Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting

Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting

Sule Alan and Elif Kubilay*

November 1, 2024

Abstract

We test the effectiveness of a behavioral program grounded in the idea that status granting and self-persuasion might yield a robust behavioral change in disadvantaged adolescents. We enlist socially connected senior middle school students with high emotional intelligence as "student-teachers" and entrust them with delivering a curriculum to their junior peers. The program empowers student-teachers, leading them to improve their social environment. It reduces disciplinary incidents and anti-social behavior among student-teachers and their friendship networks. The intervention significantly enhances the likelihood of admission to selective high schools for student-teachers, offering a cost-effective way to help disadvantaged adolescents escape neighborhood disadvantages.

JEL Codes: C93, D63, I24

Keywords: neighborhood disadvantages; adolescent empowerment; self-persuasion; school climate

---

*Alan: Brooks School of Public Policy and Department of Economics, Cornell University, USA and Bilkent University, Turkiye, salancrossley@gmail.com. Kubilay: Department of Economics, University of Essex, UK, elif.kubilay@essex.ac.uk. Funded by the European Union under ERC Advance grant INCLUSION-101018401. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. We are also grateful to the 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, Turkiye (IRB#: 2019_07_22_02). We are grateful to numerous field assistants for their efforts and care in onerous data collection. This randomized controlled trial was registered in the American Economic Association's registry for randomized controlled trials under ID AEARCTR-0009140 (Alan, 2024)

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, Mertz and Okatenko, 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 and neighborhood disadvantages abound.

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 Turkiye. Our qualitative inquiry resulted in insights consistent with Yeager, Dahl and Dweck (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. Based on these insights, we developed an empowerment program designed for implementation in disadvantaged middle schools in Southeast Turkiye. 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, Tripodi and van der Weele (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 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, Turkiye. 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 the "Reading the mind in the eyes" test, and extensive social networks elicitation. Our student-teacher sample (over 1250 7th and 8th-grade 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 when selecting them. We then randomly assigned 32 schools to treatment and 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, student-teachers were neither given any tasks nor informed 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 (5th graders) 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.

We estimate the effect of the intervention for the first academic year and show its per-

sistence into the second academic year. We present our results for the full sample, junior sample, and for our target (senior) subsamples. Our target subsamples are student-teachers, and student-teacher 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. We find that the program empowers targeted adolescents, leading them to improve their social environment. In treated schools, the probability of disciplinary flagging for high-intensity behavioral problems was significantly reduced. These positive effects primarily stem from the targeted subgroups, i.e., treated student-teachers and their networks. About 3.5% of student-teachers were flagged due to high-intensity disciplinary acts at endline in the control group. This value 2.9% for student-teachers' networks and 1.4% for the senior students who are outside student-teacher networks, 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). Against this base, we estimate a 2.4 (3.5) percentage point decline in flagging of treated student-teachers in the first (second) year of the program. Both effects are precisely estimated and imply large relative effects. The estimated effects are similarly sized and precisely estimated for treated seniors in student-teacher networks, implying significant positive spillovers.

Using the decisions in a third-party punishment game, we find that the program reduced 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 the second year. Similar to the impacts seen in disciplinary flagging, the effects within student-teacher networks are sizable and statistically significant, indicating significant spillover effects. Using elicited support networks, we estimate a significant increase in inter-grade support ties within treated schools. Consistent with this, we find that perceived behavioral norms significantly improved in treated schools, with no consistent improvement in the perception of adults.

In addition to transforming their social environment, treated student-teachers have effectively altered their academic trajectory for the better. The program increased the likelihood of admission to selective high schools among student-teachers from two consecutive cohorts. Only 9.2% (13.0%) of student-teachers secured spots in such schools in the pure control group in the end of the first (second) academic year. The program increased the likelihood of admission by 8.4 percentage points in the first year and 7.4 percentage points in the second, implying substantial relative treatment effects in both years. We find similarly sizable but

statistically weaker effects for the student-teacher networks. Given the enhanced pathway to college provided by these well-resourced high schools, these results suggest that the program was remarkably effective in helping adolescents escape from neighborhood disadvantages.

