Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting*

Sule Alan[†1] and Elif Kubilay[2]

[1]European University Institute

[2]University of Essex

January 2, 2024

Abstract

We test the effectiveness of a behavioral approach designed to empower socioeconomically disadvantaged adolescents. The approach leverages adolescents' desire for social status and is grounded in the idea that self-persuasion might yield a more robust behavioral change in challenging adolescents than direct lecturing. We enlist socially connected senior students with high emotional intelligence as "student-teachers" and entrust them with delivering an empowerment curriculum to their junior peers. Using randomized variation in program implementation, we show that this indirect targeting empowers targeted adolescents, leading them to improve their social environment. The program reduces disciplinary incidents and anti-social behavior among student-teachers and their friendship networks while fostering supportive network ties between senior and junior students. The program also lowers the tolerance for anti-social behavior, measured by the willingness to destroy unfairly gained payoffs in a third-party punishment game. Our study offers a cost-effective way to help disadvantaged adolescents escape neighborhood disadvantages.

JEL Codes: C93, D63, I24

Keywords: neighbourhood disadvantages; adolescent empowerment; school climate

---

*We are grateful to the European Research Council and J-PAL Post-Primary Education Initiative for funding the study. We thank the seminar participants in various universities and conferences for valuable comments. The project has ethics approvals from the European University Institute, Italy and Bilkent University, Turkey. We are grateful to numerous field assistants for their efforts and care in onerous data collection. The trial has been registered at the AEA Registry: AEARCTR-0009140.

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

1 Introduction

A child's social environment has a profound influence on their life chances and eventual outcomes. In disadvantaged neighborhoods with concentrated poverty, many challenges hinder children's cognitive, social, and emotional growth (Sharkey, 2010; Chyn, 2018; Chyn and Katz, 2021; Dustmann et al., 2023). Adolescence, marked by profound changes in the brain and intense emotional fluctuations, is a period of exceptional vulnerability (Dahl, 2004; Steinberg, 2008). Damaging norms and behavioral codes prevalent in disadvantaged neighborhoods can be quickly internalized by adolescents who are in the process of developing their self-concept and social identity. While education offers a means of breaking free from this vicious cycle, schools frequently mirror the very neighborhoods they serve. Schools in disadvantaged neighborhoods are typically characterized by a poor relational atmosphere that hinders the development of a healthy self-concept and emotional stability in adolescents. Nevertheless, schools remain vital in providing children with the tools to escape poverty when parental input is of low quality.

In this paper, we test the effectiveness of a behavioral approach to making schools in disadvantaged neighborhoods a better social environment for adolescents. For this, we indirectly target intellectually bright, socially influential, yet challenging adolescents by entrusting them with the task of transforming their schools and immediate surroundings. The approach was shaped through in-depth qualitative work involving repeated interactions with senior middle school students in Turkey. Our qualitative inquiry resulted in insights consistent with Yeager et al. (2018) that shows that interventions targeting adolescents tend to fail when they do not align with adolescents' desire to feel respected and be granted social status. Drawing from these insights, we developed an empowerment program designed for implementation in disadvantaged middle schools in Southeast Turkey. The program was built on two principles: First, approaching adolescents with respect by entrusting them with responsibilities will help them develop a healthy self-concept and empower them. Second, fostering self-persuasion, rather than direct lecturing, has a higher chance of achieving the desired behavioral changes in adolescents who may have limited trust in adults around them.[1]

[1]The idea of self-persuasion is that trying to persuade a person to adopt a particular belief allows one's own mind to be gradually persuaded (see, e.g., Schwardmann et al. (2022)). It is also known that if the message used for persuasion creates a certain degree of discomfort in the persuader's mind due to the inconsistency between their behavior and the conveyed message (cognitive dissonance), a resolution may transpire over time, i.e., the persuader's behavior may align with their freshly embraced convictions (see, e.g., Mullainathan and Washington (2009)).

The program required first selecting a number of emotionally intelligent and socially influential senior students, using baseline data. Labeling them as "student-teachers," we gave them the responsibility of delivering a specifically designed empowerment curriculum to their junior peers. The curriculum, coined "Our Future-Our Dream," is structured around nine topics to be delivered within an academic year in 15 to 20 weekly sessions. Topics include envisioning the ideal school and ideal human relationships, recognizing one's power to shape their social environment and becoming a decision-maker to build a better future, and understanding the perils of violence and anti-social behavior. In each session, student-teachers gave a presentation and administered in-class activities/games. Prior to a session, student-teachers thoroughly rehearsed their presentations and activities among themselves. These rehearsal sessions were designed to intensify the self-persuasion, and create subtle discomfort in student-teachers' minds if there was any inconsistency between the messages they delivered and their everyday behavior.

The study was launched in the academic year of 2021-2022, covering 65 middle schools and over 20,000 students in the province of Diyarbakir, Turkey. In October 2021, we collected our baseline data. Then, we selected our student-teachers based on an emotional intelligence test, developed by Baron-Cohen et al. (1997), known as "Reading the mind in the eyes" test, and extensive social networks elicitation. Our student-teacher sample (over 1250 students) comprises 10-15% of the population of 5th and 6th graders in the study sample with the highest score on the average of the emotional intelligence test, the number of friendship nominations received (in-degree ties), and the number of popularity nominations received. We aimed to have five to eight student-teachers per junior classroom and ensured gender balance in selecting them. We then randomly assigned 32 schools to treatment and the remaining 33 to control. To tease out a mere interaction mechanism the program delivery generates, we further randomized the control schools, assigned 16 to placebo treatment, and left the remaining 17 as pure control. In placebo schools, student-teachers delivered an unrelated curriculum consisting of doing mazes, connecting dots, and coloring tasks. In pure control schools, we did not give any task to student-teachers or inform them about their status. The first endline was conducted in April-May 2022, at the end of the 2021-2022 academic year, allowing us to assess short-term impacts. The program was re-implemented after collecting baseline data from the newcomers of the 2022-2023 academic year. We collected final endline data in April-May 2023, after two years of rotating program implementation, enabling us to assess the persistence of our short-term results. The intention was to implement the program at scale upon positive evaluation results, enlisting new student-teachers every year.

We use a rich toolkit encompassing administrative records, an incentivized behavioral game, surveys, and cognitive tests to evaluate the effectiveness of the intervention. To assess the level of high-intensity violence and anti-social behavior, we gained access to disciplinary flagging administered by the school. We developed an incentivized third-party punishment game to determine the propensity to engage in anti-social behavior for own advantage and the desire to punish such behavior. The former is designed as two players competing in a real-effort task and deciding to transfer the opponent's points to themselves to gain advantage. The latter involves an observer of the players' transfers deciding to destroy players' points. We also elicited the web of friendship and support networks. Here, our primary interest is whether the program increased support links between senior and junior students, i.e., generated positive and supportive inter-grade relationships. In addition, we used survey instruments to measure social norms and behaviors of adults and teachers, as perceived by students. To assess the degree of empowerment and explore potential mechanisms, we measured locus of control, mental well-being, self-worth, perspective-taking, impulsivity, sense of belonging, and sense of responsibility. We also assessed whether the intervention affected academic achievement using achievement tests prepared based on the national curricula and implemented in class.

We estimate the effect of the intervention for year 1 and show its persistence in year 2. We present our results for the full sample, junior sample, and for our target (senior) subsamples. Our target subsamples are student-teachers, and student-teachers' networks (those who nominate student-teachers as their friends, i.e., in-degree ties, and those whom student-teachers nominate as friends, i.e., out-degree ties). We also assess further spillover effects on seniors outside student-teacher networks. The estimated effects across the broad range of outcomes suggest that the program significantly improved treated schools' social and behavioral atmosphere. In treated schools, the probability of disciplinary flagging for high-intensity behavioral problems was significantly reduced. As conjectured, these positive effects primarily stem from the target subgroups, i.e., treated student-teachers and their networks. Incidences of high-intensity disciplinary acts were about 4 percentage points higher than the school average for the student-teachers in the control group, consistent with the findings that popularity (high social status) is correlated with troublesome behavior in disadvantaged middle schools (Luthar and McMahon, 1996; Thunfors and Cornell, 2008). The estimated program effect of 4.1 percentage points implies a statistically significant 70.7% decline in flagging in treated student-teachers relative to the student-teachers in the control group in year 1. The effect size is 3.3 percentage points in the second year, implying a statistically

significant 55% effect. The estimated effects are similarly sized and precisely estimated for treated seniors in student-teachers' networks (around 47.2 to 63.3% decline relative to their control counterparts in both years).

Using the decisions in our third-party punishment game, we find that the program diminished anti-social behavior while enhancing the tendency to punish such behavior. These effects are large and precisely estimated for the student-teachers and their networks, especially in year 2. Specifically, in the control group, student-teachers transferred, on average, 1.142 (1.157) correct answers from their opponents in year 1 (year 2). This amount was 8.5% (21%) lower in the treatment group, although the year 1 estimate does not reach statistical significance. Similarly, treated student-teachers sacrificed 29.8% more gifts of their own to punish anti-social behavior in year 2, with no significant effect in year 1. The effects on student-teacher networks are large and statistically significant, implying significant spillover effects as in disciplinary flagging.

Using elicited support networks, we find that the program increased the number of intergrade support ties within the school. Treated juniors nominated significantly more senior students as supportive schoolmates than control juniors. This is an important result, as many anti-social acts concern seniors harassing juniors in our setting. Consistent with this, we find that perceived behavioral norms significantly improved in treated schools, with no consistent improvement in the perception of adults. The positive results on perceived norms come exclusively from our target subgroups. Treated student-teachers report significantly better behavioral norms for their school in year 1, persisting into year 2. The estimates are similar in size and precision for their networks and for seniors outside student-teacher networks. We also explore whether the program negatively affected student-teachers' academic outcomes due to the time devoted to implementing it. Overall, we find no evidence of adverse effects on academic outcomes. On the contrary, we find 0.123 sd increase in student-teachers' math performance in year 1, with some evidence of persistence in year 2 (and 0.042 sd), although the latter does not reaching statistical significance.

We rule out a mere interaction mechanism using our placebo arm. For most of our outcomes, we are able to reject the equality of the treatment and placebo effects. Our rich data allow us to reveal suggestively that the program, combined with its content and its delivery style, achieves these positive results by changing the students' beliefs, attitudes and behaviors. Notably, we find substantial and statistically significant improvements in empowerment-related attributes, such as internal locus of control and mental well-being.

Again, these improvements are predominantly observed within our targeted subgroups. Additionally, we find that treated student-teachers have significantly higher perspective-taking ability, a higher sense of belonging, lower impulsivity, and a higher sense of responsibility toward world issues such as crime, violence, and environmental disasters. Striking improvements in these outcomes are also observed in student-teachers' networks, further confirming the program's robust spillover effects. Finally, our heterogeneity analysis reveals that most of our results are driven by female student-teachers and females in student-teacher networks.

Our paper offers two main contributions. One pertains to the nature of the intervention, and the other to the rich toolkit we developed to assess its effectiveness. The former emphasizes a unique behavioral targeting approach devised through in-depth qualitative inquiry. The approach leverages adolescents' desire for autonomy and social status by entrusting them with the task of assisting their younger schoolmates. This empowerment improves their social and emotional well-being, leading them to transform their social environment. In comparison to programs targeting disadvantaged youth, like Cognitive Behavioral Therapy (CBT) or change of neighborhood (Move to Opportunity), our intervention stands out for its cost-effectiveness ($11.7 per student-teacher per year) and its capacity to generate significant spillover effects. Our second contribution involves the outcome set we developed to comprehensively characterize a school's social environment, considering both objective measures and adolescents' perceptions. The consistent improvements observed across our multifaceted outcomes underscore the effectiveness of our approach. Our results, therefore, can inform policies aimed at helping adolescents escape neighborhood disadvantages.

Our paper contributes to various strands of the literature. First, it complements the broad literature on the relationship between socioeconomic environment and individual outcomes. This literature shows how socioeconomic background affects children's social and economic outcomes (Katz et al., 2001; Oreopoulos, 2003; Dahl and Lochner, 2012; Damm and Dustmann, 2014; Chetty and Hendren, 2018). A strand of this literature tests the effectiveness of various interventions aimed at helping children and young people trapped in parental and neighborhood disadvantages, such as negligence, crime, and violence. Heller (2014) tests the effects of a Chicago program in which disadvantaged youth took up summer employment and finds a significant drop in crime rates. Chetty et al. (2016) evaluate the effect of the Moving to Opportunity (MTO) experiment, which offers an opportunity to relocate to higher-income neighborhoods, on children's long-term outcomes. They find that age at the time of the move and the duration of exposure matter a lot for better outcomes. Heller et al. (2017) evaluate a program, Becoming a Man (BAM), aimed at reducing violent

crime and improving school engagement and find favorable results on arrests and graduation rates. Paluck et al. (2016) evaluate a school-based intervention involving students of high social status taking a public stance against conflict at their school and find that the intervention reduced overall levels of conflict. Finally, there are numerous studies evaluating the effects of mentoring programs designed for disadvantaged children and adolescents on outcomes including crime, achievement and socio-emotional skills (Oreopoulos et al., 2017; Guryan et al., 2021; Resnjanskij et al., 2023; Dinarte-Diaz and Egana-delSol, 2023). Our paper complements these studies by showing that empowering adolescents by entrusting them with the responsibility of transforming their social environment can yield both socially and individually beneficial results.

Second, we contribute to the growing literature on adolescent development. This literature shows that the period of adolescence is unique as the adolescent brain undergoes drastic social and cognitive changes resulting in sensitivities to the social environment. This sensitive period presents many challenges but also provides ample opportunities to offer a healthy developmental trajectory for adolescents in need (Blakemore and Mills, 2014; Dahl et al., 2018; Andrews et al., 2021). However, interventions that work for children and young adults may not work for adolescents who are in the process of developing self-identity and adapting to their social environment, especially in contexts where adult input is of low quality and neighborhood disadvantages abound (Yeager et al., 2018). With the help of extensive qualitative inquiry involving repeated interactions with hard-to-approach adolescents, we show that insights from behavioral science can help us empower them to take control of their social environment. Our in-school approach also complements the growing literature on social and emotional development in the school environment (Alan and Ertac, 2018; Alan et al., 2019; Sorrenti et al., 2020; Alan et al., 2021).

Finally, we contribute to the growing literature on peer effects. This literature strives to understand how peers influence and shape each other's academic outcomes (Sacerdote, 2001; Zimmerman, 2003; Jackson, 2008; Calvó-Armengol et al., 2009; Sacerdote, 2011). Recent studies explore effects beyond educational achievement outcomes. For example, Zárate (2023) shows the impact of socially central adolescents on their peers' social skills and academic performance. Kiessling and Norris (2023) show that peers are crucial in determining the long-term health of individuals. Leveraging the friendship ties in the classroom, Alan and Mumcu (2023) show that information dissemination among peers is vital to achieving high-quality learning and socio-emotional development. Our paper shows that interventions aimed at improving adolescents' social environment have a better chance of success if they

consider the importance of peer relationships in adolescence.

The rest of the paper is organized as follows. Section 2 summarizes the key features of the program and the context in which it was implemented. Section 3 details the evaluation design. Section 4 gives a detailed account of the toolkit we use to evaluate the program. Section 5 describes the data. Section 6 presents our main results. In Section 7, we discuss the mechanisms through which the program might have improved the social environment in schools. We conclude in Section 8.

2 Context, Intervention and Delivery

2.1 Context and Qualitative Inquiry for Program Design

The Turkish compulsory education system spans 12 years, divided into 4 years of primary school, 4 years of middle school, and 4 years of high school. In disadvantaged low-income regions, such as the focus of our study, middle schools in city centers are quite large, with numerous classrooms (up to 20 classrooms) per grade, often with crowded classes containing 40 to 50 students each. Due to Turkey's fast-changing demographic structure, the student body is smaller in remote village schools, where we typically observe two classrooms per grade level. Public middle schools in these districts and villages do not offer favorable learning conditions. These schools are typically characterized by poor student attendance, low academic achievement, and highly prevalent anti-social behavior and peer violence.

We started our qualitative work to assess middle school students' perception of their social climate and their socio-emotional health in 2019 in several out-of-sample pilot schools. Our focus on middle schools (early adolescence) was motivated by recent neuroscience findings highlighting the unique attributes of the early adolescent brain and the potential for positive behavioral changes during this period (Blakemore and Mills, 2014; Dahl et al., 2018). Our qualitative research took various iterative forms as we encountered challenges in connecting with students at first. We discovered a severe lack of trust in adults among these teens, leading to an initial disregard for our efforts.[2] We also noticed in pilot schools that not all students flagged as troublesome by the administrators were violent or anti-social. Many were clever, somewhat mischievous children with little trust in adults, acting up and frustrating

[2]Consistent with what we observed in pilot schools, in our data, about 50% of senior students state that adults do not respect them and pay little to no attention to their opinions.

their teachers and school administrators. Realizing that conventional lecture-type interventions would not work, we decided to leverage their desire to be respected and granted social status. We approached the pilot senior students with the idea of helping junior students in their school, emphasizing the vulnerability of juniors and how seniors' guidance could make them feel more belonging and safe. This idea received enthusiastic support from most senior students in our pilot schools. We then conjectured that if we could have senior students repeatedly deliver our messages to juniors for an extended period, their beliefs and behaviors would eventually align with the messages they deliver.

To test and evaluate our targeting approach, we collaborated with the provincial education authority of Diyarbakir, Turkey, to recruit middle schools in disadvantaged city districts and villages. Diyarbakir, a major city in the Southeast region of Turkey, presented an ideal environment for our project due to its demographics and socioeconomic conditions. The city has a population of about 2 million and, like similarly sized cities in Turkey, faces challenges of concentrated poverty and social unrest.

2.2 Program Content and Implementation

We aimed to target students with intellectual capabilities and social influence who can positively impact their peers. In disadvantaged middle school settings, social influence often comes with problematic behavior (Luthar and McMahon, 1996; Thunfors and Cornell, 2008). Popular students tend to gain popularity more due to their rebellious and mischievous actions rather than their academic achievements, though the latter is not unimportant. To select our student-teachers, we used a three-input algorithm, conditional on their baseline willingness to be student-teachers.[3] The first input was the student's baseline score on the Reading the Mind in the Eyes test (RMET), a measure of emotional intelligence or cognitive empathy. This test involves identifying emotional states from images of people's eyes. Cognitive empathy is shown to be correlated with fluid intelligence, importantly, with prosocial behavior and highly predictive of effective leadership (Wolff et al., 2002; Alan et al., 2023). The second and third inputs came from our baseline network data, specifically, the number of friendship nominations received (in-degree ties) and the number of nominations received as a "popular" student.

[3]At baseline, we asked all seniors whether they would be willing to teach some materials to juniors if selected. About 73.2% of the seniors responded positively to this question.

We averaged these three inputs and selected the highest-scoring students, constituting around 10-15% of the population of selected 5th and 6th-grade classrooms. We ensured gender balance among our student-teachers. After obtaining consent from both the selected students and their parents, we assigned 7th-grade student-teachers to 5th graders and 8th-grade student-teachers to 6th graders. The emotional intelligence test helped exclude students with severely violent behavior or those requiring professional help, as enlisting them as student-teachers would have raised ethical concerns. Our student-teacher sample is diverse, with some high and low academic achievers, troublemakers, and well-behaved ones. On average, student-teachers tend to have higher disciplinary issues than the rest.

To organize the messages we aimed to convey, we helped develop a curriculum containing slides, videos, posters, activities, and games to be delivered by student-teachers, using the well-being hours allocated to all middle schools by the Turkish Ministry of Education. The curriculum, named "Our Future-Our Dream," has been designed by a team of education consultants and artists, supervised by the authors. It is structured around nine topics intended for delivery throughout an academic year in weekly sessions spanning 15 to 20 weeks. The topics included concepts such as envisioning an ideal school and peer relationships, recognizing one's power to influence the social environment and become a decision-maker, thinking about the world's problems, recognizing the dangers of intolerance and violence, and more. For example, in one session, student-teachers showed a video on a profile of a bully. The movie highlights where the power of the bully comes from (followers) and how weak they become when everyone collectively disapproves of their behavior. This session is a prime example of creating discomfort in student-teachers' minds, as some of them are likely to be bullies themselves. In another session, student-teachers showed the juniors a short film on our planetary challenges and gave a presentation on a collaborative approach to generating solutions. Online Appendix Figure B1 and Table B1 illustrate all nine topics, including the 10th topic involving an exhibition of materials created throughout the academic year. All written, visual, and multimedia materials, including placebo activities, are available as a single package from the authors.

We assigned one or two interns per treatment and placebo school, depending on the size of the school, to monitor student-teachers' activities. To deliver a session, student-teachers (around 5 to 8 student-teachers per junior classroom) met with their designated intern in a designated room in the school. They practiced and understood the session's activities before delivering them to their classrooms, whether that be a treatment or a placebo session. They then delivered the session with no intern interference. Respecting the autonomy of student-

teachers in the way they deliver the session was of paramount importance in this project, and the interns were extensively trained by the authors to comply with this aspect of the program.[4] Each session (both treatment and placebo) lasted for a lecture hour, delivered once a week, except during exam weeks or significant school activities when sessions might be postponed to the following week. Following each session's implementation, every student-teacher completed a progress report using a notebook provided. This report gathered their opinion regarding the session's success, suggestions for improvement, and ways to optimize the impact. This part was also designed to signal the respect and trust extended to the student-teachers' ideas and opinions. Online Appendix I and II show some treatment and placebo implementation photos.

