Building Inter-Ethnic Cohesion in Schools:
An Intervention on Perspective Taking *

Sule Alan[1], Ceren Baysan[2], Mert Gumren[3], and Elif Kubilay[2]

[1]European University Institute
[2]University of Essex
[3]Koc University

Abstract

We evaluate the impact of an educational program that aims to build inter-ethnic cohesion in schools by developing perspective-taking ability in children. The program takes place in southeastern Turkey, a high-stakes context in which there has been a massive influx of refugees. We measure outcomes that are fundamental to economic interactions and social cohesion, including peer violence, social exclusion, and prosocial behavior. Using randomized variation in program implementation, we find that the program significantly lowers peer violence and victimization on school grounds. It also reduces social exclusion and ethnic segregation in the classroom, measured by inter-ethnic friendship ties. We find that the program is highly effective in enhancing prosocial behavior: Treated students exhibit significantly higher trust, reciprocity, and altruism toward each other. Our results suggest that well-targeted educational strategies can go a long way in building social capital, even in sociopolitically difficult circumstances.

JEL Codes: I24, I28, D24, C93

Keywords: social cohesion; education; refugee integration; social exclusion

---

*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, participants in 2020 Conference on Forced Displacement in Copenhagen for their valuable comments. We thank Enes Duysak, Ipek Mumcu, Ozge Seyrek and Melek Celik for wonderful field assistance. The trial has been registered at the AEA Registry: AEARCTR-0003974

1 Introduction

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 to communicate effectively with others, engage in efficient economic interactions, and work together in groups for a common goal. Underlying these social skills is the ability to relate to others, understand their perspectives, and accept individual differences. Public education has shown to have a critical role in instilling these virtues in children, 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, ethnic and socioeconomic segregation, and intolerance toward different cultures and religious practices, can arise under turbulent sociopolitical conditions. In 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, Easterly et al., 2006, Alesina and Ferrara, 2005, Fryer Jr and Loury, 2013, Hjort, 2014).

This paper is an experimental evaluation of an educational program that is designed to build social cohesion in schools. While applicable to any educational context in which the objective is building social capital, we put this program to the test in a high stakes context where the ethnic composition in schools has changed abruptly and significantly due to a massive influx of refugee children. Since 2016, the Turkish Ministry of Education (MoE) has adopted an active policy of placing refugee children in state schools. While providing refugee children with access to education, the general perception maintained by the locals is that this policy worsened the school climate by increasing peer violence and creating visible ethnic segregation on school grounds.[2] The social cohesion program, known as "Understanding Each Other," was born in response to these circumstances. It involves a curricular module implemented by children's own teachers. The module is designed to help teachers build cohesion in the classroom to ensure a healthy learning climate for all pupils while enabling the smooth integration of the refugee children into the host country's education system.

[1] Gradstein and Justman (2002) examines 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.

[2] UNHCR (2018) reports deterioration in the quality of education in Lebanon and Jordan following the massive influx of Syrian refugees.

The program takes a particular socio-cognitive skill, perspective-taking ability, as a core concept. Perspective-taking is a cognitive process of viewing a situation from the perspective of another person. It has been shown to invoke cognitive mechanisms in the brain, and as such, dissociated from what is generally known as empathetic concern (or emotional empathy).[3] Studies show that perspective-taking is associated with lower social aggression, higher trust, and social cooperation; see (Batson et al., 1997, Galinsky and Ku, 2004, Galinsky and Moskowitz, 2000). Higher perspective-taking ability is also related to being able to analyze social situations through slow deliberations, weighing the pros and cons of action before engaging in the act. Studies show that this type of deliberation is effective in reducing crime and violent behavior in various contexts (Blattman et al., 2017, Heller et al., 2017). Our data concur with these findings: Classes with higher average perspective-taking ability exhibit lower ethnic segregation, measured by inter-ethnic friendship ties, have a higher tendency for cooperation, and lower incidences of bullying and peer violence. Motivated by these findings, a multidisciplinary team of educators, pedagogical consultants and multimedia developers designed "Understanding Each Other" as a set of curricular activities and pedagogical tools to develop children's ability to understand each others' perspectives, their capacity to make inferences about others' intentions, goals, and motives. The designers of the program took great care to ensure that the content makes no explicit reference to ethnicity. Instead, they aimed to encourage students to exert effort to tolerate individual differences, whether they be ethnic or otherwise.

The program was implemented as a cluster randomized controlled trial. The evaluation sample includes about 7000 elementary school children, 18% of whom are refugees, from 80 elementary schools in Turkey. These schools are located in two Southeast provinces of Turkey that received a massive influx of refugees in a short period. 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. Teachers used the entire academic year of 2018-2019, about three lecture hours per week, to cover the program in the extra-curricular project hours allotted by the MoE. We sampled, on average, three classrooms per school out of an average of 15 classrooms for evaluation purposes. We collected endline data in May 2019.

While there is no universal definition of social cohesion, there are widely accepted indi-

[3] Stietz et al. (2019) show that empathy and perspective-taking recruit different neural circuits in the brain, and both capacities vary substantially between situations and people.

cators that characterize a cohesive environment. Peaceful conflict resolution, trust, and reciprocity between individuals, altruism toward the needy, tolerance of individual differences, a strong sense of belonging are to name a few[4]. To evaluate the program, we put together a multidisciplinary toolkit that characterizes the cohesiveness of the school and classroom environment along these lines. Our toolkit includes i) administrative diary logs recording high-intensity peer violence on the school ground, ii) carefully elicited friendship (network) ties to measure social exclusion and ethnic segregation, iii) incentivized lab-in-the-field experiments to measure prosocial behaviors (trust, reciprocity, cooperation, and altruism), iv) achievement tests to measure cognitive and academic ability, and v) psychometric tests to measure socio-emotional skills, social norms, and ethnic bias (Fehr and Schmidt, 1999, 2000, Boisjoly et al., 2006, Burns et al., 2015, Rao, 2019).

We find that the program is highly effective in lowering high-intensity peer violence and victimization, measured via diary logs completed by school administrators. Over ten consecutive school days, about 1.9 violent events perpetrated by children in the control group. This number more than halves (reduced to 0.65 events) in treatment schools. This substantial treatment effect is statistically significant at the 1 percent level. Given this result, we also explore whether the program has 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, the program may make 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.

The program also increases the likelihood of forming inter-ethnic friendship ties, thereby reduces ethnic segregation in the classroom. We find that treated children, hosts, and refugees alike, are significantly less likely to be socially excluded and more likely to receive emotional and academic support from their classmates. These positive effects are particularly prominent for refugee children: Refugee children in treated schools are about 7 percentage points (11 percent) more likely to form a friendship tie with a host classmate and 12 and 10 percentage points more likely to receive emotional and academic support, respectively, from host classmates relative to the refugees in untreated schools.

[4]Sociologist Emile Durkheim defines social cohesion as a characteristic of a society that shows the interdependence between individuals of that society. Such a society is characterized by the absence of conflict based on wealth, ethnicity, race, and gender and the presence of strong social bonds; Durkheim (1973).

We also estimate significant improvements in prosocial behaviors of children, measured by incentivized games. Treated children exhibit significantly more trust and reciprocity toward their classmates as well as toward anonymous peers outside of their schools. They also exhibit higher altruistic tendencies toward anonymous recipients and positively discriminate in favor of refugees in donating parts of their endowments in a dictator game: 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 revealed to be a refugee peer. We show that this heightened prosociality is welfare improving from the perspective of payoffs children received in these incentivized games: By exercising more trust and reciprocity toward their classmates, treated children collectively increase their payoff gains by about 5 percent relative to untreated children.

Overall, the program appears to be highly effective in building a cohesive classroom environment. Refugee children emerge as the primary beneficiaries of this environment. In addition to facilitating their social inclusion in their new home, the program significantly improves refugee children's ability in the language of the host country, which is an essential marker for a successful integration: Treated refugee children received 0.14 standard deviations higher score in the objective Turkish language test we implement in classrooms. Such a remarkable improvement in language ability suggests that creating a peaceful and cohesive learning environment, where inter-ethnic friendship ties are easily formed, is critical to teach migrant children the language of the host country, and as such, a prerequisite for effective integration policy.

Our exploratory analyses suggest that these positive effects stem mainly from the program's success in improving children's perspective-taking ability. Treated children report exerting higher effort to understand others' perspectives and higher capacity to tolerate individual differences. While improved perspective-taking ability emerges as a prominent channel, we explore several other possible mechanisms. In particular, we test whether the program also works through improving classroom norms, reducing ethnic bias, increasing empathetic concern, and enhancing the ability to regulate impulsive behavior. We find that in addition to enhancing the ability of perspective-taking, the program improves children's ability to self-regulate impulsivity. Treated children report a higher ability to reflect on the pros and cons of action before engaging in the act, which may have contributed to the positive results on violence and victimization; see (Kahneman, 2011, Fudenberg and Levine, 2006, Heller et al., 2017).

We show that a carefully designed program that encourages children to exert effort in understanding others' perspectives can go a long way in building social capital. To the best of our knowledge, this is the first paper that provides a cost-effective policy action to build a cohesive school environment by fostering a particular socio-cognitive skill, perspective-taking. It shows that developing perspective-taking ability in children is possible through educational tools and teacher training, and such efforts can lead to significant improvements in economically and socially vital outcomes. Besides showing the malleability of a valuable skill in children, the program's applicability to a wide range of educational contexts, in which building social capital is of necessity, is another contribution of the paper. Such a necessity may arise due to socioeconomic as well as ethnic segregation within and across schools, which we have been witnessing in many parts of the world Echenique et al. (2006). Our study relates to three bodies of literature in economics. First, it complements the research on reducing crime, violence, and ethnic segregation through behavioral interventions; see (Blattman et al., 2017, Heller et al., 2017)[5]. Second, it 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, 2013, Alan and Ertac, 2018, Alan et al., 2019, Cappelen et al., 2019, Kosse et al., 2019). Finally, the paper contributes to the literature on ethnic taste-based discrimination (see Charles and Guryan (2009), Bertrand and Duflo (2017) and references therein). It shows that inclusive educational programs that target social skills can be very effective in building tolerance toward ethnic differences, without making ethnicity an explicit subject matter.

The paper is organized as follows. Section 2 summarizes the key features of the program and the sociopolitical context in which it was implemented. Section 3 details the evaluation design and gives a detailed account of our outcome measures. Section 4 describes the data and tests for internal validity. Our main results are presented in Section 5. In Section 6, we explore treatment effect heterogeneity and potential mechanisms. We conclude in Section 7.

2 Program and the Context

Since the beginning of the Syrian Civil War in 2011, Turkey has received more than 3.5 million registered refugees. This figure is 14% of the world's refugees and makes Turkey the

[5]A strand of this literature tests "contact theory" in a similar context; see Rao (2019), Burns et al. (2015), Paluck et al. (2019) and references therein.

host country with the highest number of Syrian refugees. Currently, there are over 1 million Syrian children in Turkey. Over the past couple of years, the Turkish Ministry of Education 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 in the school ground.

The program we evaluate is designed, implemented and evaluated in this sociopolitical context. It is an educational cohesion program targeted at elementary school children. It aims to provide teachers with an easy-to-follow road map to foster cohesion in the classroom and ensure a healthy learning environment for all children. The program has well-designed curricular content. This content is comprised of a set of written and animated class activities compiled as a modular book known as "Understanding Each Other." The program takes perspective-taking as a core concept. It encourages students to understand and experience the emotions of the described subject through a variety of 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 happens to a character, followed by a similar event happening to another character. This animated material aims to emphasize the similarity of the effects of hurtful events to different people. In our sample, such an event is likely to be random across ethnic groups, so its purpose is to make sure children exert effort to take others' perspectives in general. Another example is a reading material, which is a diary extract of a refugee student. Students read what this (hypothetical) child wrote about his first day in his new school. Students then read a diary extract from a host child who wrote about this very same day of a new friend's arrival from another country. Throughout the curriculum, ethnic identity is never explicitly stated, but occasionally, as in this example, it can be inferred in some activities.

The program also includes various follow-up activities and games implemented by the teacher. For example, after watching an animated video that highlights an act of social exclusion, 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 terms of building perspective-taking ability is emphasized in the literature (Galinsky et al., 2005).[6] It is important to re-emphasize that none of the activities or curricular topics involve an explicit

[6]The general framework for the topic each week 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 can be found in the Appendix E

ethnic reference, normative or otherwise. Instead, the program encourages tolerance toward individuals (and animals, for that matter) and cherish individual differences, whether these differences be based on ethnicity or any other individual characteristics.

The program aims to increase cohesion in schools that received a significant number of refugee students in a short period of time, and where teachers demand urgent guidance to maintain the quality of the learning environment. We conjecture that the program will achieve this end by improving students' ability to take others' perspectives in social interactions, especially in responding to conflicts and anti-social behavior. While we expect that all pupils will benefit from this cohesive classroom environment, refugee children will emerge as primary beneficiaries.

3 Evaluation Design and Cohesion Outcomes

The program was implemented as a cluster randomized controlled trial. The study sample contains 222 classrooms from 80 elementary schools in Sanliurfa and Mersin, two provinces of Turkey where the refugee placement program has been in effect since 2017. The study covers approximately 7000 3rd and 4th-grade children age 8 to 12. Approximately 18% of the children in our sample are refugees. Our 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. Because the program is oversubscribed, and we aim to collect detailed data, including time-consuming incentivized tasks, we were forced to sample on average 3 classrooms per school to include in our study.

The timeline of the trial is as follows: We collected baseline data in the province of Sanliurfa in 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 by within province student absenteeism tertiles. The probability of treatment is 50%, assigning 40 schools to treatment and 40 to control. The reason we chose to stratify our randomization on absenteeism is that absenteeism is highly predictive of educational attainment and a particularly pressing concern in this part of Turkey. Many families in southeast Turkey, including now refugee families, work as seasonal agricultural workers, usually leaving their home for work in May and returning in October.

Teacher training seminars for 40 treatment schools (124 teachers) took place in November

2018. In these seminars, teachers were introduced to the concept of perspective-taking and recent evidence on the development of cognitive empathy in young children. They then participated in an intensive workshop where they studied the "Understanding Each Other" module and related activities chapter by chapter, interactively with their designated education consultants. Teachers were provided with a detailed implementation kit in hard and soft copy, explaining the particulars of the module and its accompanying activities. Teachers were expected to spend two to three lecture hours per week to cover the curriculum throughout the academic year of 2018-2019. Our field partner periodically monitored the implementation and informed us about the process. We collected endline data in May 2019. Figure 1 depicts the timeline of the trial. It is important to note that the Turkish Ministry of Education allows (and encourages) all elementary school teachers to implement socially useful extra-curricular projects for a maximum of 5 lecture hours per week. Being involved in Ministry-approved extra-curricular projects is, therefore, a common practice for Turkish elementary school teachers. During the implementation of this program, our control teachers were also engaged in various extra-curricular projects, such as environment, health, and personal hygiene-related activities. In the absence of extra-curricular projects, teachers tend to use these free hours as a supervised activity/playtime. Therefore, the program we evaluate is unlikely to crowd out core teaching activities.

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 in baseline and endline, to conduct incentivized games, achievement and psychometric tests, and to implement 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 on March 8, 2019, along with a pre-analysis plan. The paper presents analyses involving all the outcomes that were pre-specified in our analysis plan. In what follows, we give a detailed account of these outcomes, and the related hypotheses we test.

