

BUILDING SOCIAL COHESION IN ETHNICALLY MIXED SCHOOLS: AN INTERVENTION ON PERSPECTIVE TAKING*

SULE ALAN[†]
CEREN BAYSAN
MERT GUMREN
ELIF KUBILAY

Abstract

We evaluate the impact of an educational program that aims to build social cohesion in ethnically mixed schools by developing perspective-taking ability in children. The program is implemented in Turkish elementary schools affected by a large influx of Syrian refugee children. We measure a comprehensive set of outcomes that characterize a cohesive school environment, including peer violence incidents, the prevalence of inter-ethnic social ties, and prosocial behavior. Using randomized variation in program implementation, we find that the program significantly lowers peer violence and victimization on school grounds. The program also reduces the likelihood of social exclusion and increases inter-ethnic social ties in the classroom. We find that the program significantly improves prosocial behavior, measured by incentivized tasks: treated students exhibit significantly higher trust, reciprocity, and altruism toward each other as well as toward anonymous out-school peers. We show that this enhanced prosociality is welfare improving from the ex-post payoff perspective. We investigate multiple channels that could explain the results, including ethnic bias, impulsivity, empathetic concern, emotional intelligence, behavioral norms, and perspective-taking. Children's increased effort to take others' perspectives emerges as the most robust mechanism to explain our results. *JEL* Codes: I24, I28, C93

*We are grateful to IPA Peace and Recovery, J-PAL initiatives of Crime and Violence, and European Social Inclusion for funding this study. We thank seminar participants at University of Cologne, Max Planck Institute of Bonn, University of Zurich, University of Lausanne, University of Stockholm, University of Toronto, LSE, participants in the 2020 Conference on Forced Displacement in Copenhagen, 2020 ESOC Annual Meeting, and LSE Political Economy Conference for their valuable comments. We thank Enes Duysak, Ipek Mumcu, Ozge Seyrek, Melek Celik and Yusuf Agus for wonderful research and field assistance. The trial has been registered at the AEA Registry: AEARCTR-0003974.

[†]European University Institute, Bilkent University and J-PAL, salancrossley@gmail.com.

I INTRODUCTION

“...public education does not serve a public. It creates a public. The question is, what kind of public does it create? A conglomerate of self-indulgent consumers? Angry, soulless, directionless masses? Indifferent, confused citizens? Or a public imbued with confidence, a sense of purpose, a respect for learning and tolerance?”

– Postman (1996), p.18

Well-developed social skills are essential to building cohesive communities. Encompassing a wide range of behaviors and attitudes such as trust, reciprocity, and cooperation, these skills together form social capital and enable effective communication and efficient economic interactions (Putnam, 1993). Public education has been shown to have a critical role in developing social skills, and therefore reducing social distance between individuals in culturally diverse environments.¹ Although humans are better off collectively as well as individually in cohesive environments with high social capital, non-cohesive environments, characterized by violence, intolerance, and identity-based segregation can arise under turbulent sociopolitical conditions. Under such conditions, the existing social capital may be damaged, impeding economic growth, and rebuilding strategies through educational interventions may become a policy imperative (Rodrik, 1999; Alesina and Ferrara, 2005; Miguel and Gugerty, 2005; Easterly et al., 2006; Deming, 2011; Fryer Jr and Loury, 2013; Hjort, 2014; Bandiera et al., 2019; Hendren and Sprung-Keyser, 2020; Voigtlaender et al., 2020).

In this study, we evaluate an educational program designed to develop social skills and build social cohesion in schools. While applicable to any educational context in which the objective is building social capital, we evaluate this program in a high-stakes context where the ethnic composition in schools has changed due to a massive influx of refugee children. The context involves Turkish elementary schools where host students have been in contact with refugee students as their peers for an extended period, and ethnic tensions on school grounds and surrounding neighborhoods are alarmingly on the rise. The educational program we evaluate is a curricular intervention implemented by children’s own teachers against this background.

The program takes a particular socio-cognitive skill, perspective-taking ability, as a core concept. Perspective-taking is the ability to perceive others’ states of mind, understand their goals and intentions, and as such, is a product of purely cognitive processes. It is considered different from what is generally known as emotional empathy or

¹Gradstein and Justman (2002) examine the relationship between education, social cohesion, and economic growth within a theoretical framework. They show that social distance between an individual and other members affects the productivity of human capital accumulation, suggesting an important role for educational interventions when society is divided along ethnic or religious lines.

empathetic concern, and this difference has been extensively studied by cognitive psychologists and neuroscientists in recent years.² Studies show that perspective-taking is associated with lower social aggression, higher trust, and more social cooperation (e.g. see [Batson et al., 1997](#); [Galinsky and Moskowitz, 2000](#); [Galinsky and Ku, 2004](#)). High perspective-taking ability is also related to being able to analyze social situations through slow deliberations, weighing the costs and benefits of an action before engaging in the act. Studies show that this type of deliberation is a malleable skill and is effective in reducing crime and violent behavior in various contexts ([Blattman et al., 2017](#); [Heller et al., 2017](#); [Alan and Ertac, 2018](#)). Motivated by these findings, a multidisciplinary team of educators, pedagogical consultants, and multimedia developers designed a program as a set of curricular activities to develop children's ability to understand each others' perspectives and their capacity to make inferences about others' intentions, goals, and motives. These curricular activities are compiled within a book titled "Understanding Each Other." In addition to various games and reading activities that encourage students to understand others' mental states, the curricular activity set includes several animated videos emphasizing the similarity of the effects of hurtful events on different people. The program designers took great care to ensure that the content makes no explicit reference to ethnicity. Instead, they aimed to encourage students to exert effort to understand the perspective of any individual or living being, regardless of their identity.

The program was implemented as a cluster randomized controlled trial. The evaluation sample includes over 6500 elementary school children, 16% of whom are refugees, from 80 elementary schools in Turkey. These schools are located in two southeastern provinces of Turkey that received a massive influx of Syrian refugees. We deliberately chose schools that are part of the Ministry of Education's (MoE) refugee placement program since its official inception in 2016 so that all pupils in our sample had already been in inter-ethnic contact for about two academic years. After collecting detailed baseline data from all children in spring and fall 2018, 124 teachers in 40 randomly selected schools received training on the implementation of the curriculum and related class activities. Between November 2018 and May 2019, the teachers covered the program for three lecture hours per week during official extracurricular project hours available to all public elementary schools. In control schools, the extracurricular project hours remained as the status-quo, which included activities related to learning good hygienic practices, environmental awareness, and group activities involving arts and

²Perspective-taking is also referred to as cognitive empathy or theory of mind. In recent years, advances in neuroscience have enabled behavioral and cognitive scientists to distinguish these two traits on a neural network level. It has been shown that these two traits recruit different neural circuits in the brain; see [Kanske \(2018\)](#) and [Stietz et al. \(2019\)](#).

games. Therefore, the number of hours that children spent together and had contact via group activities under teacher supervision remained the same across treatment and control schools. We collected endline data in May 2019.

While there is no universal definition of social cohesion, there are widely accepted indicators that characterize a cohesive environment. These include low incidents of violence, high prevalence of inter-ethnic social ties, trust, reciprocity, and cooperation between individuals.³ To evaluate the program, we put together a multidisciplinary toolkit that measures the cohesiveness of the school and classroom environment based on these indicators. Our toolkit includes i) administrative diary logs recording high-intensity peer violence and victimization that occurred within 10 consecutive school days on school grounds, ii) teacher reports of behavioral conduct and anti-social behavior, iii) student reports of bullying and victimization experienced in a typical school day, iv) carefully elicited social networks to measure social exclusion and ethnic segregation in classrooms, v) incentivized lab-in-the-field experiments to measure prosocial behaviors (trust, reciprocity, cooperation, and altruism), vi) achievement tests to measure cognitive and academic ability, and item-response questionnaires to measure behavioral norms, ethnic bias, perspective taking, empathetic concern and impulsivity (Fehr and Schmidt, 1999, 2000; Boisjoly et al., 2006; Burns et al., 2019; Rao, 2019).

We find that the program is highly effective in lowering high-intensity peer violence and victimization on school grounds. Over ten consecutive school days, about 1.9 violent events were perpetrated by children in the control group. This number is reduced by more than 60% in treatment schools. This substantial treatment effect is statistically significant at the 1 percent level. Given this result, we also explore whether the program had an unintended effect of generating more victims. The idea behind this concern is that by encouraging children to show understanding toward their peers in a generally violent environment, such as our study site, the program may have made them more susceptible to victimization. We find, on the contrary, that the program significantly reduced the victimization of children, suggesting that by keeping children away from conflict, the program also lowered the risk of being a victim of a violent act. However, these encouraging treatment effects on the administrative records of high-intensity violence are not reflected in student reports of bullying and victimization, nor in teachers' behavioral conduct grades. We do not find a statistically significant program effect on the overall likelihood of being bullied, but detect a statistically weak decline in victimization reported by refugee children.

The program also reduces the probability of social exclusion and increases the likelihood of forming inter-ethnic social ties, thereby decreasing ethnic segregation in the

³Sociologist Emile Durkheim defines a cohesive society as a society that is free from conflict based on wealth, ethnicity, race, and gender and with strong social ties among its members (Durkheim, 1897).

classroom. We find that both hosts and refugees are significantly less likely to be socially excluded and more likely to receive emotional and academic support from their classmates in treated schools. Overall, treated children are about 6% (8%) more likely than untreated children to receive emotional (academic) support from their classmates. Moreover, refugee children in treated schools are approximately 25% (21%) more likely than those in control schools to receive emotional (academic) support from their host classmates. Finally, we show that the program reduces ethnic segregation in classrooms by 15 to 21%.

We also estimate significant improvements in prosocial behavior among children, measured using incentivized tasks. Treated children exhibit significantly more trust and reciprocity toward their classmates as well as toward anonymous peers outside of their schools. The latter might be of concern if the program inadvertently disadvantages treated children by encouraging them to trust others in a generally non-cohesive environment where such behavior might be exploited. We show that this heightened trust is welfare improving from the perspective of payoffs children received in incentivized games. By exercising more trust and reciprocity toward their classmates, treated children collectively increase their payoffs by about 5 percent relative to untreated children. Furthermore, we show that treated children are not worse off by exercising more trust toward out-school anonymous peers, but they deliberately lower their overall payoff relative to the control group by exercising more reciprocity toward out-school anonymous peers. We also find that treated children exhibit higher altruism toward anonymous recipients in a dictator game and even more so when randomly paired with an anonymous refugee recipient. Specifically, treated children are 7 percentage points more likely to make donations to an anonymous peer recipient. This effect size becomes significantly higher (10 percentage points) when the anonymous recipient is randomly revealed to be an anonymous refugee peer.

Overall, the program appears to be highly effective in building a cohesive school environment. Refugee children emerge as the primary beneficiaries of this environment. In addition to improving their social interactions with their classmates, the program significantly improves refugee children's ability in the language of the host country, which is an essential marker for successful integration. Despite the program not having an academic focus, treated refugee children received 0.13 standard deviations higher scores in the objective Turkish language test we implemented in classrooms. Such a remarkable improvement in the host country's language suggests that creating a peaceful and cohesive learning environment, where inter-ethnic support ties are easily formed, is critical for the achievement of minority children, and as such, a prerequisite for a successful integration policy (Fryer Jr and Levitt, 2004; Guryan, 2004; Echenique et al.,

2006; Card and Rothstein, 2007; Hanushek et al., 2009).

Our exploratory analyses suggest that these positive effects stem mainly from the program's effectiveness in increasing children's effort to take others' perspectives. Treated host and refugee children report exerting higher effort to understand others' perspectives and higher capacity to tolerate individual differences. While the increased effort of perspective-taking emerges as a robust channel, we explore several other possible mechanisms using self-reported measures and tests. In particular, we test whether the program also works through increasing emotional intelligence, improving behavioral norms in the classroom, reducing ethnic bias, increasing empathetic concern, and enhancing the ability to regulate impulsive behavior. We find evidence that in addition to enhancing perspective-taking, the program improves perceived behavioral norms in the classroom, but only by refugee students.

