms that had not received refugee students at the time of our data collection. Specifically, we appended our data with classrooms without refugees and added the variables "No refugee in Class" and "Teacher Ethnic Bias*No Refugee in Class" in our specification. As presented in Online Appendix Table A6, teachers' ethnic bias does not have the effect it has in mixed classrooms on socialization choices of hosts students. For all socialization outcomes, the coefficient of the variable of interest (Teacher Ethnic Bias*No Refugee in Class) is statistically insignificant, suggesting that the effect of teachers' bias on host students' socialization outcomes is due to the presence of refugee students.

5 Potential Mechanisms

What are the likely mechanisms that explain our results? Why are refugee children assigned to teachers with ethnic bias more likely to be socially excluded and subject to more peer bullying? Moreover, why do host students with such teachers tend to form more homophilic ties? Before we move on to exploring the plausible channels we conjecture, we provide evidence against a channel that poses a threat to our identification strategy: Reverse causality.

5.1 Reverse Causality

The reverse causality may be an issue if the teacher's IAT score is- rather than representing an inherent bias against refugees -influenced by her observation of refugee students. For example, if refugee students have a higher tendency to behave in an anti-social manner than the host students, teachers may form an unfavorable opinion about them, and this may be partly reflected in their IAT scores. Anti-social students may find it hard to form social ties with their classmates, rendering causal interpretation of our results difficult. To assess whether reverse causality may be driving our results, we leverage the timing of our data collection.

Despite being part of the refugee placement program, several classrooms and (even schools) had not received refugee students at the time of our data collection. This was simply because the placement program was relatively new, started slow, and gained substantial momentum after Fall 2018. Since teachers are centrally appointed to their work stations in our setting, this allows us to supplement our sample with teachers who had not yet received refugee students and compare their IAT scores with those in our sample. We do have 28 such teachers with a total of 800 (host) students in our data. Moreover, because we also have information on teachers' tenure in their current class and how long ago each refugee child was placed in the class, we can perform this comparison for various degrees of refugee exposure. If our IAT measure is influenced by the duration of exposure to refugee children in the classroom, we would expect to see a significant difference in the distribution of the IAT scores across different exposure levels relative to the no-exposure case. Figure 1 plots the estimates obtained from the regression of teachers' ethnic bias (standardized IAT score) on exposure dummies. The duration of exposure is measured as terms, which correspond to semesters in our setting. The figure shows that teachers' implicit bias is insensitive to the duration of their exposure to refugee students. All point estimates are virtually zero with wide confidence bands. We also test the equality of distributions across all exposure levels

and cannot reject equality across exposure levels; see p-values from the Kolmogorov Smirnov test shown in Figure 1.

As another piece of evidence, we note that refugee children are not more violent or anti-social than host children according to their teachers' assessments (see Table 1). We also show that even biased teachers do not assign bad behavior grades to refugee children, which goes against the argument that teachers' ethnic bias may be formed by observing refugee students' anti-social behavioral conduct. Online Appendix Figure A5 shows that there is no relationship between teachers' ethnic bias and their own assessment of children's behavioral conduct. These findings suggest that our results are unlikely to be driven by reverse causality. Instead, they imply that our IAT captures, at least partially, some deep-seated implicit prejudice against Syrians among Turkish teachers who are likely not very different from other Turkish adults with similar characteristics in this regard (Cagaptay, 2019, Starck et al., 2020).

5.2 Plausible Channels

We now explore various plausible mechanisms that might explain our results. An obvious channel to explore is a mechanism whereby the teachers' ethnic bias is transmitted to students, lowering their willingness to develop social ties with classmates from different ethnic groups. We test this channel by estimating the effect of teachers' ethnic bias on students' ethnic bias.

Another mechanism may be related directly to the teacher's behavior and the fact that he/she is in a powerful position to set behavioral norms in the classroom, i.e., he/she is a role model. A teacher with a strong prejudice against an ethnic group may adopt exclusionary practices in the classroom and neglect minority students who generally need extra attention. Host students might internalize these behaviors as norms and mimic these behaviors. To test this mechanism, we construct two measures of exclusionary teacher practices using nominations for the teacher's favorite student status and teacher's classroom seating

29

arrangement. The first measure was elicited using our network elicitation templates, where students were asked to nominate at most three classmates whom they considered to be their teacher's favorite. This measure is motivated by studies that show that students who have good relationships with their teachers are more likely to be accepted by their classmates (see, e.g. Hughes et al., 2001). This status is mostly enjoyed by host students in our data. While host students receive an average of 1.78 nominations in this category, refugee students receive 0.90 nominations; see Table 1. The second measure is more subtle than the first in terms of capturing teachers' exclusionary behaviors. Classroom seating arrangements are considered to be essential management tools, and research shows that these arrangements impact learning outcomes (Wannarka and Ruhl, 2008). A teacher with a strong prejudice against an ethnic group may choose to push the minority students to back seats and interact with them less frequently, which may influence the way students socialize with each other.