3.1 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 have overcome this difficulty by obtaining a special permit to collect these data ourselves. Our permit allows us to collect these data at the school level without using student identifiers. Our peer violence measure is the number of high-intensity disciplinary episodes that took place in the school ground in the last ten school days. Here, the term high-intensity refers to severe conflicts involving perpetrators and victims, and events that are serious 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. We choose the administrators who are not in any way involved in the program. In addition, the administrators were not approached about the diary log until after the curriculum had already been implemented and so the request was unexpected. The sheer size of the schools we work with, and the fact that many other programs are continuously in effect, made it possible to designate an independent administrator for this task.

The diary log is an electronic Excel spreadsheet that has four columns. The first column indicates the date. At the top of every sheet highlighted some classroom identifiers. These are the classrooms included in our evaluation study.[7] In the first column, the administrator was asked to record each day the number of high-intensity disciplinary events that took place in the school without referring to any particular classroom or a child. In the second column, he/she was asked to record the number of events that were perpetrated by someone from the classrooms highlighted in their diary sheet, without identifying the perpetrators. In the final column, he/she was asked to record the number of events where someone from the highlighted classrooms were victimized, again without identifying the victimized child. An example diary log is provided in Figure 4.

Because schools in our sample are very large, and we include only a few classrooms from each school into our study, we do not expect the program to have a significant impact on overall school-level peer violence. We do expect, however, that the program is effective in reducing the number of violent events perpetrated by children from treated classrooms.

[7]Turkish 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 where there are only four grade 3 classes, they would take the identifiers 3A, 3B, 3C and 3D. Students in these classrooms progress to grade 4 into classrooms 4A, 4B, 4C, and 4D.

The reason we added the last column in the diary log is 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 also collected data from children on peer violence using surveys. For this, we asked children about their experiences of bullying perpetrated by classmate(s) from their classroom as well as schoolmate(s) from outside the classroom (but in the child's school).[8] Finally, we asked teachers to rate each student's behavior using a 1 to 5 scale where one refers to very good, and five refers to very violent and anti-social behavior.

3.2 Social Inclusion and Ethnic Segregation

The prevalence of social exclusion based on a personal characteristic, such as ethnicity, is another measure of the cohesiveness of an environment. Social exclusion based on personal characteristics may lead to the formation of groups identified with such characteristics or social isolation of an individual. To construct social exclusion measures, we elicit friendship ties where we ask children to nominate up to three classmates as their close friends, and up to three classmates from whom they frequently receive emotional and academic help. We collected these data both at baseline and endline. Using these ties, we construct two sets of social exclusion measures. The first set contains our individual level outcomes. These are several binary measures indicating whether the child has formed any social tie at all in the form of friendship, emotional support and academic support in the classroom. We expect that the program increases the probability of forming these ties, i.e., lowers the probability of being socially excluded for both host and refugee children. We also expect to observe significantly more inter-ethnic ties (ties between host and refugee children) in the treatment group.

Our second set of measures concerns ethnic segregation. For this, we construct a classroom level segregation index that summarises the level of homophily in the classroom. Utilizing the idea put forward in Schelling (1969), we construct a homophily measure for each classroom as the difference between the theoretical probability of randomly formed inter-ethnic ties and their observed counterparts. To construct the former, we proceed as follows:

[8]These questions ask the number of children in the class (school) who physically and verbally bully the respondent child with the options of zero, 1, 2, 3 or more.

If all links were formed randomly, the number of links between refugee and host students would follow the Hypergeometrical 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 equal to

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

where n_R is the number of refugee students and n_H is the number of host students 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 equal to

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

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:

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

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 friendship 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 2 presents actual friendship ties from two classrooms in our data for illustration purposes. Both classrooms have similar size (27 and 30) and a similar number of refugee students (9 and 10). It can be seen clearly that classroom 1 is less ethnically segregated than classroom 2. Segregation score (ES_c constructed using the above method yields 0.04 and 0.36 for classrooms 1 and 2, respectively. Figure 3 depicts the distribution of the expected

and observed proportion of inter-ethnic ties for all three social ties (friendship, emotional support, academic support) at baseline. We observe substantial ethnic segregation for all social tie measures. We expect the program to lower the distance between expected and observed inter-ethnic ties, i.e., ethnic segregation.

3.3 Experimentally Elicited Prosocial Behaviors

An essential feature of a cohesive environment is the prevalence of pro-social behavior in social interactions. Trust, reciprocity, cooperation, and altruism are the best-known prosociality indicators studied by economists in lab and field settings. We follow the convention and elicit these indicators using incentivized decision tasks in the following manner: In every classroom, the experimenter, assisted by field assistants, first introduces himself/herself to children. Then he/she informs the children that they will be spending about three lecture hours to play four games, and fill up some surveys afterward.[9] The experimenter shows 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. He/she then tells the children that they will play four games, and in each game, they will have an opportunity to earn "tokens". These tokens can be converted to any gifts of their choice in the basket at the end of the 3-hour visit. Experimenters carefully explain to children that at the end of the visit, one game will be randomly selected for the classroom, and everyone will receive the tokens they earn from that particular game, i.e., tokens will 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 play a version of a dictator game, and then fill up a survey.

Trust game (Berg et al., 1995) involves two participants that are anonymously paired. We designed this game to have two versions played within-subject. In both versions, children are endowed with four tokens. In the first version, which we refer to as "in-group", each child is paired with an anonymous classmate. A child is either a sender or a receiver. 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 returning

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

some amount of the now-tripled tokens to the sender, which may also be zero.

We design this game using 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 if they assume the role of the receiver.[10] The latter 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). Children re-play this game with a modification whereby their anonymous pair is an unknown student from another school. We refer to this version as "out-group."[11] Full instructions of the trust game are in the Appendix C. The amount of tokens sent is our measure of "trust," and the amount of tokens sent back is our measure of "reciprocity." We expect that the program increases trust and reciprocity in children, both toward in and out-group members, and relatively more toward the in-group members.

Cooperation game, which is a modified version of one-shot prisoner's dilemma game, also involves two participants to be anonymously paired. We similarly design this game and have in-group and out-group 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 of the card and the color of her pair's card. The payoff scheme is given in Table B1. As can be seen in the table, choosing the green card is the decision to cooperate. We refer to the binary choice of a green card as "cooperative" action and expect that the program increases the probability of cooperative action.

The reason we design two versions for trust and cooperation games is to explore possible unintended 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. In our context, such a disadvantage would manifest itself as treated children collecting, on average smaller numbers of tokens than control in out-group games. We will explore this possibility (non-optimality) using simulated payoffs later in the text. Full instructions for trust and cooperation games are given in the Appendix C.

After playing these four games, children play 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

[10] See Harbaugh et al. (2003) for a similar setup.

[11] We collected decisions from an out-of-sample school to have an out-group decision data.

variation to this game: The 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 take into account recipients' ethnic identity when deciding to donate. We expect that the program increases the tendency to donate to both host and refugee children. We also expect that treated children are less likely to be sensitive to the recipient's ethnicity.

3.4 Secondary Cohesion Outcomes

In addition to incentivized decisions that measure prosociality, we collected data on perspective taking, empathetic concern, impulsivity, and ethnic bias using surveys. The primary motivation to collect these outcomes is to substantiate our conjectured mechanism as well as to detect/rule out other potential channels. Children are asked to use 4-item response categories to answer these questions. We also collected descriptive classroom norms in the endline to assess whether the treatment improves cohesion by improving classroom norms. For this, children are asked descriptive norm questions regarding their classroom and given 4-item response categories. The full set of questions we use to measure empathetic concern, impulsivity, ethnic bias, perspective-taking and classroom norms are presented in the Appendix.

3.5 Achievement Outcomes

A healthy school environment is essential to ensure academic achievement. We conjecture that the successful integration of refugee children without disturbing the quality of education for host children requires a cohesive classroom environment. To test whether the program facilities 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 verbal ability of host and refugee children.

In the next section, we describe our data and present some relevant correlations we

observe at baseline. The latter also motivates our conjectured theory of change.

4 Data

4.1 Descriptive Statistics and Internal Validity

Before assigning treatment status, we visited all 80 schools (222 classrooms) and collected detailed baseline data on demographics, self-reported experiences of bullying, cognitive and emotional empathy, ethnic bias, and cognitive function using Raven's progressive matrices. We also measured cooperation and altruism using incentivized games at baseline.

Table 1, presents the balance of baseline variables across treatment status. Note first that about 18 percent of our sample consists of refugee children. The first panel presents the balance in individual-level characteristics collected directly from children using surveys, tests, and incentivized games. The second panel presents classroom and teacher characteristics, and the last panel shows the balance in school characteristics. 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. Because it was not logistically possible to include all classrooms in our study in a given school, we randomly selected an average of 3 classrooms from each school. The table shows no significant imbalance in any of the variables except for the proportion of students who reported to be physically bullied by their classmates (significant at 5 percent level) and the number of study classrooms (significant at the 10 percent level). We will show that including either covariate in the analysis does not change the results.

Before moving onto presenting the effects of the treatment, we present some correlational evidence that sets the stage for our results and motivates our theory of change. We first note that the proportion children bears no relation to the self-reported peer violence in schools (Figure 5). We next note that the relationship between measured cognitive empathy and violence is, as expected, negative at baseline. Figure 6 depicts the relationship between the mean level of cognitive empathy in a classroom and the average number of bullies reported per child in that classroom (ranges between zero to 3). Table 2 presents some baseline conditions for refugee children. We use our control group to estimate the probabilities of experiencing regular bullying, having friends, and receiving regular support from classmates. It is fairly clear from this table that refugee children are significantly more likely to be socially

excluded and subject to bullying. They are about 4 percentage points more likely to report experiences of bullying, 7 percentage points less likely to have a friend in their classroom, 13 and 11 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 concerning refugee status to see if the program improves the social disadvantages of the refugee children we document here.

5 Results

To test the null hypothesis that the program had no impact on cohesion outcomes detailed in Section 3, we estimate average treatment effects conditioning on baseline covariates that are predictive of the outcome of interest and the strata fixed effects:

$$y_{is} = \alpha_0 + \alpha_1 T_s + X_{is}^{'}\gamma + \text{Other}_{is} + \delta_b + \varepsilon_{is} \tag{1}$$

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 potentially predictive of the outcome y. The latter include school and class size, gender, standardized cognitive score (measured by Raven's Progressive Matrices), and the outcome variable collected at baseline. $Other_{is}$ 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. We present all our results without covariates in the Appendix A.

The program requires teachers to cover all weekly topics throughout the academic year. Even though the participation was voluntary and the program was oversubscribed, compliance in the sense of program implementation may not be perfect on the part of the teacher. To assess compliance with the program, we asked treated teachers to report their estimated degree of program implementation at endline. Specifically, we asked them to mark their estimated degree of execution throughout the academic year using an unmarked 10cm line, which gives us a continuous measure of program implementation intensity, albeit subjective. Figure 7 depicts the distribution of the reported implementation intensity. Treated teachers report covering about 60 percent of the program on average, with about 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 Ministry of Education in the middle of the academic year, which is quite common in our study site. Given this noncompliance, the estimated $\hat{\alpha}_1$ should be interpreted as the average intent to treat effect (ITT). In what follows, we will focus on ITTs.

It is important to note before we present our results that in carrying out our empirical analyses, we try to remain very close to our pre-analysis plan. Unless we indicate otherwise, presented analyses and related outcomes are pre-specified. However, we present analyses that were not pre-specified but shaped mainly by the feedback we received in seminars and conferences. We consider these exploratory.

5.1 Treatment Effects on Peer Violence and Victimization

Table 3 presents the estimated treatment effects on the number of high-intensity violent episodes recorded in 10-day dairy logs. Recall that the study sample covers, on average, about three classrooms in each school. The first two columns use only the study sample. The first column presents the estimated treatment effect on the number of episodes perpetrated by children from study classrooms. The second column presents the effect on the number of episodes that victimized children from study classrooms. The third column presents the estimated treatment effect on the number of school-wide episodes, and finally, the last column presents the effect on non-study classrooms, therefore representing the spillover effects of the program.

As can be seen in Column 1, the program significantly reduced the number of violent events perpetrated by children from study classrooms. There are, on average, 1.88 events recorded in 10 days in which perpetrators were from study classes in the control group. The treatment effect of -1.23 fewer events implies a substantial (about 65 percent) decline. Note also that despite the small sample size in this analysis (80 schools), this effect is estimated with very high precision p-value=0.007).

The second column in the table shows that the program also significantly reduced the number of events that victimized children from study classrooms. While the number is 1.50 events in the control group, it is about 50 percent lower (0.75 fewer events) in the treatment group (p-value=0.044). 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 lowered the risk of being a

victim in a conflict by keeping children away from conflict.

Considering the sheer size of the schools, and the fact that we could only include a small number of classrooms in our study, the effect of the program on the entire school is striking. We estimate a substantial treatment effect on the overall number of violent events in the school. As can be seen in column 3, the average number of violent episodes in ten days is 7.83 in control schools, and the program generates 2.3 fewer events. This effect is not precisely estimated, but we note the large effect size. Similarly, the last column, which removes the events that involve study classrooms, shows a large (about 19 percent), albeit imprecise effect on peer violence: non-treated classrooms in treated schools record one less violent event in 10 days. These results are suggestive of spillover effects of the program within schools.[12]

The program appears to be highly effective in reducing high-intensity peer violence on the school ground. A natural question now is whether this is reflected in student and teacher reports of bullying and bad behavior. Recall that we also asked students about their experiences of bullying at baseline and endline. We also asked teachers to rate each student's general behavior. For the former, we construct a binary outcome, which takes the value of 1 if the child reports being bullied by his/her schoolmates, zero otherwise. The latter is constructed as a standardized behavior score. For this, we asked the teacher to rate each student on three antisocial behavior categories using a 1 to 5 scale with 1 referring to outstanding behavior, 5 to violent and antisocial behavior. We then aggregated the scores and standardized them to have a mean zero and variance of 1.

It is worth noting that, while we do expect the program to be effective in reducing violence and anti-social behavior in the school ground, our conjecture for the self-reports is somewhat ambiguous. As we also mentioned in our PAP, the idea behind this ambiguity is that we cannot rule out the possibility that the program may raise awareness or make previously unnoticed (marginally bad) behavior more salient for children and teachers. Our results are consistent with this line of reasoning. Table 4 presents the estimated effects on self-reported bullying experience and teacher-reported behavior scores. In both cases, we estimate small positive but imprecise treatment effects (1.5 percentage points, implying a 2 percent effect).

[12]We did receive some anecdotal testimonies from some teachers that they provided their interested colleagues with some of the material to be implemented in their classrooms.

5.2 Treatment Effects on Social Inclusion and Ethnic Segregation

We now answer the question of whether the program leads to the formation of more friendship ties between hosts and refugees and lower social exclusion and ethnic segregation, controlling for the proportion of refugees in the classroom. Panel 1 in Table 5 presents the estimated marginal effect of the program on these three binary outcomes for all children. We observe that most children (94 percent) report to have at least one friend in their classroom. The program has a small but significant effect on the probability of having at least one friend: a child in a treatment classroom is 0.9 percentage points more likely to have a friend, and this effect is significant at 10 percent level. The results of emotional and academic support (column 2 and 3) are also positive. Treated children are 5 percentage points more likely to have at least one classmate from which they receive emotional support, and 6 percentage points more likely to have at least one classmate from which they receive academic support.