Our contribution is twofold. First, we show that a carefully designed curricular program that encourages perspective-taking in social situations can go a long way in building a cohesive school environment. Our results show that fostering this important socio-cognitive skill is possible in the classroom environment, and doing so can lead to significant improvements in economically and socially vital outcomes. Our research design allows us to show that such improvements are likely to bring significant welfare gains. Second, the program is applicable to a wide range of educational contexts in which rebuilding social capital is of necessity. Such a necessity may arise in challenging sociopolitical conditions where social segregation in various domains emerges, and public education becomes an ideal policy sphere to intervene.

Our study relates to several bodies of literature. First, it complements the research on reducing crime and violence through behavioral interventions and policy changes (Lochner and Moretti, 2004; Blattman et al., 2017; Heller et al., 2017; Alan and Ertac, 2018). The educational program we evaluate cultivates a particular socio-cognitive skill, perspective-taking, which also reduces violence. Moreover, our study shows that increasing perspective-taking ability, while serving as a violence reduction tool, also enhances prosociality in children and facilitates social inclusion. Second, this study is relevant to the literature that tests the "contact hypothesis." There are experimental and quasi-experimental studies that test the contact hypothesis by evaluating interventions that facilitate inter-group contact through various activities (Bazzi et al., 2019; Burns et al., 2019; Paluck et al., 2019; Lowe, 2020; Mousa, 2020) or a policy change in which poor students were enrolled into elite private schools (Gould et al., 2004; Rao, 2019). Our study complements this literature by testing whether fostering a particular socio-cognitive skill in the classroom environment can improve peer relations in a context where inter-group contact is already high, and there is evidence of social exclusion and

maltreatment. Third, our study advances the literature on perspective-taking related experiments, most of which have taken place in a lab. The existing experimental field studies have focused on short, priming-type interventions concerning adults and where between-group interactions are rare (Broockman and Kalla, 2016; Adida et al., 2018; Kalla and Broockman, 2020a). The authors find positive results, but the measured outcomes are limited to self-reported attitudes or a one-time anonymous action. Bruneau and Saxe (2012) evaluate the importance of perspective-taking in two contexts of ongoing conflict, but in a lab, and they only measure self-reported attitudes and beliefs. A recent study by Kalla and Broockman (2020b) compares the effectiveness of narrative strategies involving analogic perspective-taking, vicarious perspective-giving, and perspective-getting in reducing exclusionary attitudes toward unauthorized immigration. Drawing on two canvassing experiments, the study points to perspective-getting, which is hearing about the experiences of outgroup members, as the most effective strategy. Our intervention can be thought of as a strategy that combines these three strategies within a curriculum. Finally, our study contributes to the growing literature on the development of socio-emotional skills by providing causal evidence on the malleability of perspective-taking ability in young children (Heckman et al., 2006; Deming, 2009; Heckman et al., 2013; Alan and Ertac, 2018; Alan et al., 2019; Cappelen et al., 2020; Eisner et al., 2020; Kosse et al., 2020).

While conducted in a specific context, the study has far-reaching implications. In the core of all inter-group conflicts lies the reluctance to make an effort to see the other's perspective, refusal to see the "other" as someone similar to self in many respects, and understand their motivations. There is a tendency to instead engage in dehumanization (e.g. see Harris and Fiske, 2009; Bruneau et al., 2018). Social psychologist Gordon Allport argues that living life together with those considered "others" can make people more tolerant. However, history suggests that mere contact may not be enough to have inter-group cohesion. In fact, it can be harmful, especially for historically disadvantaged groups. Our study attempts to offer an institutional solution to the profoundly structural problem of social cohesion by leveraging the power of public education. Fostering perspective-taking ability in children can help us build our social capital in a sustainable way and increase the size of the pie in a world where groups tend to engage in zero-sum behavior.

The paper is organized as follows. Section II summarizes the key features of the program and the sociopolitical context in which it was implemented. Section III details the evaluation design and gives a detailed account of our outcome measures. Section IV describes the data and tests for internal validity. Our main results are presented in Section V. In Section VI, we explore potential mechanisms. We conclude in Section VII.

II PROGRAM AND THE CONTEXT

Since the beginning of the Syrian Civil War in 2011, Turkey has received more than 4 million refugees. This figure is 14% of the world's refugees and makes Turkey the host country with the highest number of Syrian refugees. Currently, there are over 1 million Syrian children in Turkey. Over the past few years, the Turkish MoE has been facing enormous challenges in placing refugee children into state schools. Teachers and school administrators urgently require proper training and guidance to facilitate cohesion among host and refugee students and to cope with increasing ethnic segregation and peer violence on school grounds.

The program we evaluate is designed, implemented, and evaluated in this sociopolitical context. It is an educational cohesion program targeted at 3rd and 4th-grade elementary school children. This group was targeted because extensive extra-curricular hours are available only in elementary schools. Younger children were not included because the program required some basic reading ability, and, in this socioeconomic group, there are reading difficulties through 2nd grade. The program aims to provide teachers with an easy-to-follow curriculum to build cohesion in the classroom and ensure a healthy learning environment for all children. The curriculum content comprises written and animated class activities compiled as a modular book known as "Understanding Each Other."

The curriculum takes perspective-taking as the core concept. It encourages students to understand and experience the emotions of the described subject through various reading and visual materials. For example, in an animated video, children see several adverse events (e.g., falling while running after a ball and hurting knees) that happen to a character, followed by a similar event happening to another character. Such an event occurs randomly across ethnic groups, so the purpose of this animated material is to emphasize the similarity of the effects of hurtful events on different people. Other activities included reading materials, such as a diary extract of a hypothetical student who arrives at a new school. Students then read a diary extract from another child who writes about a new friend's arrival from another country. Another example is a guessing game where children try to understand their friends' mental state, starting from simple emotions like happiness and anger and moving toward more subtle and sophisticated emotions and thoughts. Throughout the curriculum, ethnic identity is never explicitly stated, but occasionally, as in the diary extract example, it can be inferred in some activities.

The program also includes various activities and games implemented by the teacher. For example, after watching an animated video that highlights an act of social exclusion or malfeasance toward animals, children are asked to guess what the characters in the

video must be feeling and fill up thinking balloons. The effectiveness of this type of deliberation in building perspective-taking ability is emphasized in the psychology literature (Galinsky et al., 2005).⁴ Instead of making any explicit ethnic reference, normative or otherwise, the program encourages tolerance toward individuals (and animals, for that matter) and cherishing individual differences.

III EVALUATION DESIGN AND COHESION OUTCOMES

The program was implemented as a cluster randomized controlled trial. The study sample contains 222 classrooms (teachers) from 80 elementary schools in Sanliurfa and Mersin, two provinces of Turkey where the refugee placement program has officially been in effect since 2016. The study covers over 6500 3rd and 4th-grade children ages 8 to 12. Approximately 16% of the children in our sample are refugees. The sample schools are very large, which is typical of our study site. The average number of 3rd and 4th-grade classrooms per school in our sample is 15. We had to limit our sample to an average of 3 out of 15 3rd and 4th-grade classrooms/teachers despite additional teacher demand for the program. This constraint was due to detailed data collection that involved implementing time-consuming incentivized tasks and social network elicitation. Since the program was implemented at the school-level, the statistical power of the experiment is predominantly driven by the number of schools rather than classrooms or students.

In selecting the sample of teachers, the school headteachers first received paperwork explaining that the MoE had approved a program called "Understanding Each Other," and that all 3rd and 4th-grade teachers could sign up for the program. All the teachers who wanted to participate in the program then met us during our initial visit to the school for baseline data collection. The number of volunteer teachers in most schools exceeded the number of classrooms we could sample. We randomly selected a subset of volunteer teachers in those schools.

The timeline of the trial is as follows: we collected baseline data in the province of Sanliurfa in April-May 2018 and in the province of Mersin in October 2018. We then conducted the randomization at the school level. We stratified our randomization by province and within province tertiles of school-level student absenteeism. We stratified the randomization by absenteeism to increase the power of our design as absenteeism

⁴The program was created as part of a private university's philanthropic efforts. The general framework for each week's topic was provided by a multidisciplinary team of pedagogy consultants and a group of elementary school teachers under the supervision of the R&D division in the Ministry of Education. More details about the content of the curriculum and example of class activities can be found in Online Appendix H.

is highly predictive of educational attainment and is a particularly pressing concern in this part of Turkey. The ex-ante probability of treatment is set to 50%, assigning 40 schools to treatment and 40 to control.

Teacher training seminars for 40 treatment schools (124 teachers) took place in the first week of November 2018. In these seminars, teachers were introduced to the concept of perspective-taking and its importance for children's cognitive and socio-emotional development. They then participated in an intensive workshop where they studied the "Understanding Each Other" module and related activities chapter-by-chapter and interactively with their designated education consultants. Teachers were provided with a detailed implementation kit, in hard and soft copy, explaining the module's particulars and accompanying activities. Teachers were expected to spend three lecture hours per week to cover the curriculum throughout the academic year of 2018-2019, which effectively runs from mid-October to mid-May in our schools. This time-frame corresponds to when many families in southeastern Turkey, now including refugee families, work as seasonal agricultural workers. The program was tailored to accommodate this reality so that teachers could cover the curricular activities and allow us sufficient time to conduct an endline. Our field partner periodically monitored the implementation and informed us about the process. We collected endline data in May 2019. Figure I depicts the timeline of the study.

The Turkish MoE allows (and encourages) all elementary school teachers to implement socially beneficial extra-curricular projects for a maximum of 5 lecture hours per week. Being involved in Ministry-approved extra-curricular projects is common practice for Turkish elementary school teachers. Participation in a Ministry-approved project is voluntary on the part of teachers, but not students. Once a project is approved by the MoE, and the teacher signs up to implement it, her students are required to participate. The program we evaluate is one such MoE-approved program. As mentioned, during the implementation of this program, the extra-curricular hours remained as status-quo in our control schools, e.g., learning good hygiene practices, and environmental awareness. In the absence of extra-curricular projects, teachers tend to use these free hours for supervised arts and games activities. Therefore, the program we evaluate did not crowd out core teaching activities. More importantly, because these 5 extra-curricular hours are mandated to be used under teachers' supervision, the number of hours children have contact with their peers and teachers remained the same across treatment and control classrooms.

Both baseline and endline data collection were carried out by the research team, assisted by locally recruited and trained field assistants. We spent about 3 lecture hours in each classroom, both at baseline and endline, to conduct incentivized games, tests,

and surveys. Data from children were collected using pen and paper. Teachers were not present in the classroom during data collection. They were in isolated rooms, completing their paper-based surveys. Coding and digitizing the data took about three months after the completion of endline fieldwork.

The trial was registered at the AEA Registry along with a pre-analysis plan (PAP). Unless we indicate otherwise, presented analyses and related outcomes were specified in our PAP. In what follows, we give a detailed account of these outcomes and the related hypotheses we test.

III.A. Peer Violence, Victimization and Antisocial Behavior

Peer violence and victimization are outcomes of primary interest in this study. However, such events are not officially recorded until middle school in Turkey to avoid unnecessary labeling of students at young ages. We overcame this difficulty by obtaining a special permit to collect these data ourselves from administrators. Our permit allowed us to collect these data at the school-level without referring to any particular student. This administrative peer violence measure is the number of high-intensity disciplinary episodes that took place on school grounds in the ten school days following our endline visit. Here, the term high-intensity refers to severe conflicts involving perpetrators and victims severe enough to reach school administrators and/or involve parents.

We collected these data by providing a designated school administrator with a 10-day diary log. In considering this record-keeping process, we leveraged the fact that most of our schools are very large, and we sampled only a small subset of classrooms in each grade. In addition, the schools employ multiple non-teaching administrative staff. This, along with the fact that several extra-curricular projects run in tandem in these schools at any given time, allowed us to approach an administrator who had no previous exposure or knowledge of the intervention with this unexpected request. The designated administrators in both treatment and control schools were given a minimum amount of information regarding the purpose of this exercise. We chose a 10-day period because of logistical constraints. Recall that the endline was conducted in May 2019 because of the issue of absenteeism induced by seasonal migration mentioned earlier in Section III. Choosing a relatively short time period for the diary log avoided overlapping with the start of mass absenteeism.