Figure 2 plots the coefficient estimates of the effect of teachers' ethnic bias on student ethnic bias, the number of nominations for the status of teacher's favorite, and the probability of sitting at the back corner of the classroom. We plot these estimates separately for host and refugee students to highlight the differential impacts. It is evident in this picture that the bias transmission mechanism is an important channel. The striking finding, however, is that this transmission holds only for refugee students. It appears that students who are exposed to teachers with ethnic prejudice are likely to develop biases toward those whom they consider as out-group. This finding also squares well with the implications of Currarini et al. (2009), and our results that refugee students are more likely to nominate host students as friends in classrooms with biased teachers.

The mechanism of teachers' exclusionary practices is directly evident in favorite student nominations and seating arrangements. As shown in Figure 2, refugee students are significantly less likely to enjoy the status of "teacher's favorite" in classrooms with biased teachers. Specifically, a one standard deviation increase in teacher bias is associated with 0.21 fewer

nominations for refugee students for the status of the teacher's favorite. The effect is null for host students. Consistent with this result, ethnically biased teachers tend to push refugee students to the back corners of the classroom. Refugee students are 7 percentage points more likely to sit at back desks in classrooms with biased teachers. The effect is again null for host students.

All said, the transmission of ethnic prejudice from teachers to refugee students and teachers' exclusionary classroom practices internalized by students are likely to drive our results. The former may be particularly relevant for our homophily results and the fact that refugee students appear to be particularly keen to form social ties with their host classmates rather than refugee classmates. It appears that teachers can transmit their ethnic attitudes to refugee children, making them self-conscious about their own ethnicity. The latter might partly explain why refugee students are more likely to be socially excluded and more likely to fall victim to peer bullying in classrooms with biased teachers. If these mechanisms are at work, it is then not surprising to observe that refugee students in classrooms with biased teachers fall further behind host students academically.

6   Conclusion

We show that teachers' implicit bias against a minority group significantly affects students' socialization choices in the classroom, particularly their inter-ethnic relationships. For identification, we exploit a setting where the ethnic composition in schools changed due to a massive refugee influx, and students, including refugees, are randomly assigned to classrooms. We find that teachers' ethnic bias, measured by an Implicit Association Test, significantly lowers the prevalence of social ties between host and refugee children, creates socially segregated classrooms, puts refugee children at a higher risk of peer violence, and hinders refugee children's progress in learning the host country's language. Biased teachers' exclusionary classroom practices internalized by students appear as a likely mechanism driving these results.

Given the importance of the childhood period for developing socio-emotional and cognitive skills, our results imply that the type of teachers children are exposed to in ethnically diverse schools can have significant societal impacts. Our results suggest that ethnic prejudice can breed in schools, marginalize minority children, and deprive the native children of realizing the benefits of ethnic diversity. To the extent that these harmful effects persist into adolescence and adulthood, they may damage communities' cohesiveness by increasing the risk of ethnic tension and conflict. This study shows that ethnically diverse schools are obvious grounds to take preemptive measures against these adverse social outcomes. Some of these measures may target teachers, either at the extensive or the intensive margin. For the former, teacher recruitment processes can be more attuned to identifying potential ethnic biases in candidates. This can be done by incorporating bias assessment tests, such as the IAT, in evaluation processes. The latter pertains to targeting the existing stock of teachers through training. Making them aware of their implicit biases is one way, but another, perhaps a more constructive way could be to offer them evidence-backed training on inclusive classroom practices and the benefits of these practices for migrant and native pupils.