We then estimate the impact of the program on the level of ethnic segregation in the classroom. Recall that we calculate the level of ethnic segregation by constructing the difference between the ties expected to be formed at random and the actual ties. Panel 2 in Table 5 presents the estimated treatment effects on ethnic segregation. Each measure includes links based on friendship, emotional support and academic support. Note that the measure is obtained as a classroom level outcome. We estimate a significant increase in inter-ethnic friendship and academic support ties here as well. Our estimate amounts to a 20% and 25% decline in ethnic segregation for friendship and academic support respectively and the estimate is significant at the 5 percent level. While the size is similar (18%), the effect based on emotional support is less precisely estimated.

Taken together, these results strongly suggest that the program was highly effective in making children socially inclusive and in reducing ethnic segregation in the classroom. We now turn to investigate whether the program was effective in cultivating children's social skills.

5.3 Treatment Effects on Trust, Reciprocity, Cooperation and Altruism

Panel 1 in Table 6 presents 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-group), and to an anonymous receiver outside the respondent's school (out-group). About 1.38 and 1.45 tokens were sent in the control group to an anonymous in-group and out-group receivers,

respectively. This surprising difference is significant at 10 percent level, implying children in our study site exhibit higher trust to the out group. 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 0.29 extra tokens to an anonymous in-group peer, and about 0.22 more tokens to an anonymous out-group peer. While the effect size in the case of out-group is smaller, the difference is not statistically significant (p-value=0.14).

Table 7 presents the estimated program effects on reciprocity toward an in-group member and an out-group member. Here, because reciprocity was elicited based on all four scenarios of receiving 1, 2, 3 and 4 tokens, we construct four dependent variables as "the ratio of the tokens sent back to the sender." For example, if in the case of receiving 1 token (raised up to 3 by the experimenter), the child stated that he/she would send back 2, the value of the variable "Reciprocity 1" for this child would be 2/3. Similarly, say the same child stated that he/she would send back 4 of the tokens back in case of receiving 3 (raised up to 9 by the experimenter), the value of the variable "Reciprocity 3" would be 4/9. This re-construction makes it easier to compare the effect sizes across four scenarios.

We find that children exhibit high reciprocity toward their classmates in general. Between 36 to 43 percent of the tokens received were sent back to the sender in the control group. The estimated treatment effects are high both in terms of size and precision: treated children sent back about 5 to 6 percentage points more tokens to the anonymous sender in their classroom relative to the control. This corresponds to about 13-15 percent higher reciprocity toward in-group members relative to the control group. Results on out-group reciprocity is very similar: here, we observe again that children exhibit high reciprocity toward out group peers. They sent back about 38 to 42 percent of the tokens they received from these peers. The estimated effect sizes range between 4 and 6 percentage points, and highly precisely estimated.

Taken together, these results strongly suggest that the program increased trust and reciprocity among children not only toward their classmates, but also toward peers they do not know. Before moving on to other incentivized outcomes, we ask the question whether the latter results, i.e. becoming more trusting is a good thing for children. It is easy to predict that classrooms where students exhibit more trust and reciprocity increase the size of the gift pie and end up earning more gifts than children in the control group, therefore higher trust and reciprocity in an in-group setting is efficient from the ex-post payoff perspective. The question is, was it efficient for treated student to exhibit such high levels of trust and

reciprocity to strangers in an environment that can be described as non-cohesive?

We designed the out-group experiment to be able to answer this question. Whether the program make students too trusting and disadvantages them in based on ex-post payoff calculations. We acknowledge that considering disadvantage from an ex-post payoff perspective assumes away utility considerations and therefore it is simplistic. An individual may derive utility from exhibiting trust toward others even if he/she does not expect this trust to be reciprocated. However, we believe that we can still learn a great deal about the optimality of trusting behaviour in generally non-cohesive environments from ex-post payoff distributions. To assess whether the program disadvantaged children by inadvertently encouraging them to exhibit "too much" trust, we estimate its effect on ex-post payoff gains. To do this, we first calculate expected payoff gains from both in-group and out-group trust games. For in-group payoffs, we use the within-class empirical distribution of decisions, for out-group payoffs, we use the empirical distributions of decisions of the control group.

Consider the in-group exercise. The expected payoff P^s_{ic} of child i in class c in the case of being a sender is:

$$P^s_{ic} = E - S_i + E_c(R_j|S_i)$$

where E is 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^r_{ic} of i in the case of being a receiver is:

$$P^r_{ic} = E + mS_j - R_i(S_j)$$

where $R_i(S_j)$ is the number of tokens sent back to the sender, and the experiment multiplier $m = 3$. Note that conditional on sender's decision, the receiver's decision to reciprocate is strategic, and does not involve uncertainty. Given student i has a 50 percent chance of being a sender or a receiver, then her expected overall payoff P_{ic} is:

$$P_{ic} = 0.5P^s_{ic} + 0.5P^r_{ic} \tag{2}$$

To calculate the payoffs for senders and receivers in the out-group game, we use the empirical distribution of decisions in the control group in child $i's$ district.

Panel 2 in Table 6 presents the estimated treatment effects on expected payoffs for in-group and out-group trust games, respectively. In addition to overall payoffs (Equation 2), we also present results for each role, both sender and a receiver. Note first that in both games, children in the control group ended up gaining 5.38 and 5.46 tokens in in-group and out-group games, respectively (see Columns 3 and 6). Given the lower bound is 4 tokens, this number indicates that they increased the size of the pie by exhibiting some trust. Not surprisingly, treated students ended up gaining about 0.28 more tokens by trusting and reciprocating more in the in-group game (Column 3). However, they ended up with 0.13 fewer tokens than control in the out-group game. Even though these results seem to indicate a disadvantage on the part of the treated children, a closer look at the results reveal an interesting justification for this outcome. We do not estimate a statistically significant payoff difference in the case of being a sender. However, we estimate a large and significant difference in case of being a receiver: Treated receivers gave up about 0.24 more gifts to reciprocate their out-group senders. Note that reciprocity is a strategic decision that does not involve uncertainty. The fact that treated children deliberately lowered their payoffs to reciprocate the out-group senders implies that the lower overall amount of payoffs they obtained (column 6), cannot be considered sub-optimal. Rather, it implies that program increased the tendency to reciprocate the kindness children receives from out-group peers.

Table 8 presents results for our cooperation outcome. Recall that we also implemented this game for both the in and out-groups. The outcome of interest is the propensity to cooperate (choosing green card). The table presents the estimated average marginal effects from a logit regression where the binary independent variable takes the value of 1 if the child choose to cooperate (choosing green card), zero otherwise (choosing orange card). The table shows that 52 percent of children in the in-group version, and 50 percent in the out-group version chose to cooperate in the control group. While we estimate a statistically significant effect of the treatment for the in-group (4 percentage points), the estimated effect for the out-group is zero. We explore the optimality of cooperation in a similar way to the trust game. Table 8 presents the results. Treated children received 0.14 more tokens than control by exhibiting in-group cooperation. Not surprisingly, we estimate a zero payoff difference across treatment status with respect to out-group cooperation.

Our final incentivized cohesion indicator is altruism, measured by a dictator game. As

explained in Section 3.3, 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 regarding the recipient's ethnicity. Here, in addition to estimating the impact of the program 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:

$$y_{is} = \alpha_0 + \alpha_1 T_s + \alpha_2 S_i + \alpha_3 T_s * S_i + X_{is}^{'}\gamma + \delta_b + \varepsilon_{is} \tag{3}$$

where y_{is} is either the probability to donate or the fraction of the endowment donated, T_s is the treatment indicator for school s, S_i is the indicator that the child i received the donation question with the explicit reference to the recipient's ethnicity. In this specification, estimated coefficient $alpha_1$ is the treatment effect on donation to an anonymous recipient. The estimated coefficient $alpha_3$ is the additional donation the treated children make to an anonymous Syrian child.

Table 9 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 7.4 percentage point more likely to donate their endowment and this value increases by another 3.3 percentage points if the anonymous recipient is revealed to be a Syrian refugee child. The latter effect is significant only at the 10 percent level. Similarly, treated children donate 5.6 percentage points more tokens to an anonymous recipient but the fraction of endowment donated does not significantly increase when the anonymous recipient's ethnicity is revealed.

To summarise, the program appears to have a significant impact on prosocial attributes in children. Treated children exhibit higher trust and reciprocity toward their peers, they tend to cooperate more, and they tend to show higher level of altruism.

6 Heterogeneity in Treatment Effects and Potential Mechanisms

Before exploring the potential mechanisms through which the program might generate these positive results, we explore heterogeneity in treatment effects. We pre-specified three domains in our PAP: refugee status, gender, and cognition. We do not detect any significant heterogeneity concerning gender and cognitive function in most of our outcomes. How-

ever, we detect a notable treatment effect heterogeneity concerning refugee status. Table 10 presents estimated treatment effects on self-reported bullying and teacher reports of antisocial behavior separately for refugee and host children. We estimate a statistically significant difference in treatment effect with respect to self-reports with treated refugee children being less likely to report bullying. We do not estimate heterogeneity in teacher reports.

Table 11 presents estimated treatment effects on social exclusion separately for host and refugee children. The table reveals that while the program is effective in mitigating the social exclusion of all children, host, and refugees alike, the effects on refugee children are particularly strong. The program does not affect the probability of having a friend in the classroom for host children, but it increases the probability of a refugee child befriending at least one classmate by about 5 percentage points. Similarly, the program is highly effective in children's likelihood of receiving emotional support from their classmates, but more so for refugee children. Specifically, the program increases the likelihood of receiving emotional support by 4.3 and 9.3 percentage points for hosts and refugees, respectively, and the difference is statistically significant at the 10 percent level. As for the academic support, the program seems to be equally effective for both host and refugee children.

Because we are particularly interested in social ties between host and refugee children, we also estimate the effect of the program on the likelihood of a refugee child forming ties with host children. Panel 2 of Table 11 presents these results. The previous picture almost repeats itself in this table: The program increases refugee children's likelihood of having a host friend by 7 percentage points. This effect translates into a 11 percent increase over the baseline value of 61%. In addition, we see that refugee children's likelihood of receiving emotional support from a host student increases by almost 12 percentage points. We do not estimate any statistically significant heterogeneous effects concerning academic support outcome.

In terms of trust and reciprocity, we detect heterogeneity only in the out-group games. While we estimate a significant increase in out-group trust in host children, we estimate null effects for refugees. These results also hold for reciprocity. It appears that while the program strengthens trust and reciprocity among classmates, refugee children show reluctance when it comes to trusting (and reciprocating to) children they do not know. We do not estimate any notable heterogeneous effect of the program with respect to cooperation and altruism. We provide detailed heterogeneity results, including those concerning gender and cognitive ability, in the Appendix B.

All in all, 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 and, iii) increasing trust, reciprocity, and cooperation among students, as well as their altruism toward one another. The positive results on pro-social behaviors are not limited to behaviors toward close friends but extend to anonymous 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 with host children and receive emotional and academic support from their classmates. Perhaps it is not surprising that we estimate a striking improvement in refugee children's test scores in Turkish language. Table 14 presents the estimated treatment effects on math and Turkish test scores for both host and refugee children. Note that these tests were prepared according to the Turkish national curricula, and implemented by the research team in classrooms in the absence of teachers. While we estimate null effects on both Turkish and math scores for host children, we estimate a large and significant program effect on Turkish test scores for refugee children: The effect is about 0.14 standard deviations and significant at the 5% level. We do not estimate a significant treatment effect on math scores of refugee children.

We now turn to identifying (and ruling out) possible channels through which the program generates these positive results. Although we try to substantiate our claims using data, we caution that the analyses in what follows remain largely suggestive.

The objective of the program is to increase children's ability to take other's perspectives, especially in cases of conflict or social exclusion. The module provides children with examples of different social situations and teaches 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 (fictitious) individuals. Given its strong and repeated emphasis on this type of deliberation, we conjecture that the program may increase cohesion through increasing children's ability to take others' perspectives.

While we hypothesize this particular mechanism to be mainly at play, we acknowledge that there may be other channels through which the program increased cohesion. For example, the program may increase cohesion by also changing classroom norms regarding what is acceptable or unacceptable behavior. It may also do so by invoking children's compassion toward others. Yet, as another mechanism, the program may reduce conflict and victimization by teaching children to better manage their impulsivity, a behavior that is often responsible

for the escalation of conflicts and can be affected by engaging in deeper deliberation. Finally, the program may improve cohesion by increasing tolerance toward individual differences or changing beliefs, thereby reducing ethnic bias.

To assess the degree to which these channels play a role, in addition to perspective-taking, we collected information on classroom norms, empathetic concern (compassion), ethnic bias, and impulsivity using item-response questions. The full set of questions we use to construct each measure is given in the Appendix D. Figure 9 depicts the estimated treatment effects and a 95 percent confidence interval on perspective taking, classroom norms, compassion, impulsivity, and ethnic bias, controlling for baseline values of each measure. The corresponding regression results are presented in Table 15. The message that emerges from these results is quite clear. Treated children report having about 0.27 standard deviation higher level of perspective-taking ability relative to control. We also estimate a smaller but statically significant decline in reported impulsivity (about 0.07 standard deviation with a 5 percent significance level). This result is consistent with our school-level violence and victimization findings. We find a small positive effect on classroom norms, and this effect is significant at the 10 percent level. We do not estimate any significant change in ethnic bias or the level of empathetic concern, although both estimated coefficients have intuitive signs.

Taken together, we interpret these results as the program's effectiveness in increasing cohesion in the school ground by mainly improving children's perspective-taking ability and reducing their impulsivity. We acknowledge that the potential channels we test here are by no means an exhaustive list of all possible channels. Nevertheless, the overall evidence seems strong to conclude that the program has been quite successful in achieving its objective of building cohesion in the classroom.

7 Conclusion

We evaluate the effectiveness of a unique educational program that aims to foster cohesion in the classroom. The program is implemented in schools where the ethnic composition has changed rapidly due to a fast influx of refugee children. The program involves a full year curricular module to be covered by elementary school teachers for at least 3 hours per week.

We evaluate the program with respect to a wide range of cohesion indicators, including peer violence and victimization, social exclusion, and prosocial behaviors. We find that the

program significantly lowers high-intensity peer violence and victimization in school grounds. It also reduces social exclusion and lowers the ethnic segregation in the classroom. We also find that treated children exhibit higher trust and reciprocity toward their peers, cooperate more, and show higher altruistic tendencies. Finally, we show that the program leads to a significant improvement in the refugee children's ability in the language of the host country.

The results of the study are promising in terms of their external validity. While the participation was voluntary, in practice, the program was oversubscribed. In all participating schools, the majority of the teachers were eager to join the program, compelling us to choose the participant teachers randomly. Therefore we are reasonably confident that the program would show a similar success at scale in Turkey. However, the program probably needs to be re-tested in other countries that have similar migration issues, such as Lebanon and Jordan, as well as some European countries, before scaling up. Given its low-cost nature (already developed materials), and the ease of implementation, such re-testing should be straightforward.

The major caveat at the moment is that our results are short-term. It is crucial from the policy point view that these positive impacts persist after students progress into middle school and begin interacting with new classmates in a different environment. We will conduct another follow-up fieldwork in May 2020, so evaluate the longer-term effects of the program.

