Building Social Cohesion in Ethnically Mixed 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

October 2020

Abstract

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

JEL Codes: I24, I28, C93

Keywords: social cohesion; social exclusion; ethnic segregation; perspective taking

---

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

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

1 Introduction

> *"...public education does not serve a public. It creates a public. The question is, what kind of public does it create? A conglomerate of self-indulgent consumers? Angry, soulless, directionless masses? Indifferent, confused citizens? Or a public imbued with confidence, a sense of purpose, a respect for learning and tolerance?"*
>
> – Postman 1996, p.18

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

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

[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.

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).[2] Studies show that perspective-taking is associated with lower social aggression, higher trust, and social cooperation (e.g. see Batson et al., 1997, Galinsky and Moskowitz, 2000, Galinsky and Ku, 2004). High perspective-taking ability is also related to being able to analyze social situations through slow deliberations, weighing the costs and benefits of an action before engaging in the act. Studies show that this type of deliberation is a malleable skill, and is effective in reducing crime and violent behavior in various contexts (Heller et al., 2017, Blattman et al., 2017, Alan and Ertac, 2018). Motivated by these findings, a multidisciplinary team of educators, pedagogical consultants, and multimedia developers designed a program as a set of curricular activities to develop children's ability to understand each others' perspectives and their capacity to make inferences about others' intentions, goals, and motives. The program designers took great care to ensure that the content makes no explicit reference to ethnicity. Instead, they aimed to encourage students to exert effort to understand the perspective of any individual, or living being, regardless of their identity.

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

[2] 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.

We collected endline data in May 2019.

While there is no universal definition of social cohesion, there are widely accepted indicators that characterize a cohesive environment. These include low incidents of violence, high prevalence of inter-ethnic social ties, trust, reciprocity, and cooperation between individuals.[3] To evaluate the program, we put together a multidisciplinary toolkit that measures the cohesiveness of the school and classroom environment based on these indicators. Our toolkit includes i) administrative diary logs recording high-intensity peer violence and victimization on school grounds, ii) carefully elicited social networks to measure social exclusion and ethnic segregation in classrooms, 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 item-response questionnaires to measure behavioral norms, ethnic bias, perspective taking, empathetic concern and impulsivity (Fehr and Schmidt, 1999, 2000, Boisjoly et al., 2006, Burns et al., 2019, Rao, 2019).

We find that the program is highly effective in lowering high-intensity peer violence and victimization on school grounds. Over ten consecutive school days, about 1.9 violent events were perpetrated by children in the control group. This number is reduced by more than 60% in treatment schools. This substantial treatment effect is statistically significant at the 1 percent level. Given this result, we also explore whether the program had an unintended effect of generating more victims. The idea behind this concern is that by encouraging children to show understanding toward their peers in a generally violent environment such as our study site, the program may have made them more susceptible to victimization. We find, on the contrary, that the program significantly reduced the victimization of children, suggesting that by keeping children away from conflict, the program also lowered the risk of being a victim of a violent act.

The program also reduces the probability of social exclusion and increases the likelihood of forming inter-ethnic social ties, thereby reducing ethnic segregation in the classroom. We find that both hosts and refugees are significantly less likely to be socially excluded and more likely to receive emotional and academic support from their classmates in treated schools. Overall, treated children are about 6% (8%) more likely than untreated children to receive emotional (academic) support from their classmates. Moreover, refugee children in treated schools are approximately 25% (21%) more likely than those in control schools to receive

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

emotional (academic) support from their host classmates. Finally, we show that the program reduces ethnic segregation in classrooms by 15 to 21%.

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

Overall, the program appears to be highly effective in building a cohesive school environment. Refugee children emerge as the primary beneficiaries of this environment. In addition to improving their social interactions with their classmates, the program significantly improves refugee children's ability in the language of the host country, which is an essential marker for successful integration. Despite that the program did not have an academic focus, treated refugee children received 0.13 standard deviations higher scores in the objective Turkish language test we implemented in classrooms. Such a remarkable improvement in the host country's language suggests that creating a peaceful and cohesive learning environment, where inter-ethnic support ties are easily formed, is critical for the achievement of minority children, and as such, a prerequisite for a successful integration policy (Fryer Jr and Levitt, 2004, Guryan, 2004, Echenique et al., 2006, Card and Rothstein, 2007, Hanushek et al., 2009).

Our exploratory analyses suggest that these positive effects stem mainly from the program's effectiveness in increasing children's effort to take others' perspectives. Treated chil-

dren report exerting higher effort to understand others' perspectives and higher capacity to tolerate individual differences. While the increased effort of perspective-taking emerges as a likely channel, we explore several other possible mechanisms using self-reported measures. In particular, we test whether the program also works through improving behavioral norms, reducing ethnic bias, increasing empathetic concern, and enhancing the ability to regulate impulsive behavior. We find some evidence, albeit weak, that in addition to enhancing perspective-taking, the program improves behavioral norms in the classroom and children's ability to self-regulate impulsivity.

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

Our study relates to several bodies of literature. First, it complements the research on reducing crime and violence in schools through behavioral interventions and policy changes (Lochner and Moretti, 2004, Heller et al., 2017, Alan and Ertac, 2018). Our study is unique in that it evaluates an educational program that cultivates a particular socio-cognitive skill, perspective-taking, that has not yet been tested in a large-scale field setting with inter-ethnic dynamics. Second, this study is relevant to the literature that tests the "contact hypothesis." There are experimental and quasi-experimental studies that test the contact hypothesis by evaluating interventions that facilitate inter-group contact through various activities (Bazzi et al., 2019, Burns et al., 2019, Paluck et al., 2019, Lowe, 2020, Mousa, 2020) or a policy change in which poor students were enrolled into elite private schools (Gould et al., 2004, Rao, 2019). Our study complements this literature by testing whether fostering a particular socio-cognitive skill in the classroom environment can improve peer relations in a context where inter-group contact is already high, and there is evidence of social exclusion and maltreatment. Third, our study advances existing perspective-taking related field experiments. Previous studies have focused on short, priming-type interventions concerning adults in settings where between-group interactions are rare and measured outcomes are mostly limited to self-reported beliefs and do not involve any explicit interactions between groups (Broock-

man and Kalla, 2016, Adida et al., 2018). Finally, our study contributes to the growing literature on the development of socio-emotional skills by providing causal evidence on the malleability of perspective-taking ability in young children (Heckman et al., 2006, Deming, 2009, Heckman et al., 2013, Alan and Ertac, 2018, Alan et al., 2019, Cappelen et al., 2020, Eisner et al., 2020, Kosse et al., 2020).

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

The program we evaluate is designed, implemented, and evaluated in this sociopolitical context. It is an educational cohesion program targeted at 3rd and 4th-grade elementary school children. It aims to provide teachers with an easy-to-follow curriculum to build cohesion in the classroom and ensure a healthy learning environment for all children. The curriculum content comprises written and animated class activities compiled as a modular book known as "Understanding Each Other." The curriculum takes perspective-taking as the core concept. It encourages students to understand and experience the emotions of the described subject through various reading and visual materials. For example, in an animated video, children see several adverse events (e.g., falling while running after a ball and hurting knees) that happen to a character, followed by a similar event happening to another character. This animated material aims to emphasize the similarity of the effects of hurtful events on different people. Such an event occurs randomly across ethnic groups, so its

purpose is to make sure that children exert effort to take others' perspectives in all relevant social situations. Another example includes reading material, such as a diary extract of a hypothetical student who arrives at a new school. Students then read a diary extract from another child who writes about a new friend's arrival from another country. 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 activities and games implemented by the teacher. For example, after watching an animated video that highlights an act of social exclusion or malfeasance toward animals, children are asked to guess what the characters in the video must be feeling and fill up thinking balloons. The effectiveness of this type of deliberation in building perspective-taking ability is emphasized in the psychology literature (Galinsky et al., 2005).[4] Instead of making any explicit ethnic reference, normative or otherwise, the program encourages tolerance toward individuals (and animals, for that matter) and cherishing individual differences.

3 Evaluation Design and Cohesion Outcomes

The program was implemented as a cluster randomized controlled trial. The study sample contains 222 classrooms (teachers) from 80 elementary schools in Sanliurfa and Mersin, two provinces of Turkey where the refugee placement program has been in effect since 2016. The study covers over 6500 3rd and 4th-grade children ages 8 to 12. Approximately 16% of the children in our sample are refugees. The sample schools are very large, which is typical of our study site. The average number of 3rd and 4th-grade classrooms per school in our sample is 15. Since the program was oversubscribed, and we aimed to collect detailed data that involves implementing time-consuming incentivized tasks and social network elicitation, we sampled on average 3 classrooms (teachers) per school.

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

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

by absenteeism to increase the power of our design as absenteeism is highly predictive of educational attainment and is a particularly pressing concern in this part of Turkey. Many families in southeastern Turkey, now including refugee families, work as seasonal agricultural workers, usually leaving their homes for work in May and returning in October. The ex-ante probability of treatment is set to 50%, assigning 40 schools to treatment, and 40 to control ex-post.

Teacher training seminars for 40 treatment schools (124 teachers) took place in the first week of November 2018. In these seminars, teachers were introduced to the concept of perspective-taking and its importance for children's cognitive and socio-emotional development. They then participated in an intensive workshop where they studied the "Understanding Each Other" module and related activities chapter-by-chapter and interactively with their designated education consultants. Teachers were provided with a detailed implementation kit, in hard and soft copy, explaining the module's particulars and accompanying activities. Teachers were expected to spend three lecture hours per week to cover the curriculum throughout the academic year of 2018-2019. 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.