In treatment schools, we deliberately emphasized details like student-teachers wearing t-shirts with the project logo, discussing the session content before delivery, and critically evaluating the completed session when writing progress reports. These actions aimed at compelling student-teachers to think deeply about the content. Our conjectured behavioral change depended on them embracing the project, feeling trusted, empowered, and responsible, and eventually subscribing to the messages they were asked to deliver to juniors. Thus, we anticipate improvements, primarily within our student-teachers and, because of their powerful social status, within their social networks. Although junior students were not our central targets, we also expect positive changes in their behavior as they may absorb the delivered content.

3 Evaluation Design and Timeline of the Study

We recruited 65 middle schools hosting over 20,000 students in the province of Diyarbakir, Turkey. These schools varied in size and type. Some were very large, with many classrooms for each grade. Twenty four of them were located in distant villages, and 21 were categorized as religious schools (Imam Hatips). All religious schools, some located in villages, some in inner-city districts, follow the national curriculum, with additional teaching of Islam and Arabic, leading to an extra lecture hour per school day.

We conducted our first baseline in October 2021 by visiting each school in person and collecting data via tablets. We collected data from all students in the school if the school

[4]Ensuring this implementation protocol was adhered to, we regularly met with interns, and they submitted weekly progress reports via an online platform.

has at most three classrooms per grade level. For larger schools, we randomly picked three 5th and three 6th-grade classrooms as our junior targets. To choose our student-teachers, however, we had to span the entire senior population in a given school regardless of its size, administer the emotional intelligence test, and collect social network data. Seniors who were not in student-teachers' networks helped us assess further spillover effects. In very large schools, we randomly selected two or three 7th and 8th-grade classrooms for this purpose.[5] This intensive data collection required spending an entire school day in a school by 3 to 4 field team members, assisted by 6 to 8 locally recruited field assistants.

After baseline and selecting student-teachers based on the algorithm mentioned above, we randomly assigned 32 schools to treatment, 33 to control. Among the 33 control schools, we randomly assigned 16 to placebo control and 17 to pure control. In placebo schools, chosen student-teachers conducted unrelated activities with no particular content in their assigned junior classes, again, with the hands-off monitoring by their assigned interns, in the same intensity (one lecture hour per week throughout the academic year). These activities included solving mazes, connecting dots to draw animal shapes, and coloring.[6] The purpose of this arm is to rule out a mechanism that the estimated effects stem from the senior-junior-intern interactions that our delivery method creates. By replicating the nature of the interaction in the treatment arm exactly, we are able to assess its role in influencing the outcomes we consider. The acceptance rate of the student-teacher role was 100% at post-randomization, both in treatment and placebo schools.[7]

We conducted our first endline in April-May 2022. In October 2022, we visited all schools again, collected baseline data from newly arrived 5th graders, and conducted a new network elicitation for the entire 7th-grade population. The latter was in 6th grade in the previous academic year. We chose new student-teachers among these 7th graders using the same 3-input algorithm and assigned them to the new 5th graders. The idea of this design is that when scaled up, it would rotate every year so that a once-junior student can have a

[5]In schools with at most 3 classrooms per grade, we automatically have seniors outside student-teachers' networks, as we covered the entire student body in these schools.

[6]The length of the placebo treatment and the interactions it generated were similar to the treatment: Student-teachers met and rehearsed the activities before the session and wrote post-session reports.

[7]Due to a small coding error in calculating the RMET score at baseline, 159 student-teachers were wrongly selected, corresponding to 12.5% of our student-teacher sample. The distribution of these students is balanced across treatment status, with 75 in treatment, 43 in placebo, and 41 in pure control. We did not want to disappoint treatment and placebo students by excusing them after announcing their roles and receiving their and their parent's consent. Instead, we added 159 new student-teachers to our sample. Our analyses are robust to dropping or controlling for wrongly selected student-teachers.

chance to become a student-teacher when she is in the 7th grade. In the second run of the intervention, we did not re-enlist the 8th graders (previously 7th graders) as they had already completed their task. In the second year, the program ran between 7th and 5th graders, but the activities were visible to all students as before. At this point, our previous 8th graders had already gone to different high schools in or outside the region. It is important to note that during the second year of the implementation, a devastating earthquake hit the region (February 6, 2023). The province of Diyarbakir was one of the affected regions. Because the Ministry kept the schools closed for about six weeks to use the buildings for earthquake relief, the program paused until March 1, 2023. Upon re-opening, we resumed the implementation activities. We conducted the second endline in April-May 2023 and completed the trial. Figure 1 depicts the timeline of the study and its rotating nature.

4 Outcomes

We use a comprehensive set of outcomes that characterize students' social environment in their school. Our toolkit includes administrative records, an incentivized game, and surveys. Except for our incentivized behavioral task, we collected all our outcomes, both baseline, and endline, in both years of the project. In addition to our primary outcomes that describe the social climate in the school, we collected individual outcomes using surveys and cognitive tests to assess the extent of socio-emotional improvements and possible undesired effects. Below, we explain the primary outcomes we use to describe the social and relational environment in the school.

4.1 Social and Relational Outcomes

Our first order interest is improving the school's relational atmosphere. Considering the nature of the intervention and its delivery method, we consider the improvement in students' social and emotional well-being as the most plausible pathway to achieve such an atmosphere.

4.1.1 Disciplinary Flagging

In a healthy school environment, acts of extreme violence are not expected. To assess whether the intervention affected the probability of extreme behavioral issues, we use administrative data on disciplinary flagging, which is an official record keeping of high-intensity behavioral

issues by the school administration. Flagging is actively updated throughout the year. A student can be flagged and then unflagged several times within the same academic year. Our data is a snapshot of these records at the time of our endline. We expect that the intervention will reduce the probability of disciplinary flagging, especially in our target subgroups.

4.1.2 Tolerance for Anti-Social Behavior: A Third- Party Punishment Game

We expect fewer anti-social and unfair acts between schoolmates in a healthy school environment. We also expect stronger backlash toward such acts when they occur. A costly third-party punishment game is ideal for us to explore these behaviors in an incentive-compatible way in our setting. Therefore, as part of our toolkit, we designed a novel third-party punishment game to quantify the tendency to engage in an unfair/anti-social behavior and the tendency to punish such behavior.

Our game involves randomly forming student groups of three within the classroom and assigning two of them the role of "player" and the other "observer." Students did not know which role they would take at the outset and were told this would be determined at the end of the game. Therefore, they were to play the game first by assuming the role of players, and their decisions and outcomes would be recorded. Then, they were to change roles and play as observers, and those decisions and outcomes would also be recorded. At the end of the session, those who were chosen as players would receive their gifts according to points they earned as players, and those selected as observers would receive their gifts according to points they earned as observers.[8]

Players first compete against each other in a real-effort task. The task involves typing as many meaningless 5-character password sequences as possible, containing lower and upper case letters and numbers in 1.5 minutes. The player who types the most at the end of 1.5 minutes receives 1 gift point per correct password. The other player receives zero points regardless of the number of correct passwords they typed. In case of a tie, a player is chosen as a winner at random and receives 1 gift point per correct password, and the other player receives zero points. Each gift point can be converted into an actual gift we brought to the

[8]This implementation method, referred to as strategy method, is theoretically equivalent to playing the game by splitting the sample at the outset (Brandts and Charness, 2011). The advantage of this method in our context is that it preserves the power of our design, as we do not have to split the sample into players and observers to estimate the treatment effects on decisions. The method also allows us to simulate payoffs easily as we can observe everyone's decisions as a player and observer.

classroom and showed the students before introducing the game.[9] After students completed the task, before knowing their winning status against their opponent, they were informed that all players would receive 2 bonus gift points so that even the ones who lost would end up with two points worth of gifts at the very least. They were then informed that they could use these two bonus points to transfer their opponents' correct answers to themselves to increase (decrease) their (their opponent's) chances of winning. They were informed that each transfer, however, would cost one bonus point, so they could transfer a maximum of 2 correct answers since they have only 2 bonus points. Players were repeatedly told that the transfer decision was theirs, so they could choose zero transfer, costing nothing, one transfer, costing one bonus point, or two transfers, costing two bonus points. They were also reminded that their opponent had the same transfer options.

Additionally, before they made their transfer decision, students were informed that the observer in their group would see how many correct answers each player transferred from the other player, and if she desired, she could destroy correct answers. She could, of course, choose not to destroy, destroy one or two from each player, so the maximum she could destroy from a player is 2. We gave numerous examples of different situations to ensure students fully understood the consequences of possible actions.[10] Students were then asked to make their transfer decisions, which amounted to transferring from the other player zero, 1, or 2 correct answers. After eliciting these decisions, they were asked to make a guess about the action of their opponent, i.e., whether they believed the opponent transferred zero, 1, or 2 from them. The action of costly transfer is our incentive-compatible measure of anti-social behavior, as it represents unfair competition in our setting. Elicited beliefs about the opponent's action tell us about the perceived relational climate, as it quantifies the anti-social behavior expected from classmates.

After performing as players, students were asked to assume the role of observers and reminded about the possible actions of the observer. They were told that each observer had an endowment of 6 gift points. Their role is to decide whether to destroy the answers of the players in their group and, if so, which player and how many. Students were told that destruction was costly for the observer. To destroy one correct answer, the observer had to give up one of her 6 points, and she could destroy up to 4 correct answers (2 from each

[9]These are small but attractive gift items of value to our target group, including key chains and play cards of famous football teams, attractive stationary items, notebooks, pens, etc.

[10]The implementation of this task took an entire lecture hour. We did not elicit decisions before ensuring the students fully understood the game.

player). They were told over and over again that they did not have to destroy, and if they did not, they would keep their 6 points. We elicited the decisions of the observers using a strategy method again. For this, we went through every possible transfer behavior and elicited the punishment decision, starting from the situation that neither player transferred any, then player 1 transferred 1, the other did not transfer any, then each transferred 1, and so on. Note that there are 6 possibilities to consider due to the symmetry between two players. We present detailed instructions for the game, as it was implemented in the classroom, as well as the screenshots of the game in the Online Appendix V.

Given the structure of the game, the payoff P of player i is

$$P_i = 2 + X_i + Y_i - Y_j - Z_i, \tag{1}$$

and the payoff of the observer is

$$P_o = 6 - Z_i(Y_i, Y_j) - Z_j(Y_i, Y_j), \tag{2}$$

where

- X_i: Number of correct answers by player i
- Y_i: Number of correct answers transferred by player i from player j, and $Y_i \in \{0, 1, 2\}$
- Y_j: Number of correct answers transferred by player j from player i, and $Y_j \in \{0, 1, 2\}$
- $Z_i(Y_i, Y_j)$: Number of correct answers of player i destroyed by the observer, and $Z_i \in \{0, 1, 2\}$
- $Z_j(Y_i, Y_j)$: Number of correct answers of player j destroyed by the observer, and $Z_j \in \{0, 1, 2\}$

We consider three social outcomes using the decisions in this game. First is the number of correct answers transferred, i.e., Y_i, representing anti-social/unfair behavior toward a classmate. The second is the anti-social behavior expected from classmates (players' guess about the action of their opponents). The final one is the cost incurred to punish players. For this, we define an aggregate cost of punishment metric using asymmetric transfer cases. The aggregate cost of punishment of players i and j in cases of asymmetric transfers:

$$C_o = \sum_{(Y_i, Y_j) \in Y} (Z_i(Y_i, Y_j) - Z_j(Y_i, Y_j)), \tag{3}$$

where $Y_i, Y_j \in \{0, 1, 2\}$ and $Y = \{(Y_i, Y_j) | Y_i \neq Y_j\}$.

We expect treated students, especially our target subgroups, to transfer less, expect more fair behavior from their classmates, and exhibit a higher tendency to punish unfair actions (stronger desire to punish anti-social behaviour).

4.1.3 Social Networks and Perceived Social Environment

To understand the social relationships in the school further, we elicited social networks at both baseline and endline. For this, we asked students to nominate (i) at most 3 schoolmates as close friends and (ii) at most 3 schoolmates who provide emotional support, allowing the two domains to overlap. Our primary interest is the latter as we would like to assess whether the intervention increased the prevalence of support from our target senior groups to juniors, i.e., inter-grade (directional) support ties. For this, we estimate treatment effects on support ties directed to senior subgroups by juniors. We expect that the intervention will increase these support ties. Finally, we collected perceived behavioral norms and perceptions of adult behaviors using item-response questionnaires. For each domain, we construct a standardized index using the relevant item response questions.

4.2 Individual Outcomes: Socio-emotional Well-Being and Achievement

We conjecture that the treatment will improve the social relationships and make the school a better learning environment by improving targeted seniors' social and emotional well-being. Given the nature of the targeting, we expect significant improvement in empowerment indicators. One of these indicators is the internal locus of control. Locus of control refers to an individual's belief about the extent to which they can control or influence events in their lives (Rotter, 1966). Individuals with an internal locus of control are inclined to believe in their own ability to influence the outcomes in their lives. This belief in personal efficacy is often associated with higher self-worth and better mental well-being (Kesavayuth et al., 2022). It has been shown in social psychology and recently in the economics literature that individuals with an internal locus of control are more prosocial toward others and act more responsibly toward their physical environment because they believe that their actions can make a difference (Midlarsky and Midlarsky, 1973; Bierhoff et al., 1991; Andor et al., 2022). To assess the extent of improvement in these empowerment outcomes, we measured internal locus of control, self-worth, mental well-being, and sense of responsibility at baseline and

endline. We also measured perspective-taking, impulsivity, and sense of belonging to the school both at baseline and endline, as the program may also have affected these attributes. Items for each index constructed for perceived social environment and individual outcomes are given in the Online Appendix IV.

Finally, although the program did not target achievement outcomes, we explore its effects on test scores. These outcomes could be affected positively through the improvement in the social environment. Moreover, improvements in social and emotional well-being may have led to higher motivations to study and, in turn, to higher academic achievement. On the other hand, the program was time-consuming, especially for the student-teachers. Student-teachers had to read, prepare, and practice the material before delivering their sessions. Even though interns were instructed not to crowd out any core lesson when they visited their schools, some study time could, in principle, be devoted to the project activities. To assess the program's impact on achievement, we administered in-class math and Turkish tests prepared based on the national curricula at baseline and endline in both years.

5 Data and Empirical Analysis

We collected data from about 18,000 students from 65 schools by visiting each school in person. Our administrative data on disciplinary flagging cover over 27,000 officially registered students. Demographic information and some indicators of home environment were collected only at baseline. We also collected fluid IQ using Raven's progressive matrices (Raven and Court, 1998) (only at baseline) and emotional intelligence using Reading the Mind in the Eyes test (Baron-Cohen et al., 1997). We collected all outcomes, except for the incentivized third-party punishment game and official disciplinary flagging, both at baseline and endline.[11] In the first year of the trial, we selected 633 7th-grade student-teachers, providing us with 3184 friends of theirs (in-degree+out-degree ties) and 636 8th-grade student-teachers, giving us 3055 friends. To assess further spillover effects on seniors outside student-teacher networks, we collected data from 4893 randomly chosen senior students. These students are the ones with no out or in-degree ties with student-teachers. About 48% of our student-teachers are female. Table 1 compares the characteristics of student-teachers with other senior students in year 1. As can be seen, most characteristics are starkly different. Our student-teachers

[11]We did not collect official flagging at baseline because flagging starts later in the term. Instead, we collected self-reported experiences of bullying and anti-social acts from all students at baseline.

have higher cognitive scores. This is expected as they were selected partly based on the RME test, which is a cognitive test and is highly correlated with fluid IQ. Note that they are not different from non-selected seniors in terms of their perceived behavioral norms and their impulsivity.

5.1 Internal Validity

Table 2 illustrates the balance across three treatment arms in year 1. We also provide the balance for juniors, seniors and finally for student-teachers and their network (in-degree ties) in the Online Appendix (see tables B2, B3, and B4). Overall, the randomization worked well and we observe no noteworthy imbalance across treatment status in any of the outcomes.

5.2 Empirical Model

We estimate the average treatment effects of the program on outcomes of interest by conditioning on baseline covariates and randomization strata (district) fixed effects. In our main specification, we pool the placebo and pure control and estimate the following (benchmark) model:

$$y_{ics} = \alpha_0 + \alpha_1 T_s + X_{ics}^{'}\beta + \delta_d + \varepsilon_{ics}, \tag{4}$$

where y_{ics} is the outcome of interest for child i in classroom c, school s. T_s is the binary treatment indicator, which equals one if school s is in the treatment group and zero otherwise (placebo+pure control), and X'_{ics} is a vector of student-level observables, including gender, age, Raven's score, and RMET score. We also control for school type fixed effects. δ_d represents district (strata) fixed effects.

To tease out the effect of the senior-junior interaction created by the program implementation, we also estimate the following (full) model:

$$y_{ics} = \alpha_0 + \alpha_1 T_s + \alpha_2 P_s + X_{ics}^{'}\beta + \delta_d + \varepsilon_{ics}, \tag{5}$$

where P_s is a binary indicator for placebo treatment, which equals one if school s is in the placebo group and zero otherwise.

We had full compliance in this trial in the first year, i.e., all sessions were completed in all treatment and placebo schools. In the second year, we lost a treatment school due

to a significant structural change that made it difficult for us to implement the program. Nevertheless, given the near-perfect compliance, the estimated α_1 and α_2 can be considered an average treatment effect on our study population. Note that the selection of new student-teachers in year 2 is post-treatment selection. In year 2, the proportion of the 7th graders who wanted to be student-teachers was about 5 percentage points higher than in year 1 in the treatment group. This value statistically differs from placebo and pure control (p-values, 0.064 and 0.006, respectively). Therefore, we exclude these student-teachers when we condition our sample on student-teachers in year 2. This implies analyzing only grade 8 student-teachers (previous grade 7 student-teachers) in year 2 when assessing the program's persistent effects on student-teachers.[12] We cluster standard errors at the school level in all analyses and provide wild bootstrapped p-values. Figure A1 presents our results on targeted subgroups without covariates in the Appendix. A Romano-Wolf correction due to the multiple tests we conducted is presented in the Appendix (see tables A1 and A2).[13]

5.3 Absenteeism as an Outcome and as a Threat to Internal Validity

We work in a high absenteeism setting. In this setting, it is common that on a given school day, about 15-20 percent of students are absent from school. Absenteeism among senior students is typically higher, and it goes up before and after major religious holidays and toward the end of the academic year. Teachers in Turkey are required to record attendance. Every classroom has its own A3 size attendance record book, and teachers record absent students before the lecture begins. We collected these hand-written records for several weeks of different months to have a full picture of overall absenteeism in the academic year.[14] In year 1, the overall rate of absenteeism is 14.5%. It is slightly higher (14.6%) for seniors than for juniors (14%). The absenteeism is very similar in year 2. We find no treatment effect on absenteeism, neither in year 1 nor in year 2.

At the time of the endline in year 1, the percentage of students who were present at baseline but not at endline was 20.6%. This value is significantly higher (26.4%) in year 2

[12] Our year 2 results for student-teachers remain materially the same if we include the new student-teachers.

[13] Following Kling et al. (2007), we construct summary indices by aggregating relevant outcomes. We use five outcome groups: i) disciplinary flagging, ii) experimental outcomes, ii) outcomes related to social environment, iii) network outcomes and iv) socio-emotional outcomes.

[14] Specifically, we picked week 3 in December, weeks 1 and 2 in January, February, and March, took pictures of teacher records of absent students (recorded by their school numbers), and then digitized these records to merge with our main dataset.

(p-value for the difference < 0.001). Importantly, the absenteeism at endline is not correlated with treatment status (p-value = 0.473 in year 1, 0.294 in year 2), ensuring the internal validity of our results.[15] Note also that we have a primary outcome that is not affected by absenteeism on a given day, that is, administrative records of disciplinary flagging. Another outcome, which is less affected by absenteeism, is network in-degree ties, i.e., supportiveness nominations students receive. The latter is because a student present at the time of data collection could nominate any schoolmate whether or not their nominee is present in the school. However, we may miss potential ties between absent students (missing networks). Given that absenteeism was not related to the treatment, our results remain internally valid.

6 Results

In all presentations, the first set of results comes from our benchmark model, where we compare treatment with placebo+pure control (Equation 1), and the second from our full model (Equation 2), along with a test of equality between treatment and placebo effects. We present the results, first for the full sample (Panel 1), then for juniors (Panel 2), corresponding to grades 5 and 6, and for seniors (Panel 3), corresponding to grades 7 and 8. We then split the senior sample further and show the results for our targeted subgroups: student-teachers, seniors in student-teacher networks (student-teachers' in-degree ties), and finally, the seniors outside student-teacher networks. The latter two inform us about the spillover effects of the program. Student-teacher networks inevitably include some student-teachers as well. We remove those when we estimate treatment effects on student-teacher networks.