References

Alan, S., Boneva, T., and Ertac, S. (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, S. and Ertac, S. (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*, 126(5):1865–1911.

Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, 43(3):762–800.

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., and Highberger, L. (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.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142.

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

Blattman, C., Jamison, J. C., and Sheridan, M. (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):1165–1206.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or antipathy? the impact of diversity. *American Economic Review*, 96(5):1890–1905.

Burns, J., Corno, L., and La Ferrara, E. (2015). Interaction, prejudice and performance. evidence from south africa. *Unpublished Working Paper*.

Cappelen, A. W., List, J. A., Samek, A., and Tungodden, B. (2019). The effect of early childhood education on social preferences.

Charles, K. and Guryan, J. (2009). *Taste-based discrimination*.

Durkheim, E. (1973). Le suicide: Etude de sociologie. Technical report.

Easterly, W., Ritzen, J., and Woolcock, M. (2006). Social cohesion, institutions, and growth. *Economics & Politics*, 18(2):103–120.

Echenique, F., Fryer Jr, R. G., and Kaufman, A. (2006). Is school segregation good or bad? *American Economic Review*, 96(2):265–269.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fehr, E. and Schmidt, K. M. (2000). Fairness, incentives, and contractual choices. *European Economic Review*, 44(4-6):1057–1068.

Fryer Jr, R. G. and Loury, G. C. (2013). Valuing diversity. *Journal of political Economy*, 121(4):747–774.

Fudenberg, D. and Levine, D. K. (2006). A dual-self model of impulse control. *American economic review*, 96(5):1449–1476.

Galinsky, A. D. and Ku, G. (2004). The effects of perspective-taking on prejudice: The moderating role of self-evaluation. *Personality and Social Psychology Bulletin*, 30(5):594–604.

Galinsky, A. D., Ku, G., and Wang, C. S. (2005). Perspective-taking and self-other overlap: Fostering social bonds and facilitating social coordination. *Group Processes & Intergroup Relations*, 8(2):109–124.

Galinsky, A. D. and Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, 78(4):708.

Gradstein, M. and Justman, M. (2002). Education, social cohesion, and economic growth. *American Economic Review*, 92(4):1192–1204.

Harbaugh, W. T., Krause, K., and Vesterlund, L. (2003). Trust in children. *Trust and reciprocity: interdisciplinary lessons from experimental research*, pages 302–322.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482.

Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S., and Pollack, H. A. (2017). Thinking, fast and slow? some field experiments to reduce crime and dropout in chicago. *The Quarterly Journal of Economics*, 132(1):1–54.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4):1899–1946.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., and Falk, A. (2019). The formation of prosociality: causal evidence on the role of social environment. *Forthcoming at the Journal of Political Economy*.

Paluck, E. L., Green, S. A., and Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2):129–158.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review*, 109(3):774–809.

Rodrik, D. (1999). Where did all the growth go? external shocks, social conflict, and growth collapses. *Journal of economic growth*, 4(4):385–412.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2):488–493.

Stietz, J., Jauk, E., Krach, S., and Kanske, P. (2019). Dissociating empathy from perspective-taking: Evidence from intra-and inter-individual differences research. *Frontiers in Psychiatry*, 10.

UNHCR (2018). Global appeal 2018-2019. *Available at: http://www.unhcr.org/publications/fundraising/5a0c05027/unhcr-globalappeal2018-2019-full-report.html*.

8 Tables

Table 1: Balance at Baseline

Panel 1

	Control Mean	Difference
Male (=1)	0.51	-0.00
Age in Months	106.36	0.32
Refugee (=1)	0.20	-0.02
Working Mother	0.27	0.02
Working Father	0.86	-0.01
Proportion of Parents Reading a Book	0.44	0.01
Computer at Home	0.41	0.00
Internet at Home	0.53	-0.01
Spoken Language in Family (Turkish)	0.55	-0.01
Spoken Language in Family (Kurdish)	0.27	0.02
Spoken Language in Family (Arabic)	0.18	-0.02
Proportion of Verbally Bullied by Peers in Classroom	0.79	0.02
Proportion of Physically Bullied by Peers in Classroom	0.64	0.04**
Proportion of Verbally Bullied by Peers in School	0.75	0.02
Proportion of Physically Bullied by Peers in School	0.61	0.02
Standardized Donation	-0.00	-0.02
Proportion of Cooperation	0.54	0.03
Standardized Cognition	0.00	-0.07
Standardized Cognitive Empathy	-0.00	-0.02
Perspective Taking	0.00	-0.00
Standardized Ethnic Bias	0.00	-0.03
Empathic Concern	0.01	-0.02
Impulsivity	-0.02	0.04

Panel 2

	Control Mean	Difference
Class Size	33.84	0.45
Teacher Age in Years	34.53	0.24
Proportion of Male Teacher	0.42	-0.01
Teacher Experience in Years	10.09	0.51
Proportion of Tenured Teacher	0.89	0.02
Standardized Teacher Cognition	-0.01	0.18
Standardized Teacher Cognitive Empathy	0.03	0.02

Panel 3

	Control Mean	Difference
School Size	483.30	3.30
Number of Total Classrooms	14.65	-0.39
Number of Study Classrooms	2.50	0.68*

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Panel 1 presents the balance of individual level variables collected from children using surveys, tests and incentivized games. Panel 2 presents the balance of classroom and teacher characteristics, and Panel 3 school characteristics

Table 2: Baseline Conditions for Refugee Children

	Experience Bullying	Having Friend	Emotional Support	Academic Support
Refugee (=1)	0.042**	-0.066***	-0.127***	-0.108***
	(0.02)	(0.01)	(0.02)	(0.02)
Observations	6335	6135	6135	6135

Reported estimates are marginal effects obtained from logit regressions. 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. Binary dependent variables are column 1: reported being bullied, column 2: reported to have a friend, column 3: reported to have a classmate providing emotional support when needed, column 4: reported to have academic help by a classmate when needed.

Table 3: Official Records of Peer Violence and Victimization

	Perpetrated	Victimized	Total Events	Spillover
Treatment	-1.227***	-0.752**	-2.349	-1.121
	(0.44)	(0.36)	(1.92)	(1.78)
Strata FE	✓	✓	✓	✓
Control Mean	1.88	1.5	7.83	5.95
Observations	80	80	80	80

Reported estimates are obtained from ordinary least squares (OLS) regressions. Robust standard errors are in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. The outcome variables in this table are obtained from 10-day diary logs filled by designated school administrators. The dependent variable in column 1 is the number of violent events occurred in 10 days, perpetrated by a student from study classes, in column 2 the number of violent events in which victimized children are from study classes. The dependent variable in column 3 is the total number of school-wide violent events that took place within 10 school days. The final column removes the events perpetrated by study classes from total school-wide events, and estimates the effect on non-study classes. All regressions control for randomization strata.

Table 4: Student and Teacher Reports of Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Bullying
Treatment	0.015	0.076
	(0.02)	(0.06)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Mean	0.79	-0.02
Observations	6332	6031

The dependent variable in column 1 is a dummy which equals to 1 if the student reports physical and verbal bullying. The column presents estimated marginal effects from a logit regression. The dependent variable in column 2 is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. Reported estimates in column 2 are obtained from ordinary least squares (OLS) regressions. 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. All regressions control for randomization strata and baseline covariates. Baseline covariates include cognitive ability (Raven's Matrices), cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 5: Treatment Effects on Social Inclusion and Ethnic Segregation in the Classroom

Panel 1: Social Inclusion

	Friendship	Emotional Support	Academic Support
Treatment	0.009*	0.050***	0.058***
	(0.00)	(0.01)	(0.01)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Mean	0.94	0.79	0.73
Observations	6663	6663	6663

Panel 2: Ethnic Segregation

	Friendship	Emotional Support	Academic Support
Treatment	-0.022**	-0.016	-0.025**
	(0.01)	(0.01)	(0.01)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Mean	0.11	0.09	0.10
Observations	222	222	222

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. In Panel 1, reported estimates are average marginal effects obtained from OLS regressions. Binary dependent variables are column 1: reported to have a friend, column 2: reported to have a classmate providing emotional support when needed, column 3: reported to have academic help by a classmate when needed. Regressions control for randomization strata and include baseline covariates. These are IQ (Raven's score), cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. In Panel 2, reported estimates are obtained from ordinary least squares (OLS) regressions. Dependent variables are class-level segregation scores obtained from reported friendship, emotional support and academic support ties. Regressions control for randomization strata and baseline covariates. Baseline covariates include grade, experimenter dummies, class size, school size and district dummies.

Table 6: Treatment Effects on Trust

Panel 1: Number of Tokens Sent

	In-Group Trust	Out-Group Trust
Treatment	0.286***	0.225***
	(0.05)	(0.05)
Strata FE	✓	✓
Baseline Covariates	✓	✓
p-val (In-Group=Out-Group)	0.14	
Control Mean	1.38	1.45
Observations	6473	6509

Panel 2: Payoffs from Trust Game

	In-Group Payoffs			Out-Group Payoffs		
	Sender	Receiver	Overall	Sender	Receiver	Overall
Treatment	0.306***	0.260***	0.280***	-0.012	-0.238***	-0.126**
	(0.06)	(0.07)	(0.05)	(0.05)	(0.07)	(0.05)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
Control Mean	4.30	6.47	5.38	4.28	6.63	5.46
Observations	6473	6409	6358	6509	6378	6369

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. 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 Panel 2, the dependent variables are expected payoff gains from the trust game. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 7: Treatment Effects on Reciprocity

	In-Group				Out-Group			
	1	2	3	4	1	2	3	4
Treatment	0.059***	0.060***	0.046***	0.051***	0.058***	0.042***	0.049***	0.059***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Strata FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓	✓	✓
Control Mean	0.43	0.39	0.36	0.36	0.42	0.41	0.38	0.38
Observations	6493	6494	6489	6478	6498	6498	6481	6428

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables in column 1-4 are fraction of tokens reciprocated back in case of receiving 1 token, 2 tokens, 3 tokens, and 4 tokens from an anonymous classmate, respectively. Dependent variables in column 5-8 are fraction of tokens reciprocated back to an anonymous out-of-school peer. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 8: Treatment Effects on Cooperation

Panel 1: Cooperation

	In-Group Cooperation	Out-Group Cooperation
Treatment	0.044***	-0.002
	(0.02)	(0.01)
Strata FE	✓	✓
Baseline Covariates	✓	✓
p-val (IGC=OGC)	0.01	
Control Mean	0.52	0.50
Observations	6565	6571

Panel 2: Payoffs from Cooperation Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.141***	-0.013
	(0.05)	(0.04)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Mean	4.57	4.50
Observations	6565	6571

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Standard errors are in parentheses and clustered at the school level (unit of randomization). Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. In Panel 1, reported estimates are average marginal effects obtained from logit regressions where the binary dependent variable takes the value 1 if the child choose to cooperate (green card), and zero otherwise. In Panel 2, the dependent variables are expected payoff gains from the cooperation game

Table 9: Treatment Effects on Altruism

	Willingness to Donate	Fraction Donated
Treatment	0.074***	0.056***
	(0.02)	(0.02)
Ethnic Reference	-0.014	0.007
	(0.01)	(0.01)
Treatment*Ethnic Reference	0.033*	0.012
	(0.02)	(0.02)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Mean	0.70	
Observations	6574	6574

Reported estimates in column 1 are average marginal effects obtained from a logit regression, in column 2 are obtained from an Ordinary Least Square regression. 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. 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 randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 10: Heterogeneous Treatment Effects on Self and Teacher Reported Antisocial Behavior

	Student Self Report		Teacher Report	
	Host	Refugee	Host	Refugee
Treatment	0.025	-0.037	0.071	0.065
	(0.02)	(0.03)	(0.06)	(0.10)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
P-val for Heterogenous Effect		0.04		0.94
Control Mean	0.77	0.86	-0.03	0.01
Observations	5310	1022	5123	908

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. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 11: Heterogeneous Treatment Effects on Social Inclusion

Panel 1: All Social Ties

	Friendship		Emotional Support		Academic Support	
	Host	Refugee	Host	Refugee	Host	Refugee
Treatment	0.002	0.051***	0.043***	0.093***	0.056***	0.072***
	(0.00)	(0.02)	(0.01)	(0.03)	(0.01)	(0.03)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
P-val for Heterogenous Effect		0.008		0.110		0.516
Control Mean	0.96	0.86	0.82	0.66	0.75	0.64
Observations	5575	1088	5575	1088	5575	1088

Panel 2: Social Ties with Host Children

	Host Frienship		Host Emotional Support		Host Academic Support	
	Host	Refugee	Host	Refugee	Host	Refugee
Treatment	0.002	0.070*	0.043***	0.118***	0.055***	0.100***
	(0.01)	(0.04)	(0.01)	(0.04)	(0.02)	(0.03)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
P-val for Heterogenous Effect		0.085		0.056		0.167
Control Mean	0.95	0.61	0.80	0.46	0.72	0.46
Observations	5575	1088	5575	1088	5575	1088

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. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 12: Heterogeneous Treatment Effects on Trust

	In-Group Trust		Out-Group Trust	
	Host	Refugee	Host	Refugee
Treatment	0.297***	0.259**	0.297***	0.056
	(0.06)	(0.10)	(0.06)	(0.09)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
P-val for Heterogenous Effect		0.71		0.05
Control Mean	1.28	1.89	1.28	1.98
Observations	5436	1037	5436	1044

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. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 13: Heterogeneous Treatment Effects on Cooperation

	In-Group		Out-Group	
	Host	Refugee	Host	Refugee
Treatment	0.050***	0.011	0.004	-0.034
	(0.02)	(0.02)	(0.01)	(0.02)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
P-val for Heterogenous Effect		0.20		0.18
Control Mean	0.52	0.54	0.49	0.49
Observations	5509	1056	5513	1058

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. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 14: Heterogeneous Treatment Effects on Test Scores

	Turkish		Math	
	Host	Refugee	Host	Refugee
Treatment	0.003	0.141**	-0.000	0.010
	(0.04)	(0.05)	(0.06)	(0.07)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
Observations	5502	1063	5508	1065

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables are standardized factors constructed using relevant item-set questions so all coefficient estimates are standard deviation effects. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table 15: Potential Mechanisms

	Descriptive Norms	Ethnic Bias	Perspective Taking	Impulsivity	Compassion
Treatment	0.093*	-0.050	0.270***	-0.068**	0.027
	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)
Strata FE	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓
Observations	6341	6055	6167	5680	5921

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables are standardized factors constructed using relevant item-set questions so all coefficient estimates are standard deviation effects. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