The diary log is an electronic Excel spreadsheet. An example is provided in Figure E.1 in the Online Appendix. At the top of the spreadsheet, specific classroom identifiers are highlighted. These are the classrooms included in our evaluation study.⁵ The

⁵Turkish schools assign a classroom identifier to each classroom using a grade-level identifier and a letter of the alphabet starting with A. For example, in a school with only four grade 3 classes, they would

spreadsheet has four columns. The first column indicates the date. In the second column, the administrator was asked to record the number of high-intensity disciplinary events in the school at the end of each day without referring to any particular classroom or child. In the third column, the administrator was asked to record the number of events that were perpetrated by someone from the classrooms highlighted at the top of their diary sheet, without identifying the perpetrators. In the final column, the administrator was asked to record the number of events where someone from the highlighted classrooms was victimized, again without identifying the victimized child. We added a measure of victimization to the diary log to establish whether the program has an unintended effect of generating more victims. The idea behind this concern is that because the program encourages children to be more understanding of others in a generally non-cohesive and violent environment, it may make them more vulnerable to perpetrators. We do not expect the program to have a significant impact on overall school-level peer violence since we sampled only a few classrooms from each school. We do expect, however, the program to be effective in reducing the number of violent events perpetrated by children from treated classrooms.

We also collected data from children on peer violence using surveys, both at baseline and endline. For this, we asked children about their experiences of bullying perpetrated by their classmate(s) as well as schoolmate(s) from outside the classroom, but in the child's school. The bullying questionnaire contains two parts, with three questions in each. In the first part, students were asked to report the number of classmates who regularly hurt them (i) verbally, (ii) physically, and (iii) by ridicule. In the second part, they answered the same questions for schoolmates. Finally, we asked teachers to rate each student's behavioral conduct using a 1 to 5 scale, where one refers to very good, and five refers to very violent and anti-social behavioral conduct. This measure is available only at endline. All questions are provided in the Online Appendix G.

Note that our diary-log measure of violence is very different from the reports we collected from students and behavioral conduct grades assigned by teachers. The diary measure is about high-intensity episodes, severe enough to draw the administration's attention, and has a specific time component. The latter measures notionally contain all events, and they entail no time component. We collected a self-reported measure of bullying in order to i) make a comparison to the administrator data in case of experimenter demand effects and ii) document any baseline discrepancies in self-reported bullying faced by refugee versus host students.

take the identifiers 3A, 3B, 3C, and 3D. Students in these classrooms progress to grade 4 into classrooms 4A, 4B, 4C, and 4D.

III.B. Social Exclusion and Ethnic Segregation

The prevalence of social exclusion based on personal characteristics, such as ethnicity, is another measure of the level of cohesiveness of an environment. Social exclusion based on personal characteristics may lead to the formation of groups identified with such characteristics (segregation) or social isolation of an individual. To construct social exclusion and ethnic segregation measures, we elicited social networks in classrooms. To do this, we provided children with a user-friendly paper template and asked for nominations of up to three classmates in each of the three categories of social ties: friendship, emotional support, and academic support, allowing for overlaps across categories. For emotional and academic support, the exact wordings are "classmates who help you when you feel sad" and "classmates who help you with homework," respectively. Before we began our elicitation, we told children that they could also nominate friends who were absent that day.⁶ We collected these data both at baseline and endline.

Using elicited ties, we construct two sets of outcomes. The first set constitutes our individual-level social exclusion measures. These include binary measures indicating whether the child nominates at least one classmate, i.e., he/she has formed any social tie at all in the form of friendship, emotional support, and academic support in the classroom. We also consider the number of in-degree ties, which is the number of nominations received by the child in each category. We expect that the program increases the probability of forming social ties, i.e., lowers the likelihood of being socially excluded for both host and refugee children.

Our second set of outcomes concerns ethnic segregation. For this, we construct a classroom-level segregation index that summarizes the degree of inter-ethnic ties in the classroom. Utilizing the idea put forward in [Schelling \(1969\)](#), we construct an ethnic segregation measure for each classroom as the difference between the expected proportion of inter-ethnic links, based on the theoretical probability of randomly formed inter-ethnic ties, and the observed proportion of inter-ethnic links. To construct the former, we proceed as follows: If all links were formed randomly, the number of links between refugee and host students would follow the hypergeometric distribution. Specifically, for a refugee student who nominates $x \in \{1, 2, 3\}$ classmates, the probability of forming $y \leq x$ links with host students would be equal to

$$p_R(x, y) = \frac{\binom{n_H}{y} \binom{n_R-1}{x-y}}{\binom{n_R+n_H-1}{x}},$$

where n_R is the number of refugee students, and n_H is the number of host students

⁶We gave detailed examples of how to fill up the template before starting the elicitation and made sure children fully understood the procedures.

in a given classroom. Analogously, for a host student, who nominates x students, the probability of forming $y \leq x$ links with refugee students would be equal to

$$p_H(x, y) = \frac{\binom{n_R}{y} \binom{n_H-1}{x-y}}{\binom{n_R+n_H-1}{x}}.$$

Of course, if a student nominates no friends, $p_i(x, y) = 0$ where $i \in \{R, H\}$.

We then calculate the probability of forming inter-ethnic ties for each classroom under the assumption that links were formed at random:

$$\mu = \frac{\sum_{x=1}^3 \sum_{y=1}^x [n_R(x)p_R(x, y)y + n_H(x)p_H(x, y)y]}{\sum_{x=1}^3 x[n_R(x) + n_H(x)]},$$

where $n_R(x)$ and $n_H(x)$ denote, respectively, the number of refugee and host students who nominated x students. Then, we calculate the observed frequency of inter-ethnic ties based on the actual nominations in each classroom:

$$\tilde{\mu} = \frac{e_{RH} + e_{HR}}{e_{HR} + e_{RH} + e_{HH} + e_{RR}},$$

where e_{ij} denotes the number of edges from students with ethnicity i to students with ethnicity j and $i, j \in \{R, H\}$. Our measure of ethnic segregation ES_c in classroom c is:

$$ES_c = \mu_c - \tilde{\mu}_c.$$

Figure II depicts the cumulative distribution of the expected and observed proportion of inter-ethnic ties for all three categories of social ties (friendship, emotional support, academic support) at baseline. We observe substantial ethnic segregation for all three social tie categories at baseline. We expect the program to lower the distance between expected and observed inter-ethnic ties, i.e., classroom-level ethnic segregation.⁷

III.C. Experimentally Elicited Prosocial Behaviors

An essential feature of a cohesive environment is the prevalence of prosocial behavior in social interactions. Trust, reciprocity, cooperation, and altruism are the best-known prosociality indicators studied by economists in lab and field settings. We followed the convention and elicited these indicators using incentivized decision tasks in the following manner: In every classroom, the leading experimenter, assisted by field assis-

⁷We provide an illustration of our segregation index in the Online Appendix; see Figure E.2.

tants, informed the children that they would be playing four games.⁸ The experimenter showed the children a basket full of small gifts that are of value to them. These are small attractive stationery items, balls, key chains, hairpins, and more. The experimenter then told the children that in each of the four games, they would have an opportunity to earn “tokens.” The children were informed that these tokens could be converted to any gifts of their choice in the basket at the end of the visit, and more tokens meant more gifts. The experimenters carefully explained to children that one game would be randomly selected for the classroom at the end of the visit, and everyone would receive the tokens they earn from that particular game, i.e., tokens would not accumulate game after game. These four games are two versions of a trust game and two versions of a cooperation (prisoner’s dilemma) game. After these games, children played a version of a dictator game.

The trust game involves two participants that are anonymously paired (Berg et al., 1995). We designed this game to have two versions played within-subject. In the first version, which we refer to as “in-class,” each child is paired with an anonymous classmate. In the second version of the trust game, each child is instead paired with an unknown student from another school. We refer to this version as “out-school.”

In both versions, children are endowed with four tokens and there are two roles: a sender or a receiver. The game was designed using a strategy method such that students make decisions on how much to send if they assume the role of a sender and how much to send back (reciprocate) if they assume the role of the receiver.⁹ The sender must decide how many of his/her tokens to send to his/her anonymous classmate (the receiver). The amount the sender chooses to send, which may also be zero, is tripled by the experimenter and then given to the receiver. The receiver makes a similar choice â returning some amount of the now-tripled tokens to the sender, which may also be zero. The choice of the receiver is elicited for all 4 cases: the case of receiving 1 (tripled to 3), 2 (tripled to 6), 3 (tripled to 9), and 4 (tripled to 12). The amount of tokens sent is our measure of “trust,” and the amount of tokens sent back is our measure of “reciprocity.” More specifically, for each decision, we calculate the fraction of the tokens sent back and measure reciprocity for each child by taking the average of all four fractions. We expect the program to increase trust and reciprocity in children, both toward their classmates and out-school peers.

The cooperation game, which is a modified version of a one-shot prisoner’s dilemma game, also involves two participants to be anonymously paired. We similarly design this

⁸Children were also told that they are allowed to not participate in these activities, and even if they do participate, they can stop participating at any time they wish to do so. In practice, all students who were present on the day of the visit participated in the incentivized tasks.

⁹See Harbaugh et al. (2003) for a similar setup. Also, see Brandts and Charness (2011) for a review of papers that use the strategy method.

game and have in-class and out-school versions. Children are endowed with three gift tokens for this game. The game involves choosing a card that is either green or orange. A child's payoff depends on both the color she chooses and the color her pair chooses. The payoff scheme is given in Figure F.1 in the Online Appendix. If both choose orange, each remains with their initial endowments of 3 tokens. If they choose different colors, the one that chooses orange triples her tokens to 9, and her pair loses all her tokens. If both choose green, both double their endowments to 6. The cooperative action is to choose the green card. We refer to the binary choice of the green card as the "cooperative" action and expect that the program increases the probability of the cooperative action.

The reason we designed in-class as well as out-school versions for the trust and cooperation games is to explore possible welfare effects of the program. From the ex-post payoff perspective, while it may be optimal to trust and cooperate in a cohesive environment, such actions may disadvantage trusting/cooperating individuals in a generally non-cohesive environment where such behavior may be exploited. In our context, such a disadvantage would manifest itself as treated children collecting fewer tokens on average than children in the control group, especially in out-school games. We will explore this possibility by constructing expected payoffs using the empirical distribution of decisions.

After playing these four games, children played a dictator game. For this, students were given four tokens and asked whether they would like to donate some of their tokens to an anonymous child from another school we did not visit. We added a between-subject variation to this game: A random half of a given classroom received a question where the anonymous recipient's ethnicity was not referenced. The other half received a question where the anonymous recipient was stated as a Syrian refugee child. With this design, we can estimate the effect of the treatment on altruism and assess whether the treated children are more or less likely to consider recipients' ethnic identity when deciding to donate. We expect that the program increases the tendency to donate to both host and refugee children. Full instructions and procedures for the trust, cooperation and dictator games are given in Online Appendix F.

III.D. Self-Reported Outcomes for Mechanism Search

We also collected data from children on perspective taking, empathetic concern, impulsivity, and ethnic bias using item-response questions, both at baseline and endline. The primary motivation to collect these outcomes is to substantiate our conjectured mechanism as well as to detect or rule out other potential channels. We also measured descriptive classroom norms, but this data was only collected at endline. For this, children were asked item-response questions regarding their classmates' general behavioral

conduct. All of these self-reported questions are presented in Online Appendix G.

III.E. Achievement Outcomes

A healthy school environment is essential to ensure academic achievement. To test whether the program facilitated the integration of refugees without hurting host children academically, we implemented math and Turkish language tests in classrooms both at baseline and endline. We prepared these tests separately for 3rd and 4th graders, based on the national curricula. Because the program had no academic focus, we did not specify these outcomes in our PAP. Nevertheless, we analyze these data and present the estimated treatment effects on standardized math and Turkish verbal ability of host and refugee children. We also have access to teacher-assigned grades. However, they are of limited use for us as they are given on a narrow scale of 1 to 3, and teachers tend to grade to a known distribution (Alan et al., 2019).