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. 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, impulse control, 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-teacher networks, further confirming the program's robust spillover effects.

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 and alter their academic trajectory for the better. In comparison to programs targeting disadvantaged youth, such as Cognitive Behavioral Therapy (CBT), family relocation (Move to Opportunity), or mentorship initiatives, our intervention stands out for its cost-effectiveness, at just \$20.02 per student-teacher per year (increasing to \$54.60 when adjusted for Purchasing Power Parity, PPP), and its capacity to generate significant spillover effects relative to these other programs. 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 diverse 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 out-

comes. This literature shows how socioeconomic background affects children's social and economic outcomes (Katz, Kling and Liebman, 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, Hendren and Katz (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, Shepherd and Aronow (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. Shinde et al. (2018) and Shinde et al. (2020) are complementary to our work and reinforce our message by highlighting the strong link between school climate and health outcomes of adolescents. 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, Brown and Lavecchia, 2017; Guryan et al., 2021; Dinarte-Diaz and Egana-delSol, 2023; Resnjanskij et al., 2024). 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, Ahmed and Blakemore, 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, Dahl and Dweck, 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, Boneva and Ertac, 2019; Alan et al., 2021; Sorrenti et al., 2024).

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, Patacchini and Zenou, 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 (2024) find 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 dynamics in adolescence.[2]

The remainder of the paper is organized as follows. Section I summarizes the key features of the program and the context in which it was implemented. Section II details the evaluation design. Section III gives a detailed account of the toolkit we use to evaluate the program. Section IV describes the data. Section V presents our main results. In Section VI, we discuss the mechanisms through which the program might have improved the social environment in schools and enhanced the academic achievement of targeted groups. In section VII, we discuss cost-effectiveness and potential scale-up issues, and we conclude in Section VIII.

I Context, Intervention and Delivery

A 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 our study site, middle schools in city centers are quite large, with numerous

[2]Our study also complements a strand of the role model literature showing that role models might improve empowerment outcomes; see, for example, Nguyen (2008), Porter and Serra (2020), and Kipchumba et al. (2021).

classrooms (up to 20) per grade, often with crowded classes containing 40 to 50 students each. Due to Turkiye's fast-changing demographic structure, student populations are 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 that highlight 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.[3] 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 their teachers and school administrators. Realizing that conventional lecture-type interventions may not work, we decided to leverage adolescents' strong 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 a greater sense of belonging and safety. This idea received enthusiastic support from most senior students in our pilot schools. We 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. We also conjecture that by entrusting them with such an important role, we empower them and cultivate a sense of agency to work toward a better future for themselves.

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

[3]Consistent with our observations in pilot schools, in our data, nearly half of senior students state that adults do not respect them and pay little to no attention to their opinions.

challenges of concentrated poverty and social unrest.

B 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 is often associated with rebellious 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, which stood at 73.2%. 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, Pescosolido and Druskat, 2002; Alan et al., 2023). Moreover, a significant and negative correlation between emotional intelligence and violent behavior is well-documented in social psychology (Richardson, Green and Lago, 1998). 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. This input is motivated by the research that shows that certain individuals within a group are more influential and persuasive, making targeting them an efficient and cost-effective strategy (Kempe, Kleinberg and Tardos, 2003; Galeotti and Goyal, 2009; Paluck, Shepherd and Aronow, 2016; Banerjee et al., 2019; Galeotti, Golub and Goyal, 2020).

We averaged these three inputs and selected the highest-scoring senior students among the willing ones, constituting around 10-15% of the student population in selected 5th and 6th-grade classrooms. The 10-15% amounted about 5 to 8 student-teachers per junior classroom. We ensured gender balance in our student-teacher sample. 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.[4] Our student-

[4]It was possible that despite this filtering, an unsuitable student could still be included in the student-teacher pool. However, our IRB protocol mitigated this possibility. Our protocol required us to run the

teacher sample is diverse, with both high and low academic achievers, troublemakers, and well-behaved students.