9 Figures

Figure 1: Timeline of the Study

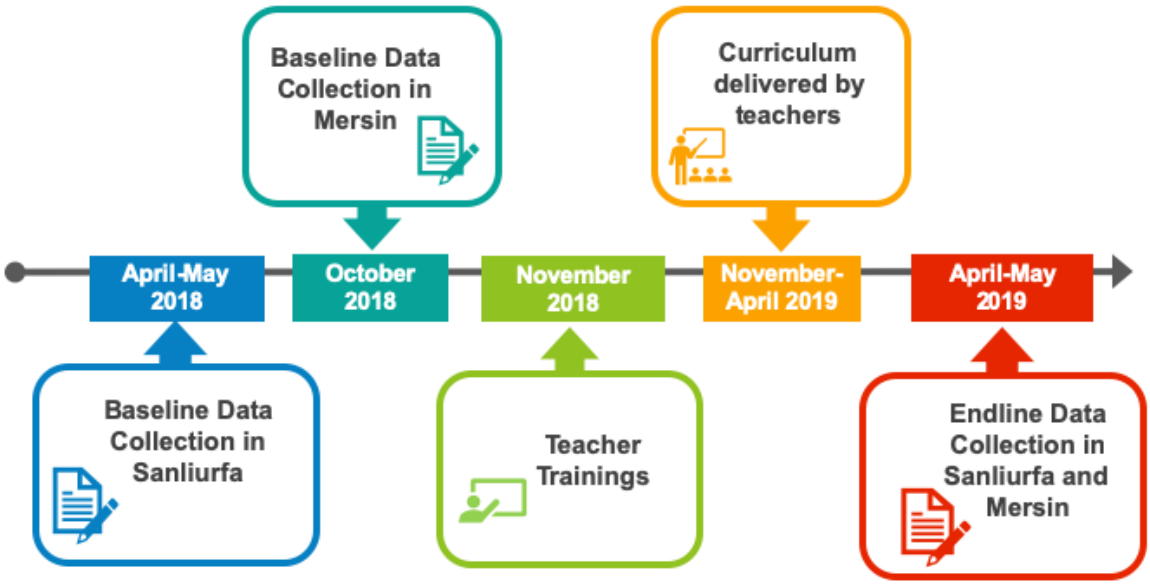
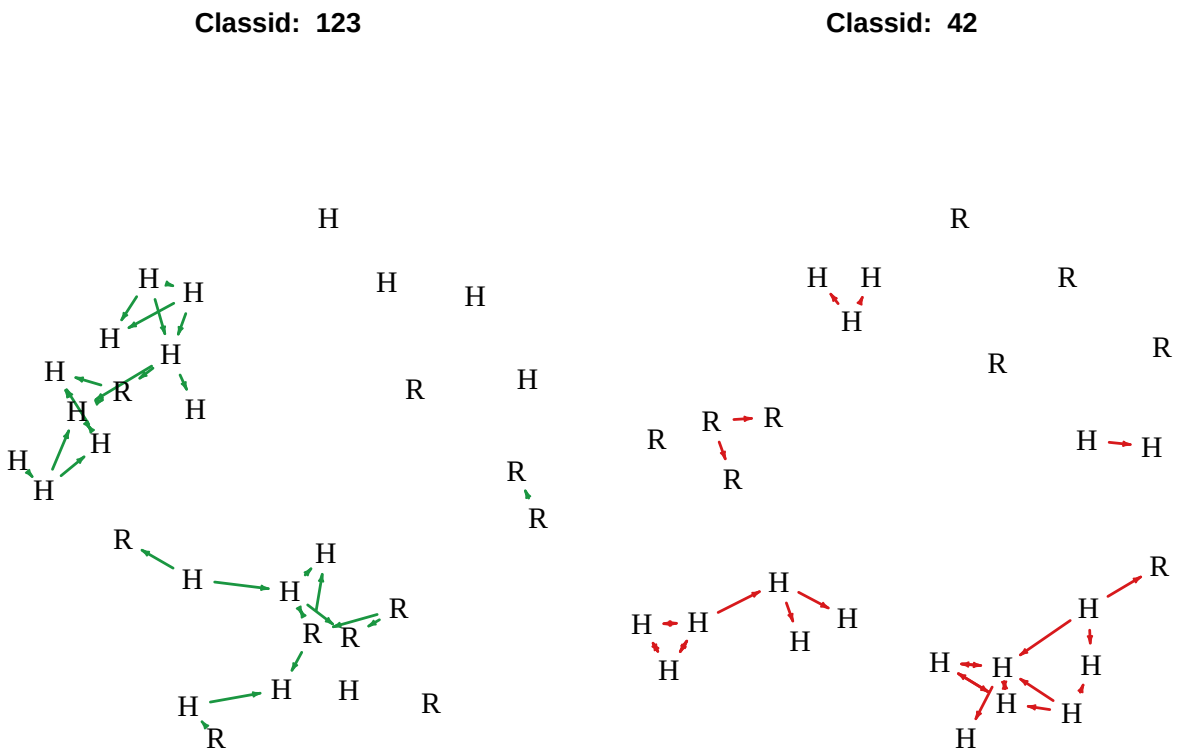


Figure 2: Ethnic Segregation: An Illustration of Two Classrooms



Each letter denotes a node (student). Letters H and R refer to host and refugee child, respectively. Nominations of the friendship are shown with directional edges between nodes. Segregation score is calculated as described in Section 3.2, with higher a number referring to higher segregation.

Figure 4: The Diary Log

School:	**Study Classrooms:**		
	Disciplinary events will be recorded below by the school administration		
	The number of high-intensity disciplinary events	Has there been a student from a study classroom who perpetrated a disciplinary event? If yes, how many?	Has there been a student from a study classroom who was victimized in a disciplinary event? If yes, how many?
April 12, 2019			
April 13, 2019	WEEKEND	WEEKEND	WEEKEND
April 14, 2019	WEEKEND	WEEKEND	WEEKEND
April 15, 2019			
April 16, 2019			
April 17, 2019			
April 18, 2019			
April 19, 2019			
April 20, 2019	WEEKEND	WEEKEND	WEEKEND
April 21, 2019	WEEKEND	WEEKEND	WEEKEND
April 22, 2019			
April 23, 2019	NATIONAL HOLIDAY	NATIONAL HOLIDAY	NATIONAL HOLIDAY
April 24, 2019			
April 25, 2019			
April 26, 2019			

Figure 3: Baseline Inter-Ethnic Tie Share Distribution

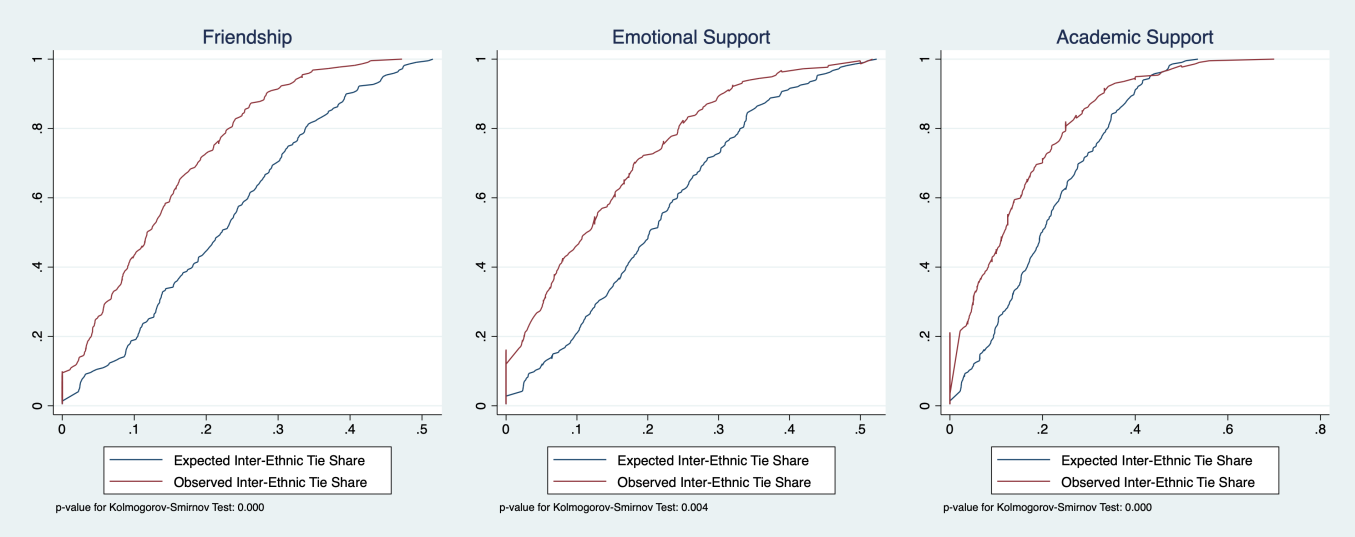
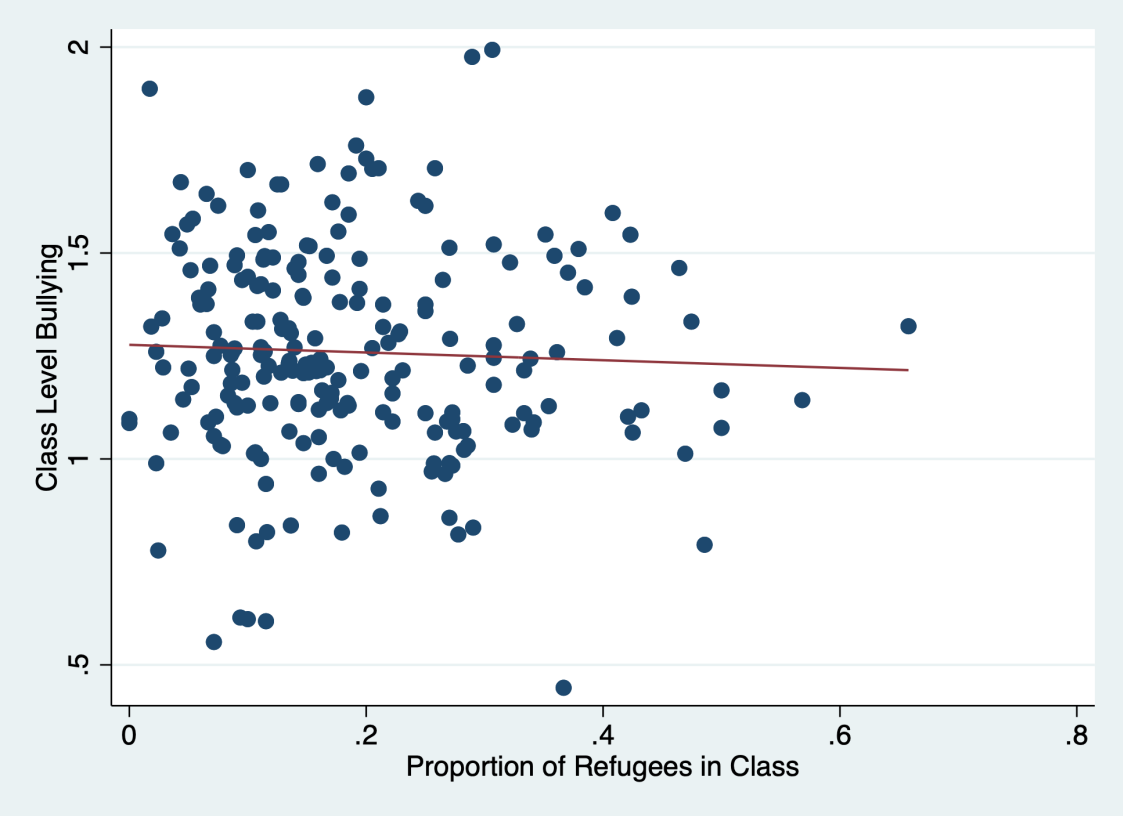
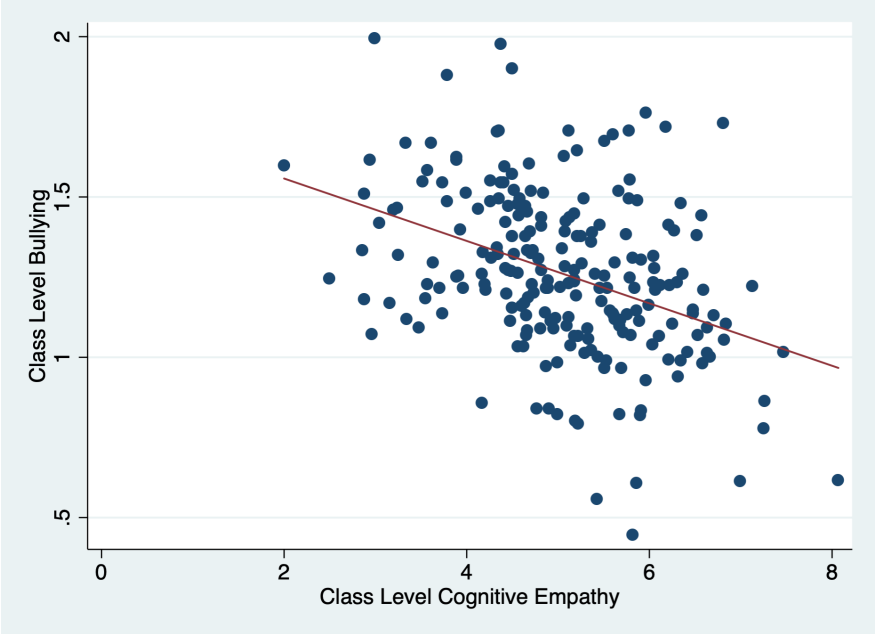


Figure 5: Proportion of Refugees and Peer Violence in the Classroom



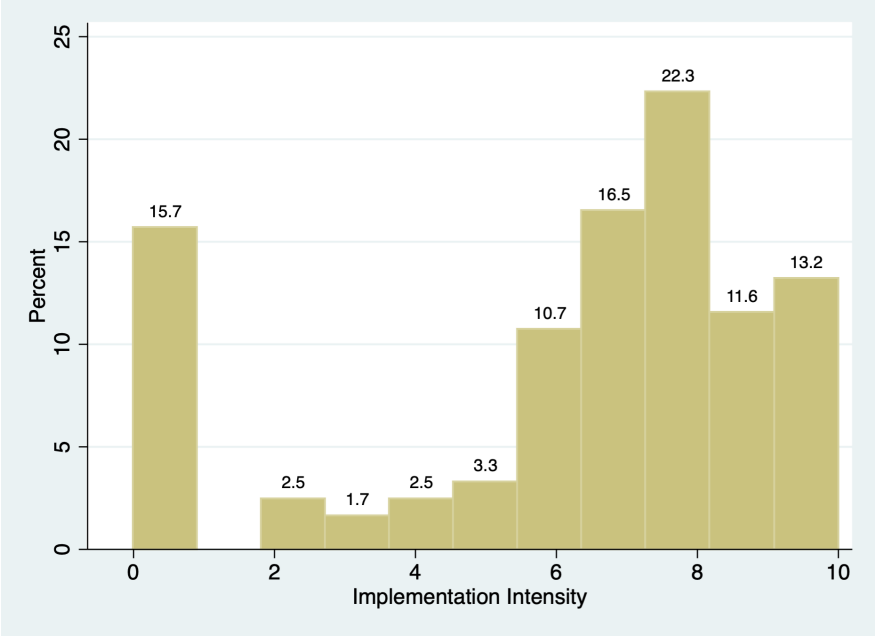
Y-axis is the average of number of bullies, X-axis is the proportion of refugees in a classroom. Values are calculated using individual level baseline data.