Our study covers a particular country with particularly odd sociopolitical circumstances. Therefore, the evidence we provide likely lacks external validity. In particular, our results may not be generalizable to countries with more settled migrant populations. However, we believe that the population concerned in this study is an important population to focus on for both policy and humanitarian reasons. Human migratory flows are currently at an unprecedented scale. The majority of these flows concern forced displacements, turning into long-term settlements in many countries. The mass refugee influx from the global South is far from being the problem of a single country. While many countries are struggling to accommodate refugees by making amendments to their education systems, the eventual success (or failure) of these efforts will have tremendous implications for all countries that are at risk of facing the influx of displaced people in the future. Therefore, while conducted in Turkey, the results we document offer lessons of global value.

References

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking." *The Quarterly Journal of Economics* (March 2021).

Alan, Sule, Teodora Boneva, and Seda Ertac. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *The Quarterly Journal of Economics* 134 (August 2019): 1121–1162.

Alan, Sule and Seda Ertac. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (June 2018): 1865–1911.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti. Revealing Stereotypes: Evidence from Immigrants in Schools. Working Paper 25333, National Bureau of Economic Research, December 2018.

Alesina, Alberto and Eliana La Ferrara. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43 (September 2005): 762–800.

Arbatli, Cemal Eren, Quamrul H. Ashraf, Oded Galor, and Marc Klemp. "Diversity and Conflict." *Econometrica* 88 (2020): 727–797.

Avitzour, Eliana, Adi Choen, Daphna Joel, and Victor Lavy. On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes. Technical Report w27818, National Bureau of Economic Research, September 2020.

Baron-Cohen, S., T. Jolliffe, C. Mortimore, and M. Robertson. "Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome." *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 38 (October 1997): 813–822.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong. "Unity in Diversity? How Intergroup Contact Can Foster Nation Building." *American Economic Review* 109 (November 2019): 3978–4025.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel N. Posner, Bertil Tungodden, and Kelly Zhang. "Ethnically Biased? Experimental Evidence from Kenya." *Journal of the European Economic Association* 18 (February 2020): 134–164.

Bergh, Linda Van Den, Eddie Denessen, Lisette Hornstra, Marinus Voeten, and Rob W. Holland. "The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap." *American Educational Research Journal* (June 2010).

Bietenbeck, Jan. "Teaching Practices and Cognitive Skills." *Labour Economics* 30 (2014): 143–153.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. "Empathy or Antipathy? The Impact of Diversity." *American Economic Review* 96 (December 2006): 1890–1905.

Buhs, Eric S., Gary W. Ladd, and Sarah L. Herald. "Peer Exclusion and Victimization: Processes That Mediate the Relation between Peer Group Rejection and Children's Classroom Engagement and Achievement?." *Journal of Educational Psychology* 98 (2006): 1–13.

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

Cagaptay, Soner. "Growing Anti-Syrian Sentiment in Turkey." *The Washington Institute* (May 2019).

Carlana, Michela. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *The Quarterly Journal of Economics* 134 (August 2019): 1163–1224.

Carrell, Scott E., Mark Hoekstra, and James E. West. "The Impact of College Diversity on Behavior toward Minorities." *American Economic Journal: Economic Policy* 11 (November 2019): 159–182.

Coleman, James. "Relational Analysis: The Study of Social Organizations with Survey Methods." *Human Organization* 17 (December 1958): 28–36.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica* 77 (2009): 1003–1045.

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist. "Ethnic Diversity and Preferences for Redistribution." *Journal of Political Economy* 120 (2012): 41–76.

Figlio, David and Umut Özek. "Unwelcome Guests? The Effects of Refugees on the Educational Outcomes of Incumbent Students." *Journal of Labor Economics* 37 (February 2019): 1061–1096.

Figlio, David and Umut Özek. "Cross-Generational Differences in Educational Outcomes in the Second Great Wave of Immigration." *Education Finance and Policy* 15 (October 2020): 648–674.

Forscher, Patrick S., Calvin K. Lai, Jordan R. Axt, Charles R. Ebersole, Michelle Herman, Patricia G. Devine, and Brian A. Nosek. "A Meta-Analysis of Procedures to Change Implicit Measures." *Journal of Personality and Social Psychology* 117 (2019): 522–559.

Gershenson, Seth. "Linking Teacher Quality, Student Attendance, and Student Achievement." *Education Finance and Policy* 11 (April 2016): 125–149.

Gershenson, Seth, Cassandra M. D Hart, Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge. The Long-Run Impacts of Same-Race Teachers. Working Paper 25254, National Bureau of Economic Research, November 2018.