The Turkish MoE allows (and encourages) all elementary school teachers to implement socially beneficial extra-curricular projects for a maximum of 5 lecture hours per week. Being involved in Ministry-approved extra-curricular projects is common practice for Turkish elementary school teachers. During the implementation of this program, our control teachers were also engaged in various extra-curricular projects, typically related to the environment, health, personal hygiene, financial awareness and more. In the absence of extra-curricular projects, teachers tend to use these free hours for supervised arts and games activities. Therefore, the program we evaluate did not crowd out core teaching activities. More importantly, because these 5 extra-curricular hours are mandated to be used under teachers' supervision, the number of hours children have contact with their peers and teachers remained the same across treatment and control classrooms.

Both baseline and endline data collection were carried out by the research team, assisted by locally recruited and trained field assistants. We spent about 3 lecture hours in each classroom, both at baseline and endline, to conduct incentivized games, tests, and surveys. Data from children were collected using pen and paper. Teachers were not present in the classroom during data collection. They were in isolated rooms, completing their paper-based

surveys. Coding and digitizing the data took about three months after the completion of endline fieldwork.

The trial was registered at the AEA Registry along with a pre-analysis plan (PAP). Unless we indicate otherwise, presented analyses and related outcomes are pre-specified in our PAP. 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 overcame this difficulty by obtaining a special permit to collect these data ourselves from administrators. Our permit allows us to collect these data at the school-level without referring to any particular student. Our peer violence measure is the number of high-intensity disciplinary episodes that took place on school grounds in the last ten school days following our endline visit. 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 chose 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 extra-curricular programs are continuously in effect, made it possible for us to designate an independent administrator for this task.

The diary log is an electronic Excel spreadsheet. At the top of the spreadsheet, specific classroom identifiers are highlighted. These are the classrooms included in our evaluation study.[5] The spreadsheet has four columns. The first column indicates the date. In the second column, the administrator was asked to record the number of high-intensity disciplinary events in the school at the end of each day without referring to any particular classroom or a

[5]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 with 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.

child. In the third column, the administrator was asked to record the number of events that were perpetrated by someone from the classrooms highlighted at the top of their diary sheet, without identifying the perpetrators. In the final column, the administrator was asked to record the number of events where someone from the highlighted classrooms was victimized, again without identifying the victimized child. We added a measure of victimization to the diary log to establish whether the program has an unintended effect of generating more victims. The idea behind this concern is that because the program encourages children to be more understanding of others in a generally non-cohesive and violent environment, it may make them more vulnerable to perpetrators. An example diary log is provided in Figure 2.

We sampled only a few classrooms from each school to include in our study because most of our schools have a large number of students and classrooms. Therefore, we do not expect the program to have a significant impact on overall school-level peer violence. We do expect, however, that the program to be effective in reducing the number of violent events perpetrated by children from treated classrooms.

We also collected data from children on peer violence using surveys, both at baseline and endline. For this, we asked children about their experiences of bullying perpetrated by classmate(s) as well as schoolmate(s) from outside the classroom (but in the child's school).[6] Finally, we asked teachers to rate each student's behavioral conduct using a 1 to 5 scale where one refers to very good, and five refers to very violent and anti-social behavior. This measure is available only at endline.

3.2 Social Exclusion and Ethnic Segregation

The prevalence of social exclusion based on personal characteristics, such as ethnicity, is another measure of the level of cohesiveness of an environment. Social exclusion based on personal characteristics may lead to the formation of groups identified with such characteristics (segregation) or social isolation of an individual. To construct social exclusion and ethnic segregation measures, we elicited social networks in classrooms. To do this, we provided children with a user-friendly paper template and asked for nominations of up to three classmates in three categories of social ties: friendship, emotional support, and academic support, allowing for overlaps across categories. For emotional and academic support, the

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

exact wordings are "classmates who help you when you feel sad" and "classmates who help you with homework," respectively. Before we begin our elicitation, we told children that they could also nominate friends who were absent that day.[7] We collected these data both at baseline and endline.

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

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

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

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

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

[7] We designed the template as three boxes to write one classmate in each box for each category. We gave detailed examples of how to fill up the template before starting the elicitation and made sure children fully understood the procedures.

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 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 3 illustrates observed friendship ties from two classrooms in our data. Both classrooms have similar sizes (29 and 28) and a similar number of refugee students (5 and 6). It can be seen clearly that classroom 1 is more ethnically segregated than classroom 2. The segregation index scores, ES_c, are 0.22 and 0.01 for classrooms 1 and 2, respectively. Figure 4 depicts the cumulative distribution of the expected and observed proportion of inter-ethnic ties for all three categories of social ties (friendship, emotional support, academic support) at baseline. We observe substantial ethnic segregation for all three social tie categories at baseline. We expect the program to lower the distance between expected and observed inter-ethnic ties, i.e., classroom-level ethnic segregation.

3.3 Experimentally Elicited Prosocial Behaviors

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

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

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

We design this game using a strategy method such that students make decisions on how much to send if they assume the role of a sender, and how much to send back if they assume the role of the receiver.[9] 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-school." The amount of tokens sent is our measure of "trust," and the amount of tokens sent back is our measure of "reciprocity." We expect the program to increase trust and reciprocity in children, both toward their classmates and out-school peers.

Cooperation game, which is a modified version of one-shot prisoner's dilemma game, also

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

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

involves two participants to be anonymously paired. We similarly design this game and have in-class and out-school versions. Children are endowed with three gift tokens for this game. The game involves choosing a card that is either green or orange. A child's payoff depends on both the color she chooses, and that of her pair chooses. The payoff scheme is given in Figure 5. As can be seen, choosing the green card is the decision to cooperate. We refer to the binary choice of the green card as "cooperative" action and expect that the program increases the probability of cooperative action. Full instructions for trust and cooperation games are given in Online Appendix D.

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

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

3.4 Self-Reported Cohesion Outcomes

We also collected data from children on perspective taking, empathetic concern, impulsivity, and ethnic bias using item-response questions, both at baseline and endline. The primary motivation to collect these outcomes is to substantiate our conjectured mechanism as well as to detect/rule out other potential channels. We also collected data on descriptive classroom norms at endline to assess whether the treatment improves cohesion by improving behavioral

norms in the classroom. For this, children were asked item-response questions regarding their classmates' general behavioral conduct. The questions used to measure perspective-taking, impulsivity, empathetic concern, ethnic bias, and behavioral norms are presented in Online Appendix E.

3.5 Achievement Outcomes

A healthy school environment is essential to ensure academic achievement. To test whether the program facilitated the integration of refugees without hurting host children academically, we implemented math and Turkish language tests in classrooms both at baseline and endline. We prepared these tests separately for 3rd and 4th graders, based on the national curricula. Because the program had no academic focus, we did not specify these outcomes in our PAP. Nevertheless, we analyze these data and present the estimated treatment effects on standardized math and Turkish verbal ability of host and refugee children.

4 Data

4.1 Descriptive Statistics and Internal Validity

Before randomizing, we visited all 80 schools (222 classrooms) and collected detailed baseline data on demographics, self-reported experiences of bullying, perspective-taking, empathetic concern, impulsivity and ethnic bias. We also implemented math and Turkish language tests, measured fluid cognitive ability using Raven's progressive matrices (Raven et al., 2004), and emotional intelligence using Reading the Mind in the Eyes Test (Eyes Test) (Baron-Cohen et al., 2001). The latter is commonly used to measure individual differences in theory of mind and shown to be weakly related to cognitive empathy and emotion perception and strongly related to vocabulary (Olderbak et al., 2015). Finally, we elicited social networks and measured cooperation and altruism using incentivized games at baseline. Except for fluid cognitive ability (Raven's score), all these outcomes were also collected at endline. We added two versions of trust games, a modified dictator game and behavioral norms questionnaire to our endline inventory. We also collected important baseline information from teachers. In addition to standard demographics, we tested teachers' fluid cognitive ability and emotional intelligence with the same tests we use for children (Raven's Progressive Matrices and Eyes Test).

Table 1 presents the balance of baseline variables across treatment status. The first panel presents the balance in student characteristics. The second panel presents classroom and teacher characteristics, and the last panel shows the balance in school characteristics. Note first that about 16 percent of our sample consists of refugee children at baseline. The table shows no significant imbalance in any of the variables except for the proportion of students who reported being bullied by their classmates (significant at 10 percent level). As shown in Panel 3, the schools in our sample are of considerable size. The average number of 3rd and 4th-grade classrooms is about 15, with approximately 500 students.

Table 2 presents some baseline descriptive statistics for refugee children. The table shows that refugee children are significantly more likely to be socially excluded and subject to regular bullying. They are about 4.2 percentage points more likely to report experiences of bullying, 6.3 percentage points less likely to have a friend in their classroom, 12 and 10 percentage points less likely to receive emotional and academic support from their classmates. In what follows, we will present the effect of the program on the cohesion outcomes we detailed above for all children. We will also present heterogeneity results by refugee status to see if the program benefits host and refugee children differently.

5 Results

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

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

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 predictive of the outcome y. The latter includes age, gender, refugee status, Raven's score, Eyes Test score, the outcome variable collected at baseline and a dummy for developmentally challenged students.[10] We also control for class size, school size, and district fixed effects. $Other_{is}$ captures other variables (for particular outcomes) that might be added for specific regressions, and δ_b are

[10]The Turkish MoE has an active policy to place a small number of students with some learning disabilities in classrooms to facilitate inclusion. These students, if present, were identified by teachers for us before we commenced our data collection. We took great care to include these children in the activities and often assigned an assistant to exclusively help them. Approximately 5 percent of our sample consists of these students, and this proportion is balanced across treatment status (p-value=0.528).

strata fixed effects. The estimated $\hat{\alpha}_1$ is the average treatment effect. We present all our results without covariates in Online Appendix B. Because we test multiple hypotheses using a wide range of outcomes, we also provide Romano-Wolf p-values for our main outcomes in Online Appendix A.