6.1 Treatment Effects on Disciplinary Flagging

Table 3 presents estimated treatment effects on the probability of being flagged as a behaviorally challenging student by the school administration. Note first that the incident is quite small in the control group in the full sample, as flagging is done only for really difficult cases in middle schools. Only 2.1% (3.3%) of the student body is flagged as challenging in the full sample in year 1 (year 2). As seen in the table, we estimate a significant treatment effect on the probability of being flagged. Considering the low rates of incidents, the relative effect

[15] A high rate of absenteeism toward the end of the academic year is typical for Southeast Turkey due to the seasonal agricultural worker mobility. The higher rate in Year 2 was due to the fact that after the February 6 earthquake, MoE lifted the attendance requirement in the affected regions for the academic year.

sizes are substantial. We estimate the effect sizes of about 51.4% for juniors, though only in the second year. We estimate a 1.3 and 1.7 percentage point decline in the first year and second years, respectively, for the senior students. These estimates correspond to a 51.5 to 61.9% treatment effect. The program's effect on flagging appears to come mainly from the senior students.

Table 4 shows that the rate of being flagged is significantly higher among student-teachers (about 6% in both years) in the control group. This indicates that our student-teachers, while cognitively able and socially central, are often considered troublemakers in the school. Seniors in student-teacher networks have a lower flagging rate (3-3.6%), but still higher than the seniors outside of student-teacher networks. We estimate a remarkable decline in disciplinary flagging of student-teachers and students in their friendship networks. The effect size is 70.7% in year 1 and persistent (55%) in year 2 for student-teachers. They are similarly sized (63.3%, and 47.2% in year 1 and year 2, respectively) for student-teacher networks. We also estimate a significant spillover effect (44.4% and 38.9% respectively for years 1 and 2) among those outside student-teacher networks, albeit less precisely estimated. We also note an increase in flagging in our placebo group in year 1, but this finding does not repeat in year 2.

6.2 Treatment Effects on Anti-Social Behavior and Tolerance for Anti-Social Behavior

Our third-party punishment game aims to capture the propensity to engage in behavior that harms a peer for one's own advantage and tolerance for such anti-social action. We expect both decisions to capture some aspects of the school's social climate and correlate with the individual's cognitive, social, and emotional skills. Therefore, before exploring the estimated treatment effects on the decisions in the game, we provide evidence on the predictive validity of the decision to transfer correct answers from an opponent and the cost incurred by punishing transfers.

Figure 2 Panel 1 shows the distribution of the transfer behavior for control and treatment. We observe that the most prominent transfer behavior was to transfer two correct answers from the opponent. Note the visible difference between treatment and control in zero transfer behavior. Panel 2 presents the distribution of incurred punishment costs (the total number of correct answers destroyed) for each transfer type, again across treatment and control.

The striking difference between treatment and control emerges in cases where the transfers were unequal. Treated students tend to punish the players who transfer more than their opponents, i.e., cases of (1,0), (2,0), and (2,1). These cases likely trigger a sense of injustice and invoke costly punishment behavior.

Figure 3 presents associations between transfer and punishment behavior, two decisions made in the game, with the indicators of social environment, and socio-emotional and cognitive skills. The figure is generated for the control group only by pooling two years of data together. As can be seen clearly in Panel 1, transfer behavior correlates negatively with positive indicators of social climate and socio-emotional skills. The amount of transfer is positively correlated with impulsivity and negatively correlated with cognitive skills. Associations regarding the costly punishment behavior are even more pronounced. Undertaking costly punishment of unequal transfers is positively correlated with cognitive ability (both fluid and crystallized), positive social climate, internal locus of control, mental well-being, perspective taking, self-worth, sense of belonging and sense of responsibility, and negatively correlated with impulsivity. These strong correlations suggest that the decisions in this game capture significant aspects of the social climate in the school as well as the socio-emotional well-being of adolescent students.

6.2.1 Treatment Effects on Transfer and Punishment Decisions

Table 5 presents the estimated treatment effects on the transfer and punishment decisions on the third-party punishment game. The results indicate lower transfers and higher punishment in treated classrooms in both years, but the estimates are much larger and more precise in the second year. Considering the full sample, control students transferred 1.113 points from their opponents on average and destroyed 1.489 points in cases of asymmetric transfers. Treated students transferred 0.03 and 0.16 fewer points in the first and second year, respectively, with only the second-year estimates statistically significant at the 1% level. The effects are not heterogeneous across subgroups but seem larger and more precise for the seniors. Table 6 zooms into senior subgroups. Here, we see that treated student-teachers transferred 8.5% (21%), seniors in student-teacher networks transferred 1.27% (21.9%), and seniors who are out of student-teacher networks transferred 3.39% (20.5%) fewer points from their peers relative to their counterparts in the control group in year 1 (year 2). Estimated effects are weaker in year 1 but larger and highly significant in year 2. We also reject the equality between treatment and placebo effects in year 2, ruling out the pure interaction channel for

these effects.

Consistent with the significant decline in anti-social behavior, treated students are willing to incur substantially higher costs to implement justice when they observe unequal transfers. For the full sample (Table 5), the effect size is 10.3% in year 1, significant at the 5% level, and 30.6% in year 2, significant at 1%. These effects are strong and statistically significant for both juniors and seniors, with larger point estimates for senior subgroups, especially in the second year. Table 6 shows that while we do not detect a significant effect in year 1, treated student-teachers were willing to sacrifice about 0.483 points to punish unequal transfers in year 2, implying an effect size of 29.8%. Again, the spillover effects are also large and significant. Both seniors in and outside student-teacher networks in treatment schools transferred less and sacrificed more points to punish unequal transfers, with all year 2 effects statistically significant at the 1% level. Specifically, in year 2, we estimate 22.9% and 27.6% more punishment in treated student-teacher networks and seniors outside of student-teacher networks, respectively.

In sum, consistent with the estimated effects on disciplinary flagging, the program reduced the tendency to act in an anti-social manner and lowered the tolerance for such behavior, especially among the targeted subgroups. The incentivized nature of the third-party punishment game allows us to construct interesting classroom-level outcomes that are economically relevant and estimate how the intervention affected these outcomes. The below simulation exercise allows us to use what we learn from schools as social units to understand the possible causal effects of a positive relational environment on pure economic outcomes.

6.2.2 The Effect of Positive Social Climate on Economic Prosperity

Extensive literature has explored the relationship between a positive group culture and productivity, primarily in workplace settings (Ostroff, 1992; Judge et al., 2001; Srivastava et al., 2018; Guadalupe et al., 2020). This connection extends to more general discussions of the link between social cohesion and economic prosperity (Rawls, 1971; Coleman, 1990; Putnam, 1993; Gradstein and Justman, 2002). Because we are interested in building a cohesive school environment through improving adolescent socio-emotional well-being, we can leverage our incentivized game to assess the economic disparities across distinct social groups (i.e., classrooms/schools) with different levels of cohesion. Specifically, our students earned actual gifts based on their decisions and the decisions of their peers. We investigate variations

in per-capita earnings and earnings inequality across treatment categories by considering each classroom as a community (a social unit) and simulating payoffs. A positive (more pro-social) social environment, in our case, is characterized by fewer transfers and more willingness to penalize transfers. Holding other factors constant, if the treatment leads to a higher tendency to punish (destroy payoffs), we anticipate lower per-capita payoffs and lower inequality in treatment classrooms, representing an economically inefficient outcome. However, if norm compliance is more prevalent in treated classrooms, implying a reduced need for punishment, the effect of the treatment on per-capita earnings and inequality is ambiguous.

We combined both years to do our simulation analysis. We first simulate the benchmark per-capita payoff and Gini coefficient for each classroom, where no one transfers and punishes. Recall that this corresponds to the first stage of the game, where two players compete against each other, with only one of them winning. We then simulate the payoffs based on the decisions of transfers and punishment for each classroom. Figure 4 presents a comparison of aggregate economic indicators between treatment and control groups under (i) the ideal scenario of no transfers and no punishment and (ii) the realistic scenario of both transfers and punishment. Unsurprisingly, we detect no significant differences between treatment and control classrooms in the ideal world with no transfers and punishment for both grade levels. Under the possibility of third-party punishment, in a more cohesive positive culture (treated classrooms), anti-social behavior is less prevalent, leading to reduced instances of punishment. Therefore, despite their increased tendency to destroy wealth in case of unequal transfers, treated communities (classrooms) are wealthier than the less cohesive (control) communities (see Panel (a) in Figure 4). Consistent with the results in Table B5, treated senior classrooms received 3.55% higher per-capita payoffs (0.165 more per-capita gift points) than control classrooms despite their significantly higher willingness to destroy payoffs. Note, also, that there is no notable change in the Gini coefficient. We also observe 2.69% higher per-capita payoffs for treated juniors, but this effect does not reach statistical significance.

6.3 Treatment Effects on Social Support Networks

In a healthy school environment, we anticipate not only positive interactions among peers within the same grade but also supportive relationships between students in upper and lower grades. In disadvantaged schools, however, it is common to observe seniors abusing their power and mistreating juniors. When we collected our network data, we allowed students to

nominate friends and support providers from any classroom (including their own) and any grade (upper and/or lower). This broad elicitation was to assess whether the intervention generated new connections across classrooms and grade levels. As a backdrop to our analysis, we find about 78.4% (81.4%) of all friendship links and 77% (80.4%) of support links are within-grade links in the control group in year 1 (in year 2). Predictably, we found no effects on friendship links, and they are very well-formed at this age. We did not expect the intervention to have an effect on friendship ties, especially between juniors and seniors. However, we did expect the intervention to increase inter-grade support connections, particularly from the perspective of junior students. Table 7 presents the estimated treatment effects on the total number of support links (sum of in-degree and out-degree ties) directed to target subgroups. Panel 1 presents links directed to student-teachers, and Panel 2 presents links to student-teacher networks.

First, note that the average number of support ties between student-teachers and juniors is very low in the control group (0.193 and 0.244 in year 1 and year 2, respectively), as opposed to the number of ties between seniors (4.98 and 4.11 in year 1 and year 2, respectively). Given this low base, we estimate that the number of support links directed to student-teachers from juniors went up by 56% in year 1, and by 51.6% in year 2, and both estimates are statistically significant. As in previous outcomes, we can rule out the pure interaction mechanism for this outcome, as the equality between placebo and treatment is rejected for both years (see columns 3 and 7 in Panel 1). Interestingly, we observe positive treatment effects on the number of support links directed to student-teachers from other seniors, which rules out the possibility of increased senior support of juniors at the expense of support among seniors. Note, however, that while the placebo treatment did not increase junior-senior ties, it increased links directed to student-teachers from seniors. Panel 2 repeats the above analysis for student-teacher networks, where we examine links directed to seniors in student-teacher networks. Here, we also observe increased links from juniors, albeit only in year 2.

6.4 Treatment Effects on Perceived Social Environment

Our targeted survey questions provide additional evidence of the program's effect on the relational environment in the school. To further describe the school climate, we constructed a summary index of behavioral norms, using item response statements, such as "My schoolmates/classmates trust each other." We combined these items with expected anti-social be-

havior from classmates in the third-party punishment game, as the latter is also informative about perceived behavioral norms. We also constructed an index measuring perceived adult behavior, using items such as "adults respect me and care about my opinions" and "teachers treat me unfairly." Table 8 presents the estimated effects on these standardized indices where the control mean is normalized to zero. We estimate significant improvements in behavioral norms in both years 1 and 2 in the full sample. We also estimate significant effects on seniors' perception of adults in year 1. However, these effects seem to dissipate in year 2. As in other outcomes, these positive effects are driven by target subgroups. Zooming into target subgroups, Table 9 shows that perceived behavioral norms improved significantly among student-teachers and their friends. We also estimate significant improvements in these metrics among seniors outside of student-teacher networks.

7 Treatment Effect Heterogeneity and Potential Mechanisms

Entrusting socially influential and emotionally intelligent adolescents with the responsibility of improving their school's climate led to significant improvements in the relational climate of the school. Consistency of the estimated treatment effects across a large number of outcomes indicates that our targeting approach was successful, and the improvements primarily came from the groups we meant to target: student-teachers and their networks. Recall also that for most of our outcomes, we were able to rule out a mere interaction mechanism generated by the program delivery through our placebo arm. We claim that the content of the empowerment curriculum and the way it was delivered were responsible for generating the positive changes we estimate. By repeatedly deliberating and conveying the positive messages provided in the curriculum, student-teachers became empowered and felt responsible for their social environment. As a result, they became more pro-social, avoided actions that triggered disciplinary flagging, extended their support to their junior schoolmates, and improved their school's climate.

Our next question is then: What exactly did this intervention change in these adolescents? Our conjecture is that repeatedly conveying positive messages about how to make a positive impact on the social environment improves their socio-emotional well-being. To investigate whether our data support this conjecture, we examined various attributes that could be affected by the intervention, including attributes indicating empowerment, such as internal locus of control, mental well-being, self-worth, and sense of responsibility. We also explore

changes in perspective-taking, impulsivity, and sense of belonging, as the program content could also affect these attributes. While not a comprehensive list, these attributes together describe, at least partially, the markers of adolescent socio-emotional well-being.

Our findings, illustrated in Figure 5, reveal striking improvements in all these attributes among student-teachers and their immediate networks. Consistent with our main results, while we also estimate some positive effects among juniors, the most substantial improvements were consistently observed in the senior subgroups, particularly among student-teachers and their networks. For student-teachers, we estimate a 0.396 sd increase in internal locus of control, a 0.177 sd improvement in mental well-being, a 0.26 sd increase in perspective-taking ability, and a 0.294 sd reduction in impulsivity in year 1. We also estimate a significant 0.231 sd increase in the sense of belonging to the school and a 0.274 sd increase in the sense of responsibility for world issues. The effects are similar for student-teacher networks with slightly smaller sizes. Furthermore, we detected improvements in the internal locus of control and the sense of responsibility even among seniors outside the student-teachers' networks. We observe that these effects persist in year 2 for both student-teachers and their immediate networks.

We also detected a notable gender heterogeneity in treatment effects. Figure 6 presents treatment effects on all outcomes, separately for female and male student-teachers in both years. Figure 7 repeats the same figure for student-teacher networks. As it is clear from both figures, while we observe some positive effects on male student-teachers, such as disciplinary flagging, most effects are driven by female student-teachers and females in student-teacher networks. To understand why the program was more effective for female student-teachers, we explored our baseline data further. Firstly, the baseline willingness to be a student-teacher was significantly higher among female seniors than males (p-value=0.001). Figure A2 shows the differences in attributes that may be necessary for successful program implementation between female and male student teachers at baseline. Most notably, female student-teachers scored 0.353 sd higher than males in the emotional intelligence (RME) test. They also appear to have a higher sense of responsibility and belonging, higher perspective-taking, and internal locus of control but similar impulsivity and mental well-being as male student-teachers at baseline. All of these suggest that female student-teachers may have embraced the program and put more effort into implementing it than male student-teachers, explaining the heterogeneous treatment effects.

In addition to the intended positive effects, we explored potential unintended conse-

quences of the program. Concerns were raised about the demanding nature of being a student-teacher and possible crowding out of study time. Table 10 and 11 present math and Turkish test results. As can be seen, if anything, we see some improvement in math. However, this improvement is only in the first year. While we estimate a non-trivial treatment effect (0.042sd), the estimate does not reach statistical significance. Note, however, that we have significantly less power to detect a cognitive improvement in year 2. Perhaps more importantly, we implemented the same achievement tests in both years based on the grade-level national curriculum. However, in year 2, schools were closed for about six weeks due to the devastating earthquake that hit the region on February 6, 2023. The teachers informed us that they could only partially cover the required curriculum. Our second-year achievement results should be viewed in light of this caveat.

Finally, as another unintended effect, we explored the possibility that instead of acting responsibly, student-teachers may act in a manner to abuse the power given to them. We rule out this possibility by estimating the effect on an index constructed using item-response questions about power abuse and narcissistic tendencies. We estimate null effects for both years (p-values 0.988 and 0.285 in years 1 and 2, respectively).

Overall, our results suggest that targeting disadvantaged adolescents in an innovative way through an empowerment curriculum improved their socio-emotional well-being and their awareness of the world and opportunities around them. These changes, in turn, contributed to a more positive school environment characterized by reduced violence and increased peer support, all without adversely affecting academic achievement. Importantly, the positive impacts extended beyond the entrusted adolescents to their friendship networks and even to those outside their networks, amplifying the program's overall impact on the school climate.

8 Conclusion

We estimate the effectiveness of a behavioral program aimed at empowering socioeconomically disadvantaged adolescents to improve their social environment and socio-emotional well-being. The program involves selecting a number of emotionally intelligent, socially influential, yet slightly challenging senior students, labeling them as "student-teachers," and giving them the responsibility of delivering a specifically designed empowerment curriculum to their junior peers. The program was implemented first in the 2021-2022 academic year and then repeated in the 2022-2023 academic year in Diyarbakir, Turkey. The evaluation

study covered 65 middle schools, with 32 schools randomly assigned to treatment and 33 to control. To rule out a mere interaction mechanism, we further randomized the control schools, assigned 16 to placebo treatment, and left the remaining 17 as pure control.

Using a rich toolkit encompassing administrative records, an incentivized behavioral game, surveys, and cognitive tests, we found that this indirect targeting reduced disciplinary incidents and anti-social behavior while fostering supportive network ties between senior and junior students. The intervention also lowered the tolerance for anti-social behavior and enhanced the willingness to penalize such behavior. We also found that the intervention significantly improved the socio-emotional well-being of the targeted students, supporting our mechanism claims.

The program is highly cost-effective and has generated significant positive spillovers. The first-year implementation involved both grade 7 and 8 students and incurred $12,000 in printing costs and $3000 in distribution costs. In the second year, only grade 7 students were targeted, so the printing costs were $6500, and the distribution costs remained $3000. Calculating the most conservative cost-per-student, where we consider only student-teachers and ignore spillover effects, yields about $11.7 per student per year. When we include student-teacher networks in the calculation, the cost per student goes down to $2.4 per student per year. These values indicate that compared to programs such as CBT (about $267 per participant in Blattman et al. (2017)), MTO (counseling cost of $5071 per family who took up a voucher), and BAM (on average around $2046 per participant per year), the intervention we evaluate is highly cost-effective. Note that the costs are adjusted to 2023 prices.

One caveat about our study relates to the external validity of the results. The program was implemented in a low income region of a large middle-income country. One might be concerned that some of the facts we highlight regarding neighborhood disadvantages and school climate issues may not be relevant to contexts outside Turkey. However, there are two reasons why our results might be relevant beyond our setting. First, our approach to challenging adolescents is informed by the recent literature on adolescent development, which is unlikely to be country-specific. Second, adolescents are vulnerable to bad environmental influences in socioeconomically disadvantaged contexts. The behavioral challenges we encountered in our schools are likely similar to those faced in both developed and developing countries. Therefore, our study provides lessons that likely extend to similarly underprivileged settings.

References

Alan, S., Baysan, C., Gumren, M., and Kubilay, E. (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics*, 136(4):2147–2194.

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., Corekcioglu, G., Kaba, M., and Sutter, M. (2023). DP18465 Female Leadership and Workplace Climate. https://cepr.org/publications/dp18465.

Alan, S. and Ertac, S. (2018). Fostering Patience in the Classroom: Results from Randomized Educational Intervention. *Journal of Political Economy*, 126(5):1865–1911.

Alan, S. and Mumcu, I. (2023). DP17601 Nurturing Childhood Curiosity to Enhance Learning: Evidence from a Randomized Pedagogical Intervention. https://cepr.org/publications/dp17601.

Andor, M. A., Cox, J., Gerster, A., Price, M., Sommer, S., and Tomberg, L. (2022). *Locus of Control and Prosocial Behavior*. Number no. w30359 in NBER Working Paper Series. National Bureau of Economic Research, Cambridge, Mass.

Andrews, J. L., Ahmed, S. P., and Blakemore, S.-J. (2021). Navigating the Social Environment in Adolescence: The Role of Social Brain Development. *Biological Psychiatry*, 89(2):109–118.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, 38(7):813–822.

Bierhoff, H. W., Klein, R., and Kramp, P. (1991). Evidence for the Altruistic Personality from Data on Accident Research. *Journal of Personality*, 59(2):263–280.

Blakemore, S.-J. and Mills, K. L. (2014). Is Adolescence a Sensitive Period for Sociocultural Processing? *Annual Review of Psychology*, 65(1):187–207.

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.

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.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer Effects and Social Networks in Education. *The Review of Economic Studies*, 76(4):1239–1267.

Chetty, R. and Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4):855–902.

Chyn, E. (2018). Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children. *American Economic Review*, 108(10):3028–3056.

Chyn, E. and Katz, L. F. (2021). Neighborhoods Matter: Assessing the Evidence for Place Effects. *Journal of Economic Perspectives*, 35(4):197–222.

Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press.

Dahl, G. B. and Lochner, L. (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review*, 102(5):1927–1956.

Dahl, R. E. (2004). Adolescent Brain Development: A Period of Vulnerabilities and Opportunities. Keynote Address. *Annals of the New York Academy of Sciences*, 1021(1):1–22.

Dahl, R. E., Allen, N. B., Wilbrecht, L., and Suleiman, A. B. (2018). Importance of investing in adolescence from a developmental science perspective. *Nature*, 554(7693):441–450.

Damm, A. P. and Dustmann, C. (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review*, 104(6):1806–1832.

Dinarte-Diaz, L. and Egana-delSol, P. (2023). Preventing Violence in the Most Violent Contexts: Behavioral and Neurophysiological Evidence from El Salvador. *Journal of the European Economic Association*, page jvad068.

Dustmann, C., Mertz, M., and Okatenko, A. (2023). Neighbourhood Gangs, Crime Spillovers and Teenage Motherhood. *The Economic Journal*, 133(653):1901–1936.