Glover, Dylan, Amanda Pallais, and William Pariente. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *The Quarterly Journal of Economics* 132 (August 2017): 1219–1260.

Gradstein, Mark and Moshe Justman. "Education, Social Cohesion, and Economic Growth." *American Economic Review* 92 (September 2002): 1192–1204.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74 (June 1998): 1464–1480.

Greenwald, Anthony G., Mahzarin R. Banaji, and Brian A. Nosek. "Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects." *Journal of Personality and Social Psychology* 108 (2015): 553–561.

Hidalgo-Cabrillana, Ana and Cristina Lopez-Mayan. "Teaching Styles and Achievement: Student and Teacher Perspectives." *Economics of Education Review* 67 (December 2018): 184–206.

Hughes, Jan N., Timothy A. Cavell, and Victor Willson. "Further Support for the Developmental Significance of the Quality of the Teacher–Student Relationship." *Journal of School Psychology* 39 (2001): 289–301.

Jackson, C. Kirabo. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy* 126 (June 2018): 2072–2107.

Jackson, Matthew O. *Social and Economic Networks.* Princeton University Press, 2008.

Lane, Kristin A., Mahzarin R. Banaji, Brian A. Nosek, and Anthony G. Greenwald. "Understanding and Using the Implicit Association Test: IV: What We Know (So Far) about the Method." *Implicit Measures of Attitudes.* . New York, NY, US: The Guilford Press, 2007. 59–102.

Lemm, Kristi M., Kristin A. Lane, David N. Sattler, Saera R. Khan, and Brian A. Nosek. "Assessing Implicit Cognitions with a Paper-Format Implicit Association Test." *The Psychology of Modern Prejudice.* . Hauppauge, NY, US: Nova Science Publishers, 2008. 123–146.

Liu, Jing and Susanna Loeb. "Engaging Teachers Measuring the Impact of Teachers on Student Attendance in Secondary School." *Journal of Human Resources* 56 (March 2021): 343–379.

Lowe, Matt. Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. Technical Report CESifo Working Paper No. 8089, 2020.

Marmaros, David and Bruce Sacerdote. "How Do Friendships Form?." *The Quarterly Journal of Economics* 121 (February 2006): 79–119.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *The Quarterly Journal of Economics* 134 (August 2019): 1627–1673.

Mousa, Salma. "Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq." *Science* 369 (August 2020): 866–870.

Paluck, Elizabeth Levy, Seth A. Green, and Donald P. Green. "The Contact Hypothesis Re-Evaluated." *Behavioural Public Policy* 3 (November 2019): 129–158.

Papageorge, Nicholas W., Seth Gershenson, and Kyung Min Kang. "Teacher Expectations Matter." *The Review of Economics and Statistics* 102 (May 2020): 234–251.

Pool, Steven W. "The Relationship of Job Satisfaction With Substitutes of Leadership, Leadership Behavior, and Work Motivation." *The Journal of Psychology* 131 (May 1997): 271–283.

Rao, Gautam. "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools." *American Economic Review* 109 (March 2019): 774–809.

Raven, J, J. C Raven, and J. H Court. *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio, TX: Pearson, 2004.

Schneider, Wolfgang, Frank Niklas, and Sandra Schmiedeler. "Intellectual Development from Early Childhood to Early Adulthood: The Impact of Early IQ Differences on Stability and Change over Time." *Learning and Individual Differences* 32 (2014): 156–162.

Starck, Jordan G., Travis Riddle, Stacey Sinclair, and Natasha Warikoo. "Teachers Are People Too: Examining the Racial Bias of Teachers Compared to Other American Adults." *Educational Researcher* 49 (May 2020): 273–284.

Vargas, Patrick T., Denise Sekaquaptewa, and William von Hippel. "Armed Only with Paper and Pencil: "Low-Tech" Measures of Implicit Attitudes." *Implicit Measures of Attitudes.* . New York, NY, US: The Guilford Press, 2007. 103–124.

Wannarka, Rachel and Kathy Ruhl. "Seating Arrangements That Promote Positive Academic and Behavioural Outcomes: A Review of Empirical Research." *Support for Learning* 23 (2008): 89–93.