IV DATA: DESCRIPTIVE STATISTICS AND INTERNAL VALIDITY

Before the randomization procedure, we visited all 80 schools (222 classrooms) and collected detailed baseline data on demographics, self-reported experiences of bullying, perspective-taking, empathetic concern, impulsivity, and ethnic bias. We also implemented math and Turkish language tests, measured fluid cognitive ability using Raven’s progressive matrices (Raven et al., 2004), and captured emotional intelligence using Reading the Mind in the Eyes Test (Eyes Test) (Baron-Cohen et al., 2001). The latter is commonly used to measure individual differences in theory of mind and shown to be weakly related to cognitive empathy and emotion perception and strongly related to emotion-based vocabulary (Olderbak et al., 2015). Finally, we elicited social networks and measured cooperation and altruism using the incentivized games at baseline. Except for fluid cognitive ability (Raven’s score), all these outcomes were also collected at endline. We added two versions of the trust game, a modified dictator game, and behavioral norms questionnaire to our endline inventory. We also collected important baseline information from teachers. In addition to various demographic information, we collected the following information from teachers: fluid cognitive ability (Raven’s Progressive Matrices), emotional intelligence (Eyes Test), and teaching styles (modern vs. traditional, warm vs. authoritarian, etc.). We collected these to explore possible treatment effect heterogeneity based on teacher characteristics. Table E.1 in the Online Appendix shows all our student outcomes and whether they were collected at baseline,

endline, or both.

Table I presents the balance of baseline variables across treatment status. The first panel presents the balance in student characteristics. The second panel presents balance in classroom and teacher characteristics, and the last panel shows the balance in school characteristics. Note first that about 16 percent of our sample consists of refugee children at baseline. The table shows no significant imbalance in any of the variables except for the proportion of students who reported being bullied by their classmates (significant at 10 percent level). As shown in Panel 3, the schools in our sample are of considerable size. The average number of 3rd and 4th-grade classrooms is about 15, with approximately 500 students. We also provide a balance table where we restrict our sample to students who were present both baseline and endline; see Table E.2 in the Online Appendix.

Figure III depicts the estimated ethnicity gaps at baseline with respect to self-reported bullying and social exclusion. The figure shows that refugee children are significantly more likely to be socially excluded and subject to regular bullying. They are about 4.0 (7.0) percentage points more likely to report experiences of bullying by their classmates (out-class schoolmates), 8.4 percentage points less likely to have a friend in their classroom, and 12.8 and 10.3 percentage points less likely to receive emotional and academic support from their classmates. In what follows, we will present the effect of the program on the cohesion outcomes we detailed above for all children. We will also present heterogeneity results by refugee status to see if the program benefits host and refugee children differently.

V RESULTS

V.A. Empirical Specification

We estimate the effect of the program on our cohesion outcomes using the empirical specification below:

$$y_{is} = \alpha_0 + \alpha_1 T_s + X'_{is} \gamma + \text{Other}_{is} + \delta_b + \varepsilon_{is}, \quad (1)$$

where y_{is} is the outcome of interest for child i in school s . T_s is the binary treatment indicator, which equals one if school s is in the treatment group and zero otherwise, and X'_{is} is a vector of observables for student i in school s that are predictive of the outcome y . The latter includes age, gender, refugee status, Raven's score, Eyes Test score, the outcome variable collected at baseline and a dummy for developmentally challenged

students.¹⁰ We also control for class size, school size, and district fixed effects. $Other_{is}$ captures other variables (for particular outcomes) that might be added for specific regressions, and δ_b are strata fixed effects. The estimated $\hat{\alpha}_1$ is the average treatment effect.

In the case of missing baseline data and complete endline data, we impute missing covariate values. For this, we replace missing values with their mean or median values. Our results are not sensitive to imputing the missing covariate values. More involved multiple imputation techniques also do not yield different results. As a robustness check, we present all our results without covariates in Online Appendix B. We also find that attrition is independent of treatment status (p-value=0.948), and the p-value for the estimated treatment effect on overall student absenteeism at endline is 0.414. These results are also robust to estimating the treatment effect on missing data for each separate outcome (see Table E.3 in the Online Appendix). Finally, because we test multiple hypotheses using a wide range of outcomes, we also provide Romano-Wolf p-values for our main outcomes in Online Appendix A.

The program requires teachers to cover all weekly topics throughout the academic year. Even though participation was voluntary and the program was oversubscribed, compliance in terms of actual implementation, i.e., coverage of the curriculum, may not have been perfect. To assess this, we asked treated teachers to report their estimated degree of curriculum coverage at endline. Specifically, we asked them to mark their coverage estimate using an unmarked 10-centimeter line, which gives us a continuous measure of program implementation intensity, albeit subjective. Figure E.3 in the Online Appendix depicts the distribution of the reported implementation intensity. Treated teachers report having covered about 60 percent of the program on average, with approximately 16 percent of teachers reporting no coverage at all. We were informed that low or no implementation is mainly due to teachers being involuntarily relocated by the MoE in the middle of the academic year. Such turnover is quite common in our study site. Given this imperfect compliance, the estimated $\hat{\alpha}_1$ should be interpreted as the average intent to treat effect (ITT). In what follows, we will focus on ITTs. We provide local average treatment effect (LATE) estimates for our primary outcomes in Online Appendix D.

¹⁰The Turkish MoE has an active policy to place a small number of students with some learning disabilities in classrooms to facilitate inclusion. These students, if present, were identified by teachers for us before we commenced our data collection. We took great care to include these children in the activities and often assigned an assistant to help them exclusively. Approximately 6 percent of our sample consists of these students, and this proportion is balanced across treatment status (p-value=0.528).

V.B. Treatment Effects on Peer Violence and Victimization

Panel 1 in Table II presents the estimated effects of the program on the number of high-intensity violent episodes recorded in the 10-day diary logs. Recall that the study sample covers, on average, three classrooms in each school. The first two columns use only the study sample. The first column presents the estimated treatment effect on the total number of episodes perpetrated by children from study classrooms within 10 school days. The second column shows the effect on the total number of episodes in which victimized children were from study classrooms. The third column presents the estimated effect for the whole school, and finally, the last column presents the estimated effect on non-study classrooms, therefore representing the program's spillover effect.

As can be seen in column 1, the program significantly reduced the number of violent events perpetrated by treated children. There are, on average, 1.88 events recorded in 10 days in control group classrooms. The treatment effect of 1.21 fewer events implies a substantial decline (about 64%). The second column in the table shows that the program also significantly reduced the number of events that victimized treated children. While the total number of victimizing events is 1.50 in control schools, it is about 50 percent lower (0.75 fewer events) in treatment schools. This result ensures that the program did not generate the undesired effects we mentioned earlier. That is, it did not make treated children more susceptible to victimization. Instead, the results suggest that the program, by keeping children away from conflict, lowered the risk of being a victim in a conflict. This result is consistent with the large raw positive correlation of 80% between the number of events perpetrated by someone and the number of events in which someone was victimized in the study classrooms.

Considering the sheer size of the schools, the program's effect on the entire school is striking. We estimate a substantial decline in the overall number of violent events in treated schools. As shown in column 3, the average number of violent episodes in ten days is 7.83 in control schools, and the program lowers this by 2.4 episodes. This effect is not precisely estimated, but the effect size is large. Note also that almost half of this overall reduction is coming from non-treated classrooms (column 4). These results are suggestive of spillover effects of the program within schools. There may be multiple channels contributing to this spillover effect, but given our context, three channels seem particularly intuitive: First, the program was oversubscribed, so that many teachers were turned down. It is plausible that some of these excluded teachers managed to get some of the program materials from their colleagues. Second, participating classrooms were provided with many visual materials such as posters and pamphlets for the participant children. These materials were quite visible and very likely observed by children's peers in nearby classrooms on the same floor. Finally, it is plausible that

spillovers are due to changes in interactions among students. For example, a bully in one classroom may team up with bullies and can influence their behavior. This would be consistent with the findings of a study by [Paluck et al. \(2016\)](#).

The program appears to be highly effective in reducing high-intensity peer violence and victimization in schools. A natural question is whether this is reflected in student reports of bullying and teacher reports of anti-social behavior. Recall that we asked students about their experiences of bullying at baseline and endline and that we also asked teachers to rate each student's general behavioral conduct at endline using a 1 to 5 grading scale. For the former, we construct a binary outcome, which takes the value of 1 if the child reports being bullied by his/her classmates (or out-class schoolmates), and zero otherwise. The latter is constructed as a standardized behavioral conduct score assigned by teachers with larger values indicating bad behavioral conduct. Given the differences between the administrative and self-reported measures of violence that was discussed in Section III.A., finding different treatment effects is entirely conceivable and was discussed in our PAP. It is plausible that treatment lowered the propensity to get involved in a severe conflict while also leading to higher sensitivity to anti-social behavior on the part of students and teachers. Our results are consistent with this line of reasoning. Panel 2 of Table II presents the estimated effects on self-reported bullying experience and teacher-reported behavior scores. In both cases, effect sizes are positive and imprecisely estimated.

Panel 3 of presents Table II estimated effects on self-reported bullying and teacher reports of behavioral conduct for hosts and refugees separately. Here, we see evidence, albeit weak, of differential treatment effects concerning the probability of being bullied by peers. While the program has no impact on host children's likelihood of being bullied by a classmate or a schoolmate, treated refugee children report lower likelihood of bullying than untreated refugee children. While the effects on refugee children are not statistically significant, the results indicate a treatment effect heterogeneity for the likelihood of being bullied by classmates, significant at the 5% level. The magnitude of this heterogeneous effect is large enough to close the gap observed in the control group and at baseline. We find no evidence of differential treatment effects concerning reported bullying by out-class schoolmates or teachers' behavioral conduct grades.

V.C. Treatment Effects on Social Exclusion and Ethnic Segregation

We now investigate the program's effect on social networks in the classroom, social exclusion, and ethnic segregation. Panel 1 in Table III presents the estimated average marginal effects of the program on the probability of having a friend and having a classmate who provides emotional or academic support. Note that most children (95%) in

the control group report having at least one friend in their classroom. The program has no effect on the probability of having a friend. However, treatment effects on emotional and academic support ties (columns 2 and 3) are positive and significant at the 1% level. Treated children are 4.6 percentage points (5.8%) more likely to have at least one classmate from which they receive emotional support and 5.6 percentage points (7.6%) more likely to have at least one classmate from which they receive academic support.

Panel 2 in Table III presents the estimated effects on the number of in-degree ties. These ties refer to the number of nominations a child receives in each category. Note first that an average child in the control group receives 2.35 friendship nominations, 1.76 emotional support nominations, and 1.51 academic support nominations. While the treatment has no effect on the number of friendship nominations received, it has significant effects on the number of nominations received in emotional and academic support categories. Treated children receive, on average, 0.10 (5.9%) more nominations than control children as emotional support providers for their classmates. Similarly, they receive about 0.13 (8.3%) more nominations as academic support providers. These results strongly suggest that the intervention increases the prevalence of support among classmates in personal and academic matters.

We then estimate the impact of the program on the level of ethnic segregation in the classroom. Recall that our ethnic segregation measure is constructed as the difference between the proportion of inter-ethnic ties expected to be formed at random and its observed counterpart at the classroom level. Panel 3 in Table III presents the estimated treatment effects on ethnic segregation, based again on friendship, emotional support, and academic support ties. We estimate a sizeable decline in ethnic segregation in classrooms. Our estimates amount to a 14.5% decline in segregation based on friendship ties, 16.6% reduction based on emotional support ties, and 21% reduction based on academic support ties relative to control classrooms. The estimated effects on friendship ties and academic support ties are statistically significant at the 10% level. While the estimated effect size is similar, the effect for emotional support ties is imprecisely estimated.

Table IV presents estimated treatment effects on social exclusion separately for host and refugee children. Panel 1 shows that the program is effective in mitigating the social exclusion of all children, hosts and refugees alike. The effect on refugee children is particularly strong with respect to the probability of having a friend (column 1). While the program does not affect the probability of having a friend for host children, it increases the probability of refugee children befriending at least one classmate by about 4.2 percentage points. We do not estimate statistically significant heterogeneity con-

cerning in-degree ties (Panel 2). Focusing only on ties formed with host children, Panel 3 of Table IV largely confirms the results presented in Panels 1 and 2. The program increases refugee children's likelihood of having a host friend by 6.6 percentage points, but this effect does not reach statistical significance. We estimate a statistically significant heterogeneous treatment effect concerning emotional support ties with hosts (at 10% level), but not concerning academic support ties. Results in Panel 3 suggest that refugee children in treated schools are approximately 25% (21%) more likely than those in control schools to receive emotional (academic) support from their host classmates.