Gradstein, M. and Justman, M. (2002). Education, Social Cohesion, and Economic Growth. *American Economic Review*, 92(4):1192–1204.

Guadalupe, M., Kinias, Z., and Schloderer, F. (2020). Individual Identity and Organizational Identification: Evidence from a Field Experiment. *AEA Papers and Proceedings*, 110:193–198.

Guryan, J., Christenson, S., Cureton, A., Lai, I., Ludwig, J., Schwarz, C., Shirey, E., and Turner, M. C. (2021). The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check & Connect Program. *Journal of Policy Analysis and Management*, 40(3):841–882.

Heller, S. B. (2014). Summer jobs reduce violence among disadvantaged youth. *Science*, 346(6214):1219–1223.

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.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Judge, T. A., Thoresen, C. J., Bono, J. E., and Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127(3):376–407.

Katz, L. F., Kling, J. R., and Liebman, J. B. (2001). Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment. *The Quarterly Journal of Economics*, 116(2):607–654.

Kesavayuth, D., Binh Tran, D., and Zikos, V. (2022). Locus of control and subjective well-being: Panel evidence from Australia. *PLoS ONE*, 17(8):e0272714.

Kiessling, L. and Norris, J. (2023). The Long-Run Effects of Peers on Mental Health. *The Economic Journal*, 133(649):281–322.

Kling, J., Liebman, J., and Katz, L. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Luthar, S. S. and McMahon, T. J. (1996). Peer reputation among inner-city adolescents: Structure and correlates. *Journal of Research on Adolescence*, 6(4):581–603.

Midlarsky, E. and Midlarsky, M. (1973). Some determinants of aiding under experimentally-induced stress. *Journal of Personality*, 41(2):305–327.

Mullainathan, S. and Washington, E. (2009). Sticking with Your Vote: Cognitive Dissonance and Political Attitudes. *American Economic Journal: Applied Economics*, 1(1):86–111.

Oreopoulos, P. (2003). The Long-Run Consequences of Living in a Poor Neighborhood*. *The Quarterly Journal of Economics*, 118(4):1533–1575.

Oreopoulos, P., Brown, R. S., and Lavecchia, A. M. (2017). Pathways to Education: An Integrated Approach to Helping At-Risk High School Students. *Journal of Political Economy*.

Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. *Journal of Applied Psychology*, 77(6):963–974.

Paluck, E. L., Shepherd, H., and Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571.

Putnam, R. (1993). The prosperous community: Social capital and public life. *The American Prospect*, 13(Spring, Vol. 4.).

Raven, J. C. and Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary Scales*, volume 759. Oxford pyschologists Press Oxford.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Resnjanskij, S., Ruhose, J., Wiederhold, S., Woessmann, L., and Wedel, K. (2023). Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor-Market Prospects. *Journal of Political Economy*.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(1):1–28.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116(2):681–704.

Sacerdote, B. (2011). Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier.

Schwardmann, P., Tripodi, E., and van der Weele, J. J. (2022). Self-Persuasion: Evidence from Field Experiments at International Debating Competitions. *American Economic Review*, 112(4):1118–1146.

Sharkey, P. (2010). The acute effect of local homicides on children's cognitive performance. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11733–11738.

Sorrenti, G., Zölitz, U., Ribeaud, D., and Eisner, M. (2020). The Causal Impact of Socio-Emotional Skills Training on Educational Success.

Srivastava, S. B., Goldberg, A., Manian, V. G., and Potts, C. (2018). Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations. *Management Science*, 64(3):1348–1364.

Steinberg, L. (2008). A Social Neuroscience Perspective on Adolescent Risk-Taking. *Developmental review : DR*, 28(1):78–106.

Thunfors, P. and Cornell, D. (2008). The Popularity of Middle School Bullies. *Journal of School Violence*, 7(1):65–82.

Wolff, S. B., Pescosolido, A. T., and Druskat, V. U. (2002). Emotional intelligence as the basis of leadership emergence in self-managing teams. *The Leadership Quarterly*, 13(5):505–522.

Yeager, D. S., Dahl, R. E., and Dweck, C. S. (2018). Why Interventions to Influence Adolescent Behavior Often Fail but Could Succeed. *Perspectives on Psychological Science*, 13(1):101–122.

Zárate, R. A. (2023). Uncovering Peer Effects in Social and Academic Skills. *American Economic Journal: Applied Economics*, 15(3):35–79.

Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The Review of Economics and Statistics*, 85(1):9–23.

Tables

Table 1: Comparison of Student-Teachers with Other Senior Students

	Non Student-Teachers Mean	Student-Teachers Mean	p-value [NST = ST]
Student Demographics:			
Male	0.511	0.516	0.505
Age (months)	153.295	152.880	0.069
Refugee	0.020	0.021	0.813
No. Siblings	3.619	3.818	0.696
Computer at Home	0.307	0.279	0.264
Internet at Home	0.603	0.588	0.067
Willingness University	0.951	0.974	0.000
Social Environment:			
Behavioral Norms	-0.047	-0.022	0.443
Perceived Adult Behavior	-0.010	0.051	0.166
Experienced Anti-social Behavior	-0.068	-0.211	0.000
Having a Friend	0.835	1.000	0.000
Friendship Ties (in-degree)	2.964	7.593	0.000
Popularity (in-degree)	1.907	8.352	0.000
Socio-Emotional Skills:			
Locus of Control	0.064	0.194	0.000
Mental Well-being	0.001	0.037	0.255
Perspective Taking	0.077	0.333	0.000
Impulsivity	0.052	0.072	0.615
Sense of Belonging	0.020	0.312	0.000
Sense of Responsibility	0.019	0.131	0.000
Cognitive Skills:			
Math Score	-0.097	0.031	0.000
Turkish Score	-0.141	0.049	0.000
Fluid IQ (Raven)	0.175	0.414	0.000
Emotional Intelligence (RMET)	0.113	0.508	0.000

Reported statistics use the baseline data collected in year 1. All cognitive test scores and survey measures are standardized to have a zero mean and a unit standard deviation. Reported p-values are obtained by controlling for district fixed effects and clustering standard errors at the school level (unit of randomization).

Table 2: Balance at Baseline - Full Sample in Year 1

	N	Control Mean	Placebo Mean	Treatment Mean	p-value [T = C]	p-value [T = P]	p-value [C = P]
Student Demographics:							
Male	22847	0.500	0.512	0.514	0.163	0.763	0.356
Age (months)	22871	142.933	142.646	142.306	0.401	0.514	0.853
Refugee	22441	0.019	0.028	0.027	0.078	0.844	0.173
No. Siblings	17853	3.847	3.869	3.744	0.839	0.697	0.839
Computer at Home	17855	0.266	0.252	0.288	0.748	0.906	0.901
Internet at Home	17855	0.550	0.511	0.572	0.756	0.684	0.877
Willingness University	17855	0.939	0.944	0.937	0.562	0.229	0.559
Social Environment:							
Behavioral Norms	22871	0.019	0.032	0.060	0.310	0.618	0.775
Perceived Adult Behavior	22871	3.946	3.964	3.938	0.566	0.323	0.533
Experienced Anti-social Behavior	17832	0.018	0.049	-0.001	0.756	0.541	0.697
Having a Friend	22443	0.810	0.780	0.807	0.792	0.119	0.144
Friendship Ties (in-degree)	22443	2.805	2.650	2.818	0.989	0.128	0.246
Socio-Emotional Skills:							
Locus of Control	22871	-0.025	-0.016	-0.005	0.785	0.895	0.733
Mental Well-being	22871	-0.002	0.004	0.027	0.351	0.673	0.874
Perspective Taking	22871	-0.032	-0.068	-0.020	0.961	0.530	0.573
Impulsivity	22871	-0.005	-0.003	-0.025	0.458	0.401	0.821
Sense of Belonging	22871	-0.007	-0.001	0.026	0.378	0.582	0.810
Sense of Responsibility	22871	-0.014	-0.064	-0.067	0.153	0.742	0.362
Cognitive Skills:							
Math Score	22871	0.015	0.033	0.104	0.267	0.685	0.696
Turkish Score	22871	0.023	0.025	0.104	0.344	0.612	0.860
Fluid IQ (Raven)	22871	-0.064	-0.089	-0.025	0.730	0.738	0.952
Emotional Intelligence (RMET)	22871	-0.048	-0.078	-0.035	0.862	0.764	0.873

The table presents the balance of student-level variables using the baseline data collected in year 1. All cognitive test scores and survey measures are standardized to have a zero mean and a unit standard deviation. Letter C indicates the pure control group, P and T placebo, and treatment groups, respectively. Reported p-values are obtained by controlling for district fixed effects and clustering standard errors at the school level (unit of randomization).

Table 3: Treatment Effects on Disciplinary Flagging

Panel 1: Full Sample

	Year 1		Year 2	
	Disciplinary Flagging		Disciplinary Flagging	
	Benchmark Model	Full Model	Benchmark Model	Full Model
Treatment	-0.011**	-0.010**	-0.017***	-0.011**
	(0.004)	(0.005)	(0.006)	(0.005)
Placebo		0.002		0.015
		(0.006)		(0.012)
Control Mean	0.021	0.018	0.033	0.026
p-value [TR = P]	.	0.029	.	0.019
Wild Bootstrap p-value	0.017	0.065	0.008	0.051
Observations	27018	27018	27841	27841

Panel 2: Juniors (Grades 5 and 6)

	Year 1		Year 2	
	Disciplinary Flagging		Disciplinary Flagging	
	Benchmark Model	Full Model	Benchmark Model	Full Model
Treatment	-0.007	-0.009	-0.018**	-0.011
	(0.005)	(0.007)	(0.008)	(0.007)
Placebo		-0.004		0.014
		(0.007)		(0.014)
Control Mean	0.020	0.021	0.035	0.028
p-value [TR = P]	.	0.372	.	0.059
Wild Bootstrap p-value	0.211	0.262	0.020	0.101
Observations	10797	10797	10818	10818

Panel 3: Seniors (Grades 7 and 8)

	Year 1		Year 2	
	Disciplinary Flagging		Disciplinary Flagging	
	Benchmark Model	Full Model	Benchmark Model	Full Model
Treatment	-0.013***	-0.010*	-0.017***	-0.012**
	(0.004)	(0.005)	(0.006)	(0.006)
Placebo		0.008		0.015
		(0.007)		(0.012)
Control Mean	0.021	0.016	0.033	0.025
p-value [TR = P]	.	0.004	.	0.022
Wild Bootstrap p-value	0.005	0.050	0.016	0.100
Observations	16221	16221	17023	17023

The table presents the estimated treatment effects on official disciplinary flagging records at the time of the endline. The binary dependent variable equals one for students flagged as having extreme behavioral issues and zero otherwise. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for gender, age in months, Raven's score, Eye's Test score, school type fixed effects, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 4: Treatment Effects on Disciplinary Flagging - Senior Subgroups

Panel 1: Student-Teachers

	Year 1		Year 2	
	Disciplinary Flagging		Disciplinary Flagging	
	Benchmark Model	Full Model	Benchmark Model	Full Model
Treatment	-0.041***	-0.024**	-0.033**	-0.035**
	(0.010)	(0.011)	(0.014)	(0.016)
Placebo		0.039**		-0.003
		(0.017)		(0.021)
Control Mean	0.058	0.035	0.060	0.056
p-value [TR = P]	.	0.000	.	0.095
Wild Bootstrap p-value	0.000	0.062	0.029	0.057
Observations	1269	1269	565	565

Panel 2: Seniors in Student-Teacher Networks

	Year 1		Year 2	
	Disciplinary Flagging		Disciplinary Flagging	
	Benchmark Model	Full Model	Benchmark Model	Full Model
Treatment	-0.019***	-0.020**	-0.017**	-0.014*
	(0.007)	(0.009)	(0.007)	(0.008)
Placebo		-0.003		0.005
		(0.011)		(0.011)
Control Mean	0.030	0.029	0.036	0.034
p-value [TR = P]	.	0.041	.	0.061
Wild Bootstrap p-value	0.015	0.044	0.032	0.085
Observations	5258	5258	2381	2381

Panel 3: Seniors outside Student-Teacher Networks

	Year 1		Year 2	
	Disciplinary Flagging		Disciplinary Flagging	
	Benchmark Model	Full Model	Benchmark Model	Full Model
Treatment	-0.008*	-0.006	-0.014*	-0.011
	(0.004)	(0.005)	(0.007)	(0.008)
Placebo		0.005		0.006
		(0.006)		(0.012)
Control Mean	0.018	0.014	0.036	0.032
p-value [TR = P]	.	0.062	.	0.103
Wild Bootstrap p-value	0.076	0.272	0.080	0.223
Observations	4893	4893	4510	4510

The table presents the estimated treatment effects on official disciplinary flagging records at the time of the endline for senior subgroups. The binary dependent variable equals one for students flagged as having extreme behavioral issues and zero otherwise. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for gender, age in months, Raven's score, Eye's Test score, school type fixed effects, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 5: Treatment Effects on Anti-Social behavior and Punishment

Panel 1: Full Sample

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment
Treatment	-0.030	0.154**	-0.020	0.156**	-0.160***	0.399***	-0.135***	0.416***
	(0.028)	(0.059)	(0.037)	(0.064)	(0.032)	(0.056)	(0.038)	(0.057)
Placebo			0.022	0.006			0.057	0.038
			(0.036)	(0.083)			(0.038)	(0.060)
Control Mean	1.113	1.489	1.097	1.497	1.105	1.302	1.078	1.309
p-value [TR = P]	.	.	0.131	0.066	.	.	0.000	0.000
Wild Bootstrap p-value	0.323	0.020	0.622	0.035	0.000	0.000	0.003	0.000
Observations	16348	16344	16348	16344	15660	15657	15660	15657

Panel 2: Juniors (Grades 5 and 6)

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment
Treatment	-0.036	0.160**	-0.014	0.134	-0.079**	0.343***	-0.044	0.334***
	(0.040)	(0.073)	(0.053)	(0.082)	(0.038)	(0.081)	(0.045)	(0.092)
Placebo			0.047	-0.055			0.076*	-0.021
			(0.052)	(0.121)			(0.041)	(0.081)
Control Mean	1.115	1.402	1.081	1.426	1.074	1.180	1.030	1.196
p-value [TR = P]	.	.	0.143	0.081	.	.	0.002	0.000
Wild Bootstrap p-value	0.426	0.042	0.822	0.154	0.066	0.000	0.381	0.004
Observations	8114	8113	8114	8113	7733	7733	7733	7733

Panel 3: Seniors (Grades 7 and 8)

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment
Treatment	-0.028	0.156**	-0.030	0.185**	-0.238***	0.454***	-0.219***	0.498***
	(0.031)	(0.077)	(0.038)	(0.077)	(0.041)	(0.068)	(0.047)	(0.075)
Placebo			-0.006	0.068			0.046	0.106
			(0.043)	(0.099)			(0.053)	(0.080)
Control Mean	1.111	1.574	1.111	1.563	1.135	1.420	1.121	1.410
p-value [TR = P]	.	.	0.513	0.274	.	.	0.000	0.000
Wild Bootstrap p-value	0.438	0.074	0.462	0.041	0.000	0.000	0.000	0.000
Observations	8234	8231	8234	8231	7927	7924	7927	7924

The table presents the estimated treatment effects on the decisions in the third-party punishment game. The dependent variables are i) the number of tokens transferred and ii) the punishment cost incurred in the case of asymmetric transfers. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, school type fixed effects, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 6: Treatment Effects on Anti-Social Behavior and Punishment - Senior Subgroups

Panel 1: Student-Teachers

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment
Treatment	-0.097	0.028	-0.081	0.008	-0.243**	0.483**	-0.263**	0.399*
	(0.059)	(0.121)	(0.072)	(0.135)	(0.092)	(0.209)	(0.117)	(0.235)
Placebo			0.034	-0.044			-0.039	-0.170
			(0.079)	(0.185)			(0.148)	(0.288)
Control Mean	1.142	1.952	1.120	1.993	1.157	1.623	1.170	1.734
p-value [TR = P]	.	.	0.108	0.763	.	.	0.065	0.041
Wild Bootstrap p-value	0.148	0.832	0.339	0.958	0.015	0.043	0.040	0.134
Observations	989	989	989	989	353	353	353	353

Panel 2: Seniors in Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment
Treatment	-0.014	0.164**	-0.021	0.182**	-0.246***	0.376***	-0.209***	0.450***
	(0.032)	(0.075)	(0.039)	(0.082)	(0.064)	(0.134)	(0.074)	(0.159)
Placebo			-0.016	0.044			0.082	0.166
			(0.049)	(0.100)			(0.081)	(0.139)
Control Mean	1.100	1.569	1.114	1.553	1.119	1.642	1.065	1.581
p-value [TR = P]	.	.	0.909	0.167	.	.	0.000	0.044
Wild Bootstrap p-value	0.674	0.052	0.639	0.064	0.000	0.023	0.014	0.024
Observations	3966	3965	3966	3965	1354	1353	1354	1353

Panel 3: Seniors outside Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment	Transfers	Punishment
Treatment	-0.038	0.203*	-0.047	0.239*	-0.235***	0.373***	-0.197***	0.396***
	(0.041)	(0.116)	(0.052)	(0.124)	(0.056)	(0.100)	(0.058)	(0.103)
Placebo			-0.019	0.082			0.094*	0.057
			(0.045)	(0.129)			(0.053)	(0.115)
Control Mean	1.120	1.397	1.109	1.394	1.149	1.351	1.118	1.340
p-value [TR = P]	.	.	0.483	0.264	.	.	0.000	0.012
Wild Bootstrap p-value	0.404	0.115	0.419	0.104	0.001	0.003	0.003	0.002
Observations	2630	2629	2630	2629	2293	2291	2293	2291

The table presents the estimated treatment effects on the decisions in the third-party punishment game for senior subgroups. The dependent variables are i) the number of tokens transferred and ii) the punishment cost incurred in the case of asymmetric transfers. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, school type fixed effects, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 7: Treatment Effects on Support Networks

Panel 1: Support Ties directed to Student-Teachers

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	From Juniors	From Seniors	From Juniors	From Seniors	From Juniors	From Seniors	From Juniors	From Seniors
Treatment	0.108***	0.251	0.123***	0.507*	0.126*	-0.107	0.128*	0.625*
	(0.038)	(0.246)	(0.043)	(0.275)	(0.065)	(0.330)	(0.076)	(0.354)
Placebo			0.034	0.585**			0.005	1.534***
			(0.054)	(0.288)			(0.069)	(0.525)
Control Mean	0.193	4.979	0.197	5.026	0.244	4.110	0.235	3.469
p-value [TR = P]	.	.	0.076	0.788	.	.	0.088	0.047
Wild Bootstrap p-value	0.010	0.379	0.017	0.141	0.098	0.783	0.170	0.142
Observations	1269	1269	1269	1269	565	565	565	565

Panel 2: Support Ties directed to Seniors in Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	From Juniors	From Seniors	From Juniors	From Seniors	From Juniors	From Seniors	From Juniors	From Seniors
Treatment	-0.026	0.008	-0.029	0.105	0.075**	-0.106	0.069**	0.021
	(0.019)	(0.142)	(0.021)	(0.154)	(0.032)	(0.213)	(0.030)	(0.267)
Placebo			-0.008	0.241			-0.014	0.273
			(0.025)	(0.158)			(0.036)	(0.267)
Control Mean	0.196	3.834	0.185	3.810	0.158	3.087	0.144	2.759
p-value [TR = P]	.	.	0.401	0.436	.	.	0.058	0.296
Wild Bootstrap p-value	0.252	0.955	0.247	0.560	0.052	0.680	0.037	0.957
Observations	5242	5242	5242	5242	2168	2168	2168	2168

The table presents treatment effects on the number of support ties formed within the school. Panel 1 presents ties directed to student teachers and Panel 2 to student-teacher networks. The dependent variable in columns 1 and 3 is the total number of support ties formed between student-teachers and juniors (directed from juniors to student-teachers). The dependent variable in columns 2 and 4 is the total number of support ties formed between student-teachers and other seniors (directed from seniors to student-teachers). Panel 2 replicates Panel 1 for student-teacher networks. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline outcomes, gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, experimenter dummies, school type fixed effects, school size, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 8: Treatment Effects on Perceived Social Environment

Panel 1: Full Sample

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior
Treatment	0.049*	0.054*	0.062*	0.069**	0.068**	0.001	0.093**	0.021
	(0.029)	(0.030)	(0.032)	(0.032)	(0.031)	(0.025)	(0.037)	(0.029)
Placebo			0.030	0.035			0.064	0.051
			(0.031)	(0.030)			(0.047)	(0.038)
p-value [TR = P]	.	.	0.335	0.343	.	.	0.473	0.373
Wild Bootstrap p-value	0.140	0.128	0.106	0.058	0.045	0.968	0.035	0.510
Observations	16289	16309	16289	16309	15623	15693	15623	15693

Panel 2: Juniors (Grades 5 and 6)

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior
Treatment	-0.017	0.008	-0.011	0.031	-0.013	0.001	-0.014	0.022
	(0.036)	(0.050)	(0.042)	(0.052)	(0.032)	(0.030)	(0.037)	(0.037)
Placebo			0.014	0.050			-0.001	0.051
			(0.042)	(0.043)			(0.047)	(0.040)
p-value [TR = P]	.	.	0.558	0.726	.	.	0.772	0.416
Wild Bootstrap p-value	0.667	0.899	0.820	0.627	0.727	0.977	0.736	0.592
Observations	8080	8092	8080	8092	7713	7758	7713	7758