7 Tables

Table 1: Refugee and Host Student Characteristics

	Host		Refugee		
	Mean	SD	Mean	SD	P-value
Student Demographics:					
Male	0.51	(0.50)	0.49	(0.50)	0.46
Age in Months	104.31	(8.80)	115.86	(14.71)	0.00
SES Indicators:					
Working Mother	0.28	(0.45)	0.36	(0.48)	0.00
Working Father	0.89	(0.31)	0.76	(0.43)	0.00
Computer at Home	0.44	(0.50)	0.34	(0.48)	0.00
Internet at Home	0.54	(0.50)	0.58	(0.49)	0.10
Cognitive & Behavioral Outcomes:					
Raven Score	0.00	(1.00)	0.02	(1.01)	0.73
Cognitive Empathy Score	0.00	(1.00)	-0.45	(0.95)	0.00
Math Score	0.01	(1.00)	-0.36	(0.97)	0.00
Verbal Score	0.01	(1.00)	-0.50	(0.93)	0.00
Ethnic Bias	-0.01	(1.00)	0.14	(0.93)	0.01
Physically Aggressive	-0.00	(0.99)	-0.05	(1.02)	0.53
Verbally Disruptive	-0.01	(0.99)	-0.14	(0.91)	0.09
Short-Tempered	-0.00	(0.99)	-0.04	(1.00)	0.63
Proportion with learning difficulties	0.05	(0.22)	0.14	(0.34)	0.00
Network Variables:					
Friendship Ties (In-Degree)	2.30	(2.07)	1.34	(1.50)	0.00
Friendship Ties (Out-Degree)	2.33	(0.91)	1.92	(1.08)	0.00
Nominations for Teacher's Favorite	1.78	(2.91)	0.90	(1.77)	0.00
Emotional Support Ties (Received)	2.64	(1.83)	1.74	(1.63)	0.00
Academic Support Ties (Received)	2.20	(1.68)	1.57	(1.50)	0.00
Emotional Support Ties (Provided)	2.65	(1.82)	1.73	(1.61)	0.00
Academic Support Ties (Provided)	2.23	(1.83)	1.43	(1.56)	0.00
Refugee Friendship Nominations	0.11	(0.36)	0.65	(0.83)	0.00
Refugee E. Support Nominations	0.10	(0.34)	0.30	(0.59)	0.00
Refugee A. Support Nominations	0.10	(0.34)	0.30	(0.61)	0.00
Reported Bullying:					
Proportion Reports Bullying	0.89	(0.31)	0.93	(0.25)	0.00
Number of Bullies Reported	7.05	(4.84)	8.08	(4.65)	0.00

Table presents descriptive statistics of student characteristics. Cognitive & behavioral outcomes are standardized to have mean zero for host students. Network variables give the number of edges (ties) for each category and represent our individual (node) level outcomes. The last column presents the p-values obtained from the tests of equality of means.

Table 2: Teacher and Classroom Characteristics

	Mean	SD	Min	Max
Teacher Demographics:				
Male	0.38	0.49	0.00	1.00
Age	35.28	7.19	23.00	49.00
Married	0.74	0.44	0.00	1.00
Number of Children	1.15	0.97	0.00	4.00
Tenured	0.95	0.21	0.00	1.00
Years of Experience	11.64	7.04	1.00	29.00
Number of Semesters in Current Class	4.20	1.63	0.00	6.00
Classroom Variables:				
Class Size	31.71	7.50	15.00	53.00
Proportion of Refugees	0.15	0.09	0.02	0.47
Coleman Host Homophily (Friendship)	0.67	0.33	-0.15	1.00
Coleman Host Homophily (Emotional S.)	0.61	0.36	-0.27	1.00
Coleman Host Homophily (Academic S.)	0.67	0.39	-0.52	1.00
Coleman Refugee Homophily (Friendship)	-0.16	0.67	-1.00	1.00
Coleman Refugee Homophily (Emotional S.)	-0.19	0.72	-1.00	1.00
Coleman Refugee Homophily (Academic S.)	-0.30	0.73	-1.00	1.00

Table presents descriptive statistics of teacher and classroom characteristics. Classroom characteristics include class-level segregation indices: Coleman host and refugee homophily for friendship, emotional and academic support.