The program requires teachers to cover all weekly topics throughout the academic year. Even though participation was voluntary and the program was oversubscribed, compliance in terms of actual implementation, i.e., coverage of the curriculum, may not have been perfect. To assess this, we asked treated teachers to report their estimated degree of curriculum coverage at endline. Specifically, we asked them to mark their coverage estimate using an unmarked 10 centimeter line, which gives us a continuous measure of program implementation intensity, albeit subjective. Figure 6 depicts the distribution of the reported implementation intensity. Treated teachers report having covered about 60 percent of the program on average, with approximately 16 percent of teachers reporting no coverage at all. We were informed that low or no implementation is mainly due to teachers being involuntarily relocated by the MoE in the middle of the academic year. Such turnover is quite common in our study site. Given this imperfect compliance, the estimated $\hat{\alpha}_1$ should be interpreted as the average intent to treat effect (ITT). In what follows, we will focus on ITTs.

5.1 Treatment Effects on Peer Violence and Victimization

Table 3 presents the estimated effects of the program on the number of high-intensity violent episodes recorded in 10-day diary logs. Recall that the study sample covers, on average, three classrooms in each school. The first two columns use only the study sample. The first column presents the estimated treatment effect on the total number of episodes perpetrated by children from study classrooms within 10 school days. The second column shows the effect on the total number of episodes in which victimized children were from study classrooms. The third column presents the effect for the whole school, and finally, the last column presents the effect on non-study classrooms, therefore representing the program's spillover effect.

As can be seen in column 1, the program significantly reduced the number of violent events perpetrated by treated children. There are, on average, 1.88 events recorded in 10 days in control group classrooms. The treatment effect of 1.21 fewer events implies a substantial (about 64 percent) decline. The second column in the table shows that the program also significantly reduced the number of events that victimized treated children.

While the total number of victimizing events is 1.50 in control schools, it is about 50 percent lower (0.75 fewer events) in treatment schools. This result ensures that the program did not generate the undesired effects we mentioned earlier. That is, it did not make treated children more susceptible to victimization. Instead, the results suggest that the program, by keeping children away from conflict, lowered the risk of being a victim in a conflict.

Considering the sheer size of the schools, the program's effect on the entire school is striking. We estimate a substantial decline in the overall number of violent events in treated schools. As shown in column 3, the average number of violent episodes in ten days is 7.83 in control schools, and the program lowers this by 2.4 episodes. This effect is not precisely estimated but note the large effect size. Note also that almost half of this overall reduction is coming from non-treated classrooms (column 4). These results are suggestive of spillover effects of the program within schools.

The program appears to be highly effective in reducing high-intensity peer violence and victimization in schools. A natural question now is whether this is reflected in student and teacher reports of bullying and anti-social behavior. Recall that we asked students about their experiences of bullying at baseline and endline. We also asked teachers to rate each student's general behavioral conduct at endline using a 1 to 5 grading scale. For the former, we construct a binary outcome, which takes the value of 1 if the child reports being bullied by his/her peers, and zero otherwise. The latter is constructed as a standardized behavioral conduct score assigned by teachers with larger values indicating bad behavioral conduct. It is worth noting that, while we do expect the program to be effective in reducing violence and anti-social behavior, our conjecture for 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 increase awareness and 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, effect sizes are positive and imprecisely estimated.

5.2 Treatment Effects on Social Exclusion and Ethnic Segregation

We now investigate the program's effect on social networks in the classroom, in particular, social exclusion and ethnic segregation. Panel 1 in Table 5 presents the estimated (average)

marginal effects of the program on the probability of having a friend and having a classmate who provides emotional and academic support. Note that most children (95 percent) in the control group report having at least one friend in their classroom. The program has no effect on the probability of having a friend. However, treatment effects on emotional and academic support ties (columns 2 and 3) are positive and significant at the 1% level. Treated children are 4.6 percentage points (5.8%) more likely to have at least one classmate from which they receive emotional support and 5.6 percentage points (7.6%) more likely to have at least one classmate from which they receive academic support.

Panel 2 in Table 5 presents the estimated effects on the number of in-degree ties. These ties refer to the number of nominations a child receives in each category. Note first that an average child in the control group receives 2.35 friendship nominations, 1.76 emotional support nominations, and 1.51 academic support nominations. While the treatment had no effect on the number of friendship nominations received, it has significant effects on the number of nominations received in emotional and academic support categories. Treated children receive, on average, 0.10 (6%) more nominations than control children as emotional support providers for their classmates. Similarly, they receive about 0.13 (8.3%) more nominations as academic support providers. These results strongly suggest that the intervention increases the prevalence of support among classmates in personal and academic matters.

We then estimate the impact of the program on the level of ethnic segregation in the classroom. Recall that our ethnic segregation measure is constructed as the difference between the proportion of inter-ethnic ties expected to be formed at random and its observed counterpart at the classroom level. Panel 3 in Table 5 presents the estimated treatment effects on ethnic segregation based on friendship, emotional support, and academic support ties. We estimate a sizeable decline in ethnic segregation in classrooms. Our estimates amount to a 15% decline in segregation based on friendship ties and 17% (21%) reduction based on emotional (academic) support ties relative to control classrooms. The estimated effects on friendship ties and academic support ties are statistically significant at the 10% level. While the estimated effect sizes are similar, the effect based on emotional support is imprecisely estimated.

5.3 Treatment Effects on Prosocial Behavior: Trust, Reciprocity, Cooperation and Altruism

Columns 1 and 2 in Table 6 present the estimated treatment effects on trust, measured as the number of tokens sent (out of 4) to an anonymous receiver in the respondent's classroom (in-class) and to an anonymous receiver outside the respondent's school (out-school). Note first that about 1.38 and 1.45 tokens were sent in the control group to an anonymous in-class peer and out-school peer, respectively. This counter-intuitive difference is significant at the 10 percent level. This difference is entirely eliminated in the treatment group (p-value=0.95). The estimated treatment effects on trust are large and precise: Treated students sent about 0.27 extra tokens to an anonymous in-class peer and about 0.21 more tokens to an anonymous out-school peer, implying a 19% and 14% increase relative to the control group, respectively. While the effect size in the out-school case is smaller, the difference is not statistically significant (p-value=0.16).

Columns 3 and 4 present the estimated treatment effects on reciprocity toward an in-class peer and an out-school peer, respectively. Here, recall that reciprocity was elicited based on all four scenarios of receiving 1, 2, 3, and 4 tokens. We construct our dependent variable as the average of all four scenarios, that is, the average ratio of tokens sent back to the sender. Note first that about 39 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 percentage points more tokens to the anonymous sender in their classroom relative to the control. This corresponds to about 13 percent higher reciprocity toward in-class peers relative to the control group. The effect on out-school reciprocity is very similar, both in terms of size and precision.

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

Taken together, these results provide strong evidence that the program increased prosociality among children. Treated children exhibit higher trust and reciprocity toward their

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

Consider the in-class version of the trust game. 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 $m = 3$ is the experiment multiplier. Note that conditional on the sender's decision, the receiver's decision to reciprocate is strategic, and does not involve uncertainty. Given that student i has a 50 percent chance of being a sender or a receiver, her expected overall payoff P_{ic} is:

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

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

Panel 1 in Table 7 presents the estimated treatment effects on expected payoffs for in-class and out-school trust games, respectively. In addition to overall payoffs (Equation 2), we also present results for the role of a sender and receiver separately. Note first that in both games, children in the control group ended up gaining 5.38 and 5.46 tokens in in-class and out-school games, respectively (see columns 3 and 6). Given the lower bound is 4 tokens (no trust condition), this number indicates that children increased the size of the gift pie by exhibiting some trust. Not surprisingly, treated students ended up gaining on average 0.26 more tokens by trusting and reciprocating more in the in-class game (column 3). However, they ended up with 0.11 fewer tokens than control in the out-school game. Even though the latter result seems to indicate a disadvantage on the part of the treated children, a closer look at the results reveals an interesting detail for this outcome. We estimate a zero treatment effect in the case of being a sender in the out-school game (column 4). However, we estimate a large and statistically significant difference in the case of being a receiver: Treated receivers gave up about 0.22 more gifts to reciprocate to their out-school senders. Note that, contrary to the trust decision made by the sender, the reciprocity decision does not involve uncertainty. The fact that treated children deliberately lowered their payoffs to reciprocate out-school senders implies that the lower overall amount of payoffs they obtained (column 6) may be sub-optimal from the expected payoff perspective, but not necessarily from a welfare point of view if the underlying utility function incorporates other-regarding preferences. This result implies that the program increased the tendency to reciprocate the kindness children receive from out-school peers at the expense of their own payoff. We explore the cooperation decisions in a similar way. Panel 2 presents the related results. Treated children received 0.13 more tokens than children in the control group by exhibiting in-class cooperation. Not surprisingly, we estimate a zero payoff difference across treatment status for out-school 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 to the recipient's ethnicity. Here, in addition to estimating the program's impact on overall altruism, we want to assess whether the treatment affects the donation patterns based on the recipient's ethnicity. To do this, we estimate the following regression:

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

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

Table 8 presents the estimated treatment effects on the willingness to donate (average marginal effects) and the fraction of the endowment (4 tokens) donated. As the first column shows, treated children are 6.9 percentage points more likely to donate their endowment, and this value increases by another 3.4 percentage points if the anonymous recipient is revealed to be a Syrian refugee child. Similarly, treated children donate 5.3 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.

6 Heterogeneous 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 consider three heterogeneity domains: refugee status, gender, and emotional intelligence (Eyes Test score). We do not detect any significant heterogeneity concerning gender and emotional intelligence in most of our outcomes. We provide these results in Appendix C.

We do detect a notable treatment effect heterogeneity concerning refugee status in some important outcomes. Table 9 presents heterogeneous effects on self-reported bullying and teacher reports of behavioral conduct. Here, we see evidence of differential treatment effects concerning the probability of being bullied. While the program has no impact on host children's likelihood of being bullied, this outcome is significantly lower for refugee children

in the treatment group. We find no evidence of differential treatment effects concerning teachers' behavioral conduct grades.