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.[5] 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; see Figures B2 - B6 for a selection of treatment and placebo activities.

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 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 interfer-

selected student-teachers through the school administration before implementation. If the school administration had any reason to exclude a student due to suspension or violent behavior, we were required to comply and drop the student from our student-teacher pool. Fortunately, we did not encounter such a case in this study.

[5]The program's intended duration is 10 weeks if implemented weekly without interruptions. However, schools have various activities that can disrupt the schedule, such as exams, mock exams, performances for national holidays, and field trips. These events often require halting the program activities for 2 or 3 weeks. Additionally, in larger schools, a single topic might take more than one week to cover, resulting in some sessions being split into two parts.

ence. 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. Ensuring this implementation protocol was adhered to, we regularly met with interns, and they submitted weekly progress reports via the online platform we designed for them. 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. It is important to note that while there may be some similarities, our programmatic approach significantly differs from CBT activities. Unlike CBT, which focuses on skill-building exercises and the learning and unlearning of behaviors, our curriculum emphasizes standard normative messages about the dangers of violence, intolerance, and anti-social behavior. The innovation of our approach lies in its delivery method: it indirectly targets a group by giving them the responsibility to deliver these normative messages. This strategy aims to leverage their desire for status, invoking self-persuasion and cognitive dissonance to foster change.

II Evaluation Design and Timeline of the Study

We recruited 65 middle schools in the province of Diyarbakir, Turkiye. These schools varied in size and type and collectively hosted over 27,000 officially registered students.[6] 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 had 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, assess willingness to become a student-teacher, administer the emotional intelligence test, and collect social network data. Seniors who were not in student-teachers' networks helped us to assess further spillover effects. In very large schools, we randomly selected two or three 7th and 8th-grade classrooms for this purpose.[7] This intensive data collection required 3 to 4 field team members, assisted by 6 to 8 locally recruited field assistants, to spend an entire school day in each school. Data we collect via in-person visits cover about 18,000 students (those who were present during baseline visit).

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 the placebo control group and 17 to the pure control group. In placebo schools, selected student-teachers conducted unrelated activities in their assigned junior classes, with hands-off monitoring by their assigned interns at 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. Placebo student-teachers met with their interns to familiarize themselves with the activities before the session and wrote post-session reports, just like the treatment student-teachers. This arm aims to rule

[6]Our school recruitment was guided by a thorough power assessment. See our power calculations in Section IV in the Online Appendix, Tables B2 and B3.

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

out the possibility that the estimated effects stem from the senior-junior-intern interactions created by our delivery method. Therefore, we made every effort to replicate the time commitment and the degree of intern-senior and senior-junior interaction required by the treatment in the placebo schools.The acceptance rate of the student-teacher role was 100% at post-randomization, both in treatment and placebo schools.[8]

We conducted our first endline in April and 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 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 and successfully completed the program 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.

III Outcomes

We use a comprehensive set of indicators to characterize the social climate in schools. Our toolkit includes administrative records, an incentivized game, surveys, and tests. In addi-

[8]Due to a small coding error in calculating the RMET score at baseline, 159 student-teachers were incorrectly 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; see Figure B7 in Section III in the Online Appendix.

tion to our primary outcomes describing the social environment in the school, we collected individual outcomes using surveys and cognitive tests to assess the extent of academic and socio-emotional improvements as well as possible undesired effects of the program. Our primary interest is to improve the school's relational atmosphere. We consider the following social and relational outcomes to describe the schools' relational atmosphere.

A Social and Relational Outcomes

A.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 problems by the school administration. Flagging is updated actively throughout the year. A student can be flagged and then unflagged several times within the same academic year. Our data are 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.