Panel 3: Seniors (Grades 7 and 8)

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior
Treatment	0.116***	0.083***	0.135***	0.092***	0.147***	0.003	0.191***	0.022
	(0.035)	(0.022)	(0.039)	(0.028)	(0.047)	(0.032)	(0.055)	(0.039)
Placebo			0.046	0.024			0.119*	0.051
			(0.045)	(0.039)			(0.068)	(0.051)
p-value [TR = P]	.	.	0.050	0.032	.	.	0.211	0.482
Wild Bootstrap p-value	0.003	0.000	0.005	0.001	0.005	0.965	0.006	0.611
Observations	8209	8217	8209	8217	7910	7935	7910	7935

The table presents treatment effects on the perceived social environment within the school. Outcomes are standardized to have a mean zero and standard deviation of a unit. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline outcomes, gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, school type fixed effects, school size and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 9: Treatment Effects on Perceived Social Environment - Senior Subgroups

Panel 1: Student-Teachers

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior
Treatment	0.100*	0.004	0.146**	-0.028	0.335**	-0.013	0.401**	-0.052
	(0.050)	(0.054)	(0.055)	(0.057)	(0.144)	(0.090)	(0.170)	(0.123)
Placebo			0.101	-0.072			0.139	-0.080
			(0.071)	(0.081)			(0.169)	(0.145)
p-value [TR = P]	.	.	0.506	0.575	.	.	0.110	0.796
Wild Bootstrap p-value	0.068	0.950	0.020	0.652	0.034	0.895	0.043	0.724
Observations	987	989	987	989	353	356	353	356

Panel 2: Seniors in Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior
Treatment	0.112***	0.117***	0.127***	0.129***	0.147*	0.007	0.228**	0.099
	(0.037)	(0.030)	(0.042)	(0.038)	(0.081)	(0.066)	(0.097)	(0.062)
Placebo			0.037	0.030			0.189*	0.214***
			(0.049)	(0.041)			(0.102)	(0.072)
p-value [TR = P]	.	.	0.055	0.004	.	.	0.660	0.156
Wild Bootstrap p-value	0.008	0.000	0.010	0.001	0.112	0.921	0.066	0.144
Observations	3958	3960	3958	3960	1352	1359	1352	1359

Panel 3: Seniors outside Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior	Behavioral Norms	Perceived Adult Behavior
Treatment	0.146***	0.082**	0.133**	0.095**	0.125**	-0.012	0.170***	0.001
	(0.049)	(0.039)	(0.059)	(0.045)	(0.048)	(0.046)	(0.057)	(0.050)
Placebo			-0.031	0.030			0.121**	0.035
			(0.067)	(0.056)			(0.056)	(0.068)
p-value [TR = P]	.	.	0.006	0.222	.	.	0.334	0.597
Wild Bootstrap p-value	0.010	0.059	0.055	0.045	0.016	0.803	0.010	0.989
Observations	2624	2627	2624	2627	2287	2290	2287	2290

The table presents treatment effects on the perceived social environment within the school for senior subgroups. Outcomes are standardized to have a mean zero and standard deviation of a unit. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline outcomes, gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, school type fixed effects, school size and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 10: Treatment Effects on Academic Outcomes

Panel 1: Full Sample

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Math	Turkish	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.063**	0.026	0.069**	0.023	-0.005	0.030	-0.010	0.042
	(0.029)	(0.031)	(0.034)	(0.040)	(0.033)	(0.032)	(0.044)	(0.044)
Placebo			0.013	-0.007			-0.010	0.027
			(0.045)	(0.049)			(0.054)	(0.054)
p-value [TR = P]	.	.	0.162	0.444	.	.	0.998	0.714
Wild Bootstrap p-value	0.044	0.473	0.076	0.633	0.885	0.393	0.866	0.403
Observations	16352	16378	16352	16378	15726	15747	15726	15747

Panel 2: Juniors (Grades 5 and 6)

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Math	Turkish	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.050	0.046	0.040	0.033	-0.005	0.032	0.002	0.051
	(0.036)	(0.045)	(0.051)	(0.061)	(0.047)	(0.045)	(0.059)	(0.058)
Placebo			-0.023	-0.028			0.015	0.041
			(0.060)	(0.072)			(0.070)	(0.077)
p-value [TR = P]	.	.	0.127	0.241	.	.	0.819	0.870
Wild Bootstrap p-value	0.192	0.371	0.508	0.651	0.936	0.501	0.979	0.442
Observations	8108	8127	8108	8127	7767	7791	7767	7791

Panel 3: Seniors (Grades 7 and 8)

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Math	Turkish	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.084**	0.005	0.106***	0.014	0.001	0.030	-0.012	0.039
	(0.035)	(0.025)	(0.036)	(0.028)	(0.031)	(0.027)	(0.039)	(0.036)
Placebo			0.052	0.019			-0.031	0.020
			(0.050)	(0.038)			(0.048)	(0.038)
p-value [TR = P]	.	.	0.281	0.872	.	.	0.636	0.466
Wild Bootstrap p-value	0.033	0.849	0.010	0.645	0.985	0.311	0.797	0.354
Observations	8244	8251	8244	8251	7959	7956	7959	7956

The table presents treatment effects on academic test scores. The dependent variables are standardized math and Turkish verbal test scores. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline outcomes, gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, school type fixed effects, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 11: Treatment Effects on Academic Outcomes - Senior Subgroups

Panel 1: Student-Teachers

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Math	Turkish	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.123**	0.046	0.151**	0.085	0.042	0.027	-0.031	0.026
	(0.061)	(0.057)	(0.071)	(0.071)	(0.106)	(0.092)	(0.115)	(0.122)
Placebo			0.060	0.084			-0.148	-0.002
			(0.091)	(0.070)			(0.124)	(0.125)
p-value [TR = P]	.	.	0.272	0.989	.	.	0.361	0.785
Wild Bootstrap p-value	0.071	0.496	0.049	0.296	0.700	0.795	0.789	0.832
Observations	990	991	990	991	356	356	356	356

Panel 2: Seniors in Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Math	Turkish	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.097**	0.016	0.113***	0.007	-0.076	0.098*	-0.054	0.128**
	(0.042)	(0.028)	(0.039)	(0.031)	(0.051)	(0.050)	(0.059)	(0.055)
Placebo			0.038	-0.023			0.050	0.068
			(0.061)	(0.046)			(0.075)	(0.063)
p-value [TR = P]	.	.	0.250	0.482	.	.	0.132	0.337
Wild Bootstrap p-value	0.031	0.582	0.011	0.862	0.169	0.090	0.402	0.041
Observations	3971	3975	3971	3975	1368	1362	1368	1362

Panel 3: Seniors outside Student-Teacher Networks

	Year 1				Year 2			
	Benchmark Model		Full Model		Benchmark Model		Full Model	
	Math	Turkish	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.056	-0.000	0.105**	0.008	0.062	0.032	0.039	0.040
	(0.045)	(0.037)	(0.051)	(0.039)	(0.049)	(0.031)	(0.058)	(0.036)
Placebo			0.114*	0.019			-0.059	0.021
			(0.066)	(0.046)			(0.066)	(0.042)
p-value [TR = P]	.	.	0.883	0.825	.	.	0.107	0.606
Wild Bootstrap p-value	0.296	0.999	0.073	0.825	0.235	0.357	0.552	0.302
Observations	2634	2635	2634	2635	2295	2298	2295	2298

The table presents treatment effects on academic test scores for senior subgroups. The dependent variables are standardized math and Turkish verbal test scores. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline outcomes, gender, age in months, Raven's score, Eye's Test score, class size, share of boys in class, school type fixed effects, and district fixed effects. Standard errors are clustered at the school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figures

Figure 1: Intervention Timeline

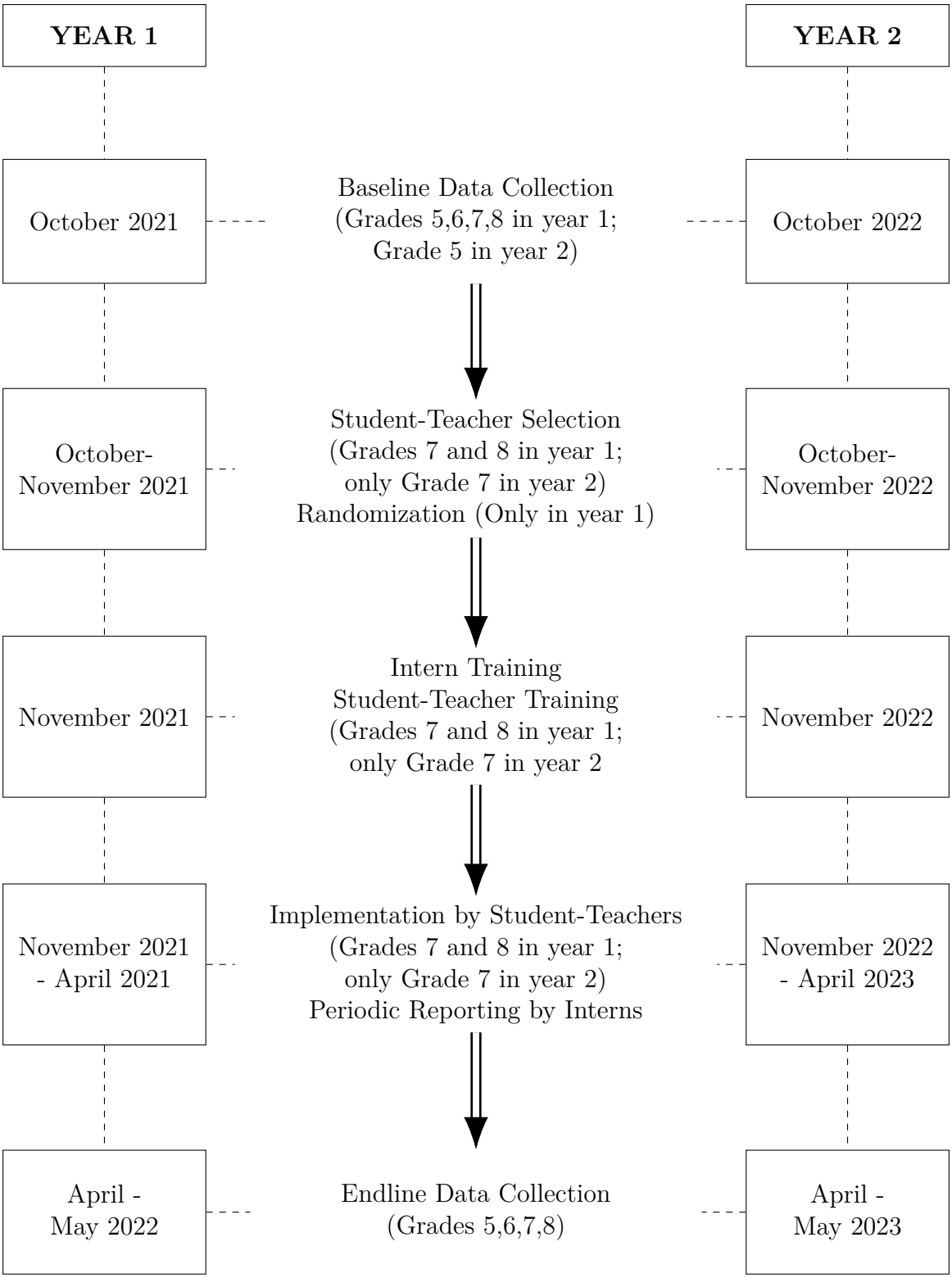
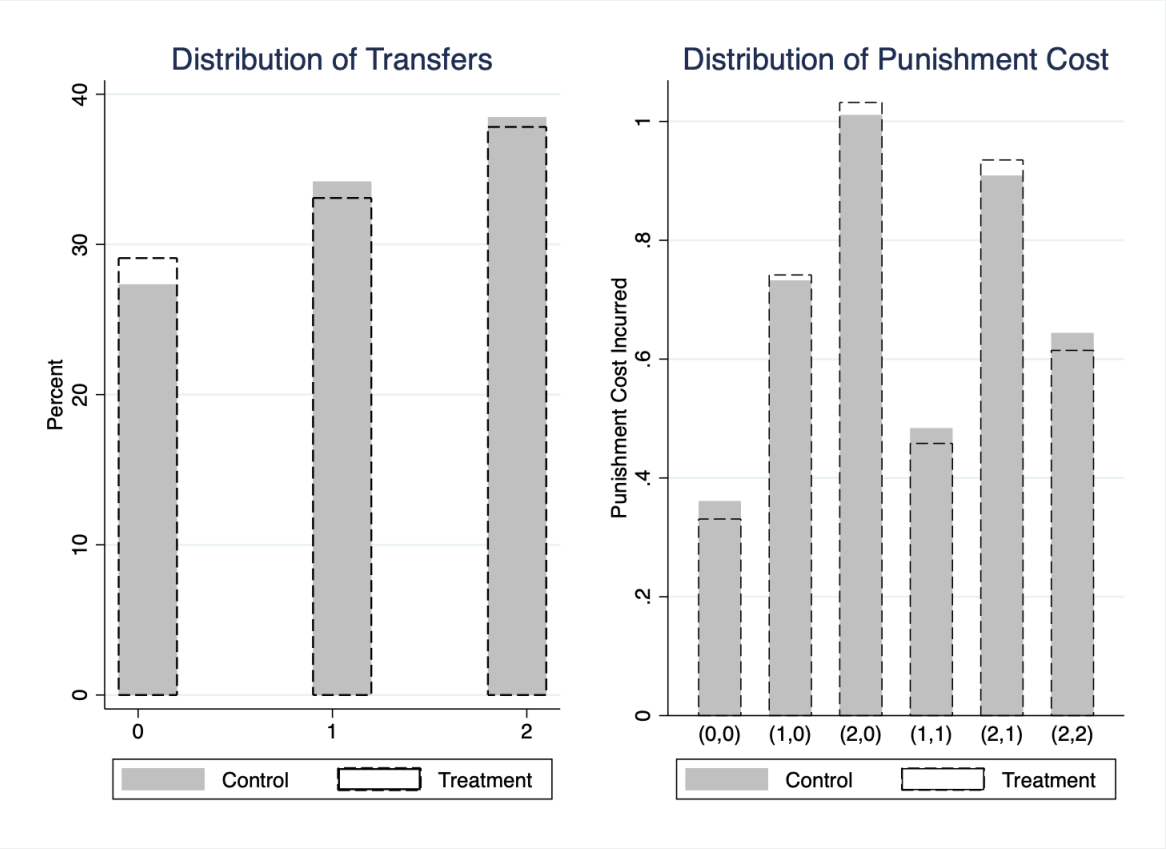
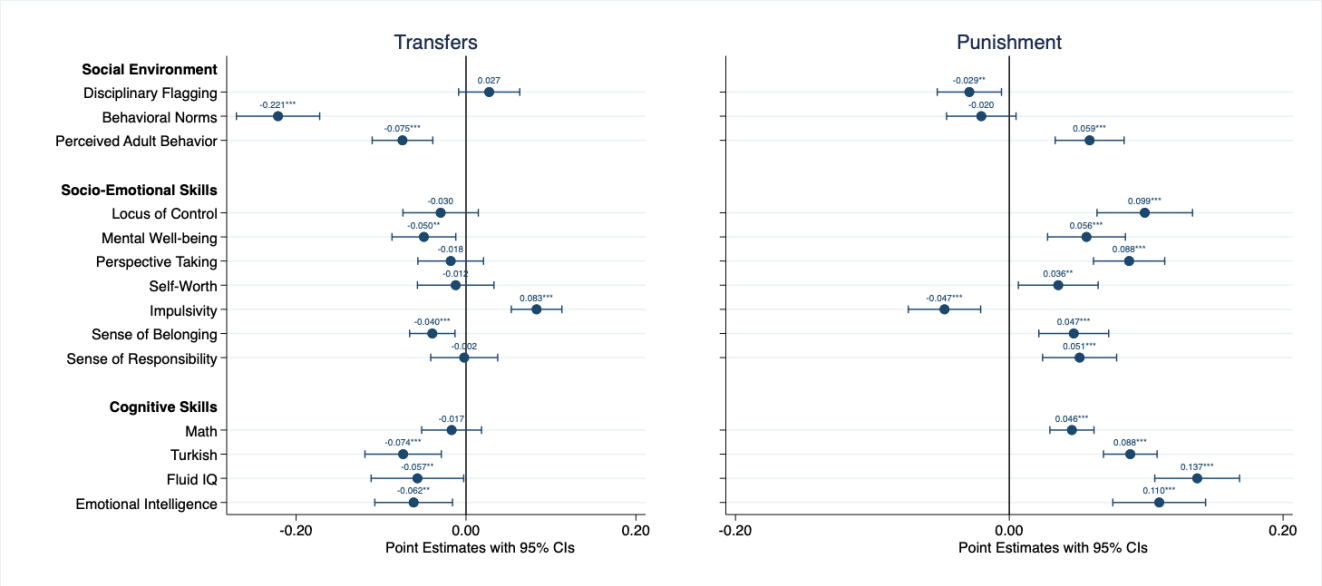


Figure 2: Distribution of Decisions in the Third-Party Punishment Game



The figure depicts the distribution of transfers (0, 1, and 2) and punishment cost incurred for each transfer scenario (0,0), (1,0), (2,0), (1,1), (2,1), and (2,2) separately for control and treatment. The analysis uses the data collected in year 1.

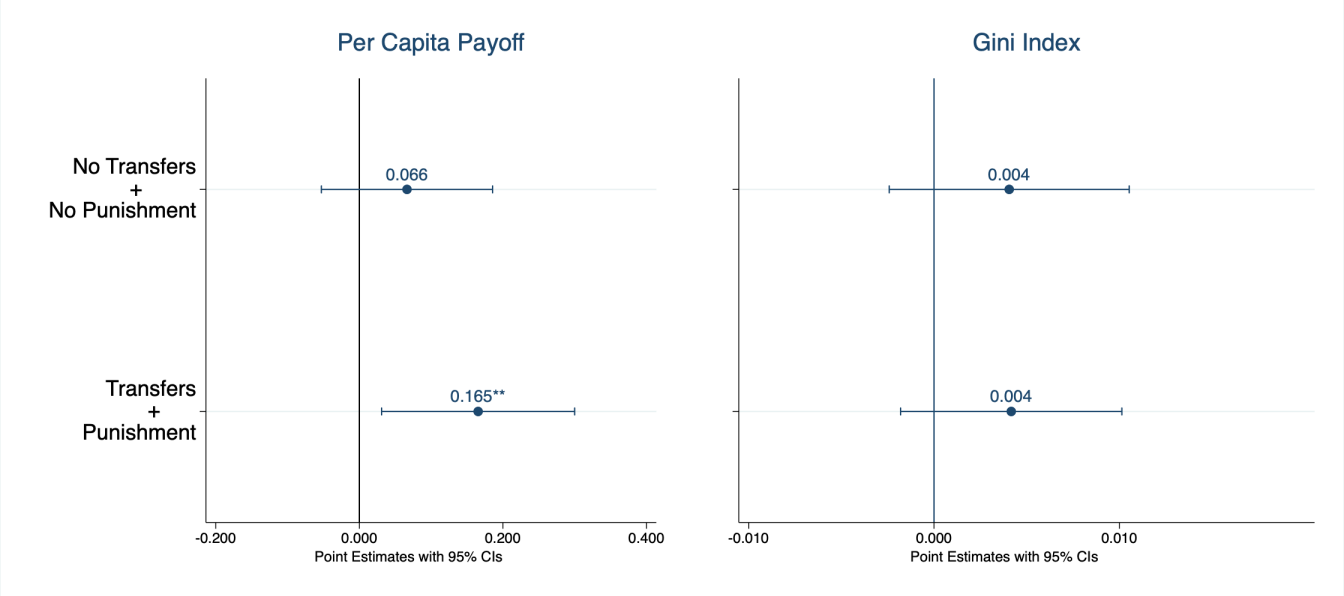
Figure 3: Predictive Validity of the Third-Party Punishment Game