Table 10 presents heterogeneous treatment effects on social exclusion. Panel 1 shows that the program is effective in mitigating the social exclusion of all children, hosts and refugees alike. The effect on refugee children is particularly strong with respect to the probability of having a friend (column 1). While the program does not affect the probability of having a friend for host children, it increases the probability of a refugee child befriending at least one classmate by about 5 percentage points. While the estimated signs for the interaction terms are all positive, suggesting treated refugees benefited from the program more than their host classmates, we do not estimate statistically significant heterogeneity concerning in-degree ties (Panel 2). Focusing only on ties formed with host children, Panel 3 of Table 10 largely confirms Panel 1 and 2 results. The program increases refugee children's likelihood of having a host friend by 7 percentage points, but this effect does not reach statistical significance. We estimate a statistically significant heterogeneous effect concerning emotional support ties with hosts (at 10% level), but not concerning academic support ties. Results in Panel 3 suggest that refugee children in treated schools are approximately 25% (21%) more likely than those in control schools to receive emotional (academic) support from their host classmates.

In terms of prosocial behavior, we detect heterogeneity only in the out-school trust and reciprocity. Table 11 shows that while we estimate a significant increase in out-school trust and reciprocity in host children, we estimate null effects for refugees. It appears that while the program strengthens trust and reciprocity among classmates, refugee children are still reluctant to trust (and reciprocate to) children they do not know. We do not estimate any statistically significant heterogeneous effect of the program with respect to cooperation (columns 5 and 6) and altruism (Table 12).

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 in the classroom and, iii) increasing trust, reciprocity, and cooperation among students, as well as their altruism toward one another. Moreover, these positive results on prosocial behaviors are not limited to behaviors toward classmates but extend to anonymous out-school peers. Even more promising is that the program seems to have lessened the social exclusion of refugee children significantly and helped them form friendship ties and receive emotional and academic support from their classmates. Given these results, perhaps it is not surprising that we estimate a striking improvement in refugee children's

test scores in the Turkish language. Table 13 presents the estimated treatment effects on math and Turkish test scores for both host and refugee children. While we estimate null effects on both math and Turkish scores for host children, we estimate a large and significant program effect on Turkish test scores for refugee children: The effect size is 0.13 standard deviations and significant at the 5% level. We do not estimate a significant treatment effect on math scores of refugee children.

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

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

While we hypothesize perspective-taking to be a likely mechanism, we acknowledge that there may be other channels. For example, the program may increase cohesion by changing classroom norms regarding acceptable and unacceptable behavior. It may also do so by invoking children's empathetic concern (compassion) toward others. Yet, as another mechanism, the program may reduce conflict and victimization by encouraging children to better manage their impulsivity, a characteristic that is often responsible for the escalation of disputes and can be controlled by engaging in deeper deliberation. Moreover, the program may improve cohesion by increasing tolerance toward individual differences, including ethnic and cultural differences, thereby reducing ethnic bias. Finally, we re-implemented the Eyes Test at endline to assess whether the program improved children's emotional intelligence. However, given the evidence on the high test-retest reliability of this test, we do not expect any improvement in children's scores over an academic year (e.g. see Fernández-Abascal et al., 2013, Vellante et al., 2013).

Figure 7 depicts the estimated treatment effects and 95 percent confidence intervals on behavioral norms, ethnic bias, perspective-taking, impulsivity, empathetic concern, and Eyes

Test scores, controlling for baseline values of each measure. Treated children report having 0.28 standard deviation higher perspective-taking relative to children in the control group, and this effect is precisely estimated. We also estimate a small decline in reported impulsivity, significant at the 10% level. We find a small positive effect on behavioral norms, which is also significant at the 10% level. We do not estimate any significant change in ethnic bias or the level of empathetic concern, although both estimated coefficients have intuitive signs. We note that our results on impulsivity and norms do not survive the multiple hypotheses correction (see Appendix Table A.1). Finally, we do not detect any improvement in Eyes Test scores.[11] In summary, the analysis suggests that increasing children's effort to take others' perspectives is an effective way to improve a wide range of cohesion outcomes in schools with ethnic segregation.

7 Concluding Remarks

We evaluate the effectiveness of a unique educational program that aims to build social cohesion in ethnically mixed schools. The program is implemented in schools where the ethnic composition has changed rapidly due to a fast influx of refugee children. The program involves covering a full year curriculum 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, ethnic segregation, and prosocial behaviors. We find that the program significantly lowers high-intensity peer violence and victimization on school grounds. It also reduces social exclusion and lowers ethnic segregation in the classroom. We also find that treated children exhibit higher trust and reciprocity toward their peers, cooperate more, and show higher altruistic tendencies. Prosocial behavior improves not only among classmates, but toward anonymous out-school peers as well. Finally, we show that the program leads to a large improvement in the refugee children's ability in the host country's language.

The results of the study are promising in terms of their external validity. While the

[11] While we do not estimate a significant effect on Eyes Test scores for host students, we estimate a 0.10 standard deviation improvement (significant at 10% level) for refugee students. We believe that this improvement is due to treated refugee students' progress in the Turkish language as this test was implemented in Turkish and requires a good command of Turkish vocabulary; see Olderbak et al. (2015) for a discussion on vocabulary and Eyes Test.

participation was voluntary, in practice, the program was oversubscribed. Most teachers were eager to join the program in all participating schools, compelling us to randomly choose among teachers. Therefore, we are reasonably confident that the program would show a similar success at scale in Turkey.

Our final remark relates to the cost-effectiveness of the program. The program is remarkably cost-effective relative to well-known educational interventions. It is difficult to gauge the individual and societal value of reducing violence, social exclusion and ethnic segregation in schools. However, even if one considers only the learning gains of refugee children with respect to the host country's language, the program can be viewed as a success. The education materials were developed as part of a private university's philanthropic endeavor, and as such, they are now a public good. The remaining program costs pertain to printing hard copy materials (the UEO book and activity kits, which are also available online), distributing the materials to schools, and conducting teacher training. For approximately 6,500 children, the printing costs were about 20,000 USD, the distribution cost 7,000 USD, and teacher training costs were about 8,000 USD. These values imply a 5.4 USD program cost per child, which is negligible compared to the cost of any known large-scale educational intervention.

References

Adida, C. L., Lob, A., and Platas, M. R. (2018). Perspective taking can promote short-term inclusionary behavior toward Syrian refugees. *PNAS*, 115(38):9521–9526.

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 a 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.

Bandiera, O., Mohnen, M., Rasul, I., and Viarengo, M. (2019). Nation-building through compulsory schooling during the age of mass migration. *The Economic Journal*, 129(617):62–109.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2):241–251.

Batson, C. 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.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2019). Unity in diversity? How intergroup contact can foster nation building. *American Economic Review*, 109(11):3978–4025.

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

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.

Brandts, J. and Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398.

Broockman, D. and Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282):220–224.

Burns, J., Corno, L., and La Ferrara, E. (2019). Interaction, prejudice and performance: Evidence from South Africa. IFS Working Papers W19/03.

Cappelen, A., List, J., Samek, A., and Tungodden, B. (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy*, 128(7):000–000.

Card, D. and Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics*, 91(11-12):2158–2184.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, 1(3):111–34.

Deming, D. J. (2011). Better schools, less crime? *The Quarterly Journal of Economics*, 126(4):2063–2115.

Durkheim, E. (1897). *Le suicide: étude de sociologie*. Alcan.

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.

Eisner, M., Ribeaud, D., Sorrenti, G., and Zölitz, U. (2020). The causal impact of socio-emotional skills training on educational success. *CEPR Discussion Paper No. DP14523*.

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.

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P., and Baron-Cohen, S. (2013). Test-retest reliability of the 'reading the mind in the eyes' test: A one-year follow-up study. *Molecular Autism*, 4(1):33.

Fryer Jr, R. G. and Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2):447–464.

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

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.

Gould, E. D., Lavy, V., and Paserman, M. D. (2004). Immigrating to opportunity: Estimating the effect of school quality using a natural experiment on Ethiopians in Israel. *The Quarterly Journal of Economics*, 119(2):489–526.

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

Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review*, 94(4):919–943.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2009). New evidence about brown v. board of education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383.

Harbaugh, W. T., Krause, K., and Vesterlund, L. (2003). Trust in children. *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, pages 302–322.

Heckman, J. 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.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3):1209–1318.

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

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., and Falk, A. (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy*, 128(2):434–467.

Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, 94(1):155–189.

Lowe, M. (2020). Types of contact: A field experiment on collaborative and adversarial caste integration. CESifo Working Paper No. 8089.

Miguel, E. and Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, 89(11-12):2325–2368.

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science*, 369(6505):866–870.

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., and Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, 6:1503.

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

Putnam, R. (1993). The prosperous community: Social capital and public life. *The American Prospect*, 13:35–42.

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

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

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. *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*, 10:126.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., and Preti, A. (2013). The "reading the mind in the eyes" test: systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, 18(4):326–354.

Voigtlaender, N., Becker, S., Grosfeld, I., Grosjean, P., and Zhuravskaya, E. (2020). Forced migration and human capital: Evidence from post-WWII population transfers. *American Economic Review*.