Figure 6: Cognitive Empathy and Peer Violence in the Classroom



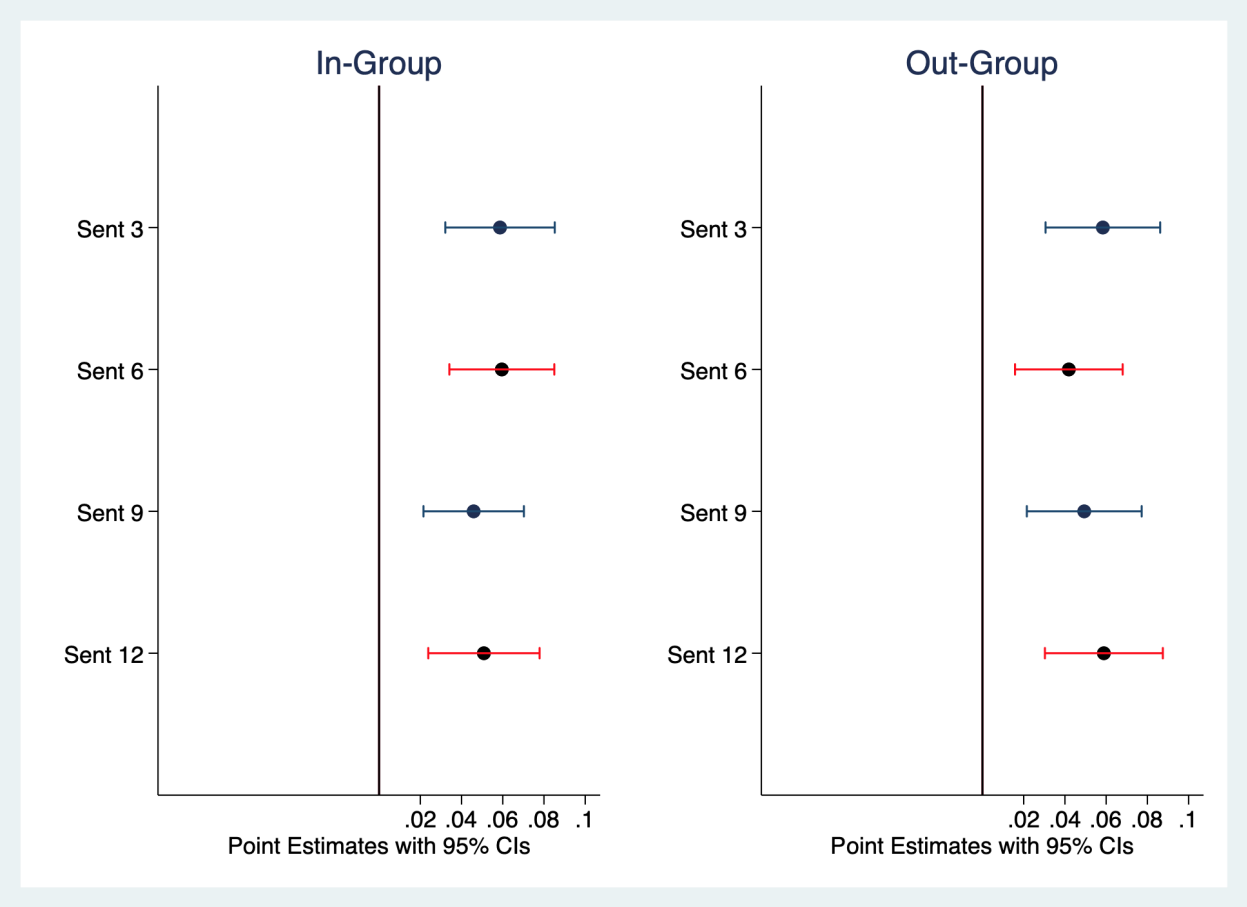
Y-axis is the average of number of bullies, X-axis is the cognitive empathy score in a classroom. Values are calculated using individual level baseline data

Figure 7: Teacher-Reported Program Implementation Intensity



Teachers were given an unmarked 10 cm line to rate their own implementation intensity. Values in this histogram are calculated by measuring (using a ruler) the distance between zero and the mark made by the teacher.

Figure 8: Treatment Effect on Reciprocity



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. Because reciprocity was elicited based on all four scenarios of receiving 1, 2, 3 and 4 tokens, we construct four dependent variables as the ratio of the tokens sent back to the sender, therefore each point estimate refers to the estimated percentage points effect for a given scenario.

Figure 9: Treatment Effect on Potential Mechanisms

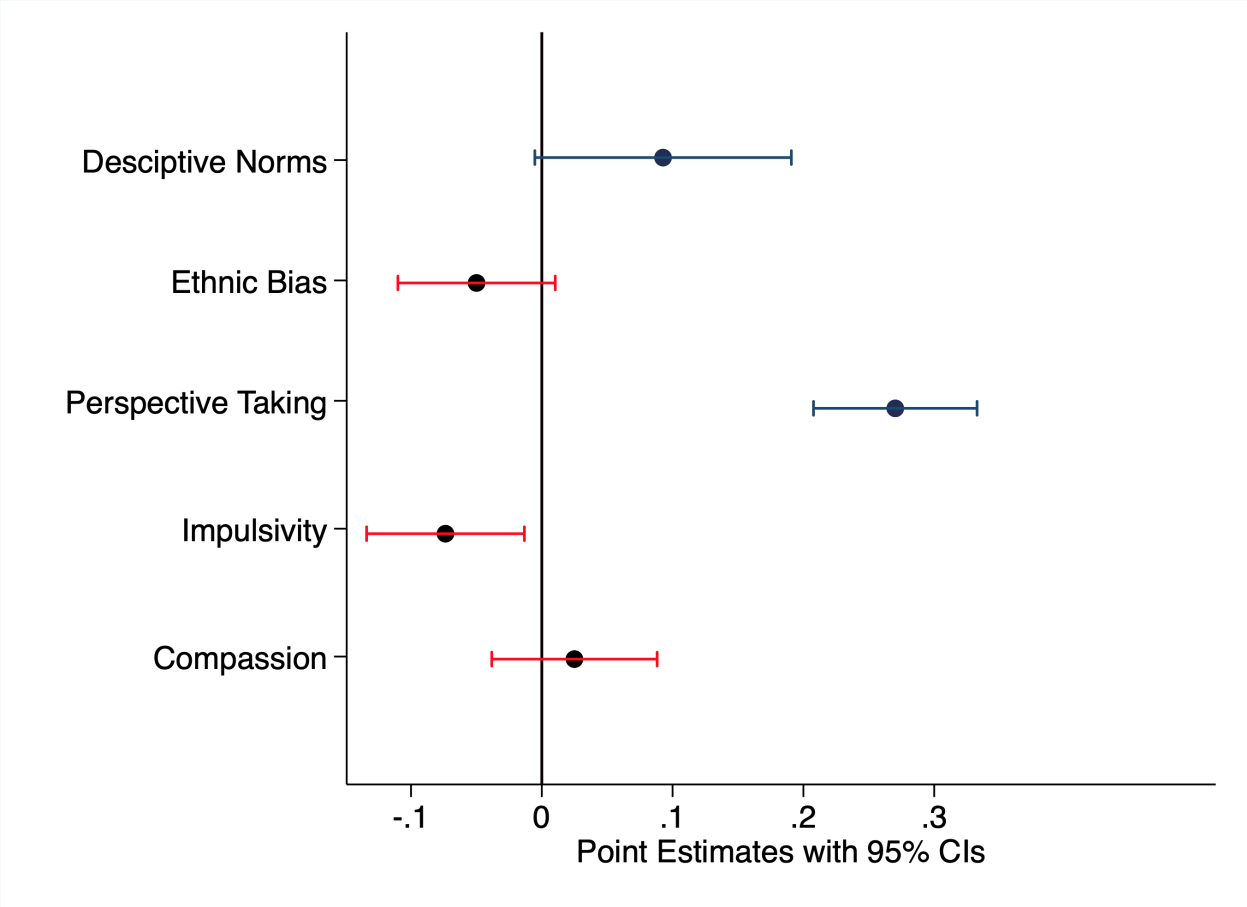


Figure Notes: The figure depicts the estimated treatment effects and their 95% confidence intervals. Confidence intervals are based on standard errors clustered at the school level (unit of randomization). The vertical line indicates a treatment effect of zero. Dependent variables are standardized factors constructed using relevant item-set questions so all coefficient estimates are standard deviation effects. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Appendix

A Main Results without Covariates

Table A1: Official Records of Peer Violence and Victimization

	Perpetrated	Victimized	Total Events	Spillover
Treatment	-0.818*	-0.686*	-3.408*	-2.591
	(0.48)	(0.40)	(1.72)	(1.59)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✗	✗	✗	✗
Control Mean	1.88	1.5	7.83	5.95
Observations	80	80	80	80

Reported estimates are obtained from ordinary least squares (OLS) regressions. Robust standard errors are in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. The outcome variables in this table are obtained from 10-day diary logs filled by designated school administrators. The dependent variable in column 1 is the number of violent events occurred in 10 days, perpetrated by a student from study classes, in column 2 the number of violent events in which victimized children are from study classes. The dependent variable in column 3 is the total number of school-wide violent events that took place within 10 school days. The final column removes the events perpetrated by study classes from total school-wide events, and estimates the effect on non-study classes. All regressions control for randomization strata.

Table A2: Student and Teacher Reports of Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Bullying
Treatment	0.007	0.050
	(0.02)	(0.06)
Strata FE	✓	✓
Baseline Covariates	✗	✗
Control Mean	0.79	-0.02
Observations	6335	6034

The dependent variable in column 1 is a dummy which equals to 1 if the student reports physical and verbal bullying. The column presents estimated marginal effects from a logit regression. The dependent variable in column 2 is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. Reported estimates in column 2 are obtained from ordinary least squares (OLS) regressions. 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. All regressions control for randomization strata.

Table A3: Treatment Effects on Social Inclusion and Ethnic Segregation in the Classroom

Panel 1: Social Inclusion

	Friendship	Emotional Support	Academic Support
Treatment	0.016**	0.059***	0.064***
	(0.01)	(0.02)	(0.01)
Strata FE	✓	✓	✓
Baseline Covariates	✗	✗	✗
Control Mean	0.94	0.79	0.73
Observations	6663	6663	6663

Panel 2: Ethnic Segregation

	Friendship	Emotional Support	Academic Support
Treatment	-0.021	-0.012	-0.020
	(0.01)	(0.01)	(0.01)
Strata FE	✓	✓	✓
Baseline Covariates	✗	✗	✗
Control Mean	0.11	0.09	0.10
Observations	222	222	222

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. In Panel 1, reported estimates are average marginal effects obtained from OLS regressions. Binary dependent variables are column 1: reported to have a friend, column 2: reported to have a classmate providing emotional support when needed, column 3: reported to have academic help by a classmate when needed. Regressions control for randomization strata.

Table A4: Treatment Effects on Trust

Panel 1: Number of Tokens Sent

	In-Group Trust	Out-Group Trust
Treatment	0.272***	0.207***
	(0.07)	(0.06)
Strata FE	✓	✓
Baseline Covariates	✗	✗
p-val (In-Group=Out-Group)	0.14	
Control Mean	1.38	1.45
Observations	6476	6512

Panel 2: Payoffs from Trust Game

	In-Group Payoffs			Out-Group Payoffs		
	Sender	Receiver	Overall	Sender	Receiver	Overall
Treatment	0.284***	0.256***	0.269***	-0.011	-0.221***	-0.117**
	(0.07)	(0.08)	(0.07)	(0.06)	(0.07)	(0.05)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✗	✗	✗	✗	✗	✗
Control Mean	4.30	6.47	5.38	4.28	6.63	5.46
Observations	6476	6412	6361	6512	6381	6372

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. 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 Panel 2, the dependent variables are expected payoff gains from the trust game. Regressions control for randomization strata.

Table A5: Treatment Effects on Reciprocity

	In-Group				Out-Group			
	1	2	3	4	1	2	3	4
Treatment	0.055***	0.056***	0.040***	0.046***	0.054***	0.035**	0.043***	0.052***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Covariates	✗	✗	✗	✗	✗	✗	✗	✗
Control Mean	0.43	0.39	0.36	0.36	0.42	0.41	0.38	0.37
Observations	6496	6497	6492	6481	6501	6501	6484	6431

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables in column 1-4 are fraction of tokens reciprocated back in case of receiving 1 token, 2 tokens, 3 tokens, and 4 tokens from an anonymous classmate, respectively. Dependent variables in column 5-8 are fraction of tokens reciprocated back to an anonymous out-of-school peer. Regressions control for randomization strata.

Table A6: Treatment Effects on Cooperation

Panel 1: Cooperation

	In-Group Cooperation	Out-Group Cooperation
Treatment	0.044***	-0.003
	(0.02)	(0.01)
Strata FE	✓	✓
Baseline Covariates	✗	✗
p-val (IGC=OGC)	0.01	
Control Mean	0.52	0.50
Observations	6568	6573

Panel 2: Payoffs from Cooperation Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.133***	-0.011
	(0.05)	(0.05)
Strata FE	✓	✓
Baseline Covariates	✗	✗
Control Mean	4.57	4.50
Observations	6568	6573

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Standard errors are in parentheses and clustered at the school level (unit of randomization). Regressions control for randomization strata. In Panel 1, reported estimates are average marginal effects obtained from logit regressions where the binary dependent variable takes the value 1 if the child choose to cooperate (green card), and zero otherwise. In Panel 2, the dependent variables are expected payoff gains from the cooperation game

Table A7: Treatment Effects on Altruism

	Willingness to Donate	Fraction Donated
Treatment	0.067***	0.050***
	(0.02)	(0.02)
Ethnic Reference	-0.017	0.004
	(0.01)	(0.01)
Treatment*Ethnic Reference	0.034*	0.013
	(0.02)	(0.02)
Strata FE	✓	✓
Baseline Covariates	✗	✗
Control Mean	0.70	
Observations	6577	6577

Reported estimates in column 1 are average marginal effects obtained from a logit regression, in column 2 are obtained from an Ordinary Least Square regression. 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. 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 randomization strata.

Table A8: Potential Mechanisms

	Descriptive Norms	Ethnic Bias	Perspective Taking	Impulsivity	Compassion
Treatment	0.086	-0.069*	0.277***	-0.096**	0.035
	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)
Strata FE	✓	✓	✓	✓	✓
Baseline Covariates	✗	✗	✗	✗	✗
Observations	6344	6057	6169	5682	5923

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables are standardized factors constructed using relevant item-set questions so all coefficient estimates are standard deviation effects. Regressions control for randomization strata.