A.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 (Fehr and Fischbacher, 2004). Therefore, as part of our toolkit, we designed a novel third-party punishment game to quantify the tendency to engage in 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 their roles at the outset and were told this would be determined at the end of the game. They first play as players, and their decisions and outcomes are recorded. Then, they play as observers, and those decisions and outcomes are also recorded. At the end of the session, players receive gifts based on their points, and observers receive gifts based on their points.[9]

[9]This implementation method, referred to as the strategy method, is theoretically equivalent to playing

Players first compete in a real-effort task, typing as many meaningless 5-character password sequences as possible in 1.5 minutes. The player who types the most receives 1 gift point per correct password; the other player receives zero points. After completing the task, before knowing their winning status, all players receive 2 bonus gift points, which they can use to transfer correct answers from their opponent to themselves, at a cost of one bonus point per transfer. This transfer decision is our measure of anti-social behavior. Before making their transfer decision, students are informed that the observer will see how many correct answers each player transferred and can choose to destroy correct answers, up to 2 from each player, at a cost of one of their own points per answer destroyed. We ensured that students fully understood the game before eliciting their decisions. The implementation of this task took an entire lecture hour.

After performing as players, students assume the role of observers, deciding whether to destroy players' answers. The observer has 6 gift points and can destroy up to 4 correct answers (2 from each player), at a cost of one point per answer destroyed. We used a strategy method to elicit these decisions. Note that a simple dictator game could measure anti-social behavior. However, we also needed a method to gauge reactions to unfair behavior, which required a third-party punishment component. We opted for a real effort task with a sabotage component rather than a simple endowment allocation to elicit a stronger sense of justice. This design allowed us to more accurately measure whether the treatment significantly heightened the sense of justice relative to the control.

We consider three social outcomes using the decisions in this game: the number of correct answers transferred, representing anti-social/unfair behavior; the anti-social behavior expected from classmates; and the cost incurred to punish players. Detailed instructions for the game and screenshots are presented in the Online Appendix Section V.

A.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,

the game by splitting the sample at the outset (Brandts and Charness, 2011).

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.

B Individual Outcomes: Academic Achievement and Socio-emotional Well-Being

The program did not directly target achievement outcomes. However, achievement could be affected positively through the improvement in the social environment and/or the improvement in socio-emotional well-being. The program could potentially stimulate targeted students to dedicate more effort to their studies by altering their beliefs regarding their control over outcomes. On the other hand, the program imposed a significant time commitment, especially for the student-teachers. Student-teachers had to read, prepare, and practice the material before delivering their sessions. We aimed to utilize well-being hours as much as possible for our intervention. When we had to use lecture time, we ensured that primary subjects critical for high school admission tests—math, Turkish, and science—were not crowded out. This required careful scheduling of intern visits, and we relied on the assistance of well-being teachers to coordinate these efforts. Despite this, some study time could, in principle, be devoted to the project activities. To assess the program's impact on test scores, we administered in-class math and Turkish tests, which were prepared based on the national curricula at baseline and endline in both years. Note that baselines were conducted at the beginning of the academic year and endlines at the end. Therefore, baseline tests were based on the previous grade level, whereas the endline tests were based on the current grade level.

In January 2024, we gained access to data on selective high school qualifications for the 2022 and 2023 cohorts.[10] After securing permission, we contacted our schools to gather the names of students admitted to the province's selective institutions in 2022 (first year) and 2023 (second year). Unfortunately, six schools no longer had data for 2022, leaving us with

[10]We acknowledge two deviations from our pre-analysis plan (PAP). Firstly, we did not explicitly specify how we would group our rich outcome set in the PAP; we only listed all the outcomes we aimed to measure. Secondly, because we were initially unsuccessful in securing permission to access high school admission test results, we did not include this critical outcome in our registry and PAP.