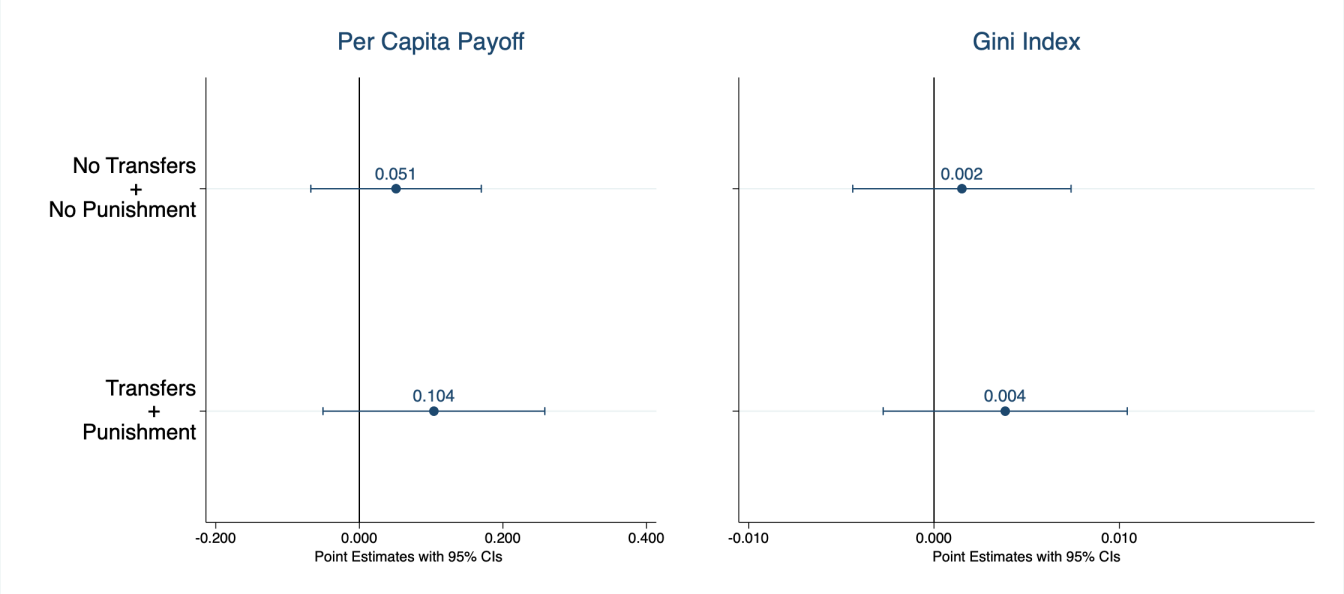
The figure depicts the OLS coefficients from the regressions of transfers and punishment cost on indicators of perceived social environment, socio-emotional well-being, and cognitive skills. The analysis combines both years of the data. Regressions control for baseline values of the corresponding outcome if available, gender, age in months, Raven's score, Eye's Test score, school type fixed effects, and district fixed effects. Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 4: Treatment Effects on Per-Capita Classroom Wealth (Payoffs) and Inequality

(a) Seniors

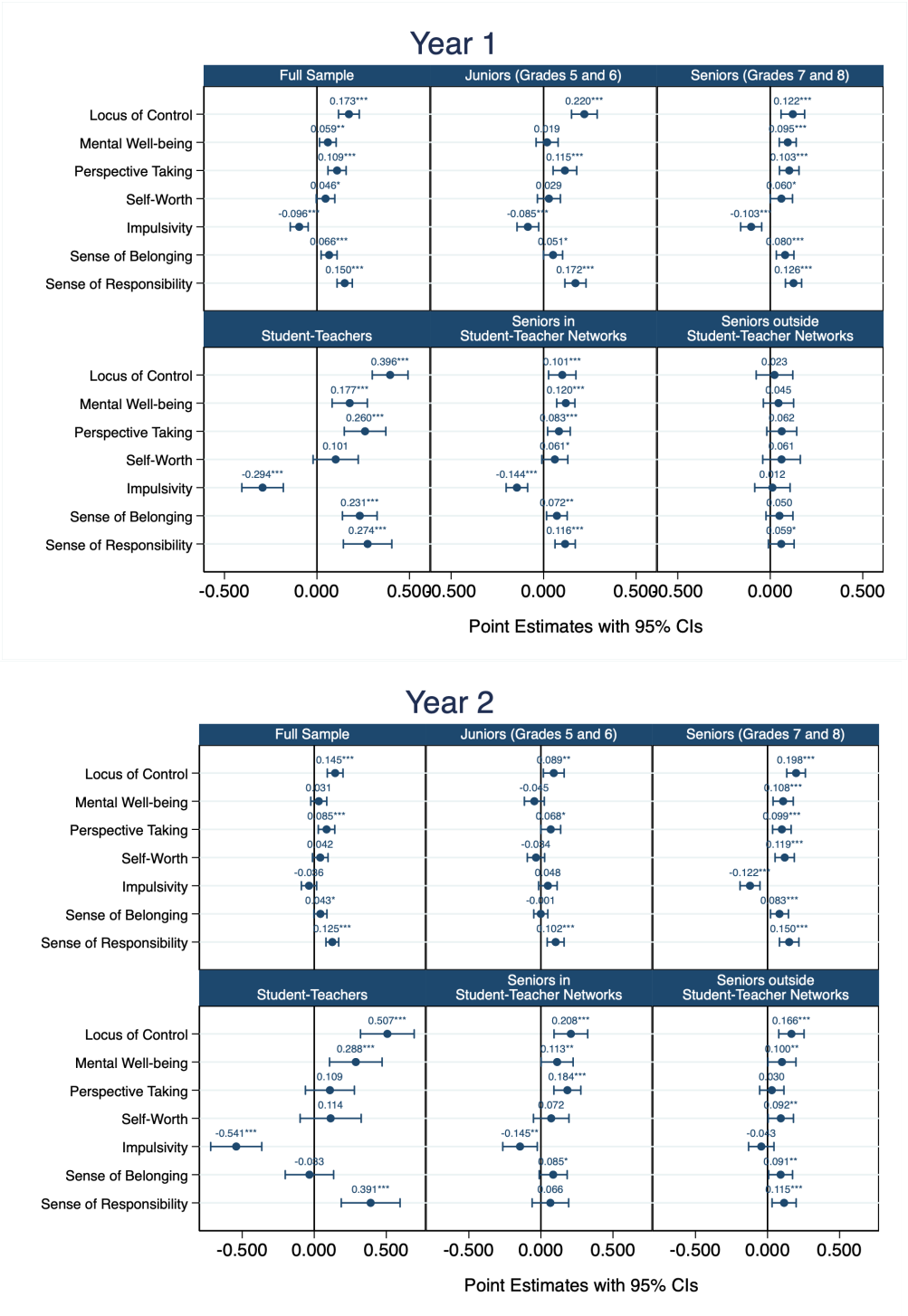


(b) Juniors



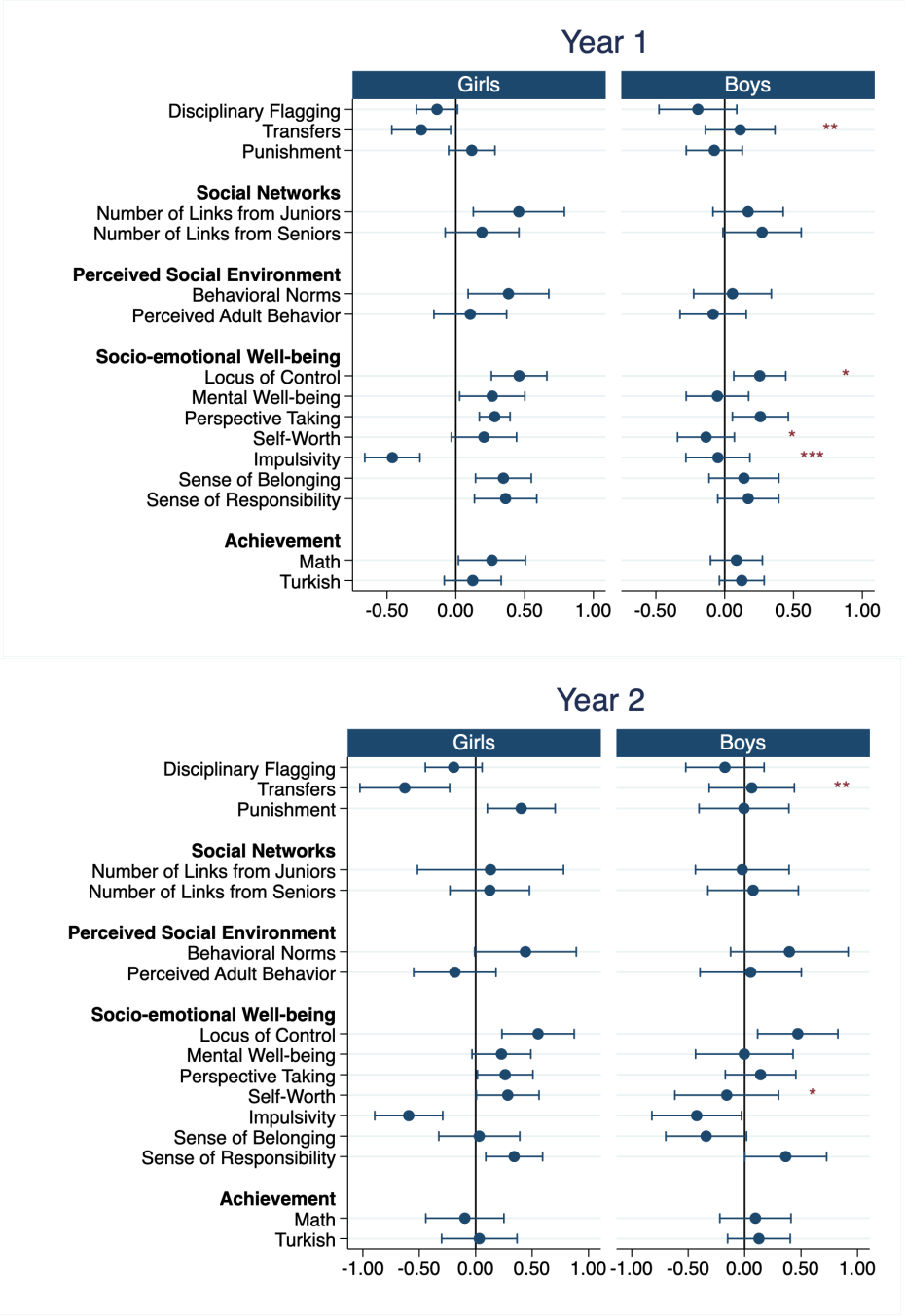
The figure depicts the estimated treatment effects on simulated per-capita classroom wealth (payoffs) and classroom Gini coefficient. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization). The vertical lines indicate a treatment effect of zero. The dependent variables are classroom per-capita payoff and Gini coefficient obtained from the third-party punishment game in the scenario where (i) no transfers and punishment are allowed and (ii) transfers and punishments are allowed. Regressions control for school type fixed effects, school size, and district fixed effects. Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 5: Treatment Effects on Socio-emotional Well-being



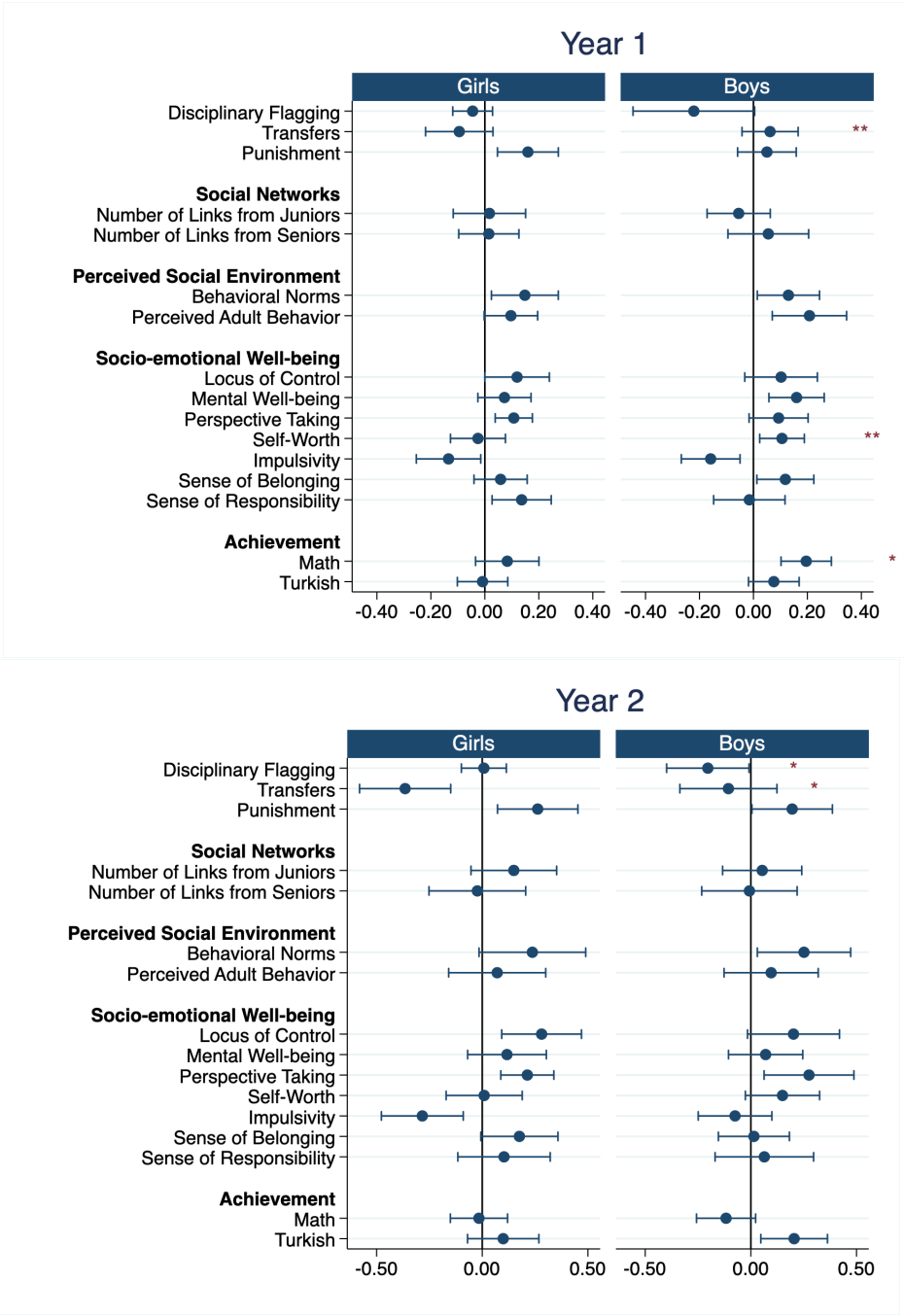
The figure depicts the estimated treatment effects on socioemotional well-being. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization). Dependent variables are standardized factors constructed using relevant item-response questions, so all coefficient estimates are in sd units. Regressions control for baseline values of the corresponding outcome when available, gender, age in months, Raven's score, Eye's Test score, school type fixed effects, and district fixed effects.

Figure 6: Gender Heterogeneity in Treatment Effects: Student-Teachers



The figure depicts the estimated treatment effects on all outcomes for female and male student-teachers. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization). Asterisks indicate the significance of the difference in estimates at the 1% ***, 5% **, and 10% * levels. Regressions control for baseline values of the corresponding outcome when available, gender, age in months, Raven's score, Eye's Test score, school type fixed effects, and district fixed effects.

Figure 7: Gender Heterogeneity: Seniors in Student-Teacher Networks



The figure depicts the estimated treatment effects on all outcomes for female and male student-teachers' networks. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization). Asterisks indicate the significance of the difference in estimates at the 1% ***, 5% **, and 10% * levels. Regressions control for baseline values of the corresponding outcome when available, gender, age in months, Raven's score, Eye's Test score, school type fixed effects, and district fixed effects.

Appendix

Tables

Table A1: Correction for Multiple Hypothesis Testing

	Original	Romano Wolf
Full Sample:		
Disciplinary Flagging	0.004	0.034
Experimental Outcomes	0.000	0.000
Social Environment Outcomes	0.038	0.034
Socio-emotional Outcomes	0.000	0.000
Juniors:		
Disciplinary Flagging	0.055	0.160
Experimental Outcomes	0.013	0.026
Social Environment Outcomes	0.989	0.790
Socio-emotional Outcomes	0.010	0.032
Seniors:		
Disciplinary Flagging	0.008	0.032
Experimental Outcomes	0.000	0.000
Social Environment Outcomes	0.009	0.016
Socio-emotional Outcomes	0.000	0.000

The table provides p-values corrected for multiple hypothesis testing using Romano-Wolf algorithm. The analysis combines both years of the data. We construct indices for each outcome group by taking the average of relevant outcomes as follows: (i) experimental outcomes: transfers and cost of punishment in the third-party punishment game, (ii) social environment: behavioral norms and perceived adult behavior, (iii) socio-emotional outcomes: Locus of control, mental well-being, perspective taking, impulsivity, sense of belonging, sense of responsibility. The number of replications is set to 500.

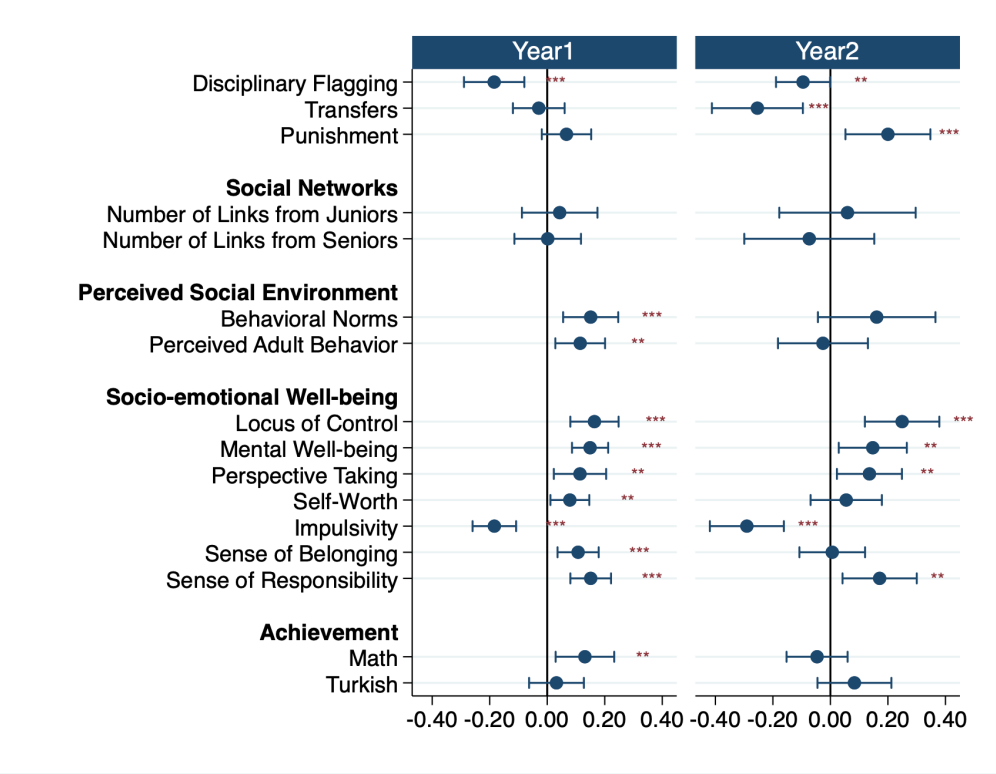
Table A2: Correction for Multiple Hypothesis Testing - Senior Subgroups

	Original	Romano Wolf
Student-Teachers:		
Disciplinary Flagging	0.007	0.012
Experimental Outcomes	0.003	0.010
Social Environment Outcomes	0.038	0.020
Network Outcomes	0.000	0.000
Socio-emotional Outcomes	0.008	0.018
Seniors in Student-Teacher Networks:		
Disciplinary Flagging	0.000	0.000
Experimental Outcomes	0.009	0.014
Social Environment Outcomes	0.000	0.000
Network Outcomes	0.047	0.186
Socio-emotional Outcomes	0.095	0.186
Seniors outside Student-Teacher Networks:		
Disciplinary Flagging	0.004	0.008
Experimental Outcomes	0.000	0.000
Social Environment Outcomes	0.000	0.000
Socio-emotional Outcomes	0.000	0.000

The table provides p-values corrected for multiple hypothesis testing using Romano-Wolf algorithm. The analysis combines both years of the data. We construct indices for each outcome group by taking the average of relevant outcomes as follows: (i) experimental outcomes: transfers and cost of punishment in the third-party punishment game, (ii) social environment: behavioral norms, and perceived adult behavior, (iii) network outcomes (only for student-teachers and their networks): total number of support ties with juniors and total number of support ties with seniors, (iv) socio-emotional outcomes: Locus of control, mental well-being, perspective taking, impulsivity, sense of belonging, sense of responsibility. The number of replications is set to 500.