V.D. Treatment Effects on Prosocial Behavior: Trust, Reciprocity, Cooperation and Altruism

Columns 1 and 2 in Panel 1 of Table V present the estimated treatment effects on trust, measured as the number of tokens sent (out of 4) to an anonymous receiver in the respondent's classroom (in-class) and to an anonymous receiver outside the respondent's school (out-school). Note first that about 1.38 and 1.45 tokens were sent in the control group to an anonymous in-class peer and out-school peer, respectively. This counter-intuitive difference is significant at the 10 percent level. This difference is entirely eliminated in the treatment group (p -value=0.95). The estimated treatment effects on trust are large and precise: Treated students sent about 0.27 extra tokens to an anonymous in-class peer and about 0.21 more tokens to an anonymous out-school peer, implying a 19.2% and 14.3% increase relative to the control group, respectively. While the effect size in the out-school case is smaller, the difference is not statistically significant (p -value=0.16).

Columns 3 and 4 of Panel 1 in Table V present the estimated treatment effects on reciprocity toward an in-class peer and an out-school peer, respectively. Here, recall that reciprocity was elicited based on all four scenarios of receiving 3, 6, 9, and 12 tokens in the trust game. We construct our dependent variable as the average of all four scenarios, that is, the average ratio of tokens sent back to the sender. Note first that about 39% of the tokens received were sent back to the sender in the control group in both the in-class and out-school versions. The estimated treatment effects are high in terms of size and precision: treated children sent back about 5 percentage points more tokens to the anonymous sender in their classroom relative to children in the control group. This corresponds to about 13 percent higher reciprocity toward in-class peers relative to the control group. The effect on out-school reciprocity is very similar, both in terms of size and precision.

Finally, the last two columns of Panel 1 in Table V present the results for the coop-

erative action in the cooperation game. Recall that this is a binary variable that takes a value of 1 if the child chooses to cooperate (choosing the green card) and zero otherwise (choosing the orange card). Observe that 52 percent of children in the in-class version and 50 percent in the out-school version decided to cooperate in the control group. While we estimate a statistically significant effect of the treatment for in-class cooperation (4.2 percentage points), we do not reject the null hypothesis for the out-school game.

Taken together, these results provide strong evidence that the program increased prosociality among children. Treated children exhibit higher trust and reciprocity toward their classmates, and toward out-school peers who they do not know. Before presenting the results on altruism, we pause here and ask an important question. Is this enhanced prosociality a desired outcome from a welfare perspective? A rigorous answer to this question requires an analysis involving well-specified utility functions. However, without invoking utility concerns, one can still infer children's welfare gains/losses using their ex-post payoffs as all our prosociality measures are obtained from incentivized tasks. It is easy to predict that classrooms where students collectively exhibit more trust and reciprocity toward each other end up earning more gifts. The question is whether the program inadvertently disadvantaged treated children from the perspective of payoffs by encouraging them to trust unknown out-school peers, who may not reciprocate their trust. We designed our out-school versions to be able to answer this question. To assess whether the program had such an unintended effect, we estimate its effect on expected payoff gains. To do this, we first calculate expected payoff gains from both the in-class and out-school trust and cooperation games. For in-class payoffs, we use the within-class empirical distribution of decisions.

Consider the in-class version of the trust game. The expected payoff P_{ic}^s of child i in class c in the case of being a sender is:

$$P_{ic}^s = E - S_i + E_c(R_j|S_i),$$

where E is the initial endowment, which is 4 gift tokens. S_i is the number of tokens i decides to send to her anonymous classmate j , which can take any integer value between zero and 4. $E_c(R_j|S_i)$ is the expected number of tokens reciprocated by j given the number of tokens sent by i . The expectation is taken using the empirical distribution of reciprocity decisions in classroom c .

The payoff P_{ic}^r of i in the case of being a receiver is:

$$P_{ic}^r = E + mS_j - R_i(S_j),$$

where $R_i(S_j)$ is the number of tokens sent back to the sender and $m = 3$ is the experi-

ment multiplier. Note that conditional on the sender’s decision, the receiver’s decision to reciprocate is strategic, and does not involve uncertainty. Given that student i has a 50 percent chance of being a sender or a receiver, her expected overall payoff P_{ic} is:

$$P_{ic} = 0.5P_{ic}^s + 0.5P_{ic}^r. \quad (2)$$

To calculate the expected payoffs for senders and receivers in the out-school version of the game, we use the empirical distribution of decisions from the control group schools in child i ’s district.

Panel 2 in Table V presents the estimated treatment effects on expected payoffs for the in-class and out-school trust games, respectively. In addition to overall payoffs (Equation 2), we also present results for the role of a sender and receiver separately. Note first that in both games, children in the control group ended up gaining 5.38 and 5.46 tokens in the in-class and out-school games, respectively (see columns 3 and 6). Given the lower bound of 4 tokens (no trust condition), this number indicates that children increased the size of the gift pie by exhibiting some trust. Treated students ended up gaining, on average, 0.26 more tokens by trusting and reciprocating more in the in-class game (column 3). However, they ended up with 0.11 fewer tokens than the students in the control group in the out-school game. Even though the latter result seems to indicate a disadvantage on the part of the treated children, a closer look at the results reveals an interesting detail for this outcome. We estimate a zero treatment effect in the case of being a sender in the out-school game (column 4). However, we estimate a large and statistically significant difference in the case of being a receiver: Treated receivers gave up about 0.22 more gifts to reciprocate to their out-school senders. Note that, contrary to the trust decision made by the sender, the reciprocity decision does not involve uncertainty. The fact that treated children deliberately lowered their payoffs to reciprocate out-school senders implies that the lower overall amount of payoffs they obtained (column 6) may be sub-optimal from the expected payoff perspective, but not necessarily from a welfare point of view if the underlying utility function incorporates other-regarding preferences. This result implies that the program increases the tendency to reciprocate the kindness children receive from out-school peers at the expense of their own payoffs. We explore the cooperation decisions in a similar way. Columns 7 and 8 of Panel 2 present the related results. Treated children received 0.13 more tokens than children in the control group by exhibiting more in-class cooperation. We estimate a zero payoff difference across treatment status for the out-school cooperation game.

Our final incentivized cohesion indicator is altruism, measured by a dictator game. As explained in Section III.C., we implemented two versions of this game using a

between-subject design. For a random half of the classroom, the ethnicity (Syrian refugee) of the anonymous receiver was revealed. The other half received no reference to the recipient’s ethnicity. Here, in addition to estimating the program’s impact on overall altruism, we want to assess whether the treatment affects the donation patterns based on the recipient’s ethnicity. To do this, we estimate the following regression:

$$y_{is} = \alpha_0 + \alpha_1 T_s + \alpha_2 S_i + \alpha_3 T_s * S_i + X'_{is} \gamma + \delta_b + \varepsilon_{is}, \quad (3)$$

where y_{is} is either the probability of donating or the fraction of the endowment donated, T_s is the treatment indicator for school s , and S_i is an indicator that child i received the donation question with the explicit reference to the recipient’s ethnicity. In this specification, the coefficient $\hat{\alpha}_1$ is the estimated treatment effect on donation to an anonymous recipient. The estimated coefficient $\hat{\alpha}_3$ is the additional donation the treated children make to an anonymous Syrian child.

Table VI presents the estimated treatment effects on the willingness to donate (average marginal effects) and the fraction of the endowment (4 tokens) donated. As the first column shows, treated children are 6.9 percentage points more likely to make a donation, and this value increases by another 3.4 percentage points if the anonymous recipient is revealed to be a Syrian refugee child. Similarly, the fraction donated to an anonymous peer is 5.3 percentage points higher among treated children, but the fraction of the endowment donated does not significantly increase when the anonymous recipient’s ethnicity is revealed.

As a final note on prosociality results, we detect heterogeneity only for out-school trust and reciprocity. As seen in Table VII, while we estimate a significant increase in out-school trust and reciprocity for host children, we estimate null effects for refugees. However, the results show that among the control group, refugees are more trusting and reciprocate more, including with out-school students. This gap is reduced among refugees in the treatment group, but only for the out-school version of the trust game. Finally, we do not estimate any statistically significant heterogeneous effect of the program with respect to cooperation (columns 5 and 6) or altruism (column 7).

V.E. Treatment Effect on Achievement

Table VIII presents the estimated treatment effects on standardized math and Turkish test scores for both host and refugee children. While we estimate null effects on both math and Turkish scores for host children, we estimate a large and significant effect on Turkish test scores for refugee children: The effect size is 0.13 standard deviations and significant at the 5% level. We do not estimate a significant treatment effect on the

math scores of refugee children.¹¹

To summarize, the program seems to be highly effective in i) reducing the frequency of high-intensity peer violence and victimization on the school ground, ii) reducing social exclusion and ethnic segregation in the classroom and, iii) increasing trust, reciprocity, and cooperation among students, as well as their altruism toward one another. Moreover, these positive results on prosocial behaviors are not limited to behaviors toward classmates but extend to anonymous out-school peers. Even more promising is that the program seems to have lessened the social exclusion of refugee children significantly and helped them form friendship ties and receive emotional and academic support from their classmates. Given these results, perhaps it is not surprising that we estimate a striking improvement in refugee children's test scores in the Turkish language.¹²

In the next section, we explore possible channels through which the program generates these promising results. Although we substantiate our claims using data, we caution that the analyses in what follows remain mostly suggestive.

VI POTENTIAL MECHANISMS

The objective of the program is to build a cohesive school environment by improving children's ability to take others' perspectives, especially in cases of conflict, maltreatment, and social exclusion. The curricular module provides children with examples of different social situations and asks them to evaluate the perspectives of the involved parties critically. Students are strongly encouraged to exert effort to understand and articulate the individual point of view in a given social context, whether or not they agree with the involved (hypothetical) individuals. Given its strong and repeated emphasis on this type of deliberation, we conjecture that the program is likely to achieve its objectives through increasing children's effort to take others' perspectives.

While we hypothesize perspective-taking to be a likely mechanism, we acknowledge that there may be other channels. For example, the program may increase cohesion by changing classroom norms regarding acceptable and unacceptable behavior. It may also do so by invoking children's empathetic concern (compassion) toward others. Yet, as another mechanism, the program may reduce conflict and victimization by encouraging children to better manage their impulsivity, a characteristic that is often responsible

¹¹Our analysis of teacher-assigned grades yield similar but imprecise results. As mentioned in Section III.E., grades are recorded on a narrow range (1 to 3). We find that the program reduces the probability of receiving a failing grade (1) both in math and Turkish for refugee students, but neither estimates reach statistical significance. The estimated effects on receiving a top grade (3) are similar.

¹²For more heterogeneity results, see our Online Appendix C and E.

for the escalation of disputes and can be controlled by engaging in deeper deliberation. Moreover, the program may improve cohesion by increasing tolerance toward individual differences, including ethnic and cultural differences, thereby reducing ethnic bias. Finally, we re-implemented the Eyes Test at endline to assess whether the program improved children’s emotional intelligence. However, given the evidence on the high test-retest reliability of this test, we do not expect any improvement in children’s scores over an academic year (e.g., see [Fernández-Abascal et al., 2013](#); [Vellante et al., 2013](#)).

Figure IV depicts the estimated treatment effects and 95 percent confidence intervals on behavioral norms, ethnic bias, perspective taking, impulsivity, empathetic concern, and Eyes Test scores. We plot the estimates separately for host and refugee students. The first thing that stands out in this picture is that perspective-taking ability is likely to be a common mechanism responsible for the strong treatment effects we estimate for both ethnicities. The estimated effects on perspective-taking are almost identical: 0.27 standard deviations for treated host children and 0.28 standard deviations for treated refugee children. Apart from perspective-taking, we observe that most of the improvements regarding other contemplated channels come solely from refugee students. Notably, we estimate a significant 0.17 standard deviation effect on behavioral norms perceived by treated refugee students without a corresponding effect on host students. This result is consistent with the finding that treated refugee students report significantly lower in-class bullying than untreated ones. We also estimate improvements in impulsivity, ethnic bias, empathetic concern, and emotional intelligence, although these effects are less precisely estimated. Interestingly, we detect an improvement in the Eyes Test scores of the treated refugee students. However, we are cautious about interpreting this as an improvement in emotional intelligence as this effect may be driven by refugee students’s progress in the Turkish language.¹³

Overall, the above analysis points to perspective-taking as a robust mechanism partly responsible for the treatment effects we estimate for both ethnicities. However, it is important to reiterate that while suggesting that perspective-taking is a likely mechanism, we remain agnostic about other potential channels that might have contributed to our results.