8 Tables

Table 1: Balance at Baseline

Panel 1: Student Characteristics

	Control Mean	Difference
Student Demographics:		
Male	0.51	-0.00
Age in Months	105.63	0.58
Refugee	0.16	-0.01
Working Mother	0.27	0.02
Working Father	0.86	-0.01
Cognitive Tests:		
Raven Score	0.00	-0.07
Eyes Test Score	0.00	-0.02
Math Score	0.00	0.01
Turkish Score	0.00	-0.01
Cohesion Indicators:		
Proportion Bullied by Peers in Classroom	0.84	0.03*
Proportion Bullied by Peers in School	0.80	0.01
Fraction Donated	-0.00	-0.02
Willingness to Donate	0.65	-0.03
Proportion Cooperate	0.54	0.03
Perspective Taking	-0.00	0.01
Empathetic Concern	0.00	-0.01
Ethnic Bias	-0.00	-0.03
Impulsivity	-0.00	0.04
Having a Friend	0.92	0.00
Having Emotional Support	0.66	-0.02
Having Academic Support	0.56	-0.01
Friendship Ties (in-degree)	1.75	0.01
Emotional Support Ties (in-degree)	1.05	-0.05
Academic Support Ties (in-degree)	0.84	-0.02

Panel 2: Classroom and Teacher Characteristics

	Control Mean	Difference
Classroom Size	33.04	1.24
Refugee Share	0.16	-0.00
Ethnic Segregation in Friendship Ties	0.09	0.01
Ethnic Segregation in Emotional Support Ties	0.08	0.01
Ethnic Segregation in Academic Support Ties	0.08	0.00
Teacher Age in Years	34.79	-0.19
Male Teacher	0.43	-0.01
Teacher Years of Experience	10.46	0.05
Tenured Teacher	0.88	-0.01
Teacher Raven Score	0.00	0.02
Teacher Eyes Test Score	-0.00	0.06

Panel 3: School Characteristics

	Control Mean	Difference
School Size (3rd and 4th grades only)	483.30	3.30
Total Number of 3rd and 4th-grade Classrooms	14.65	-0.39

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. Note that values at the school level refer only to 3rd and 4th grades. As Turkish primary education covers grades 1 to 4, the average full school size is approximately twice the values presented in this panel. All cognitive tests, donation, perspective-taking, empathic concern, ethnic bias, and impulsivity measures are standardized to have mean zero for the control group. Asterisks indicate that the difference is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 2: Baseline Conditions for Refugee Children

	Reported Bullying	Friendship	Emotional Support	Academic Support
Refugee (=1)	0.042**	-0.063***	-0.121***	-0.104***
	(0.02)	(0.01)	(0.02)	(0.02)
Refugee Mean	0.88	0.84	0.50	0.43
Observations	5638	6135	6135	6135

Reported estimates are marginal effects obtained from logit regressions. Binary dependent variables are column 1: reported being regularly bullied by peers, column 2: reported to have a friend in the classroom, column 3: reported to have a supporting classmate (emotional), column 4: reported to have a supporting classmate (academic). All regressions control for class level refugee share, class size, school size, district dummies, and randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 3: Treatment Effects of Peer Violence and Victimization-Diary Records

	Perpetrated	Victimized	Total Events	Spillover
Treatment	-1.212***	-0.749**	-2.401	-1.189
	(0.44)	(0.37)	(1.95)	(1.80)
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. 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 total number of violent events that took place in 10 days, perpetrated by a student from study classes; in column 2, the total 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 school size, number of participating classrooms in a given school, school level refugee share, district and province fixed effects and randomization strata. Robust standard errors are in parentheses. Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 4: Treatment Effects on Student and Teacher Reports of Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Behavioral Conduct
Treatment	0.013	0.070
	(0.02)	(0.06)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Mean	0.79	-0.02
Observations	6335	6034

Reported estimates are obtained from ordinary least squares (OLS) regressions. In column 1, the dependent variable is a dummy, which equals 1 if the student reports physical and verbal bullying. In column 2, the dependent variable is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. All regressions control for randomization strata and baseline covariates. Baseline covariates include relevant baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

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

Panel 1: Social Exclusion-Binary

	Friendship	Emotional Support	Academic Support
Treatment	0.003	0.046***	0.056***
	(0.00)	(0.01)	(0.01)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Mean	0.95	0.80	0.73
Observations	6642	6642	6642

Panel 2: Social Exclusion-In-degree Ties

	Friendship	Emotional Support	Academic Support
Treatment	-0.055	0.104**	0.125**
	(0.04)	(0.05)	(0.05)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Mean	2.35	1.76	1.51
Observations	6642	6642	6642

Panel 3: Ethnic Segregation

	Friendship	Emotional Support	Academic Support
Treatment	-0.016*	-0.015	-0.021*
	(0.01)	(0.01)	(0.01)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Mean	0.11	0.09	0.10
Observations	218	218	218

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

Table 6: Treatment Effects on Prosocial Behavior: Trust, Reciprocity and Cooperation

	Trust		Reciprocity		Cooperation	
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School
Treatment	0.265***	0.207***	0.050***	0.048***	0.042***	-0.002
	(0.06)	(0.05)	(0.01)	(0.01)	(0.02)	(0.01)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
p-val (In-Class=Out-School)	0.16		0.62		0.01	
Control Mean	1.38	1.45	0.39	0.39	0.52	0.50
Observations	6476	6512	6534	6523	6568	6573

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

Table 7: Treatment Effects on Expected Payoffs

Panel 1: Payoffs from Trust Game

	In-Class Payoffs			Out-School Payoffs		
	Sender	Receiver	Overall	Sender	Receiver	Overall
Treatment	0.288***	0.239***	0.260***	-0.002	-0.219***	-0.112**
	(0.06)	(0.07)	(0.06)	(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	6476	6412	6361	6512	6381	6372

Panel 2: Payoffs from Cooperation Game

	In-Class Payoffs	Out-School Payoffs
Treatment	0.132***	-0.008
	(0.04)	(0.04)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Mean	4.57	4.50
Observations	6568	6573

Reported estimates are obtained from ordinary least squares (OLS) regressions. In Panel 1, the dependent variables are expected payoffs from the trust game. In Panel 2, the dependent variables are expected payoffs from the cooperation game. Regressions control for randomization strata and baseline covariates. Baseline covariates include Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size, and district dummies. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 8: Treatment Effects on Altruism

	Willingness to Donate	Fraction Donated
Treatment	0.069***	0.053***
	(0.02)	(0.02)
Ethnic Reference	-0.016	0.006
	(0.02)	(0.01)
Treatment*Ethnic Reference	0.034*	0.012
	(0.02)	(0.02)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Mean	0.70	0.34
Observations	6577	6577

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

Table 9: Heterogeneous Treatment Effects on Student and Teacher Reports of Violence and Antisocial Behavior

	Student Reported Bullying	Teacher Reported Behavioral Conduct
Treatment	0.023	0.068
	(0.02)	(0.05)
Refugee	0.069***	-0.034
	(0.03)	(0.06)
Treatment*Refugee	-0.062**	0.006
	(0.03)	(0.09)
Strata FE	✓	✓
Baseline Covariates	✓	✓
Control Refugee Mean	0.85	0.01
Control Host Mean	0.77	-0.03
Observations	6335	6029

Reported estimates are obtained from ordinary least squares (OLS) regressions. In column 1, the dependent variable is a dummy, which equals 1 if the student reports physical and verbal bullying. In column 2, the dependent variable is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. All regressions control for randomization strata and baseline covariates. Baseline covariates include relevant baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 10: Heterogeneous Treatment Effects on Social Exclusion

Panel 1: Social Exclusion-Binary

	Friendship	Emotional Support	Academic Support
Treatment	-0.005	0.038***	0.056***
	(0.00)	(0.01)	(0.01)
Refugee	-0.080***	-0.131***	-0.097***
	(0.02)	(0.03)	(0.02)
Treatment*Refugee	0.046*	0.050	0.004
	(0.02)	(0.04)	(0.03)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Refugee Mean	0.87	0.67	0.65
Control Host Mean	0.97	0.82	0.75
Observations	6642	6642	6642

Panel 2: Social Exclusion-In-degree Ties

	Friendship	Emotional Support	Academic Support
Treatment	-0.064	0.095*	0.123**
	(0.05)	(0.05)	(0.06)
Refugee	-0.310***	-0.474***	-0.277***
	(0.09)	(0.09)	(0.07)
Treatment*Refugee	0.053	0.053	0.010
	(0.12)	(0.11)	(0.09)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Refugee Mean	1.72	1.16	1.01
Control Host Mean	2.49	1.89	1.63
Observations	6642	6642	6642

Panel 3: Social Ties with Host Children-Binary

	Host Friendship	Host Emotional Support	Host Academic Support
Treatment	-0.006	0.038**	0.055***
	(0.01)	(0.01)	(0.02)
Refugee	-0.313***	-0.299***	-0.244***
	(0.04)	(0.03)	(0.03)
Treatment*Refugee	0.071	0.080*	0.043
	(0.06)	(0.05)	(0.04)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Refugee Mean	0.62	0.47	0.47
Control Host Mean	0.96	0.81	0.73
Observations	6642	6642	6642

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

Table 11: Heterogeneous Treatment Effects on Prosocial Behavior: Trust, Reciprocity and Cooperation

	Trust		Reciprocity		Cooperation	
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School
Treatment	0.278***	0.245***	0.053***	0.054***	0.043**	0.004
	(0.06)	(0.06)	(0.01)	(0.01)	(0.02)	(0.01)
Refugee	0.578***	0.610***	0.110***	0.134***	0.006	0.022
	(0.08)	(0.07)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment*Refugee	-0.079	-0.230**	-0.017	-0.037*	-0.010	-0.034
	(0.11)	(0.10)	(0.02)	(0.02)	(0.03)	(0.03)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
Control Refugee Mean	1.89	1.98	0.48	0.51	0.54	0.53
Contol Host Mean	1.28	1.34	0.37	0.37	0.52	0.49
Observations	6476	6512	6534	6523	6568	6573

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

Table 12: Heterogeneous Treatment Effects on Altruism

	Willingness to Donate		Fraction Donated	
	Host	Refugee	Host	Refugee
Treatment	0.076***	0.043	0.058***	0.030
	(0.02)	(0.04)	(0.02)	(0.03)
Ethnic Reference	-0.024	0.022	0.005	0.009
	(0.02)	(0.03)	(0.01)	(0.02)
Treatment*Ethnic Reference	0.037	0.015	0.015	0.004
	(0.02)	(0.05)	(0.02)	(0.04)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
p-value (Host=Refugee)	0.42		0.42	
Control Mean	0.68	0.81	0.31	0.46
Observations	5513	1064	5513	1064

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

Table 13: Heterogeneous Treatment Effects on Achievement Tests

	Turkish		Math	
	Host	Refugee	Host	Refugee
Treatment	0.008	0.130**	-0.010	-0.019
	(0.04)	(0.05)	(0.06)	(0.08)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
Observations	5565	1084	5565	1084

Reported estimates are obtained from ordinary least squares (OLS) regressions. Dependent variables are standardized test scores from Turkish language and math tests. Regressions control for randomization strata and baseline covariates. Baseline covariates include relevant baseline outcomes, Raven's score, Eye Test score, gender, refugee status, age in months, a dummy variable for students who are developmentally challenged, class size, school size and district dummies. 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.