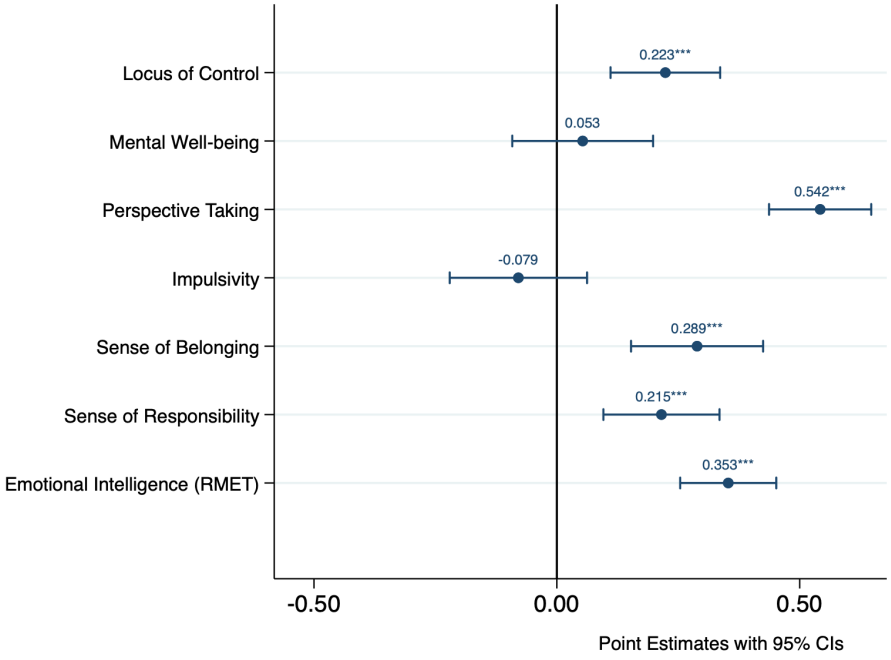
Figures

Figure A1: Treatment Effects without Covariates (Student-teachers+student-teacher networks)



The figure plots the estimated treatment effects and their 95% confidence intervals for targeted senior subgroups (student-teachers and their networks). All effect sizes are in standard deviation units. Only district fixed effects are used as covariates. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure A2: Baseline Comparison of Male and Female Student-Teachers



The figure plots the coefficient on female dummy on respective baseline characteristics and their 95% confidence intervals for the sample in year 1. All cognitive tests and survey measures are standardized. Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Online Appendix

I Intervention Content and Activity Examples

Figure B1: "Our Future-Our Dream" Curriculum Overview

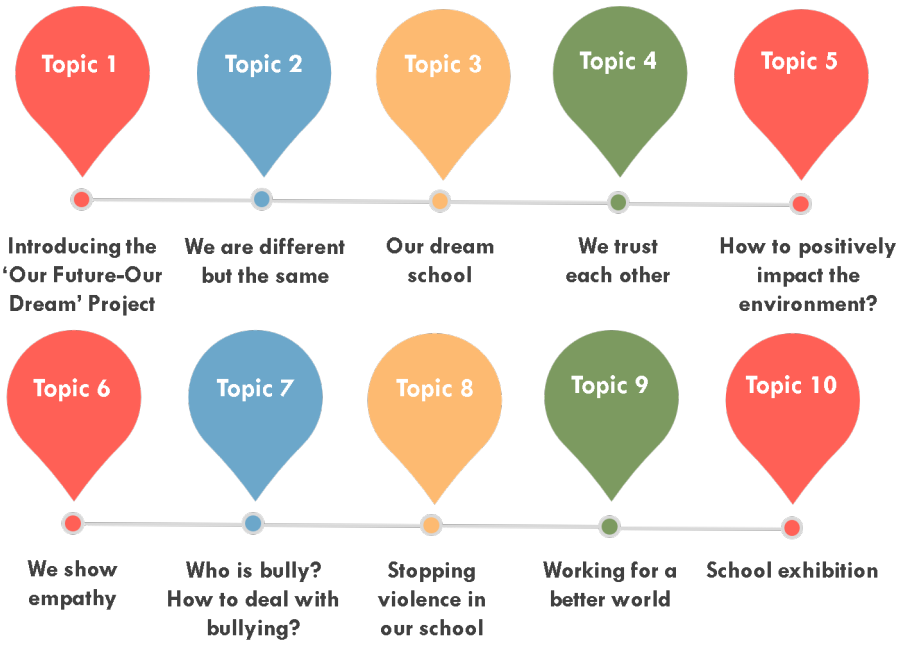
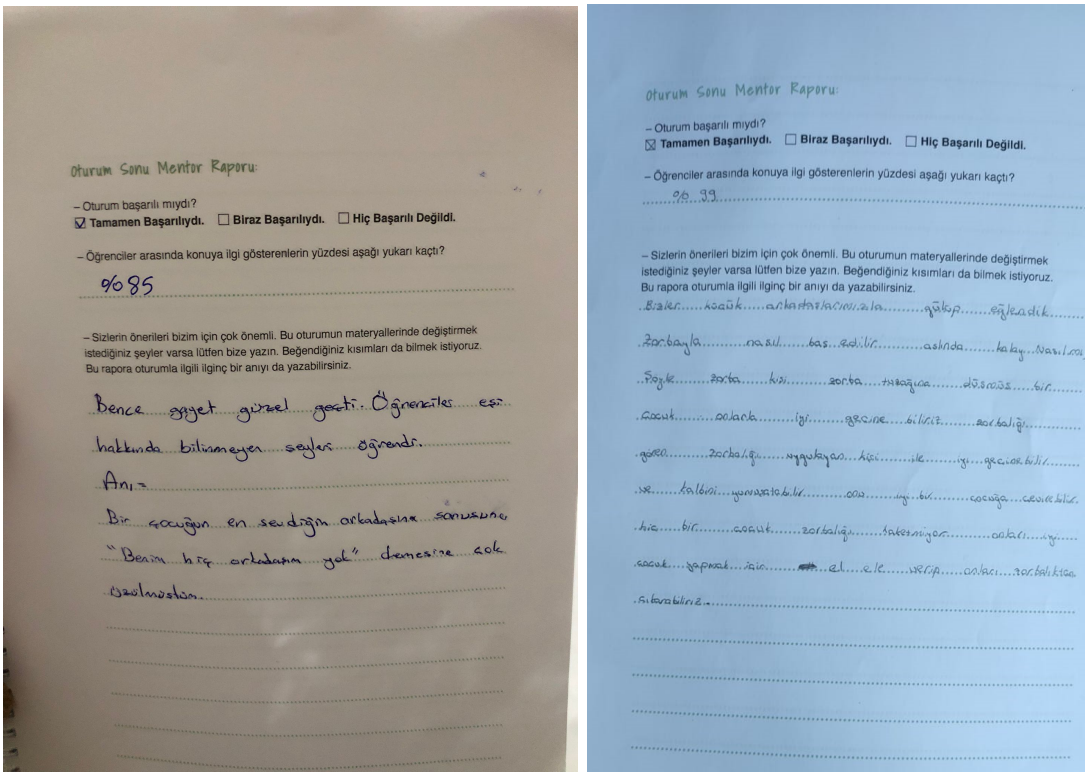


Table B1: "Our Future-Our Dream" Curriculum

TOPIC 1: INTRODUCING THE 'OUR FUTURE-OUR DREAM' PROJECT **Purpose:** Introducing student-teachers and let them announce the project to their junior classrooms. **Materials (2 weeks):** Slides, Movie, Activity, Booklet	**TOPIC 2: WE ARE DIFFERENT BUT THE SAME** **Purpose:** Conveying to junior students our shared similarities. Students learn that all individuals share similarities, they feel the pain and happiness in the same way. **Materials (2 weeks):** Activity, Booklet, Poster
TOPIC 3: OUR DREAM SCHOOL **Purpose:** Fostering a positive school environment and healthy peer relationships. Learning about one's power to transform their school. **Materials (2 weeks):** Slides, Activity, Animated Movie, Booklet, Poster preparations	**TOPIC 4: HOW TO POSITIVELY IMPACT THE ENVIRONMENT** **Purpose:** Teaching students about positive and negative impacts one can make to their social and physical environment **Materials (2 weeks):** Activity, Booklet
TOPIC 5: WE TRUST EACH OTHER **Purpose:** Understanding the importance of trust for healthy social relations **Materials (2 weeks):** Activity, Animated Movie, Booklet, Poster preparations	**TOPIC 6: WE SHOW EMPATHY** **Purpose:** Teaching students the concept of empathy and how to recognize empathetic concern **Materials (1 week):** Activity, Booklet, Poster
TOPIC 7: WHO IS THE BULLY? HOW TO DEAL WITH BULLYING? **Purpose:** Examining what may lie beneath violent and anti-social behavior. Psychology of a bully. **Materials (2 weeks):** Activity, Animated Movie, Booklet	**TOPIC 8: STOPPING VIOLENCE IN OUR SCHOOL** **Purpose:** Achieving a violence-free school, understanding the world-wide problem of peer violence and its consequences. **Materials (2 weeks):** Activity, Animated Movie, Booklet, Poster
TOPIC 9: WORKING FOR A BETTER WORLD **Purpose:** Understanding the importance of working toward cleaner, conflict-free, tolerant and sustainable world. Recognizing one's power to achieve many goals. **Materials (2 weeks):** Activity, Booklet, Poster	**TOPIC 10: EXHIBITION** **Purpose:** Preparing an exhibition using materials created during the academic year. Dissemination of project outputs to school. **Materials (1 week):** Exhibition Materials (Drawings, Posters, etc.)

Figure B2: Session Report Book (Student-Teachers)



Oturum Sonu Mentor Raporu:

– Oturum başarılı mıydı?
☑ Tamamen Başarılıydı. ☐ Biraz Başarılıydı. ☐ Hiç Başarılı Değildi.

– Öğrenciler arasında konuya ilgi gösterenlerin yüzdesi aşağı yukarı kaçtı?
%85

– Sizlerin önerileri bizim için çok önemli. Bu oturumun materyallerinde değiştirmek istediğiniz şeyler varsa lütfen bize yazın. Beğendiğiniz kısımları da bilmek istiyoruz. Bu rapora oturumla ilgili ilginç bir anıyı da yazabilirsiniz.

Bence gayet güzel geçti. Öğrencilerle eşi hakkında bilinmeyen şeyler öğrendi.

Anı:

Bir çocuğun en sevdiğin arkadaşına sarılarak "Benim hiç arkadaşım yok" demesine çok üzülmüştüm.

Oturum Sonu Mentor Raporu:

– Oturum başarılı mıydı?
☑ Tamamen Başarılıydı. ☐ Biraz Başarılıydı. ☐ Hiç Başarılı Değildi.

– Öğrenciler arasında konuya ilgi gösterenlerin yüzdesi aşağı yukarı kaçtı?
%99

– Sizlerin önerileri bizim için çok önemli. Bu oturumun materyallerinde değiştirmek istediğiniz şeyler varsa lütfen bize yazın. Beğendiğiniz kısımları da bilmek istiyoruz. Bu rapora oturumla ilgili ilginç bir anıyı da yazabilirsiniz.

Bizler küçük arkadaşlarımızla oğlup eğlendik. Zorbayla nasıl baş edilir aslında kalp Nasıl Söyle zorba kişi zorba tutuğuna düşmüş bir çocuk onlarla iyi geçine bilirit zorbalığı göre zorbalığı uygulayan kişi ile iyi geçinebilir ve kalbini yumuşatabilir. Son uyu bir çocuğa cesaretler hiç bir çocuk zorbalığı kalesmiyor onları yapmak için el ele verip onları zorbalıktan sakınabilirsiniz.

Figure B3: Examples of Children Activities

Figure B4: Booklet Covers

II Placebo Content and Activity Examples

Figure B5: Placebo Activity Booklet

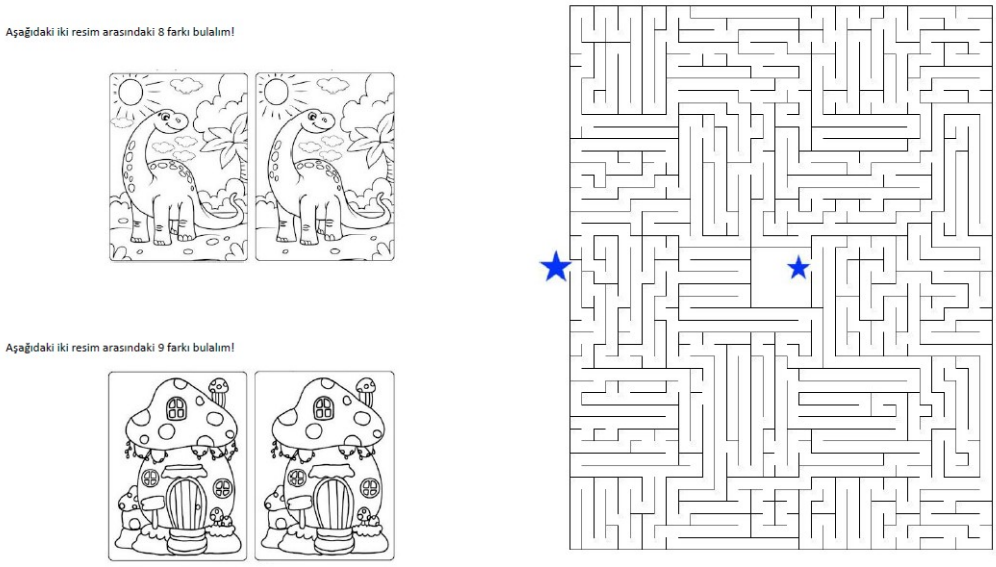


Figure B6: Examples of Children Activities (Placebo)

III Additional Tables

Table B2: Balance at Baseline - Juniors (Grades 5 and 6) in Year 1

	N	Control Mean	Placebo Mean	Treatment Mean	p-value [T = C]	p-value [T = P]	p-value [C = P]
Student Demographics:							
Male	10792	0.500	0.524	0.508	0.532	0.247	0.134
Age (months)	10806	130.732	131.273	130.492	0.511	0.263	0.516
Refugee	10539	0.023	0.030	0.030	0.141	0.893	0.208
No. Siblings	8519	3.838	3.891	3.804	0.990	0.729	0.730
Computer at Home	8520	0.264	0.247	0.290	0.650	0.657	0.966
Internet at Home	8520	0.518	0.512	0.558	0.515	0.771	0.789
Willingness University	8520	0.924	0.934	0.926	0.779	0.284	0.230
Social Environment:							
Behavioral Norms	10806	0.103	0.143	0.138	0.590	0.821	0.535
Perceived Adult Behavior	10806	3.984	3.977	3.934	0.232	0.386	0.950
Experienced Anti-social Behavior	8507	0.100	0.151	0.099	0.894	0.729	0.674
Having a Friend	10540	0.747	0.727	0.762	0.463	0.054	0.368
Friendship Ties (in-degree)	10540	2.012	1.977	2.089	0.368	0.113	0.583
Socio-Emotional Skills:							
Locus of Control	10806	-0.132	-0.107	-0.092	0.613	0.965	0.654
Mental Well-being	10806	0.044	0.017	0.044	0.973	0.731	0.724
Perspective Taking	10806	-0.176	-0.196	-0.150	0.739	0.525	0.755
Impulsivity	10806	-0.094	-0.087	-0.116	0.629	0.548	0.835
Sense of Belonging	10806	-0.034	-0.039	-0.003	0.458	0.498	0.956
Sense of Responsibility	10806	-0.097	-0.125	-0.151	0.301	0.478	0.798
Cognitive Skills:							
Math Score	10806	0.275	0.239	0.320	0.749	0.646	0.854
Turkish Score	10806	0.297	0.248	0.366	0.643	0.493	0.776
Fluid IQ (Raven)	10806	-0.262	-0.282	-0.247	0.997	0.930	0.928
Emotional Intelligence (RMET)	10806	-0.227	-0.239	-0.241	0.773	0.763	0.957

The table presents the balance of student-level variables for juniors (grades 5 and 6) using the baseline data collected in year 1. All cognitive test scores and survey measures are standardized to have a zero mean and a unit standard deviation. Letter C indicates the pure control group, P and T placebo and treatment groups, respectively. Reported p-values are obtained by controlling for district fixed effects and clustering standard errors at the school level (unit of randomization).

Table B3: Balance at Baseline - Seniors (Grades 7 and 8) in Year 1

	N	Control Mean	Placebo Mean	Treatment Mean	p-value [T = C]	p-value [T = P]	p-value [C = P]
Student Demographics:							
Male	12055	0.500	0.500	0.519	0.116	0.088	0.941
Age (months)	12065	153.268	153.559	152.871	0.646	0.229	0.618
Refugee	11902	0.015	0.025	0.024	0.089	0.806	0.232
No. Siblings	9334	3.855	3.847	3.689	0.653	0.673	0.955
Computer at Home	9335	0.267	0.258	0.287	0.889	0.866	0.797
Internet at Home	9335	0.576	0.509	0.585	0.989	0.638	0.650
Willingness University	9335	0.952	0.955	0.948	0.186	0.329	0.824
Social Environment:							
Behavioral Norms	12065	-0.052	-0.074	-0.010	0.401	0.124	0.659
Perceived Adult Behavior	12065	3.914	3.952	3.941	0.405	0.814	0.284
Experienced Anti-social Behavior	9325	-0.051	-0.052	-0.094	0.311	0.386	0.910
Having a Friend	11903	0.864	0.831	0.847	0.333	0.465	0.141
Friendship Ties (in-degree)	11903	3.469	3.288	3.463	0.836	0.230	0.256
Socio-Emotional Skills:							
Locus of Control	12065	0.067	0.071	0.073	0.922	0.805	0.923
Mental Well-being	12065	-0.041	-0.009	0.012	0.142	0.602	0.498
Perspective Taking	12065	0.091	0.055	0.096	0.890	0.638	0.583
Impulsivity	12065	0.070	0.078	0.056	0.674	0.294	0.617
Sense of Belonging	12065	0.016	0.036	0.052	0.511	0.739	0.705
Sense of Responsibility	12065	0.057	-0.006	0.008	0.160	0.786	0.134
Cognitive Skills:							
Math Score	12065	-0.205	-0.164	-0.090	0.025	0.597	0.375
Turkish Score	12065	-0.210	-0.189	-0.130	0.135	0.679	0.597
Fluid IQ (Raven)	12065	0.103	0.096	0.173	0.409	0.654	0.890
Emotional Intelligence (RMET)	12065	0.104	0.076	0.149	0.448	0.498	0.878

The table presents the balance of student-level variables for seniors (grades 7 and 8) using the baseline data collected in year 1. All cognitive test scores and survey measures are standardized to have a zero mean and a unit standard deviation. Letter C indicates the pure control group, P and T placebo and treatment groups, respectively. Reported p-values are obtained by controlling for district fixed effects and clustering standard errors at the school level (unit of randomization).

Table B4: Balance at Baseline - Student-Teachers and Seniors in Student-Teacher Networks in Year 1

	N	Control Mean	Placebo Mean	Treatment Mean	p-value [T = C]	p-value [T = P]	p-value [C = P]
Student Demographics:							
Male	6527	0.495	0.511	0.510	0.533	0.826	0.466
Age (months)	6527	152.239	152.445	152.121	0.783	0.303	0.565
Refugee	6527	0.014	0.020	0.022	0.121	0.508	0.535
No. Siblings	5769	3.801	3.839	3.684	0.804	0.617	0.794
Computer at Home	5769	0.266	0.251	0.290	0.855	0.969	0.921
Internet at Home	5769	0.580	0.489	0.594	0.940	0.338	0.378
Willingness University	5769	0.962	0.959	0.960	0.229	0.916	0.277
Social Environment:							
Behavioral Norms	6527	-0.033	-0.062	0.028	0.235	0.095	0.716
Perceived Adult Behavior	6527	3.922	3.939	3.956	0.445	0.307	0.934
Experienced Anti-social Behavior	5766	-0.088	-0.096	-0.128	0.410	0.496	0.981
Having a Friend	6527	0.959	0.951	0.954	0.584	0.411	0.199
Friendship Ties (in-degree)	6527	4.497	4.360	4.487	0.923	0.220	0.276
Socio-Emotional Skills:							
Locus of Control	6527	0.093	0.089	0.102	0.860	0.952	0.907
Mental Well-being	6527	-0.011	-0.047	0.027	0.488	0.114	0.255
Perspective Taking	6527	0.173	0.127	0.160	0.545	0.868	0.491
Impulsivity	6527	0.071	0.074	0.063	0.857	0.533	0.729
Sense of Belonging	6527	0.106	0.085	0.122	0.988	0.631	0.704
Sense of Responsibility	6527	0.092	0.044	0.052	0.206	0.880	0.229
Cognitive Skills:							
Math Score	6527	-0.205	-0.196	-0.062	0.041	0.206	0.663
Turkish Score	6527	-0.201	-0.168	-0.104	0.112	0.668	0.496
Fluid IQ (Raven)	6527	0.162	0.177	0.257	0.324	0.651	0.743
Emotional Intelligence (RMET)	6527	0.122	0.097	0.189	0.402	0.342	0.823

The table presents the balance of student-level variables for student-teachers and their networks using the baseline data collected in year 1. All cognitive test scores and survey measures are standardized to have a zero mean and a unit standard deviation. Letter C indicates the pure control group, P and T placebo and treatment groups, respectively. Reported p-values are obtained by controlling for district fixed effects and clustering standard errors at the school level (unit of randomization).