B Heterogeneous Treatment Effects: More Results

B.1 Refugeee-Host

Table B1: Student and Teacher Reported Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Bullying
Treatment	0.026	0.074
	(0.02)	(0.06)
Refugee	0.071***	-0.033
	(0.03)	(0.06)
Treatment*Refugee	-0.063**	0.008
	(0.03)	(0.09)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.189	0.434
Control Mean	0.86	0.01
Observations	6332	6031

The dependent variable in column 1 is a dummy which equals to 1 if the student reports physical and verbal bullying. The column presents estimated marginal effects from a logit regression. The dependent variable in column 2 is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. Reported estimates in column 2 are obtained from ordinary least squares (OLS) regressions. 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. All regressions control for randomization strata and baseline covariates. Baseline covariates include cognitive ability (Raven's Matrices), cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B2: Heterogeneous Treatment Effects on Trust

Panel 1: Number of Tokens Sent

	In-Group Trust	Out-Group Trust
Treatment	0.299***	0.262***
	(0.06)	(0.05)
Refugee	0.582***	0.607***
	(0.08)	(0.07)
Treatment*Refugee	-0.074	-0.222**
	(0.11)	(0.10)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.027	0.635
Control Mean	1.89	1.98
Observations	6473	6509

Panel 2: Payoffs from Trust Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.276***	-0.101***
	(0.05)	(0.03)
Refugee	-0.068	-0.236***
	(0.04)	(0.03)
Treatment*Refugee	0.022	0.044
	(0.06)	(0.04)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.000	0.119
Control Mean	5.31	5.30
Observations	6358	6369

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. In Panel 1, the dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to the anonymous classmate, and anonymous out-of-school peer, respectively. In Panel 2, the dependent variables are expected payoff gains from the trust game. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B3: Heterogeneous Treatment Effects on Reciprocity

	In-Group				Out-Group			
	1	2	3	4	1	2	3	4
Treatment	0.058***	0.062***	0.051***	0.055***	0.064***	0.047***	0.055***	0.066***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Refugee	0.111***	0.097***	0.121***	0.108***	0.133***	0.126***	0.129***	0.141***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment*Refugee	0.007	-0.012	-0.032	-0.028	-0.032	-0.030	-0.035*	-0.042**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.004	0.021	0.393	0.160	0.214	0.392	0.338	0.238
Control Mean	0.53	0.48	0.47	0.45	0.53	0.51	0.48	0.49
Observations	6493	6494	6489	6478	6498	6498	6481	6428

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables in column 1-4 are fraction of tokens reciprocated back in case of receiving 1 token, 2 tokens, 3 tokens, and 4 tokens from an anonymous classmate, respectively. Dependent variables in column 5-8 are fraction of tokens reciprocated back to an anonymous out-of-school peer. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B4: Heterogeneous Treatment Effects on Cooperation

Panel 1: Cooperation

	In-Group Cooperation	Out-Group Cooperation
Treatment	0.187**	0.015
	(0.08)	(0.04)
Refugee	0.020	0.092
	(0.10)	(0.08)
Treatment*Refugee	-0.049	-0.135
	(0.13)	(0.12)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.197	0.239
Control Mean	0.54	0.53
Observations	6565	6571

Panel 2: Payoffs from Cooperation Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.124**	-0.011
	(0.05)	(0.03)
Refugee	-0.066	-0.068
	(0.07)	(0.06)
Treatment*Refugee	0.100	0.101
	(0.09)	(0.09)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.016	0.242
Control Mean	4.51	4.41
Observations	6565	6571

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Standard errors are in parentheses and clustered at the school level (unit of randomization). Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. In Panel 1, reported estimates are average marginal effects obtained from logit regressions where the binary dependent variable takes the value 1 if the child choose to cooperate (green card), and zero otherwise. In Panel 2, the dependent variables are expected payoff gains from the cooperation game

Table B5: Treatment Effects on Altruism

	Willingness to Donate	Donation Amount
Treatment	0.424***	0.064***
	(0.12)	(0.02)
Refugee	0.563***	0.131***
	(0.16)	(0.02)
Treatment*Refugee	-0.145	-0.042
	(0.24)	(0.03)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Refugee+Treatment=0)	0.262	0.445
Control Mean	0.81	0.46
Observations	6574	6574

Reported estimates in column 1 are average marginal effects obtained from a logit regression, in column 2 are obtained from an Ordinary Least Square regression. 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. 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 randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

B.2 Gender

Table B6: Student and Teacher Reports of Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Bullying
Treatment	0.006	0.034
	(0.02)	(0.05)
Male	0.024	0.681***
	(0.02)	(0.05)
Treatment*Male	0.019	0.082
	(0.02)	(0.07)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.180	0.160
Control Mean	0.80	0.31
Observations	6332	6031

The dependent variable in column 1 is a dummy which equals to 1 if the student reports physical and verbal bullying. The column presents estimated marginal effects from a logit regression. The dependent variable in column 2 is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. Reported estimates in column 2 are obtained from ordinary least squares (OLS) regressions. 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. All regressions control for randomization strata and baseline covariates. Baseline covariates include cognitive ability (Raven's Matrices), cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B7: Trust Game

Panel 1: Treatment Effects on Trust

	In-Group Trust	Out-Group Trust
Treatment	0.279***	0.198***
	(0.07)	(0.05)
Male	-0.015	-0.000
	(0.06)	(0.04)
Treatment*Male	0.015	0.054
	(0.07)	(0.06)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.000	0.000
Control Mean	1.38	1.46
Observations	6473	6509

Panel 2: Trust Game Payoffs

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.291***	-0.083***
	(0.06)	(0.03)
Male	-0.016	0.013
	(0.03)	(0.02)
Treatment*Male	-0.021	-0.021
	(0.03)	(0.02)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.000	0.001
Control Mean	5.37	5.46
Observations	6358	6369

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. In Panel 1, the dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to the anonymous classmate, and anonymous out-of-school peer, respectively. In Panel 2, the dependent variables are expected payoff gains from the trust game. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B8: Treatment Effects on Reciprocity

	In-Group				Out-Group			
	1	2	3	4	1	2	3	4
Treatment	0.063***	0.058***	0.040***	0.050***	0.058***	0.031**	0.046***	0.052***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Male	-0.006	-0.001	-0.003	0.009	-0.001	-0.010	-0.005	-0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment*Male	-0.009	0.004	0.011	0.001	0.000	0.021*	0.006	0.013
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Strata FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.000	0.000	0.001	0.002	0.000	0.001	0.001	0.000
Control Mean	0.43	0.39	0.36	0.36	0.42	0.40	0.38	0.37
Observations	6493	6494	6489	6478	6498	6498	6481	6428

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables in column 1-4 are fraction of tokens reciprocated back in case of receiving 1 token, 2 tokens, 3 tokens, and 4 tokens from an anonymous classmate, respectively. Dependent variables in column 5-8 are fraction of tokens reciprocated back to an anonymous out-of-school peer. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B9: Treatment Effects on Altruism

	Willingness to Donate	Donation Amount
Treatment	0.484***	0.062***
	(0.15)	(0.02)
Male	-0.185**	-0.013
	(0.09)	(0.01)
Treatment*Male	-0.147	-0.012
	(0.12)	(0.02)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.005	0.003
Control Mean	0.68	0.33
Observations	6574	6574

Reported estimates in column 1 are average marginal effects obtained from a logit regression, in column 2 are obtained from an Ordinary Least Square regression. 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. 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 randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B10: Treatment Effects on Cooperation

Panel 1: Cooperation

	In-Group Cooperation	Out-Group Cooperation
Treatment	0.084	-0.000
	(0.09)	(0.07)
Male	-0.166**	0.037
	(0.07)	(0.07)
Treatment*Male	0.187*	-0.015
	(0.11)	(0.11)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.001	0.818
Control Mean	0.50	0.53
Observations	6565	6571

Panel 2: Payoffs from Cooperation Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.214***	0.000
	(0.06)	(0.05)
Male	0.119**	-0.028
	(0.05)	(0.05)
Treatment*Male	-0.146*	0.011
	(0.08)	(0.08)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Male+Treatment=0)	0.290	0.817
Control Mean	4.62	4.48
Observations	6565	6571

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Standard errors are in parentheses and clustered at the school level (unit of randomization). Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. In Panel 1, reported estimates are average marginal effects obtained from logit regressions where the binary dependent variable takes the value 1 if the child choose to cooperate (green card), and zero otherwise. In Panel 2, the dependent variables are expected payoff gains from the cooperation game

B.3 Cognition

Table B11: Student and Teacher Reports of Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Bullying
Treatment	0.008	0.084
	(0.02)	(0.07)
Low Cognition	0.000	0.154***
	(0.02)	(0.04)
Treatment*Low Cognition	0.013	-0.010
	(0.02)	(0.06)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.273	0.272
Control Mean	0.81	0.04
Observations	6332	6031

The dependent variable in column 1 is a dummy which equals to 1 if the student reports physical and verbal bullying. The column presents estimated marginal effects from a logit regression. The dependent variable in column 2 is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. Reported estimates in column 2 are obtained from ordinary least squares (OLS) regressions. 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. All regressions control for randomization strata and baseline covariates. Baseline covariates include cognitive ability (Raven's Matrices), cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B12: Heterogenous Treatment Effects on Trust

Panel 1: Number of Tokens Sent

	In-Group Trust	Out-Group Trust
Treatment	0.266***	0.200***
	(0.07)	(0.06)
Low Cognition	0.312***	0.420***
	(0.06)	(0.06)
Treatment*Low Cognition	0.036	0.046
	(0.07)	(0.07)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.000	0.000
Control Mean	1.55	1.66
Observations	6473	6509

Panel 2: Payoffs from Trust Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.222***	-0.071**
	(0.07)	(0.03)
Low Cognition	-0.165***	-0.166***
	(0.03)	(0.02)
Treatment*Low Cognition	0.062	-0.028
	(0.05)	(0.03)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.000	0.001
Control Mean	5.37	5.43
Observations	6358	6369

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. In Panel 1, the dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to the anonymous classmate, and anonymous out-of-school peer, respectively. In Panel 2, the dependent variables are expected payoff gains from the trust game. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B13: Treatment Effects on Reciprocity

	In-Group				Out-Group			
	1	2	3	4	1	2	3	4
Treatment	0.056***	0.048***	0.046***	0.039**	0.057***	0.036**	0.038**	0.048***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Low Cognition	0.087***	0.076***	0.082***	0.078***	0.083***	0.092***	0.076***	0.085***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Treatment*Low Cognition	0.005	0.019	0.001	0.020	0.003	0.010	0.020	0.019
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.000	0.000	0.002	0.000	0.000	0.003	0.001	0.000
Control Mean	0.48	0.43	0.40	0.39	0.48	0.43	0.40	0.39
Observations	6493	6494	6489	6478	6498	6498	6481	6428

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables in column 1-4 are fraction of tokens reciprocated back in case of receiving 1 token, 2 tokens, 3 tokens, and 4 tokens from an anonymous classmate, respectively. Dependent variables in column 5-8 are fraction of tokens reciprocated back to an anonymous out-of-school peer. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B14: Treatment Effects on Altruism

	Willingness to Donate	Donation Amount
Treatment	0.384***	0.060***
	(0.14)	(0.02)
Low Cognition	0.222**	0.064***
	(0.09)	(0.01)
Treatment*Low Cognition	0.038	-0.005
	(0.12)	(0.02)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.001	0.001
Control Mean	0.72	0.37
Observations	6574	6574

Reported estimates in column 1 are average marginal effects obtained from a logit regression, in column 2 are obtained from an Ordinary Least Square regression. 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. 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 randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B15: Heterogenous Treatment Effects on Cooperation

Panel 1: Cooperation

	In-Group Cooperation	Out-Group Cooperation
Treatment	0.194**	0.074
	(0.09)	(0.09)
Low Cognition	0.073	0.178*
	(0.07)	(0.10)
Treatment*Low Cognition	-0.022	-0.124
	(0.10)	(0.11)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.014	0.304
Control Mean	0.53	0.52
Observations	6565	6571

Panel 2: Payoffs from Cooperation Game

	In-Group Payoffs	Out-Group Payoffs
Treatment	0.111*	-0.057
	(0.06)	(0.07)
Low Cognition	-0.086	-0.132*
	(0.06)	(0.08)
Treatment*Low Cognition	0.036	0.094
	(0.08)	(0.08)
Strata FE	✓	✓
Baseline Controls	✓	✓
p-val (Test: Treatment*Low Cognition+Treatment=0)	0.013	0.302
Control Mean	4.54	4.45
Observations	6565	6571

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Standard errors are in parentheses and clustered at the school level (unit of randomization). Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. In Panel 1, reported estimates are average marginal effects obtained from logit regressions where the binary dependent variable takes the value 1 if the child choose to cooperate (green card), and zero otherwise. In Panel 2, the dependent variables are expected payoff gains from the cooperation game

B.4 Class Level Refugee Share

Table B16: Student and Teacher Reported Violence

	Student Self Report			Teacher Report		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.011	-0.051*	0.035	-0.000	0.006	0.098
	(0.03)	(0.03)	(0.02)	(0.12)	(0.09)	(0.09)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.76	0.81	0.79	-0.07	0.04	-0.05
Observations	2162	2192	1978	2083	2134	1814

The dependent variable in columns 1-3 is a dummy which equals to 1 if the student reports physical and verbal bullying. The column presents estimated marginal effects from a logit regression. The dependent variable in columns 4-6 is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. Reported estimates in column 2 are obtained from ordinary least squares (OLS) regressions. 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. All regressions control for randomization strata and baseline covariates. Baseline covariates include cognitive ability (Raven's Matrices), cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B17: Heterogenous Treatment Effects on Trust

Panel 1: Number of Tokens Sent

	In-Group Trust			Out-Group Trust		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.269***	0.275**	0.253***	0.247**	0.231***	0.223***
	(0.09)	(0.11)	(0.07)	(0.09)	(0.08)	(0.06)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	1.26	1.38	1.49	1.36	1.46	1.52
Observations	2204	2228	2041	2222	2239	2048

Panel 2: Payoffs from Trust Game

	In-Group Payoffs			Out-Group Payoffs		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.215**	0.263**	0.245***	-0.111**	-0.071*	-0.081*
	(0.10)	(0.11)	(0.08)	(0.05)	(0.04)	(0.04)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	5.26	5.39	5.48	5.26	5.39	5.48
Observations	2164	2196	1998	2170	2202	1997

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. In Panel 1, the dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to the anonymous classmate, and anonymous out-of-school peer, respectively. In Panel 2, the dependent variables are expected payoff gains from the trust game. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B18: Treatment Effects on Reciprocity

Panel 1: 3 Tokens Received

	In-Group Reciprocity			Out-Group Reciprocity		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.035	0.069**	0.062***	0.075**	0.048*	0.064***
	(0.03)	(0.03)	(0.01)	(0.03)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.41	0.43	0.46	0.40	0.41	0.44
Observations	2218	2234	2041	2212	2239	2047

Panel 2: 6 Tokens Received

	In-Group Reciprocity			Out-Group Reciprocity		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.054**	0.061***	0.062***	0.061**	0.037*	0.037*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.37	0.39	0.41	0.38	0.41	0.43
Observations	2214	2240	2040	2216	2238	2044

Panel 3: 9 Tokens Received

	In-Group Reciprocity			Out-Group Reciprocity		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.045*	0.049**	0.038**	0.067**	0.036*	0.045**
	(0.03)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.34	0.36	0.38	0.36	0.38	0.39
Observations	2213	2233	2043	2208	2239	2034

Panel 4: 12 Tokens Received

	In-Group Reciprocity			Out-Group Reciprocity		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.054*	0.039*	0.059***	0.072**	0.047**	0.048**
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.33	0.36	0.37	0.35	0.37	0.39
Observations	2211	2231	2036	2192	2217	2019

Reported estimates are obtained from ordinary least squares (OLS) regressions. 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. Dependent variables are fraction of tokens reciprocated back in case of receiving 1 token (column 1), 2 tokens (column 2), 3 tokens (column 3), and 4 tokens (column 4) from an anonymous classmate. Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B19: Treatment Effects on Altruism

	Willigness to Donate			Fraction Donated		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.047	0.111***	0.065**	0.042	0.085***	0.033
	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
Ethnic Reference	-0.023	-0.045*	0.028	-0.008	0.002	0.024
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment*Ethnic Reference	0.057*	0.008	0.033	0.030	0.006	0.003
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.70	0.67	0.73	0.70	0.67	0.73
Observations	2243	2267	2064	2243	2267	2064

Reported estimates in column 1 are average marginal effects obtained from a logit regression, in column 2 are obtained from an Ordinary Least Square regression. 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. 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 randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies.