VII CONCLUDING REMARKS

We evaluate the effectiveness of an educational program that aims to build social cohesion in ethnically mixed schools. The program is implemented in Turkish elemen-

¹³The effect disappears when we control for Turkish test scores; see [Olderbak et al. \(2015\)](#) for a discussion on vocabulary and the Eyes Test.

tary schools where the ethnic composition has changed rapidly due to a fast influx of refugee children. The program involves covering a curriculum implemented by teachers for at least 3 hours per week. To evaluate the program, we consider a wide range of cohesion indicators, including peer violence and victimization, social exclusion, ethnic segregation, and prosocial behaviors. We find that the program significantly lowers high-intensity peer violence and victimization on school grounds. It also reduces social exclusion and lowers ethnic segregation in the classroom. Treated children exhibit higher trust and reciprocity toward their peers, cooperate more, and show higher altruistic tendencies. Prosocial behavior improves both toward classmates and anonymous out-school peers. Finally, we show that the program leads to a considerable improvement in the refugee children's ability in the host country's language.

The program is remarkably cost-effective relative to well-known educational interventions. It is difficult to gauge the individual and societal value of reducing violence, social exclusion, and ethnic segregation in schools. However, even if one considers only the learning gains of refugee children with respect to the host country's language, the program can be viewed as a success. The education materials were developed as part of a private university's philanthropic endeavor, and as such, they are now a public good. The remaining program costs pertain to printing hard copy materials (the UEO book and activity kits, which were also made available online), distributing the materials to schools, and conducting teacher training. For 4249 officially registered treated children, the printing costs were about 20,000 USD, the distribution costs 7,000 USD, and teacher training costs were about 8,000 USD. These values imply an 8.2 USD program cost per child, which is negligible compared to the cost of any known large-scale educational intervention.

The program was implemented and evaluated in a specific context; therefore, we refrain from claiming external validity. However, noting that most teachers were eager to join the program in all participating schools, compelling us to randomly choose among teachers, we expect that the program would show a fair degree of success at scale in Turkey. While being cautious about external validity, we view the study as a significant step toward understanding the causal role of public education in building social capital. It provides much-needed evidence to design effective cohesion policies in contexts of social tensions that are rooted in ethnic diversity.

We would like to end our concluding remarks with two caveats. First, we acknowledge that a shortcoming of the present study is that we only observe short-term effects. A next step is to determine whether the benefits of the program persist. Second, even if we consider public education as the primary route to a more inclusive society, we are under no illusion that a curricular approach we promote here is the only way to achieve

this end. In addition to curricular reforms, we need fundamental changes in the culture of teaching and pedagogy while promoting social contact early in childhood. Future research would benefit from testing the effectiveness of these changes in contexts where there is a high degree of inter-ethnic contact and a need to counteract tensions.

EUROPEAN UNIVERSITY INSTITUTE, BILKENT UNIVERSITY, AND J-PAL
UNIVERSITY OF ESSEX
KOÇ UNIVERSITY
UNIVERSITY OF ESSEX

SUPPLEMENTARY MATERIAL

An Online Appendix to this article as well as the data and code replicating the tables and figures in this article can be found at *The Quarterly Journal of Economics* online.

REFERENCES

- Adida, Claire L., Adeline Lob, and Melina R. Platas (2018) “Perspective Taking can Promote Short-term Inclusionary Behavior Toward Syrian refugees,” *PNAS*, 115 (38), 9521–9526.
- Alan, Sule, Teodora Boneva, and Seda Ertac (2019) “Ever Failed, Try again, Succeed Better: Results from a Randomized Educational Intervention on Grit,” *The Quarterly Journal of Economics*, 134 (3), 1121–1162.
- Alan, Sule and Seda Ertac (2018) “Fostering Patience in the Classroom: Results from a Randomized Educational Intervention,” *Journal of Political Economy*, 126 (5), 1865–1911.
- Alesina, Alberto and Eliana La Ferrara (2005) “Ethnic Diversity and Economic Performance,” *Journal of Economic Literature*, 43 (3), 762–800.
- Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo (2019) “Nation-building through Compulsory Schooling during the Age of Mass Migration,” *The Economic Journal*, 129 (617), 62–109.
- Baron-Cohen, Simon, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb (2001) “The ‘Reading the Mind in the Eyes’ Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism,” *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42 (2), 241–251.
- Batson, C Daniel, Marina P Polycarpou, Eddie Harmon-Jones, Heidi J Imhoff, Erin C Mitchener, Lori L Bednar, Tricia R Klein, and Lori Highberger (1997) “Empathy and attitudes: Can Feeling for a Member of a Stigmatized Group Improve Feelings Toward the Group?” *Journal of Personality and Social Psychology*, 72 (1), 105.
- Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong (2019) “Unity in Diversity? How Intergroup Contact can Foster Nation Building,” *American Economic Review*, 109 (11), 3978–4025.
- Berg, Joyce, John Dickhaut, and Kevin McCabe (1995) “Trust, Reciprocity, and Social History,” *Games and Economic Behavior*, 10 (1), 122–142.
- Blattman, Christopher, Julian C Jamison, and Margaret Sheridan (2017) “Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia,” *American Economic Review*, 107 (4), 1165–1206.

- Boisjoly, Johanne, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles (2006) “Empathy or Antipathy? The Impact of Diversity,” *American Economic Review*, 96 (5), 1890–1905.
- Brandts, Jordi and Gary Charness (2011) “The Strategy versus the Direct-response Method: A First Survey of Experimental Comparisons,” *Experimental Economics*, 14 (3), 375–398.
- Broockman, David and Joshua Kalla (2016) “Durably Reducing Transphobia: A Field Experiment on Door-to-door Canvassing,” *Science*, 352 (6282), 220–224.
- Bruneau, Emile G. and Rebecca Saxe (2012) “The Power of Being Heard: The Benefits of ‘Perspective-giving’ in the Context of Intergroup conflict,” *Journal of Experimental Social Psychology*, 48 (4), 855–866.
- Bruneau, Emile, Nir Jacoby, Nour Kteily, and Rebecca Saxe (2018) “Denying Humanity: The Distinct Neural Correlates of Blatant Dehumanization,” *Journal of Experimental Psychology: General*, 147 (7), 1078–1093.
- Burns, Justine, Lucia Corno, and Eliana La Ferrara (2019) “Interaction, Prejudice and Performance: Evidence from South Africa,” IFS Working Papers W19/03.
- Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden (2020) “The Effect of Early-Childhood Education on Social Preferences,” *Journal of Political Economy*, 128 (7), 000–000.
- Card, David and Jesse Rothstein (2007) “Racial Segregation and the Black–White Test Score Gap,” *Journal of Public Economics*, 91 (11-12), 2158–2184.
- Deming, David (2009) “Early Childhood Intervention and Life-cycle Skill Development: Evidence from Head Start,” *American Economic Journal: Applied Economics*, 1 (3), 111–34.
- Deming, David J (2011) “Better Schools, Less Crime?” *The Quarterly Journal of Economics*, 126 (4), 2063–2115.
- Durkheim, Emile (1897) *Le Suicide: Étude de Sociologie*: Alcan.
- Easterly, William, Jozef Ritzen, and Michael Woolcock (2006) “Social Cohesion, Institutions, and Growth,” *Economics & Politics*, 18 (2), 103–120.
- Echenique, Federico, Roland G Fryer Jr, and Alex Kaufman (2006) “Is School Segregation Good or Bad?” *American Economic Review*, 96 (2), 265–269.

- Eisner, Manuel, Denis Ribeaud, Giuseppe Sorrenti, and Ulf Zölitz (2020) “The Causal Impact of Socio-Emotional Skills Training on Educational Success,” *CEPR Discussion Paper No. DP14523*.
- Fehr, Ernst and Klaus M Schmidt (1999) “A Theory of Fairness, Competition, and Cooperation,” *The Quarterly Journal of Economics*, 114 (3), 817–868.
- (2000) “Fairness, Incentives, and Contractual Choices,” *European Economic Review*, 44 (4-6), 1057–1068.
- Fernández-Abascal, Enrique G, Rosario Cabello, Pablo Fernández-Berrocal, and Simon Baron-Cohen (2013) “Test-retest Reliability of the ‘Reading the Mind in the Eyes’ Test: A One-year Follow-up Study,” *Molecular Autism*, 4 (1), 33.
- Fryer Jr, Roland G and Steven D Levitt (2004) “Understanding the Black-White Test Score Gap in the First Two Years of School,” *Review of Economics and Statistics*, 86 (2), 447–464.
- Fryer Jr, Roland G and Glenn C Loury (2013) “Valuing Diversity,” *Journal of Political Economy*, 121 (4), 747–774.
- Galinsky, Adam D and Gillian Ku (2004) “The Effects of Perspective-taking on Prejudice: The Moderating Role of Self-evaluation,” *Personality and Social Psychology Bulletin*, 30 (5), 594–604.
- Galinsky, Adam D, Gillian Ku, and Cynthia S Wang (2005) “Perspective-taking and Self-other overlap: Fostering Social bonds and Facilitating Social Coordination,” *Group Processes & Intergroup Relations*, 8 (2), 109–124.
- Galinsky, Adam D and Gordon B Moskowitz (2000) “Perspective-taking: Decreasing Stereotype Expression, Stereotype Accessibility, and In-group Favoritism.,” *Journal of Personality and Social Psychology*, 78 (4), 708.
- Gould, Eric D, Victor Lavy, and M Daniele Paserman (2004) “Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel,” *The Quarterly Journal of Economics*, 119 (2), 489–526.
- Gradstein, Mark and Moshe Justman (2002) “Education, Social Cohesion, and Economic Growth,” *American Economic Review*, 92 (4), 1192–1204.
- Guryan, Jonathan (2004) “Desegregation and Black Dropout Rates,” *American Economic Review*, 94 (4), 919–943.

- Hanushek, Eric A, John F Kain, and Steven G Rivkin (2009) “New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement,” *Journal of Labor Economics*, 27 (3), 349–383.
- Harbaugh, William T, Kate Krause, and Lise Vesterlund (2003) “Trust in Children,” *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, 302–322.
- Harris, Lasana T. and Susan T. Fiske (2009) “Social Neuroscience Evidence for Dehumanised Perception,” *European Review of Social Psychology*, 20 (1), 192–231.
- Heckman, James J, Rodrigo Pinto, and Peter Savelyev (2013) “Understanding the Mechanisms Through which an Influential Early Childhood Program Boosted Adult Outcomes,” *American Economic Review*, 103 (6), 2052–86.
- Heckman, James J, Jora Stixrud, and Sergio Urzua (2006) “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24 (3), 411–482.
- Heller, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A Pollack (2017) “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago,” *The Quarterly Journal of Economics*, 132 (1), 1–54.
- Hendren, Nathaniel and Ben Sprung-Keyser (2020) “A Unified Welfare Analysis of Government Policies,” *The Quarterly Journal of Economics*, 135 (3), 1209–1318.
- Hjort, Jonas (2014) “Ethnic Divisions and Production in Firms,” *The Quarterly Journal of Economics*, 129 (4), 1899–1946.
- Kalla, Joshua and David Broockman (2020a) “Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments,” *American Political Science Review*, 114 (2), 410–425.
- (2020b) “Which Narrative Strategies Durably Reduce Prejudice? Evidence from Field and Survey Experiments Supporting the Efficacy of Perspective-getting.”
- Kanske, Philipp (2018) “The Social Mind: Disentangling Affective and Cognitive Routes to Understanding Others,” *Interdisciplinary Science Reviews*, 43 (2), 115–124.
- Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk (2020) “The Formation of Prosociality: Causal Evidence on the Role of Social Environment,” *Journal of Political Economy*, 128 (2), 434–467.