9 Figures

Figure 1: Evaluation Timeline

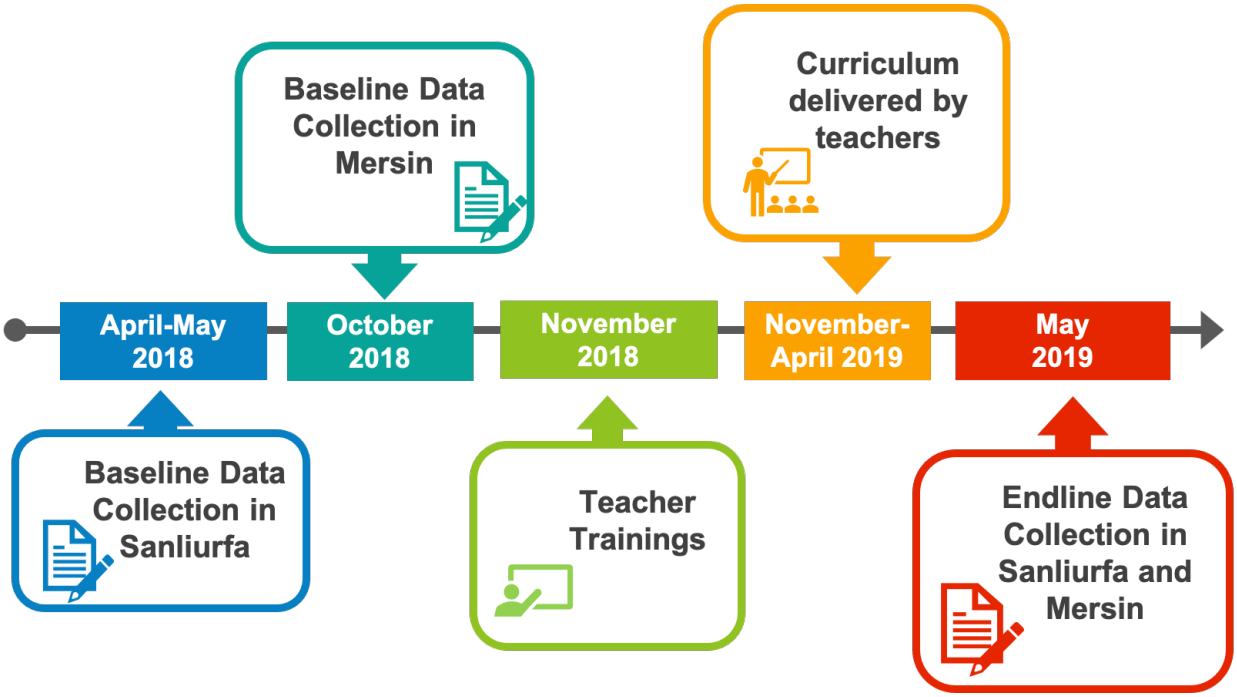
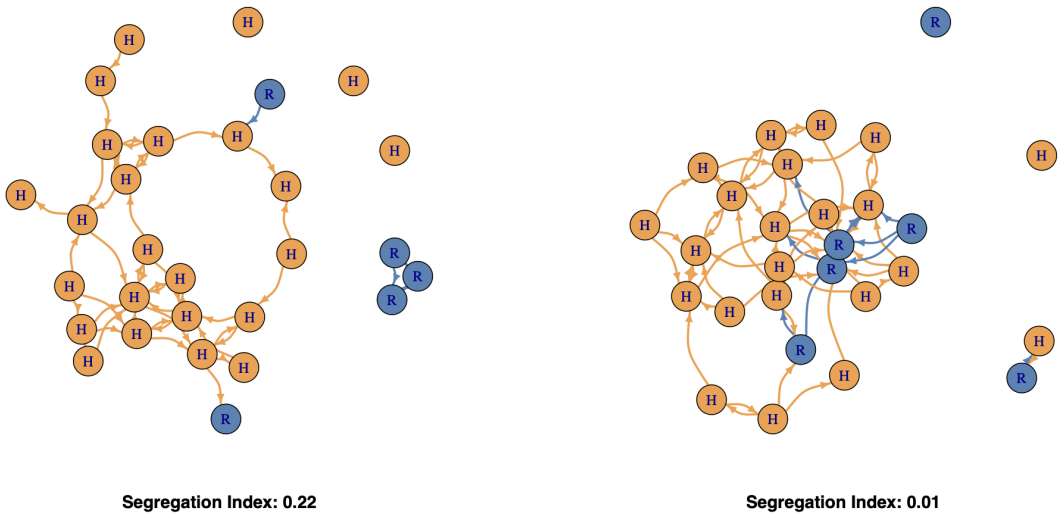


Figure 2: An Example Diary Log

School: XYZ Primary School		**Classrooms:** 3A, 3E, 4G	
Disciplinary events will be recorded daily by the designated school administrator			
	The number of high-intensity disciplinary events	Number of events perpetrated by someone in highlighted classrooms	Number of events in which someone from the highlighted classrooms was victimized
May 6, 2019			
May 7, 2019			
May 8, 2019			
May 9, 2019			
May 10, 2019			
May 11, 2019	WEEKEND	WEEKEND	WEEKEND
May 12, 2019	WEEKEND	WEEKEND	WEEKEND
May 13, 2019			
May 14, 2019			
May 15, 2019			
May 16, 2019			
May 17, 2019			

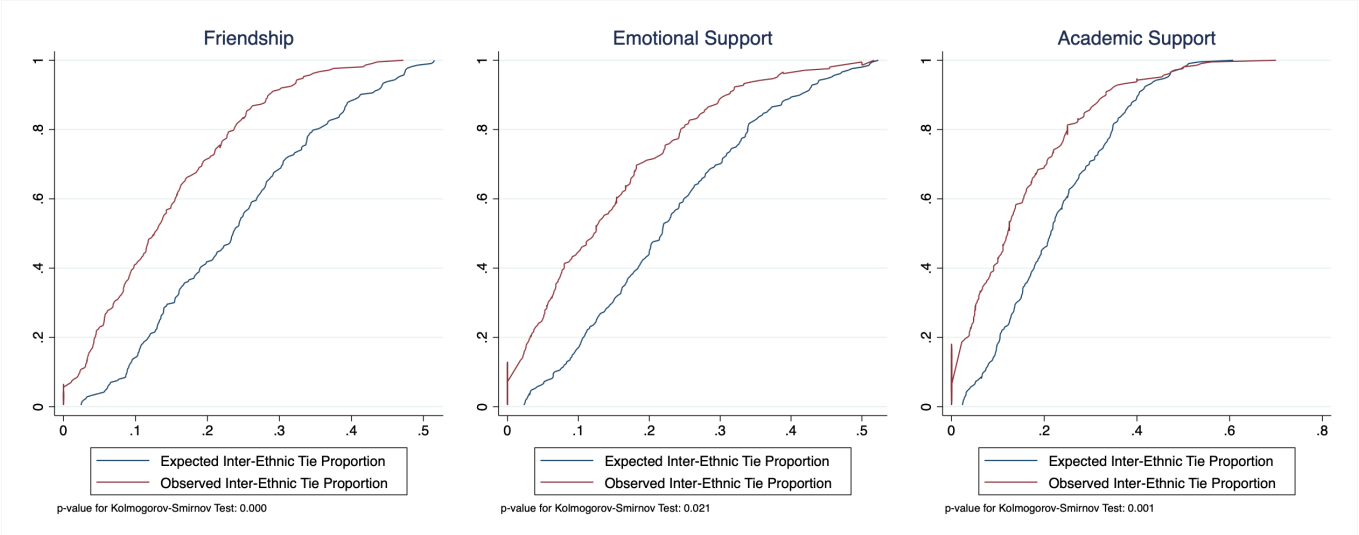
Figure 3: Ethnic Segregation: An Illustration of Two Classrooms

(a) Classroom 1 **(b)** Classroom 2



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

Figure 4: Cumulative Distribution of Expected and Observed Inter-Ethnic Ties at Baseline



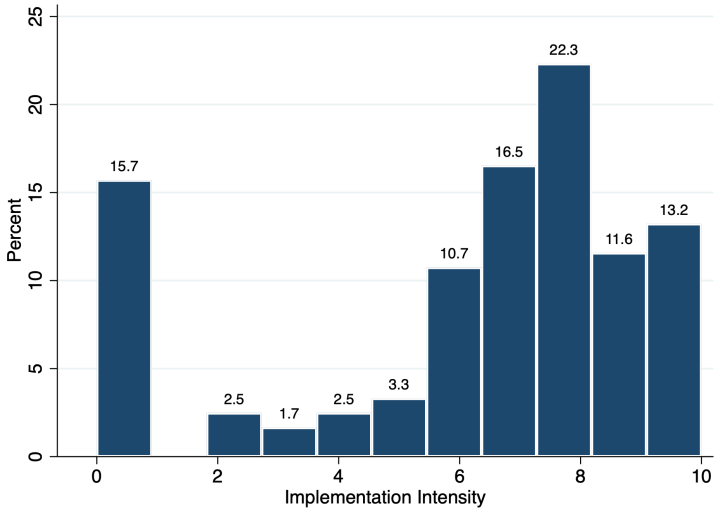
Notes: Each panel depicts the cumulative distribution of the expected proportion of inter-ethnic ties, calculated via probabilities derived from the Hypogeometrical distribution, and the observed proportion of inter-ethnic ties in classrooms for each category. P-values for the Kolmogorov-Smirnov test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