Table B20: Treatment Effects on Cooperation

Panel 1: Cooperation

	In-Group Coordination			Out-Group Coordination		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.013	0.070**	0.061***	-0.005	-0.016	0.011
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	0.52	0.52	0.53	0.49	0.50	0.51
Observations	2244	2262	2059	2248	2264	2059

Panel 2: Payoffs from Cooperation Game

	In-Group Payoffs			Out-Group Payoffs		
	Low Share	Medium Share	High Share	Low Share	Medium Share	High Share
Treatment	0.044	0.215**	0.200***	0.014	0.047	-0.033
	(0.09)	(0.10)	(0.05)	(0.06)	(0.06)	(0.06)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Control Mean	4.55	4.57	4.58	4.52	4.51	4.47
Observations	2244	2262	2059	2248	2264	2059

Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels. Standard errors are in parentheses and clustered at the school level (unit of randomization). Regressions control for randomization strata and baseline covariates. Baseline covariates include cognition, cognitive empathy, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. In Panel 1, reported estimates are average marginal effects obtained from logit regressions where the binary dependent variable takes the value 1 if the child choose to cooperate (green card), and zero otherwise. In Panel 2, the dependent variables are expected payoff gains from the cooperation game

C Instructions for Incentivized Games

We will play some fun games with you today for the next 3 lecture hours. In these games, you will make some choices, and depending on your choices, you will earn different amounts of gifts. We brought a variety of lovely gifts for you [Show the gift basket]. To get the gifts, you need to earn tokens as each gift in our basket has a different token value. The more tokens you earn, the more gifts you will be able to get.

Each game has a set of rules, but we also have an important ground rule. We ask you to make sure that you keep your choices to yourselves and never share them with anyone during the games. Do we understand?

We will play 4 games today [Write down "Game 1", "Game 2", "Game 3," and "Game 4" on the board]. At the end of our visit, we will randomly select one game, and you will receive the tokens you earned only from that game. In other words, your tokens will not accumulate game after game. Therefore, make careful choices in each game.

Trust Game

Game 1

Let us begin with our first game [Circle "Game 1"]. For this game, everyone has 4 tokens. This game is played as pairs that everyone in this class has a pair. But you cannot choose your pair. We match each of you with a student in this class, and you will not know who he/she is. Your pair could be anyone in this classroom: could be a girl or a boy, could be someone sitting in the first row, middle row, or the last row. Do we understand?

There are two roles in this game: sender and receiver. [write these words on the board]. You will either be a sender or a receiver, but you do not know your role right now. If you are the sender, then your pair is the receiver and vice versa.

Let's see what these roles mean:

Remember that everyone has 4 tokens. The sender will make the first move in this game. He can keep his tokens to himself, or he can send some of them to the receiver. He can send 0 (nothing), 1,2,3, or 4 (all of his tokens). There is no right or wrong decision here.

The tokens he sends to the receiver will triple on the way. For example, if he sends 1

token, then the receiver receives 3 tokens. If he sends 2 tokens, the receiver gets 6 tokens. [Ask the classroom] If he sends 3 tokens, how many tokens will the receiver receive? 9. If he sends 4 tokens, how many tokens will the receiver receive? 12.

Now let's see the role of the receiver in this game. The receiver will need to decide how many of the tokens he/she received from your pair to send back to him/her. How? [Give the examples below, and explain each case]

Example 1: Let's say the sender did not send any token to the receiver, what can the receiver do? As the receiver cannot send any token back, both the sender and the receiver will finish up with 4 tokens each.

Example 2: Let's say the sender sends one token [Draw a line between the sender and the receiver and write down "1" above the line]. How many tokens does the sender have left? 4-1=3. [Write down 3 under the sender and erase 4]. How many tokens does the receiver get? 3X1=3. Since the receiver gets 3 tokens, she can send back 0 (nothing), 1, 2, or 3 tokens. It is up to him/her.

- If the receiver decides to keep all her tokens, i.e., sends back 0 tokens, he will end up with 7 tokens and the sender 3 tokens.
- If the receiver sends back 1 token, he will end up with 6 tokens, and the sender 4 tokens.
- If the receiver sends back 2 tokens, he will end up with 5 tokens, and the sender will get 5 tokens.
- If the receiver sends back 3 tokens, he will end up with 4 tokens and the sender 6 tokens.

Example 3: Let's say the sender sends 2 tokens. [Repeat the examples accordingly].

Example 4: Let's say the sender sends 3 tokens. [Repeat the examples accordingly].

Example 5: Let's say the sender sends all his tokens, i.e., 4 tokens. [Repeat the examples accordingly].

[Distribute the booklets] Now write your name, surname, and classroom on the first page of the booklet. Do not start yet.

Now, suppose that you are the sender. How many tokens would you like to send? [wait until all students make their choices]. Now turn the page and choose one of the answers for the first question.

How many tokens do you think you will receive back? All of them, less than half, half, more than half. [wait until all students make their choices]

Now suppose that you are the receiver, and the sender sent you 1 token, it triples and becomes 3 tokens. How many would you send back to him/her?

Suppose the sender sends you 2 tokens, it triples and becomes 6 tokens. How many would you send back to him/her?

Suppose the sender sends you 3 tokens, it triples and becomes 9 tokens. How many would you send back to him/her?

Suppose the sender sends you 4 tokens, it triples and becomes 12 tokens. How many would you send back to him/her?

Game 2

Now we will play the second game. The second game is exactly the same as the first game except now, you will be matched with a student from another school. [Proceed in the same way as above]

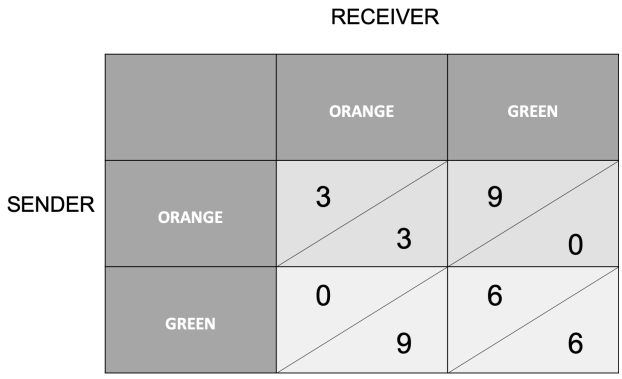
Cooperation Game

Game 3

In this game, you will again be paired with a student in this classroom. You don't know who will be your pair. Both you and your pair will make a decision simultaneously. The decision is to choose one of two cards, orange or green. Depending on the color of the card you choose and the card chosen by your pair, you will earn a different amount of tokens.

[Draw the table on the board]

Figure B1: Payment Scheme



If you choose the orange card and your pair chooses the orange card, both of you will receive 3 tokens. If you select orange and your pair chooses green, you will receive 9, your pair will receive 0. If you choose green and your pair chooses orange, you will receive 0, your pair receives 9 tokens. If both of you choose green, both will receive 6 tokens.

[Ask several questions and make sure students understand the game] [Allow students to make their choices]

Now you will make a guess about your pair's choice. [students make their choices]

Game 4

We will now play this game again. This time your pair is a student from another school. Someone you do not know.

[Students make their choices]

Dictator Game

Now, we give you 4 extra tokens. We will distribute a sheet to each of you now, and please read what is written on your sheet and make your decision regarding these 4 tokens. As soon as you make your decision, deliver the sheets to us without showing anyone.

Students were distributed sheets, and a random half of the class received sheets that read as follows:

Booklet A

How many of your 4 tokens would you like to donate to a child in another school, which

we could not visit and distribute gifts? Please choose the number of tokens you would like to donate.

The other half received sheets that read as follows:

Booklet B

How many of your 4 tokens would you like to donate to a Syrian child in another school which we could not visit and distribute gifts? Please choose the number of tokens you would like to donate.

D Survey Instruments

Perspective Taking:

- I try to understand how others feel.
- When I see someone being taken advantage of, I become protective towards him.
- I can put myself in someone else's shoes and understand how they feel.
- I can tell if a friend of mine is upset.

Impulsivity:

- I interrupt people when they are talking.
- I stop and think before I do something.
- I tend to say the first thing that comes to mind.
- I wait my turn in a game.
- I cannot help but touch things without getting permission.
- I get into trouble because I do things without thinking first.
- I can control my temper in conflict situations.
- I answer questions in class before the teacher lets me speak.

Empathetic Concern:

- When I see someone being treated unfairly, I feel very much pity for them
- I often have tender, concerned feelings for people less fortunate than me
- I feel sorry for other people when they are having problems
- My friends talk to me about their problems
- I would describe myself as a pretty soft-hearted person

Ethnic Bias:

- I don't want to be friends with kids from another country.
- I like children from other countries as much as my other friends.
- I think children from other countries often tend to start fights.
- I think children from other countries are not as smart as we are.

Descriptive Classroom Norms:

- My classmates make fun of each other.
- My classmates talk behind each other.
- My classmates hit each other and get into fights.
- My classmates make fun of students who come from other countries.
- My classmates are nice to each other.
- My classmates beat students who come from other countries.
- My classmates protect each other.

E Curriculum Content Examples and Select Activities

GENERAL PURPOSE

This program aims to develop

1. cognitive empathy, the ability to look at events and people from others perspective
2. curiosity about individual differences, understanding and respecting them
3. trust-oriented communication with their social environment
4. the ability to oppose all kinds of violence in principle

EXPECTED LEARNING OUTCOMES

Students will,

1. Enhance cognitive empathy.
2. Learn that individual differences are needed, desired, and enriching factors in our lives.
3. Be respectful for individual differences.
4. Want to establish and preserve a cohesive and inclusive culture in their groups.
5. Develop a principled attitude toward oral and physical violence in-class and in school.
6. Trust people around them and be trusted in return.
7. Develop the ability to act in harmony with team spirit.

Table B21: "Understanding Each Other" Curriculum

TOPIC 1: WHAT IS EMPATHY? **Purpose:** Introducing students to the concept of empathy **Learning outcome:** Students learn what kind of a character trait empathy is. **The Material of the Week:** Activity	**TOPIC 2: GETTING TO KNOW EMOTIONAL CUES** **Purpose:** Teaching students to recognize social cues **Learning outcome:** Students learn to make inferences from social cues. **The Material of the Week:** Activity	**TOPIC 3: DIFFERENT PEOPLE, SAME EMOTIONS** **Purpose:** Conveying students that we are all similar in our emotions **Learning outcome:** Students learn that all individuals share the emotions like pain, happiness and embarrassment. **The Material of the Week:** Video, Activity
TOPIC 4: UNDERSTANDING MY FRIEND **Purpose:** Teaching students to solve problems by adopting the perspective of another **Learning outcome:** Students learn a problem solving strategy by adopting another's point of view in a familiar scenario. **The Material of the Week:** Reading exercise	**TOPIC 5: UNDERSTANDING THE FEELINGS OF CREATURES** **Purpose:** Teaching students that animals, like humans, also need to be understood and respected **Learning outcome:** Students learn that not only humans, but also animals need to be understood and respected. **The Material of the Week:** Reading exercise	**TOPIC 6: UNDERSTANDING UNSAID THOUGHTS** **Purpose:** Fostering the ability of understanding and problem-solving in social interactions by making inferences from social cues **Learning outcome:** Students learn to understand other individuals in social situations. **The Material of the Week:** Activity
TOPIC 7: INJUSTICE AND ITS SOLUTION **Purpose:** Teaching students to exhibit a principled attitude when they witness a wrongdoing. **Learning outcome:** Students learn the importance of opposing to anti-social behaviors in principle. **The Material of the Week:** Reading exercise	**TOPIC 8: PUTTING ONESELF INTO SOMEONE ELSE'S SHOES-1** **Purpose:** Showing students two different points of views for the same situation and helping them to gain perspective. **Learning outcome:** Students learn that there could be two sides to the same story. **The Material of the Week:** Reading exercise 1	**TOPIC 9: PUTTING ONESELF INTO SOMEONE ELSE'S SHOES-2** **Purpose:** Showing students two different points of views for the same situation and helping them to gain perspective. **Learning outcome:** Students learn that there could be two sides to the same story. **The Material of the Week:** Reading exercise 2
TOPIC 10: UNDERSTANDING EMOTIONAL SIGNALS **Purpose:** Reinforcing students' understandings of social signals. **Learning outcome:** Students learn to quickly analyze anti- social situations and exhibit a principled stance. **The Material of the Week:** Activity	**TOPIC 11: DO WE KNOW EACH OTHER?** **Purpose:** Helping students to communicate with all of their friends in the class. **Learning outcome:** Students will get to know more about their classmates who were less familiar to them before. **The Material of the Week:** Activity	**TOPIC 12: BEAUTIFUL WORDS AND BEAUTIFUL EMOTIONS** **Purpose:** Teaching students the importance of positive attitudes and words for healthy social relations. **Learning outcome:** Students will learn the benefits of positive words and behavior in social interactions. **The Material of the Week:** Activity
TOPIC 13: I AM ABLE TO CONTROL MY ANGER **Purpose:** Teaching students to find constructive solutions to conflicts by controlling intense emotions. **Learning outcome:** Students will learn to cope with emotions like anger, rage and find solutions to the conflicts in a calm manner. **The Material of the Week:** Video, Activity	**TOPIC 14: WHAT KIND OF A CLASS ARE WE?** **Purpose:** Reinforcing a health classroom culture **Learning outcome:** Students will understand the importance of forming a classroom culture with a high level of tolerance. **The Material of the Week:** Video, Activity	**TOPIC 15: OUR EMPATHETIC CLASSROOM** **Purpose:** Giving awards (feedback) to students. **Learning outcome:** Students will feel proud of having built building a classroom culture. **The Material of the Week:** Activity

Figure B2: Understanding Each Other



Figure B3: Curriculum



Figure B4: Example activity from the book

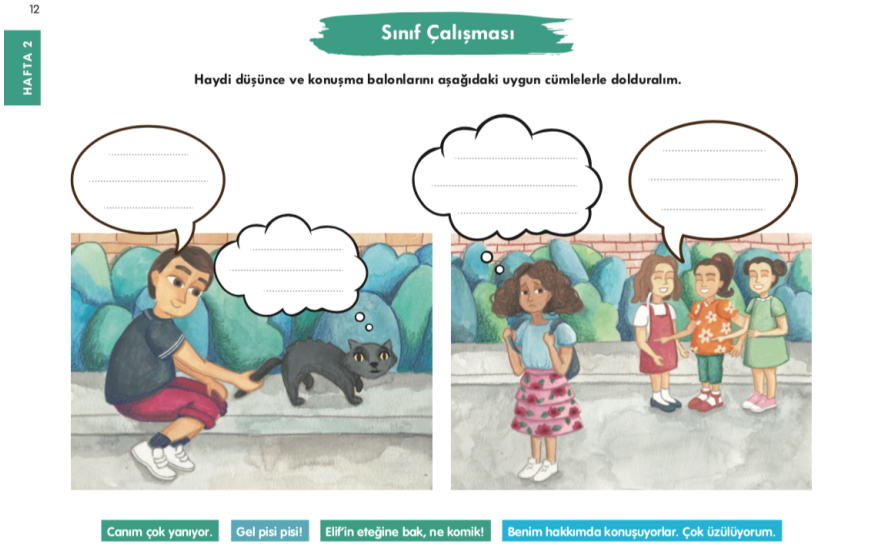


Figure B5: Example activity from the book