data from 59 schools for that cohort.[11] Fortunately, we obtained data from all 65 schools for the 2023 cohort. This is a crucial hard outcome for the purpose of our study. Selective high schools offer numerous advantages to students, including well-equipped lab and sports facilities and high-quality teaching staff. Graduates from these schools have much better chances of admission to good universities. According to the Turkish Student Selection and Placement Center, in 2019, only 8.2% of students entered college from regular public high schools in Turkiye. This value is about 50% for selective high schools. However, admission to these high schools requires high scores from a nationwide exam conducted by the Ministry of Education, open to all students who have completed grade 8 by June (the end of the academic year in Turkiye). Since our schools are situated in disadvantaged neighborhoods, gaining admission to such selective institutions implies overcoming neighborhood disadvantages and increasing opportunities for the students.

We conjecture that the treatment will improve 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, Binh Tran and Zikos, 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, Klein and Kramp, 1991; Andor et al., 2022). Moreover, it has been shown that internal locus of control is strongly associated with motivation and learning (Findley and Cooper, 1983; Hadsell, 2010). To assess the extent of improvement in these empowerment outcomes, we measured internal locus of control, mental well-being, and sense of responsibility at baseline and endline and self-worth at endline. We also measured perspective-taking, impulse control, and sense of belonging to the school both at baseline and endline, as the program may also have affected these attributes. Individual items we use to construct each index are given in

[11]Four of these schools are treatment schools, one is a placebo school, and one is a control school. In analyzing 2022 data, we conducted a conservative robustness check where we assumed no student from these four treatment schools gained admission to selective high schools. We discuss these results in Section E.

the Online Appendix VI (see tables B4 and B5).

IV Data and Empirical Analysis

We have a total of 27,052 officially registered students across our 65 schools. While we included all students in the majority of the schools, 23 schools were too large to do so. In these larger schools, we randomly sampled senior students to assess spillover effects (seniors outside student-teacher networks), resulting in 22,875 officially registered students in our study. We collected data from approximately 18,000 students who were present at school during our in-person visits at baseline. Demographic information, some indicators of home environment, fluid IQ using Raven's progressive matrices (Raven and Court, 1998) and emotional intelligence using Reading the Mind in the Eyes test (Baron-Cohen et al., 1997) were collected only at baseline. We collected all our primary outcomes at both baseline and endline, except for the incentivized third-party punishment game and official disciplinary flagging.[12] In the first year of the trial, we selected 633 7th-grade student-teachers, providing us with 2687 friends of theirs (in-degree+out-degree ties) and 636 8th-grade student-teachers, giving us 2573 friends. Our total sample for assessing further spillover effects on seniors outside student-teacher networks is 4,893. These students have 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 at baseline. As can be seen, most student-teacher characteristics are starkly different from other seniors. Our student-teachers have higher cognitive scores. This is expected as they were selected partly based on the RMET, which is a cognitive test and is highly correlated with fluid IQ. Note that they are not different from non-selected seniors regarding their perceived behavioral norms and impulse control.

A Internal Validity

Table 2 illustrates the balance across three treatment arms in the first academic year. 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 B6, B7, and B8 in Online Appendix VII). Overall, the randomization worked well, and we observe no noteworthy imbalance across

[12]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.

treatment status in any of the outcomes; see also the joint F-tests results at the bottom of the table.

B 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 estimate the following model:

$$y_{ics} = \alpha_0 + \alpha_1 T_s + \alpha_2 P_s + X_{ics}^{'}\beta + \delta_d + \varepsilon_{ics}, \tag{1}$$

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, P_s is a binary indicator for placebo treatment, which equals one if school s is in the placebo group and zero otherwise. X'_{ics} is a vector of student-level observables, including gender, age, baseline cognitive scores and the respective outcome at baseline (if collected). We also control for class size, the share of boys in the classroom, and school-type fixed effects. δ_d represents district (strata) fixed effects. We also estimate a benchmark model where we pool placebo and pure control as a combined control group. Results from this specification are presented in the Online Appendix, Sections VIII and IX.