- Lochner, Lance and Enrico Moretti (2004) “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-reports,” *American Economic Review*, 94 (1), 155–189.
- Lowe, Matt (2020) “Types of Contact: A field Experiment on Collaborative and Adversarial Caste Integration,” CESifo Working Paper No. 8089.
- Miguel, Edward and Mary Kay Gugerty (2005) “Ethnic Diversity, Social Sanctions, and Public Goods in Kenya,” *Journal of Public Economics*, 89 (11-12), 2325–2368.
- Mousa, Salma (2020) “Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq,” *Science*, 369 (6505), 866–870.
- Olderbak, Sally, Oliver Wilhelm, Gabriel Oлару, Mattis Geiger, Meghan W Brenneman, and Richard D Roberts (2015) “A Psychometric Analysis of the ‘Reading the Mind in the Eyes’ Test: Toward a Brief Form for Research and Applied Settings,” *Frontiers in Psychology*, 6, 1503.
- Paluck, Elizabeth Levy, Seth A Green, and Donald P Green (2019) “The Contact Hypothesis Re-evaluated,” *Behavioural Public Policy*, 3 (2), 129–158.
- Paluck, Elizabeth Levy, Hana Shepherd, and Peter M. Aronow (2016) “Changing Climates of Conflict: A Social Network Experiment in 56 Schools,” *Proceedings of the National Academy of Sciences*, 113 (3), 566–571.
- Postman, Neil (1996) *The End of Education: Redefining the Value of School*, New York: Vintage Books.
- Putnam, Robert (1993) “The Prosperous Community: Social Capital and Public Life,” *The American Prospect*, 13, 35–42.
- Rao, Gautam (2019) “Familiarity Does not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools,” *American Economic Review*, 109 (3), 774–809.
- Raven, J, J. C Raven, and J. H Court (2004) *Manual for Raven’s Progressive Matrices and Vocabulary Scales*, San Antonio, TX: Pearson.
- Rodrik, Dani (1999) “Where Did all the Growth Go? External Shocks, Social Conflict, and Growth Collapses,” *Journal of Economic Growth*, 4 (4), 385–412.
- Schelling, Thomas C (1969) “Models of Segregation,” *American Economic Review*, 59 (2), 488–493.

- Stietz, Julia, Emanuel Jauk, Sören Krach, and Philipp Kanske (2019) “Dissociating Empathy From Perspective-Taking: Evidence From Intra-and Inter-Individual Differences Research,” *Frontiers in Psychiatry*, 10, 126.
- Vellante, Marcello, Simon Baron-Cohen, Mariangela Melis, Matteo Marrone, Donatella Rita Petretto, Carmelo Masala, and Antonio Preti (2013) “The ‘Reading the Mind in the Eyes’ Test: Systematic Review of Psychometric Properties and a Validation Study in Italy,” *Cognitive Neuropsychiatry*, 18 (4), 326–354.
- Voiglaender, N, S Becker, I Grosfeld, P Grosjean, and E Zhuravskaya (2020) “Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers,” *American Economic Review*.

TABLES

Table I BALANCE AT BASELINE

	N	Control Mean	Treatment Mean	Difference (T-C)	SE	p-value
Student Demographics:						
Male	7487	0.511	0.509	-0.002	0.012	0.868
Age in Months	7417	105.634	106.246	0.578	0.300	0.575
Refugee	7487	0.162	0.150	-0.008	0.008	0.701
Working Mother	5405	0.274	0.293	0.019	0.012	0.277
Working Father	5266	0.864	0.857	-0.007	0.010	0.590
Cognitive Tests:						
Raven Score	6084	0.002	-0.048	-0.067	0.026	0.349
Eyes Test Score	6135	0.002	-0.002	-0.020	0.026	0.752
Math Score	6135	0.002	0.025	0.008	0.026	0.930
Turkish Score	6134	0.001	0.011	-0.009	0.025	0.911
Cohesion Indicators:						
Proportion Bullied by Peers in Classroom	5638	0.838	0.862	0.026*	0.010	0.060
Proportion Bullied by Peers in School	5676	0.797	0.803	0.010	0.011	0.446
Fraction Donated	5990	-0.001	-0.026	-0.021	0.026	0.637
Willingness to Donate	5990	0.647	0.619	-0.026	0.013	0.285
Proportion Cooperate	5918	0.536	0.565	0.028	0.013	0.124
Perspective Taking	5793	-0.000	0.010	0.006	0.026	0.816
Empathetic Concern	5796	0.001	-0.006	-0.013	0.027	0.741
Ethnic Bias	5680	-0.002	-0.030	-0.025	0.027	0.408
Impulsivity	5693	-0.001	0.031	0.039	0.027	0.301
Having a Friend	6135	0.920	0.924	0.004	0.007	0.817
Having Emotional Support	6135	0.661	0.648	-0.016	0.012	0.659
Having Academic Support	6135	0.557	0.545	-0.013	0.013	0.678
Friendship Ties (in-degree)	7487	1.754	1.787	0.015	0.045	0.824
Emotional Support Ties (in-degree)	7487	1.050	1.015	-0.052	0.032	0.467
Academic Support Ties (in-degree)	7487	0.836	0.821	-0.024	0.030	0.689

	N	Control Mean	Treatment Mean	Difference (T-C)	SE	p-value
Classroom Size	222	33.041	34.266	1.238	0.929	0.405
Refugee Share	222	0.162	0.155	-0.005	0.015	0.824
Ethnic Segregation in Friendship Ties	213	0.089	0.102	0.013	0.012	0.364
Ethnic Segregation in Emotional Support Ties	208	0.079	0.089	0.010	0.013	0.446
Ethnic Segregation in Academic Support Ties	209	0.077	0.078	0.000	0.014	0.986
Teacher Age in Years	221	34.786	34.683	-0.193	1.205	0.888
Male Teacher	222	0.429	0.419	-0.012	0.067	0.884
Teacher Years of Experience	220	10.464	10.625	0.047	1.133	0.968
Tenured Teacher	220	0.875	0.863	-0.014	0.046	0.801
Teacher Raven Score	222	0.000	0.028	0.021	0.132	0.893
Teacher Eyes Test Score	222	-0.000	0.059	0.061	0.128	0.678

	N	Control Mean	Treatment Mean	Difference (T-C)	SE	p-value
School Size (3rd and 4th grades only)	80	483.300	477.100	3.303	60.794	0.956
Total Number of 3rd and 4th-grade Classrooms	80	14.650	14.050	-0.391	1.562	0.802

Reported statistics use the baseline sample. Panel 1 presents the balance of individual-level variables collected from children. Panel 2 presents the balance of classroom and teacher characteristics and Panel 3 school characteristics. All cognitive tests and survey measures are standardized. Standard errors and p-values are obtained by controlling for randomization strata. In Panels 1 and 2, standard errors are clustered at the school level (unit of randomization). Panel 3 uses robust standard errors.

Table II TREATMENT EFFECTS ON VIOLENCE AND ANTI-SOCIAL BEHAVIOR

PANEL 1: PEER VIOLENCE AND VICTIMIZATION-DIARY RECORDS

	Perpetrated	Victimized	Total Events	Spillover
Treatment	-1.212*** (0.440)	-0.750** (0.366)	-2.401 (1.948)	-1.189 (1.804)
Control Mean	1.88	1.5	7.83	5.95
Observations	80	80	80	80

PANEL 2: STUDENT AND TEACHER REPORTS OF VIOLENCE AND ANTISOCIAL BEHAVIOR

	Bullying In-Class	Bullying Out-Class	Teacher Behavioral Grade
Treatment	0.013 (0.016)	0.001 (0.014)	0.070 (0.057)
Control Mean	0.79	0.75	-0.02
Observations	6335	6371	6034

PANEL 3: HETEROGENEOUS TREATMENT EFFECTS ON STUDENT AND TEACHER REPORTS OF VIOLENCE AND ANTISOCIAL BEHAVIOR

	Bullying In-Class	Bullying Out-Class	Teacher Behavioral Grade
Treatment (Hosts)	0.023 (0.016)	0.010 (0.015)	0.069 (0.055)
Treatment (Refugees)	-0.039 (0.028)	-0.041 (0.029)	0.075 (0.106)
p-value : Hosts=Refugees	0.039	0.107	0.952
Control Mean (Hosts)	0.77	0.73	-0.03
Control Mean (Refugees)	0.85	0.81	0.01
Observations	6335	6371	6034

Reported estimates are obtained from ordinary least squares (OLS) regressions. In Panel 1, the outcome variables are obtained from 10-day diary logs filled by designated school administrators, and they are all school-level measures. The dependent variable includes, within 10 days, the total number of violent events perpetrated by a student from study classes, the total number of violent events in which victimized children are from study classes, and the total number of school-wide violent events in columns 1, 2, and 3, respectively. The final column removes the events perpetrated by study classes from total school-wide events and estimates the effect on non-study classes. In Panel 2, the dependent variables in columns 1 and 2 are dummy variables, which equal 1 if the student reports physical and verbal bullying in the classroom and out of the classroom, respectively. In column 3, the dependent variable is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. In Panel 1, regressions control for school size, the number of participating classrooms in a given school, school level refugee share, district and province fixed effects and randomization strata, and robust standard errors are in parentheses. In Panel 2, regressions control for relevant baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. Panel 3 presents the results in Panel 2, separately for hosts and refugees. Standard errors are in parenthesis and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table III TREATMENT EFFECTS ON SOCIAL EXCLUSION AND ETHNIC SEGREGATION IN THE CLASSROOM

<i>PANEL 1: SOCIAL EXCLUSION-BINARY</i>			
	Friendship	Emotional Support	Academic Support
Treatment	0.003 (0.005)	0.046*** (0.013)	0.056*** (0.013)
Control Mean	0.95	0.80	0.73
Observations	6642	6642	6642
<i>PANEL 2: SOCIAL EXCLUSION-IN-DEGREE TIES</i>			
	Friendship	Emotional Support	Academic Support
Treatment	-0.055 (0.043)	0.104** (0.046)	0.125** (0.054)
Control Mean	2.35	1.76	1.51
Observations	6642	6642	6642
<i>PANEL 3: ETHNIC SEGREGATION</i>			
	Friendship	Emotional Support	Academic Support
Treatment	-0.016* (0.009)	-0.015 (0.011)	-0.021* (0.011)
Control Mean	0.11	0.09	0.10
Observations	218	218	218

Reported estimates are obtained from ordinary least squares (OLS) regressions. In Panel 1, binary dependent variables are column 1: reported to have a friend, column 2: reported to have a classmate providing emotional support, column 3: reported to have a classmate providing academic support. In Panel 2, the dependent variables are the number of in-degree ties. Both Panel 1 and Panel 2 regressions control for respective baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. In Panel 3, the dependent variables are class-level segregation scores. These regressions control for randomization strata and classroom level baseline covariates. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table IV HETEROGENEOUS TREATMENT EFFECTS ON SOCIAL EXCLUSION

PANEL 1: SOCIAL EXCLUSION-BINARY

	Friendship	Emotional Support	Academic Support
Treatment (Hosts)	-0.005 (0.004)	0.038*** (0.014)	0.056*** (0.014)
Treatment (Refugees)	0.042* (0.022)	0.088** (0.034)	0.060** (0.029)
p-value : Hosts=Refugees	0.053	0.176	0.890
Control Mean (Hosts)	0.97	0.82	0.75
Control Mean (Refugees)	0.87	0.67	0.65
Observations	6642	6642	6642

PANEL 2: SOCIAL EXCLUSION-IN-DEGREE TIES

	Friendship	Emotional Support	Academic Support
Treatment (Hosts)	-0.064 (0.053)	0.095* (0.050)	0.123** (0.057)
Treatment (Refugees)	-0.012 (0.100)	0.148 (0.100)	0.133 (0.090)
p-value : Hosts=Refugees	0.669	0.625	0.917
Control Mean (Hosts)	2.49	1.89	1.63
Control Mean (Refugees)	1.72	1.16	1.01
Observations	6642	6642	6642

PANEL 3: SOCIAL TIES WITH HOST CHILDREN-BINARY

	Host Friendship	Host Emotional Support	Host Academic Support
Treatment (Hosts)	-0.006 (0.007)	0.038** (0.015)	0.055*** (0.016)
Treatment (Refugees)	0.066 (0.054)	0.118*** (0.043)	0.098** (0.040)
p-value : Hosts=Refugees	0.213	0.096	0.304
Control Mean (Hosts)	0.96	0.81	0.73
Control Mean (Refugees)	0.62	0.47	0.47
Observations	6642	6642	6642