Table 3: Balance Test: Teachers' Ethnic Bias and Pre-Determined Student Characteristics

	Full	Host	Refugee
Cognition:			
Raven Score	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.03)
Cognitive Empathy Score	-0.01	-0.01	-0.01
	(0.01)	(0.02)	(0.04)
Demographics:			
Male	-0.01	-0.02	0.03
	(0.02)	(0.02)	(0.07)
Student Age in Months	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
SES Indicators:			
Working Mother	0.01	0.01	0.01
	(0.02)	(0.03)	(0.07)
Working Father	0.03	0.02	0.03
	(0.03)	(0.04)	(0.08)
Computer at Home	0.02	0.01	0.08
	(0.02)	(0.02)	(0.06)
Internet at Home	-0.02	-0.01	-0.09
	(0.02)	(0.02)	(0.07)
Classroom Characteristics:			
Proportion of Refugees	0.58	0.75	-0.65
	(1.25)	(1.27)	(1.44)
# of Semesters exposed to Refugees	-0.00	-0.01	0.05
	(0.02)	(0.02)	(0.05)
Class Size	-0.03	-0.03	-0.05
	(0.03)	(0.03)	(0.03)
N	4237	3695	542
P-Value of Joint Significance	0.224	0.693	0.869
R-Squared	0.386	0.384	0.470

Reported results are from OLS estimation. The dependent variable is the teacher's ethnic bias. Regressions control for student gender, age, cognitive ability (Raven), cognitive empathy, SES indicators (dummy variables for parents' work status, home environment), classroom characteristics (class size, proportion of refugees and number of semesters the student exposed to refugees in the classroom), grade level and school fixed effects. Standard errors clustered at class level. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 4: What Predicts Teacher's Ethnic Bias?

	(1)	(2)	(3)	(4)
Proportion of Refugees	0.00	-0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.01)
Class Size	-0.03	-0.03	-0.00	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)
Teacher Demographics:				
Male Teacher		0.00	0.11	0.26
		(0.19)	(0.19)	(0.21)
Teacher Age		0.06	0.04	0.04
		(0.04)	(0.04)	(0.04)
Years of Experience		-0.03	-0.03	-0.01
		(0.05)	(0.04)	(0.04)
Married Teacher		0.08	0.19	0.18
		(0.28)	(0.29)	(0.30)
Number of Children		0.20	0.20	0.24*
		(0.14)	(0.12)	(0.13)
Tenured Teacher		0.35	0.25	0.17
		(0.66)	(0.65)	(0.67)
Desired Work Province - % of refugee		-0.03**	-0.03**	-0.03**
		(0.01)	(0.01)	(0.01)
Teacher Cognition:				
Teacher Raven Score			-0.35***	-0.37***
			(0.10)	(0.11)
Teacher Cognitive Empathy Score			0.19*	0.17
			(0.11)	(0.11)
Teaching Styles:				
Growth Mindset				0.10
				(0.11)
Inquiry-based Pedagogy				0.20
				(0.12)
Modern Teaching				-0.02
				(0.14)
Extrinsic Motivation				0.13
				(0.16)
Warmth				-0.05
				(0.13)
Teacher Motivation:				
Competence				0.14
				(0.14)
Professional Satisfaction				0.03
				(0.12)
N	192	191	191	190
R-Squared	0.419	0.488	0.557	0.594

Reported results are from OLS estimation. The dependent variable is teacher's ethnic bias. All regressions include experimenter, grade level, school fixed effects. Standard errors clustered at class level. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 5: Teachers' Ethnic Bias and Social Exclusion

	Friendship		E.S. Received		A.S. Received		E.S. Provided		A.S. Provided	
	Host	Refugee	Host	Refugee	Host	Refugee	Host	Refugee	Host	Refugee
Teacher Ethnic Bias	0.07**	-0.24***	-0.01	-0.21***	-0.05	-0.14**	-0.02	-0.28***	-0.03	-0.22***
	(0.03)	(0.07)	(0.04)	(0.06)	(0.04)	(0.07)	(0.04)	(0.06)	(0.04)	(0.06)
N	4375	659	4375	659	4375	659	4375	659	4375	659
P-Value (Host=Refugee)	0.000		0.001		0.161		0.000		0.003	