Figure 5: Cooperation Game Payoff Scheme

	ORANGE	GREEN
ORANGE	3 / 3	0 / 9
GREEN	9 / 0	6 / 6

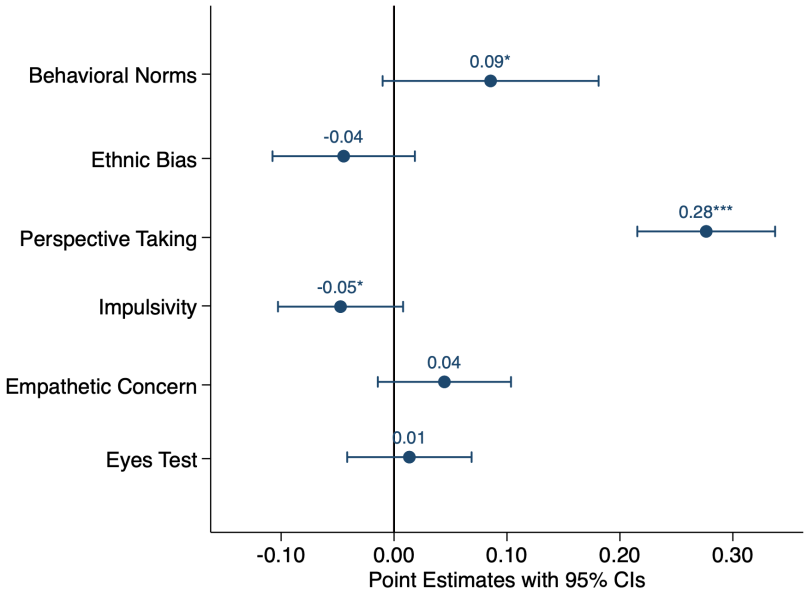
Notes: Each child is endowed with three gift tokens for this game and has an anonymous pair. The game involves the simultaneous decision to choose a card that is either green or orange. A child's payoff depends on both the color she chooses, and that of her pair chooses. If both choose orange, each remains with their initial endowments. If they choose different colors, the one that chooses orange triples her tokens, and her pair loses all. If both choose green, both double their endowments. The cooperative action is to choose the green card.

Figure 6: Teacher-Reported Program Implementation Intensity



Notes: 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 teacher's mark.

Figure 7: Treatment Effect on Potential Mechanisms



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

Appendix-For Online Publication

A Correction for Multiple Hypothesis Testing

Table A.1: Original and Romano Wolf P-Values

Panel 1: Experimental Outcomes

	Original	Romano Wolf
In-Class Trust	0.000	0.004
Out-School Trust	0.000	0.004
In-Class Reciprocity	0.000	0.004
Out-School Reciprocity	0.001	0.004
In-Class Cooperation	0.008	0.006
Out-School Cooperation	0.868	0.892
Fraction Donated	0.000	0.004
Willingess to Donate	0.000	0.004

Panel 2: Network Outcomes

	Original	Romano Wolf
Having A Friend	0.526	0.503
Having Emotional Support	0.000	0.002
Having Academic Support	0.000	0.002
Friendship Ties (in-degree)	0.207	0.503
Emotional Support Ties (in-degree)	0.028	0.261
Academic Support Ties (in-degree)	0.023	0.048

Panel 3: Survey Outcomes

	Original	Romano Wolf
Student Reported Bullying	0.405	0.593
Teacher Reported Behavioral Conduct	0.225	0.447
Behavioral Norms	0.079	0.331
Ethnic Bias	0.163	0.593
Perspective Taking	0.000	0.002
Impulsivity	0.093	0.593
Empathetic Concern	0.136	0.593

The table provides p-values corrected for multiple hypotheses testing using Romano-Wolf algorithm. The number of replications is set to 500.

B Main Results without Covariates

Table B.1: Treatment Effects of Peer Violence and Victimization-Diary Records

	Perpetrated	Victimized	Total Events	Spillover
Treatment	-1.019*	-0.757*	-3.265*	-2.247
	(0.53)	(0.44)	(1.79)	(1.61)
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. 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 total number of violent events that took place in 10 days, perpetrated by a student from study classes; in column 2, the total 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 randomization strata. Robust standard errors are in parentheses. Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B.2: Treatment Effect on Student and Teacher Reports of Violence and Antisocial Behavior

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

Reported estimates are obtained from ordinary least squares (OLS) regressions. In column 1, the dependent variable is a dummy, which equals 1 if the student reports physical and verbal bullying. In column 2, the dependent variable is standardized behavior scores based on the teacher's evaluation of each student. Higher values refer to more violent and antisocial behavior. All regressions control for randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B.3: Treatment Effects on Social exclusion and Ethnic Segregation in the Classroom

Panel 1: All Social Exclusion-Binary

	Friendship	Emotional Support	Academic Support
Treatment	0.010	0.053***	0.059***
	(0.01)	(0.02)	(0.02)
Strata FE	✓	✓	✓
Baseline Covariates	✗	✗	✗
Control Mean	0.95	0.80	0.73
Observations	6643	6643	6643

Panel 2: Social Exclusion-In-degree Ties

	Friendship	Emotional Support	Academic Support
Treatment	-0.031	0.078	0.111**
	(0.04)	(0.06)	(0.05)
Strata FE	✓	✓	✓
Baseline Covariates	✗	✗	✗
Control Mean	2.35	1.76	1.51
Observations	6643	6643	6643

Panel 3: Ethnic Segregation

	Friendship	Emotional Support	Academic Support
Treatment	-0.009	-0.006	-0.017
	(0.02)	(0.01)	(0.02)
Strata FE	✓	✓	✓
Baseline Covariates	✗	✗	✗
Control Mean	0.11	0.09	0.10
Observations	218	218	218

Reported estimates are obtained from from ordinary least squares (OLS) regressions. In Panel 1, binary dependent variables are column 1: reported to have a friend, column 2: reported to have a classmate providing emotional support, column 3: reported to have a classmate providing academic support. In Panel 2, the dependent variables are the number of in-degree ties. In Panel 3, the dependent variables are class-level segregation scores. Regressions control for randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B.4: Treatment Effects on Prosocial Behavior: Trust, Reciprocity and Cooperation

	Trust		Reciprocity		Cooperation	
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School
Treatment	0.272***	0.207***	0.049***	0.046***	0.044***	-0.003
	(0.07)	(0.06)	(0.01)	(0.02)	(0.02)	(0.01)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✗	✗	✗	✗	✗	✗
p-val (In-Class=Out-School)	0.14		0.44		0.01	
Control Mean	1.38	1.45	0.39	0.39	0.52	0.50
Observations	6476	6512	6534	6523	6568	6573

Reported estimates are obtained from ordinary least squares (OLS) regressions. The dependent variables in column 1 and column 2 are the number of tokens (out of 4) sent to an anonymous classmate and an anonymous out-of-school peer, respectively. In columns 3 and 4, the dependent variables are the average fraction of tokens sent back to an anonymous classmate and an anonymous out-of-school peer, respectively. The dependent variables in columns 5 and 6 are the binary dependent variables, which take the value 1 if the child chooses to cooperate (green card), and zero otherwise, with an anonymous classmate, and an anonymous out-of-school peer, respectively. All regressions control for randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B.5: Treatment Effects on Expected Payoffs

Panel 1: Payoffs from Trust Game

	In-Class Payoffs			Out-School 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

Panel 2: Payoffs from Cooperation Game

	In-Class Payoffs	Out-School Payoffs
Treatment	0.131**	-0.011
	(0.05)	(0.05)
Strata FE	✓	✓
Baseline Covariates	✗	✗
Control Mean	4.57	4.50
Observations	6568	6573

Reported estimates are obtained from ordinary least squares (OLS) regressions. In Panel 1, the dependent variables are expected payoffs from the trust game. In Panel 2, the dependent variables are expected payoffs from the cooperation game. Regressions control for randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B.6: 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	0.34
Observations	6577	6577

Reported estimates are obtained from ordinary least squares (OLS) regressions. The dependent variable in column 1 is the binary variable that takes the value 1 if the child donates some of her tokens, zero otherwise. The dependent variable in column 2 is the fraction of endowment (4 tokens) donated. Regressions control for randomization strata. Standard errors are in parentheses and clustered at the school level (unit of randomization). Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

C Heterogeneous Treatment Effects: Gender and Emotional Intelligence (Eyes Tests score)

C.1 Gender

Table C.1.1: Heterogeneous Treatment Effects on Self and Teacher Reported Antisocial Behavior

	Student Reported Bullying	Teacher Reported Behavioral Conduct
Treatment	0.003	0.028
	(0.02)	(0.04)
Male	0.023	0.679***
	(0.02)	(0.05)
Treatment*Male	0.020	0.084
	(0.02)	(0.07)
Strata FE	✓	✓
Baseline Controls	✓	✓
Control Male Mean	0.80	0.31
Control Female Mean	0.77	-0.37
Observations	6335	6034

Table C.1.2: Heterogeneous Treatment Effects on Social Exclusion

Panel 1: Social Exclusion-Binary

	Friendship	Emotional Support	Academic Support
Treatment	0.004	0.023	0.052***
	(0.01)	(0.01)	(0.02)
Male	-0.012*	-0.111***	-0.088***
	(0.01)	(0.01)	(0.02)
Treatment*Male	-0.001	0.047**	0.009
	(0.01)	(0.02)	(0.02)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Male Mean	0.94	0.74	0.69
Control Female Mean	0.96	0.85	0.77
Observations	6642	6642	6642