Table B5: Treatment Effects on Simulated Classroom Wealth (Payoffs) and Inequality

Panel 1: Full Sample

	Benchmark Model		Full Model	
	Per Capita Payoff	Gini Index	Per Capita Payoff	Gini Index
Treatment	0.143**	0.004	0.179**	0.007*
	(0.065)	(0.003)	(0.078)	(0.003)
Placebo			0.090	0.006*
			(0.108)	(0.004)
Control Mean	4.393	0.314	4.393	0.314
p-value [TR = P]	.	.	0.346	0.957
Wild Bootstrap p-value	0.054	0.187	0.058	0.119
Observations	1057	1057	1057	1057

Panel 2: Juniors (Grades 5 and 6)

	Benchmark Model		Full Model	
	Per Capita Payoff	Gini Index	Per Capita Payoff	Gini Index
Treatment	0.104	0.004	0.138*	0.006
	(0.077)	(0.003)	(0.082)	(0.004)
Placebo			0.078	0.005
			(0.139)	(0.004)
Control Mean	3.863	0.299	3.854	0.299
p-value [TR = P]	.	.	0.640	0.835
Wild Bootstrap p-value	0.224	0.258	0.142	0.152
Observations	336	336	336	336

Panel 3: Seniors (Grade 7 and 8)

	Benchmark Model		Full Model	
	Per Capita Payoff	Gini Index	Per Capita Payoff	Gini Index
Treatment	0.165**	0.004	0.201**	0.007*
	(0.067)	(0.003)	(0.081)	(0.004)
Placebo			0.095	0.007*
			(0.109)	(0.004)
Control Mean	4.644	0.322	4.668	0.321
p-value [TR = P]	.	.	0.257	0.888
Wild Bootstrap p-value	0.034	0.229	0.046	0.120
Observations	721	721	721	721

The table presents the treatment effects on the simulated classroom-level per capita payoff and Gini index obtained from the third-party punishment game in the presence of transfers and punishment. The analysis combines both years of the data. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for school type fixed effects, and district fixed effects. Standard errors are clustered at school level and reported in parentheses. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

IV Survey Inventories

Table B6: Student Survey Inventory - I

Inventory	Items
4-point likert scale: completely agree, agree, disagree, completely disagree	
Locus of Control	We can fix the bad things in our lives.
	I believe that I have control over the things that happen to me.
	My successes and failures are all mine.
	It is our job to fix the bad things in our school.
	We can make this school the best in the neighborhood if we work together.
	We can not realize the things we want to do for our school because adults would not let us.
Mental Well-Being	I have bad habits.
	I often get into fights and discussions.
	I feel lonely.
	I damage others' belongings.
	I get along with other kids.
	I think nobody likes me.
	I often think that everyone is against me and trying to hurt me.
	I feel inconsequential.
	I prefer to be alone rather than being with others.
	I feel unhappy.
Perspective Taking	I can put myself in someone else's shoes and understand how they feel.
	I try to understand how others feel.
	My friends talk to me about their problems.
	I can tell if a friend of mine is upset.
Self-Worth	I am proud of myself and my accomplishments.
	Sometimes I feel like I can not get anything right.
	Sometimes I feel like I am not worthy of anything.
	I generally think that I am an unsuccessful person.
Impulsivity	I tend to say the first thing that comes to my mind.
	I pay attention to the rules while playing games.
	I can control my temper when there is a fight.
	I find it difficult to sit still during lectures.
	I have a very low temper.
Sense of Belonging	My schoolmates are like my family.
	My classmates do not notice when I do not show up to school. They are never worried about me.
	I do not feel belong in my school and my classroom.
	My classmates always support me.

	My schoolmates do not care about me.

Table B7: Student Survey Inventory - II

Inventory	Items
5-point frequency scale: never, rarely, sometimes, often, always	
Sense of Responsibility	I think about serious global problems such as war and hunger.
	Environmental pollution, waste, and people's irresponsibility bother me.
	I think about the causes and consequences of global warming.
Behavioral Norms	My classmates/schoolmates talk behind each other.
	My classmates/schoolmates inform school administrators and teachers when there is a problem.
	My classmates/schoolmates threaten and hit each other.
	My classmates/schoolmates protect each other.
	My classmates/schoolmates stay out of trouble.
	My classmates/schoolmates make fun of each other
	My classmates/schoolmates are nice to each other
	My classmates/schoolmates trust each other.
Perceived Adult Behavior	Adults respect my opinions.
	When there is a problem, I can talk to adults. They listen to me.
	Adults do not take me seriously.
	Adults treat me harshly.
	My teachers grade my papers fairly.
	My teachers often treat me unfairly.
	Teachers like me.
	Teachers usually have a favourite student in the classroom and only these students can receive good grades.

V Instructions for the Third-Party Punishment Game

Hello everyone! We will play a fun game with you today. By playing this game, you will have a chance to earn gift points. Each point corresponds to a gift of equal value from our gift bag [show the gift bag to students]. The decisions you make during the game will determine the number of gift points you earn. The more points you earn, the more and nicer gifts you get!

You will see passwords that consist of letters and numbers in this game [write down the following example on the board: **eA1k7**]. What you need to do is very simple: using the tablets, you need to enter the exact same passwords in the blank spaces. Pay attention to the

lower and upper case letters. The number of correctly entered passwords will determine how many gift points you will earn! So you need to write down as many passwords as you can. Let's first work on some examples to understand better [students work on three example passwords].

Now, in this game we will randomly allocate everyone to groups of three. This means that you will be matched with two of your classmates but you will not know who they are. The people in your group could be anyone!

In each group there will be 2 PLAYERS and 1 OBSERVER, the person who observes the game. Right now, you do not know your role. You could be either of the players or the observer [draw the following diagram on the board].

PLAYER 1

OBSERVER

PLAYER 2

Suppose that you are selected as one of the players. Then you will play this password game against the other player in your group. You will have 1.5 minutes. The one who write the most of the passwords will be the winner and earn as many gift points as the number of correctly entered passwords. The other player will lose and not going to earn anything.

Let's try to understand this with some examples.

1. Example 1: Suppose that we have the following scenario:

   (a) PLAYER 1: 8 (correctly entered passwords)

   (b) PLAYER 2: 7 (correctly entered passwords)

   Who do you think would be the winner? [Ask students] How many gift points would player 1 earn? [8 points] How many gift points would player 2 get? [0 point]

2. Example 2: Suppose that we have the following scenario:

   (a) PLAYER 1: 7 (correctly entered passwords)

   (b) PLAYER 2: 8 (correctly entered passwords)

Who would be the winner? [Ask students] How many gift points would player 1 earn? [0 point] How many tokens would player 2 get? [8 points]

3. Example 3: Suppose that we have the following scenario:

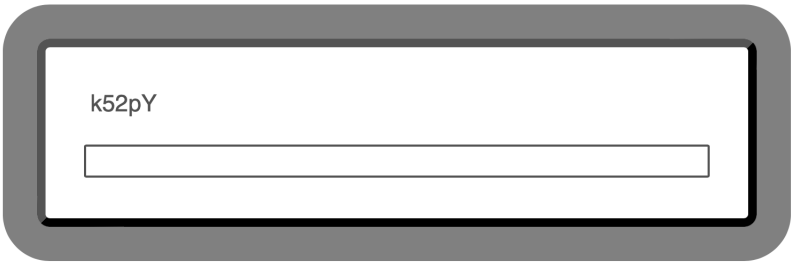
   (a) PLAYER 1: 8 (correctly entered passwords)

   (b) PLAYER 2: 8 (correctly entered passwords)

   Who would be the winner? [Ask students] Now, in this case there is a tie. When there is a tie, we will randomly select a player and announce him/her as the winner. In this case, the winner would receive 8 gift points, and the loser would receive nothing.

Have you all understood the rules? [Take questions and make sure that everyone understood the game] Then, get ready to enter the passwords! [countdown from three and start the game]

Figure B7: Password Game - An Example



[When the time is up] Ok, now the time is up. Right now, you do not know if you are the winner or not. Maybe you entered more correct passwords than the other player in your group and won the game. OR you entered less and you lost!

Now, I will give you two bonus gift points. These are are yours to keep. If you want you can use these to buy gifts. OR you can use them to do something else! [Write 2 at the upper left corner of the players to indicate the 2 bonus points as in the following:]

2
PLAYER 1

2
PLAYER 2

If you wish, you could use these bonus points to do something else: you can use these to transfer other players correctly entered passwords to yourself. However, there is a cost of transferring passwords: For each transfer you want to make, you need to spend one bonus point. For example, if you want to transfer one password, you will have to give up one bonus point, if you want to transfer two passwords, you will have to give up two bonus points. When you transfer, your opponent's correct answers will decrease, yours will increase. In other words, the your probability of winning will increase. Similarly, your opponent might or might not transfer your correct passwords to himself/herself - you don't know this. The bonus points you have not spent will be yours.

Let's briefly talk about the observer. The observer in your group will not know how many passwords are correctly entered by each player. However, the observer will see how many passwords are transferred by each player. Based on his/her observation, the observer can punish the players by destroying their correct entries.

Let's understand how transferring works with examples. Suppose that Player 1 and 2 have 8 and 7 correct entries, respectively [draw the following]

2

PLAYER 1: 8 correct

2

PLAYER 2: 7 correct

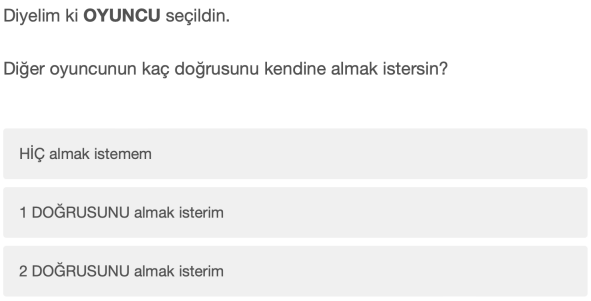
(a) Suppose that both players decide not to transfer any passwords from each other. In this case, Player 1 wins the game and Player 2 loses. In the end, Player 1 earns 10 gift points (8 from the correctly entered passwords and 2 bonus gift points) and Player 2 earns only 2 gift points (zero gift points from the game and 2 bonus points).

(b) Suppose that Player 1 does not transfer any passwords and Player 2 transfers 1 password. In this scenario Player 1's correctly entered passwords will go down to 7 and Player 2's correctly entered passwords will increase to 8. Keep in mind that Player 1 still has 2 bonus points in his pocket but Player 2 is left with only 1 bonus point since she used one of her bonus points to transfer 1 correct answer from Player 1. In this scenario, since Player 2 has more correct answers than Player 1, she wins the game and earns 9 gift points (8 from the correctly entered passwords and 1 bonus point) and Player 1 earns only 2 gift points.

(c) Now suppose that both players transfer 1 password from each other. [This scenario is explained as the ones above]

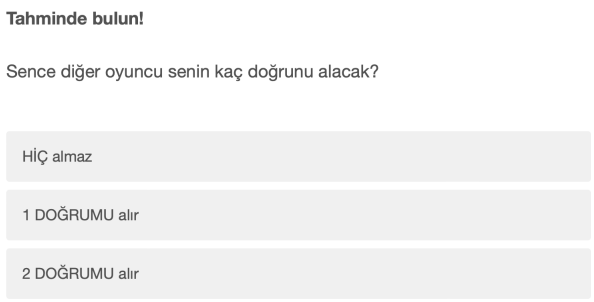
Do you have any questions? If not, let's make your choice. How many correct entries would you like to transfer from the other player? 0, 1 or 2? Enter your responses and then wait [Students enter their responses in the tablets].

Figure B8: Screen Shot of the Decision Screen 1: Transfers



Now, you all made a choice! That means that your opponent in your group also made a choice, right? Now, I would like you to guess your opponents choice. How many gift points do you think your opponent will transfer from you? 0,1 or 2? Make your guess discreetly.

Figure B9: Screen Shot of the Decision Screen: Beliefs on the Opponent's Transfer



Now, you played the game as the player, however, you might also be selected as the observer. Therefore, you will now play the game as the observer. The observer starts the game with 6 gift points. If you remember, as the observer you do not know the number of correctly entered passwords by players. However, you can see the transfers made by players. Based on this, you can destroy the correct entries of players using

his/her gift points. However, destroying correct entries comes at a cost! For each password you would like to destroy, you need to spend one gift point. For example, if you destroy one password, you are left with 5 gift points, if you destroy two passwords, you are left with 4 gift points. Do you understand? Now, let's see some examples:

(a) Suppose that you look at the transfers of the players. One likely scenario could be the following. In this scenario, since the players did not transfer anything from each other, would it make sense to destroy their correct answers?

PLAYER 1: No transfers

PLAYER 2: No transfers

(b) Another likely scenario could be the following. You look at the transfers and you see that player 1 transferred one password from the other player and player 2 has transferred no passwords. Now, you need to make a decision: if you want you can destroy correct entries of player 1, player 2 or both.

PLAYER 1: 1 transfer

PLAYER 2: No transfers

(c) Another likely scenario could be the following. You look at the transfers and you see that both player 1 and payer 2 transferred one password from each other. Now, you need to make a decision: if you want you can destroy correct entries of player 1, player 2 or both.

[Remaining scenarios are reviewed to make sure students understand them well] Do you have any questions? Now, let's start making decisions. [The scenarios are read to students one by one and students make decisions]

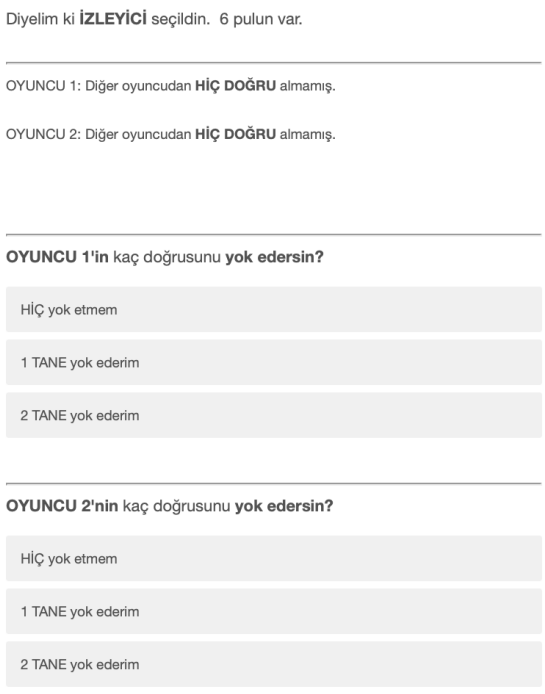
[The following scenarios are read aloud to students. Then, students make their decisions discreetly]

Scenario 1

Player 1	Zero transfer
Player 2	Zero transfer

- How many correct answers would you like to destroy from Player 1?
  - o None
  - o 1 correct answer
  - o 2 correct answers
- How many correct answers would you like to destroy from Player 2?
  - o None
  - o 1 correct answer
  - o 2 correct answers

Figure B10: Screen Shot of the Decision Screen for Scenario 1

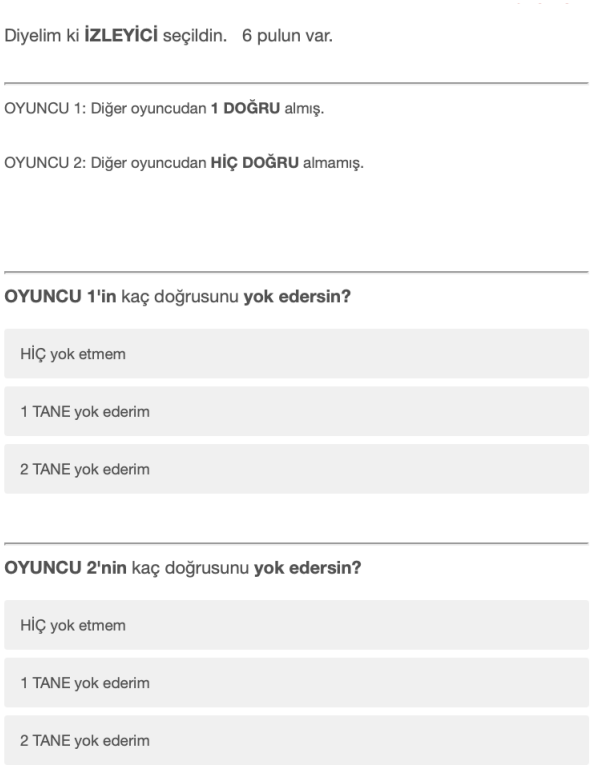


Scenario 2

Player 1	1 transfer
Player 2	Zero transfers

- How many passwords would you like to destroy from Player 1?
  - o None
  - o 1 correct answer
  - o 2 correct answers
- How many passwords would you like to destroy from Player 2?
  - o None
  - o 1 correct answer
  - o 2 correct answers

Figure B11: Screen Shot of the Decision Screen for Scenario 2

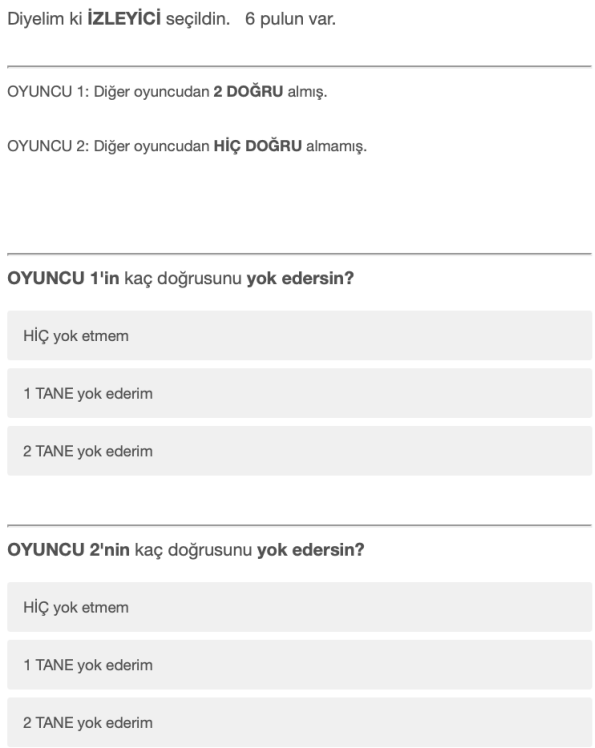


Scenario 3

Player 1	2 transfers
Player 2	No transfer

- How many passwords would you like to destroy from Player 1?
  - o None
  - o 1 correct answer
  - o 2 correct answers
- How many passwords would you like to destroy from Player 2?
  - o None
  - o 1 correct answer
  - o 2 correct answers

Figure B12: Screen Shot of the Decision Screen for Scenario 3

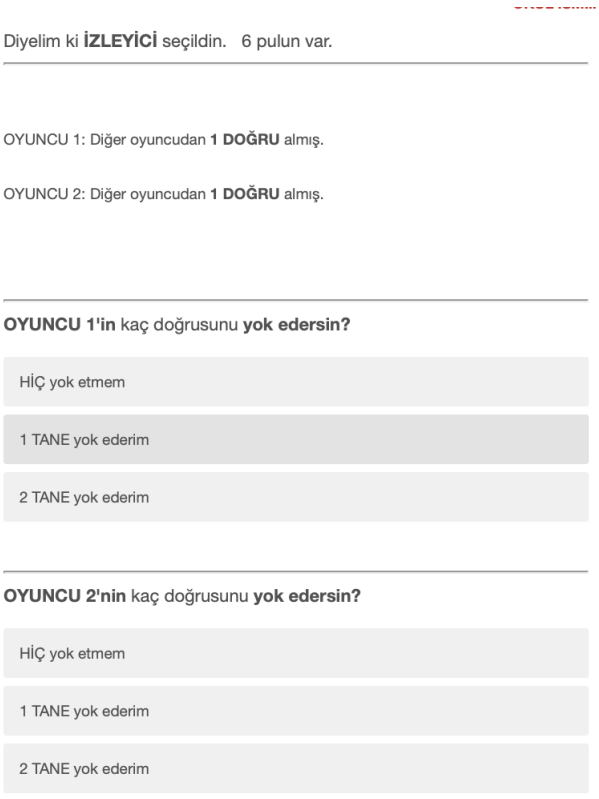


Scenario 4

Player 1	1 transfer
Player 2	1 transfer

- How many passwords would you like to destroy from Player 1?
  - o None
  - o 1 correct answer
  - o 2 correct answers
- How many passwords would you like to destroy from Player 2?
  - o None
  - o 1 correct answer
  - o 2 correct answers

Figure B13: Screen Shot of the Decision Screen for Scenario 4

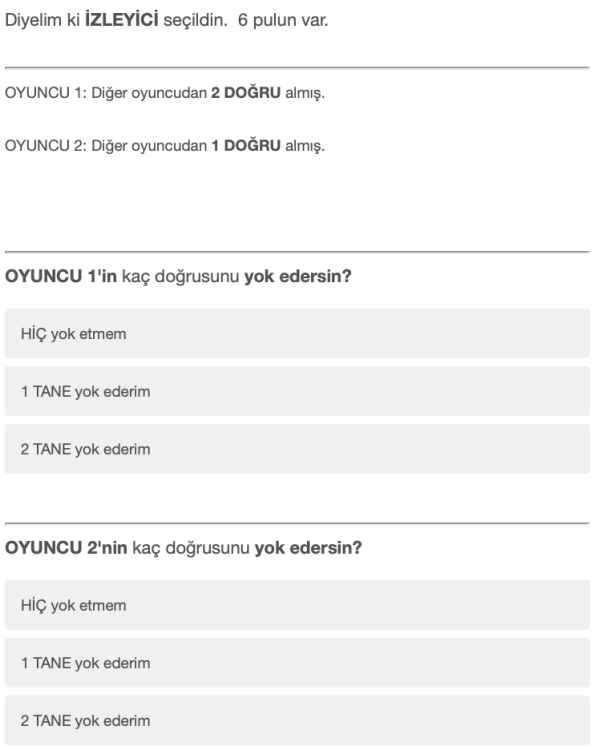


Scenario 5

Player 1	2 transfers
Player 2	1 transfer

- How many passwords would you like to destroy from Player 1?
  - o None
  - o 1 correct answer
  - o 2 correct answers
- How many passwords would you like to destroy from Player 2?
  - o None
  - o 1 correct answer
  - o 2 correct answers

Figure B14: Screen Shot of the Decision Screen for Scenario 5



Scenario 6

Player 1	2 transfers
Player 2	2 transfers

- How many passwords would you like to destroy from Player 1?
  - o None
  - o 1 correct answer
  - o 2 correct answers
- How many passwords would you like to destroy from Player 2?
  - o None
  - o 1 correct answer
  - o 2 correct answers

Figure B15: Screen Shot of the Decision Screen for Scenario 6