Reported results are from OLS estimation. Dependent variables are, Column 1 and 2: the number of friendship nominations a student receives from his/her classmates, i.e, in-degree centrality. Column 3 and 4: the total number of classmates a student claims to receive emotional support from, plus the number of such nominations she receives from her classmates. Column 5 and 6: same as column 3 and 4 for academic support. Column 7 and 8: the total number of classmates a student claims to provide emotional support to, plus the number of such nominations she obtains from her classmates. Column 9 and 10: same as Column 7 and 8 for academic support. Regressions control for student gender, age, cognitive ability, cognitive empathy, a dummy variable for students with learning difficulties, the number of semesters the student exposed to refugees in the classroom, classroom characteristics (class size and proportion of refugees), teacher characteristics (demographics, qualifications, cognitive ability, cognitive empathy, teaching styles, and motivation), grade level and school fixed effects. Standard errors clustered at class level. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 6: Teachers Ethnic Bias and Inter-Ethnic Ties

	Host Nominations		Refugee Nominations		Homophily	
	Host → Refugee	Host → Host	Refugee → Refugee	Refugee → Host	Host	Refugee
Panel 1: Friendship Ties						
Teacher Ethnic Bias	-0.02***	0.05***	-0.06	0.09	0.09***	-0.01
	(0.01)	(0.02)	(0.04)	(0.06)	(0.03)	(0.09)
Observations	4375	4375	659	659	190	171
Panel 2: Emotional Support Ties						
Teacher Ethnic Bias	-0.01	-0.02	-0.04*	-0.09**	0.06**	0.05
	(0.01)	(0.02)	(0.02)	(0.05)	(0.03)	(0.08)
Observations	4375	4375	659	659	187	159
Panel 3: Academic Support Ties						
Teacher Ethnic Bias	-0.01**	-0.01	-0.05**	0.01	0.09***	-0.04
	(0.01)	(0.02)	(0.02)	(0.05)	(0.03)	(0.08)
Observations	4375	4375	659	659	190	151

Reported results are from OLS estimation. Dependent variables are (Panel 1): the number of friendship nominations by host (Column 1 and 2) and refugee (Column 3 and 4) students and Coleman's Homophily Indices for friendship ties (Column 5 and 6), (Panel 2): the number of classmates to whom a host student claims to provide emotional support (Column 1 and 2) and from whom a refugee student claims to receive emotional support (Column 3 and 4) and Coleman's Homophily Index for emotional support ties (Column 5 and 6), (Panel 3): the number of classmates to whom a host student claims to provide academic support (Column 1 and 2) and from whom a refugee student claims to receive academic support (Column 3 and 4) and and Coleman's Homophily Index for academic support ties (Column 5 and 6). Column 1 (from host to host) and Column 2 (from host to refugee), Column 3 (refugee to refugee) and Column 4 (refugee to host). The Coleman's Homophily index is not defined for classrooms with only one refugee student at the time of the elicitation. Regressions in Column 1 to 4 control for student gender, age, cognitive ability, cognitive empathy, a dummy variable for students with learning difficulties, the number of semesters the student exposed to refugees in the classroom, classroom characteristics (class size, proportion of refugees), teacher characteristics (demographics, qualifications, cognitive ability, cognitive empathy, teaching styles, and motivation), grade level and school fixed effects. Regressions in Column 5 and 6 control for the proportion of refugees, proportion of male students, average age, class size, teacher characteristics (demographics, qualifications, cognitive ability, cognitive empathy, teaching styles, and motivation), grade level, and school fixed effects. Standard errors clustered at class level. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.
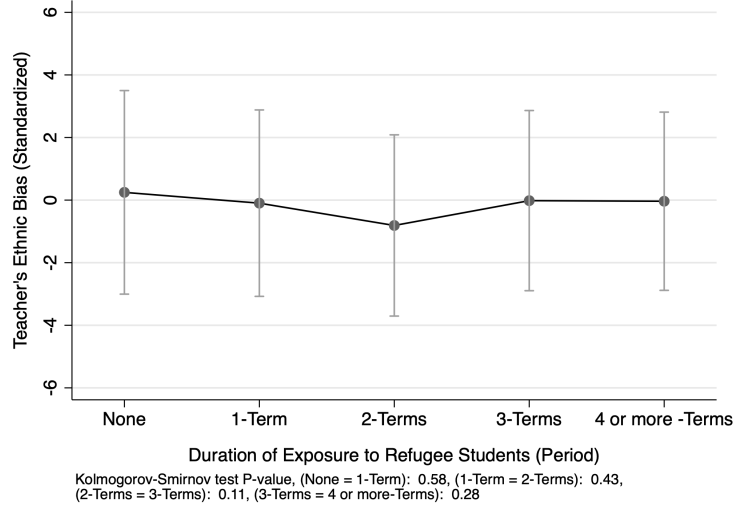
Table 7: Teachers' Ethnic Bias on Bullying and Student Achievement