Panel 2: Social Exclusion-In-degree Ties

	Friendship	Emotional Support	Academic Support
Treatment	-0.085	0.125**	0.136*
	(0.06)	(0.06)	(0.07)
Male	-0.085*	-0.316***	-0.317***
	(0.05)	(0.05)	(0.06)
Treatment*Male	0.059	-0.040	-0.021
	(0.07)	(0.07)	(0.08)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Male Mean	2.28	1.53	1.30
Control Female Mean	2.43	1.99	1.74
Observations	6642	6642	6642

Panel 3: Social Ties with Host Children-Binary

	Host Friendship	Host Emotional Support	Host Academic Support
Treatment	0.004	0.034*	0.063***
	(0.01)	(0.02)	(0.02)
Male	-0.010	-0.092***	-0.079***
	(0.01)	(0.01)	(0.01)
Treatment*Male	0.004	0.034	-0.002
	(0.01)	(0.02)	(0.02)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control Male Mean	0.90	0.70	0.65
Control Female Mean	0.90	0.79	0.72
Observations	6642	6642	6642

Table C.1.3: Heterogeneous Treatment Effects on Prosocial Behavior: Trust, Reciprocity and Cooperation

	Trust		Reciprocity		Cooperation	
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School
Treatment	0.255***	0.179***	0.048***	0.042***	0.018	0.000
	(0.07)	(0.05)	(0.01)	(0.01)	(0.02)	(0.02)
Male	-0.018	0.001	-0.001	-0.005	-0.042**	0.009
	(0.05)	(0.04)	(0.01)	(0.01)	(0.02)	(0.02)
Treatment*Male	0.020	0.056	0.004	0.012	0.047*	-0.003
	(0.07)	(0.06)	(0.01)	(0.01)	(0.03)	(0.03)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
Control Male Mean	1.38	1.46	0.39	0.39	0.50	0.51
Contol Female Mean	1.39	1.45	0.39	0.40	0.54	0.49
Observations	6476	6512	6534	6523	6568	6573

Table C.1.4: Heterogeneous Treatment Effects on Altruism

	Willingness to Donate		Fraction Donated	
	Female	Male	Female	Male
Treatment	0.079***	0.062**	0.061***	0.041**
	(0.03)	(0.03)	(0.02)	(0.02)
Ethnic Reference	-0.002	-0.027	0.021	-0.011
	(0.02)	(0.02)	(0.02)	(0.01)
Treatment*Ethnic Reference	0.018	0.045	-0.001	0.024
	(0.03)	(0.03)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
p-value (Female=Male)	0.37		0.37	
Control Mean	0.72	0.68	0.35	0.33
Observations	3260	3317	3260	3317

C.2 Emotional Intelligence (EI)-Eyes Test Scores

Below analysis define Low EI as below median scores in Eyes Test, and High EI as median and above.

Table C.2.1: Heterogeneous Treatment Effects on Self and Teacher Reported Antisocial Behavior

	Student Reported Bullying	Teacher Reported Behavioral Conduct
Treatment	0.011	0.055
	(0.02)	(0.06)
High EI	-0.042*	-0.048
	(0.02)	(0.04)
Treatment*High EI	0.008	0.047
	(0.02)	(0.06)
Strata FE	✓	✓
Baseline Controls	✓	✓
Control High EI Mean	0.72	-0.13
Control Low EI Mean	0.82	0.03
Observations	6335	6034

Table C.2.2: Heterogeneous Treatment Effects on Social Exclusion

Panel 1: Social Exclusion-Binary

	Friendship	Emotional Support	Academic Support
Treatment	0.006	0.057***	0.066***
	(0.01)	(0.02)	(0.02)
High EI	0.024***	0.059***	0.037**
	(0.01)	(0.02)	(0.02)
Treatment*High EI	-0.009	-0.032	-0.030
	(0.01)	(0.03)	(0.02)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control High EI Mean	0.99	0.88	0.77
Control Low EI Mean	0.93	0.76	0.71
Observations	6642	6642	6642

Panel 2: Social Exclusion-In-degree Ties

	Friendship	Emotional Support	Academic Support
Treatment	-0.000	0.117**	0.100
	(0.06)	(0.06)	(0.07)
High EI	0.391***	0.472***	0.286***
	(0.08)	(0.08)	(0.07)
Treatment*High EI	-0.183	-0.067	0.051
	(0.11)	(0.10)	(0.10)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control High EI Mean	3.12	2.44	2.11
Control Low EI Mean	2.00	1.44	1.24
Observations	6642	6642	6642

Panel 3: Social Ties with Host Children-Binary

	Host Friendship	Host Emotional Support	Host Academic Support
Treatment	0.010	0.065***	0.077***
	(0.01)	(0.02)	(0.02)
High EI	0.049***	0.076***	0.059***
	(0.01)	(0.02)	(0.02)
Treatment*High EI	-0.013	-0.042	-0.044**
	(0.02)	(0.03)	(0.02)
Strata FE	✓	✓	✓
Baseline Covariates	✓	✓	✓
Control High EI Mean	0.97	0.85	0.74
Control Low EI Mean	0.86	0.69	0.65
Observations	6642	6642	6642

Table C.2.3: Heterogeneous Treatment Effects on Prosocial Behavior: Trust, Reciprocity and Cooperation

	Trust		Reciprocity		Cooperation	
	In-Class	Out-School	In-Class	Out-School	In-Class	Out-School
Treatment	0.270***	0.209***	0.044***	0.034**	0.055***	-0.005
	(0.06)	(0.06)	(0.01)	(0.02)	(0.02)	(0.01)
High EI	-0.215***	-0.180***	-0.053***	-0.059***	0.008	-0.030
	(0.06)	(0.06)	(0.01)	(0.01)	(0.02)	(0.02)
Treatment*High EI	-0.005	0.004	0.019	0.023	-0.037	0.013
	(0.07)	(0.07)	(0.01)	(0.02)	(0.03)	(0.02)
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓
Control High EI Mean	1.38	1.46	0.39	0.39	0.50	0.51
Contol Low EI Mean	1.39	1.45	0.39	0.40	0.54	0.49
Observations	6476	6512	6534	6523	6568	6573

Table C.2.4: Heterogeneous Treatment Effects on Altruism

	Willingness to Donate		Fraction Donated	
	Low EI	High EI	Low EI	High EI
Treatment	0.061**	0.086***	0.049**	0.064***
	(0.03)	(0.03)	(0.02)	(0.02)
Ethnic Reference	-0.032	0.015	-0.005	0.027**
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*Ethnic Reference	0.045*	0.011	0.020	-0.005
	(0.02)	(0.03)	(0.02)	(0.02)
Strata FE	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓
p-value (Female=Male)	0.46		0.48	
Control Mean	0.71	0.67	0.36	0.29
Observations	4388	2189	4388	2189

D Instructions for Incentivized Games

We will play some fun games with you today. 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 at the end of our visit.

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.

D.1 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]

D.2 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]

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.

Table D.2.1: Payment Scheme

	ORANGE	**GREEN**
ORANGE	3 3	0 9
GREEN	9 0	6 6

[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]

D.3 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.

E Survey Instruments

4-point likert scale: completely agree, agree, disagree, completely disagree	
Instrument	**Items**
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 for 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.

4-point likert scale: completely agree, agree, disagree, completely disagree	
Empathetic Concern	When I see someone being treated unfairly, I feel sorry for them.
	I often have tender 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 do not want to be friends with children who come from another country.
	Children who come from other countries are not as smart as us.
	I like children who come from other countries as much as I like friends from here.
	Children who come from other countries are more aggressive.
Behavioral 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.

F Curriculum Content Examples and Select Activities

General Purpose	**Program Objectives**
	Evaluating social situations from other people's perspectives
	Respecting and cherishing individual differences
	Peaceful communication with others
	Standing against all kinds of violence and maltreatment toward living beings
Expected Learning Outcomes	**Students will**
	Have enhanced perspective-taking ability
	Learn that individual differences are needed, desired and enriching factors in our lives
	Be respectful for individual differences
	Want to establish and preserve a cohesive and inclusive classroom culture
	Develop a principled attitude toward oral and physical violence in class and in school
	Trust people around them and be trusted in return
	Develop the ability to act in harmony with team spirit

Table F.0.4: "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 .1: Understanding Each Other

Figure .2: Curriculum

Figure .3: Example activity from the book

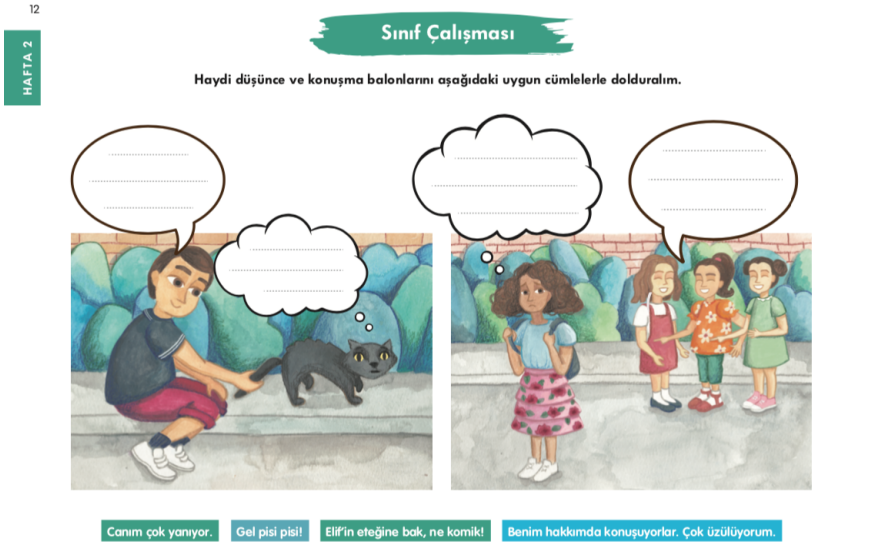


Figure .4: Example activity from the book