Reported estimates are obtained from ordinary least squares (OLS) regressions. In Panel 1, binary dependent variables are column 1: reported to have a friend, column 2: reported to have a classmate providing emotional support, column 3: reported to have a classmate providing academic support. In Panel 2, the dependent variables are the number of in-degree ties. In Panel 3, the dependent variables are the same as Panel 1, but only for ties with host children. All regressions control for relevant baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table V TREATMENT EFFECTS ON PROSOCIAL BEHAVIOR AND EXPECTED PAYOFFS

PANEL 1: TRUST, RECIPROCITY AND COOPERATION

	Trust		Reciprocity		Cooperation	
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School
Treatment	0.265*** (0.058)	0.207*** (0.050)	0.050*** (0.013)	0.048*** (0.014)	0.042*** (0.015)	-0.002 (0.009)
p-val	0.156		0.616		0.014	
Control Mean	1.38	1.45	0.39	0.39	0.52	0.50
Observations	6476	6512	6534	6523	6568	6573

PANEL 2: PAYOFFS FROM TRUST AND COOPERATION GAMES

	Trust In-Class			Trust Out-School			Cooperation	
	Sender	Receiver	Overall	Sender	Receiver	Overall	In-Class	Out-School
Treatment	0.288*** (0.064)	0.239*** (0.071)	0.260*** (0.059)	-0.002 (0.048)	-0.219*** (0.067)	-0.112** (0.046)	0.132*** (0.045)	-0.008 (0.043)
Control Mean	4.30	6.47	5.38	4.28	6.63	5.46	4.57	4.50
Observations	6476	6412	6361	6512	6381	6372	6568	6573

Reported estimates are obtained from ordinary least squares (OLS) regressions. In Panel 1, the dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to an anonymous classmate and an anonymous out-of-school peer, respectively. In columns 3 and 4, the dependent variables are the average fraction of tokens sent back. The dependent variables in columns 5 and 6 are the binary dependent variables, which take the value 1 if the child chooses to cooperate (green card), and zero otherwise, with an anonymous classmate, and an anonymous out-of-school peer, respectively. In Panel 2, the dependent variables are expected payoffs from the trust game and the cooperation game. Regressions control for respective baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table VI TREATMENT EFFECTS ON ALTRUISM

	Willing to Donate	Fraction Donated
Treatment	0.069*** (0.024)	0.053*** (0.016)
Ethnic Reference	-0.016 (0.017)	0.006 (0.010)
Treatment*Ethnic Reference	0.034* (0.020)	0.012 (0.015)
Control Mean	0.70	0.34
Observations	6577	6577

Reported estimates are obtained from ordinary least squares (OLS) regressions. The dependent variable in column 1 is the binary variable that takes the value 1 if the child donates some of her tokens, zero otherwise. The dependent variable in column 2 is the fraction of endowment (4 tokens) donated. Regressions control for respective baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table VII HETEROGENEOUS TREATMENT EFFECTS ON PROSOCIAL BEHAVIOR

	Trust		Reciprocity		Cooperation		Altruism
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School	Willing to Donate
Treatment (Hosts)	0.278*** (0.061)	0.245*** (0.055)	0.053*** (0.014)	0.054*** (0.015)	0.044** (0.018)	0.004 (0.011)	0.095*** (0.024)
Treatment (Refugees)	0.198* (0.103)	0.015 (0.088)	0.035* (0.019)	0.017 (0.019)	0.034 (0.026)	-0.030 (0.026)	0.044 (0.035)
p-value : Hosts=Refugees	0.453	0.023	0.413	0.051	0.767	0.246	0.149
Control Mean (Hosts)	1.28	1.34	0.37	0.37	0.52	0.49	0.68
Control Mean (Refugees)	1.89	1.98	0.48	0.51	0.54	0.53	0.81
Observations	6476	6512	6534	6523	6568	6573	6577

Reported estimates are obtained from ordinary least squares (OLS) regressions. The dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to an anonymous classmate and an anonymous out-of-school peer, respectively. In columns 3 and 4, the dependent variables are the average fraction of tokens sent back. The dependent variables in columns 5 and 6 are the binary dependent variables, which take the value 1 if the child chooses to cooperate (green card), and zero otherwise, with an anonymous classmate, and an anonymous out-of-school peer, respectively. The dependent variable in column 7 is the binary variable that takes the value 1 if the child donates some of her tokens, zero otherwise. All regressions control for relevant baseline outcomes, Raven score, Eyes Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table VIII HETEROGENEOUS TREATMENT EFFECTS ON ACHIEVEMENT TESTS

	Turkish		Math	
	Host	Refugee	Host	Refugee
Treatment	0.008 (0.041)	0.130** (0.051)	-0.010 (0.060)	-0.019 (0.076)
p-value : Hosts=Refugees	0.037		0.911	
Observations	5565	1084	5565	1084

Reported estimates are obtained from ordinary least squares (OLS) regressions. Dependent variables are standardized test scores from Turkish language and math tests. Regressions control for relevant baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, district dummies, and randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

FIGURES

Figure I EVALUATION TIMELINE

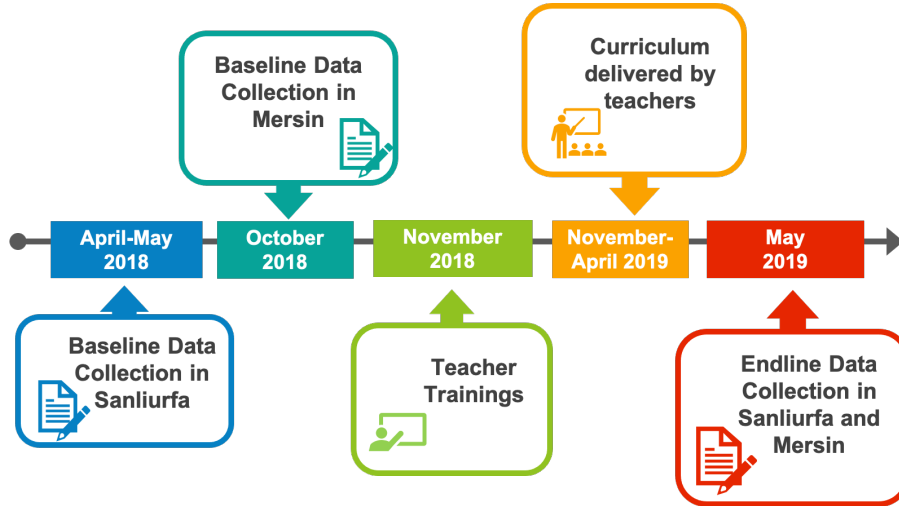
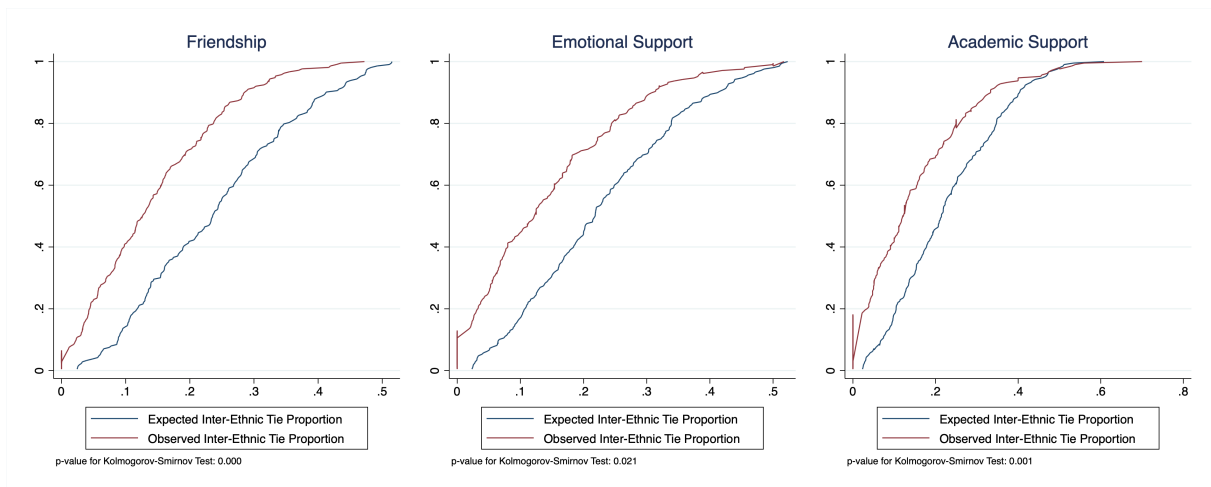
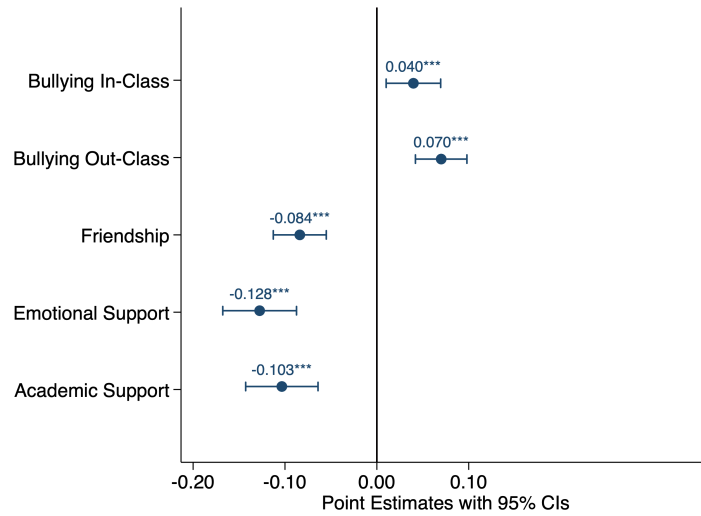


Figure II CUMULATIVE DISTRIBUTION OF EXPECTED AND OBSERVED INTER-ETHNIC TIES AT BASELINE



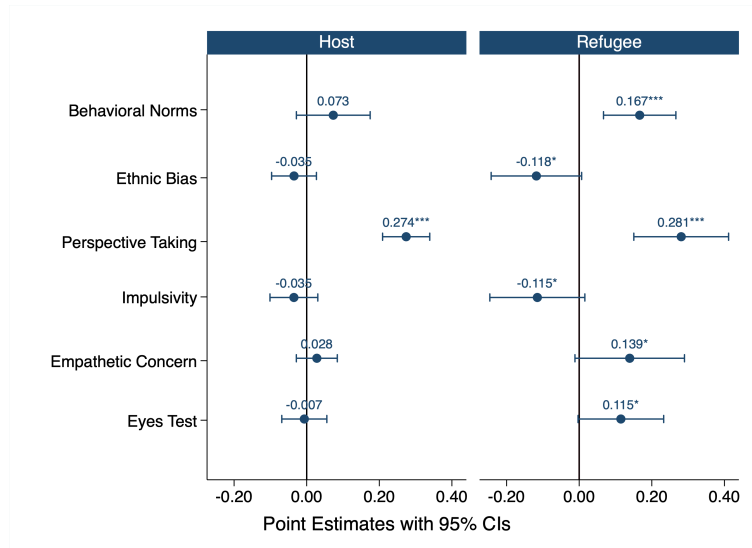
Notes: Each panel depicts the cumulative distribution of the expected proportion of inter-ethnic ties, calculated via probabilities derived from the Hypogeometrical distribution and the observed proportion of inter-ethnic ties in classrooms for each category. All figures are based on the baseline sample. P-values for the Kolmogorov-Smirnov test of equality of distributions are given at the bottom of the figures.

Figure III BASELINE CONDITIONS FOR REFUGEE CHILDREN



Notes: The figure depicts the ethnicity gaps at baseline in respective self-reports. Estimates are obtained from OLS regressions of respective binary measures on refugee dummy, controlling for class level refugee share, class size, school size, district dummies, and randomization strata. Confidence intervals are based on standard errors clustered at the school level (unit of randomization).

Figure IV POTENTIAL MECHANISMS



Notes: The figure depicts the estimated treatment effects and their 95% confidence intervals. Confidence intervals are based on standard errors clustered at the school level (unit of randomization). The vertical line indicates a treatment effect of zero for each group. Dependent variables are standardized factors constructed using relevant item-set questions, so all coefficient estimates represent standard deviation effects. Corresponding regressions control for randomization strata and baseline covariates, including baseline values of respective outcomes if available.