	Probability Bullied		Number of Bullies		Turkish		Mathematics	
	Host	Refugee	Host	Refugee	Host	Refugee	Host	Refugee
Teacher Ethnic Bias	-0.00	0.05**	0.10	0.46**	-0.01	-0.15***	-0.01	-0.07
	(0.01)	(0.02)	(0.10)	(0.18)	(0.02)	(0.05)	(0.02)	(0.05)
N	4053	337	4278	633	4375	659	4375	659
P-Value (Host=Refugee)	0.030		0.050		0.003		0.162	

The first two columns present the average marginal effects from probit regressions. The dependent variable is a binary indicator that takes the value 1 if the student reports to be bullied regularly by his/her peers and zero otherwise. The remaining columns present OLS results. Dependent variables are the number of bullies a student reports (Column 3 and 4), standardized scores obtained from the Turkish (Column 5 and 6) and math (Column 7 and 8) tests implemented in classrooms. Regressions control for student gender, age, cognitive ability, cognitive empathy, a dummy variable for students with learning difficulties, the number of semesters the student exposed to refugees in the classroom, classroom characteristics (class size and proportion of refugees), teacher characteristics (demographics, qualifications, cognitive ability, cognitive empathy, teaching styles, and motivation), grade level and school fixed effects. Standard errors clustered at class level. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.
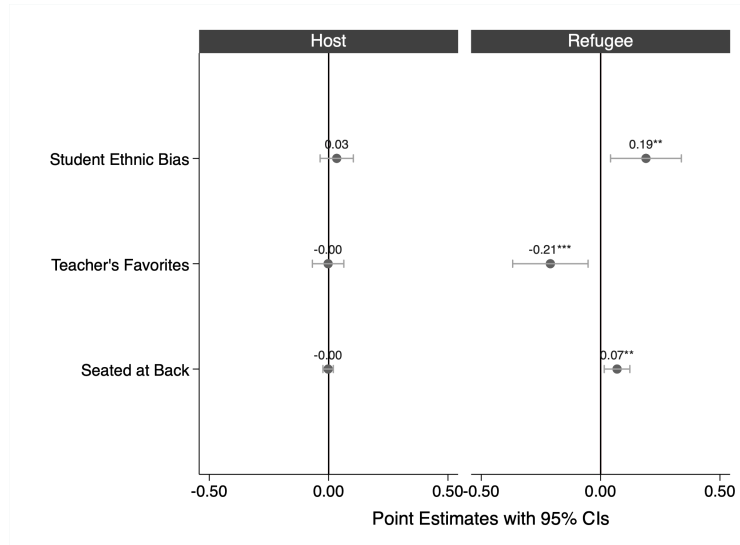
8 Figures

Figure 1: The Effect of Exposure to Refugees on Teachers' Ethnic Bias

Kolmogorov-Smirnov test P-value, (None = 1-Term): 0.58, (1-Term = 2-Terms): 0.43,
(2-Terms = 3-Terms): 0.11, (3-Terms = 4 or more-Terms): 0.28

The figure plots the estimated effect of the length of exposure to refugee students in the classroom on teachers' IAT scores. The estimates are obtained by supplementing the sample with classrooms without refugee students and running the regression of teachers' IAT score on exposure dummies. The regression controls for classroom characteristics (class size and proportion of refugees), teacher characteristics (demographics, qualifications, cognitive ability, cognitive empathy, teaching styles, and motivation), grade level and and school fixed effects. 95% confidence intervals are based on standard errors clustered at the class level. P-values for the Kolmogorov-Smirnov test of equality of distributions across terms are given at the bottom of the figure.

Figure 2: Potential Mechanisms

The figure depicts the estimated effects of teachers' ethnic bias on student ethnic bias (standardized), the number of nominations received as teacher's favorite student, and the probability of sitting at a desk in the back of the classroom. All regressions control for student gender, age, cognitive ability, cognitive empathy, a dummy variable for students with learning difficulties, the number of semesters the student exposed to refugees in the classroom, classroom characteristics (class size and proportion of refugees), teacher characteristics (demographics, qualifications, cognitive ability, cognitive empathy, teaching styles, and motivation), grade level and school fixed effects. 95% confidence intervals are based on standard errors clustered at class level. The vertical line indicates an effect of zero.