All sessions were completed in all treatment and placebo schools in both academic years. In the second academic year, we lost one 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 the second academic year necessarily involves a post-treatment selection as all students, except for the newcomers were exposed to the program for one year. 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 academic year 2. This implies that we analyze only grade 8 student-teachers (those who were grade 7 student-teachers previously) in the second academic year, representing the program's persistent effects on student-teachers.

We cluster standard errors at the school level in all analyses and provide wild boot-

strapped p-values at the bottom of all tables and figures. The latter is particularly relevant for our subgroup analyses, where the sample size becomes low. We provide all estimated treatment effects without covariates in Figure A1 in the Appendix. A Romano-Wolf correction due to the multiple tests we conducted is presented in the Appendix (see table A1).[13]

C 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 due to high seasonal worker movements in the region. Teachers in Turkiye 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 the first year of the program, the overall absenteeism rate was 14.5%. It was slightly higher for seniors (14.6%) compared to juniors (14%). The absenteeism rate was very similar in the second year. We found no treatment effect on absenteeism in either the first or the second year.

At the time of the endline in the first academic year, the percentage of students who were present at baseline but not at endline was 20.6%. This value was significantly higher (26.4%) in the second year (p-value for the difference < 0.001). Importantly, the absenteeism at endline is not correlated with treatment status (p-value $= 0.473$ in the first year, 0.294 in the second year), ensuring the internal validity of our results.[15] Note also that we have two primary administrative outcomes that are not affected by absenteeism on a given day:

[13]Following Kling, Liebman and Katz (2007), we construct summary indices by aggregating relevant outcomes. We use three outcome groups: i) social climate (disciplinary flagging, decisions on the third-party punishment game and survey outcomes of behavioral norms and adult perceptions, and network outcomes), ii) socio-emotional outcomes (locus of control, mental well-being, perspective-taking, self-worth, impulse control, sense of belonging, sense of responsibility), and iii) achievement outcomes (test scores and admission to selective high schools.

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

[15]A high rate of absenteeism toward the end of the academic year is typical for Southeast Turkiye due to the seasonal agricultural worker mobility. Table B15 presents the characteristics of absent students at endline for all subgroups. Males, older students those with lower cognitive abilities are more likely to be absent. Tables B16 - B20 in Online Appendix X show the balance of the characteristics of absentees across treatment and control for each year of the program.

administrative records of disciplinary flagging for all registered students and admission to selective high schools for grade 8 students. Nevertheless, given that absenteeism was not related to the treatment, our results on non-administrative outcomes also remain internally valid.

V Results

In all exhibits, we present the results for the full sample, juniors corresponding to grades 5 and 6, and seniors corresponding to grades 7 and 8, for academic years 1 and 2. We then split the senior sample further and show the results for our targeted subgroups: student-teachers, seniors in student-teacher networks (student-teacher networks), and finally, the seniors outside student-teacher networks. The latter two groups inform us about the spillover effects of the program. Student-teacher networks inevitably include some student-teachers as well. We exclude these student-teachers when we estimate treatment effects on student-teacher networks. The results for the first year of the program (academic year 1) represent short term effects, whereas academic year 2 represents persistence of the effects one year after.

A Treatment Effect on Social Climate: 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 1.8% (2.6%) of the student body is flagged as challenging in the full sample in year 1 (year 2). Against this small base, we estimate a significant treatment effect on the probability of being flagged, indicating substantial relative effect sizes. Specifically, we estimate a 1 and 1.2 percentage point decline in the first and second years, respectively, for the senior students with no statistically significant effect for juniors. These estimates correspond to a 48 to 62.5% relative treatment effects.

Table 4 shows that the rate of being flagged is significantly higher among student-teachers (3.5% and 5.6% in year 1 and 2, respectively) in the control group. This indicates that our student-teachers, while cognitively able and socially central, are often considered trouble-makers in the school. Seniors in student-teacher networks have a lower flagging rate (2.9-3.4%

in the control group), 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 68.6% in year 1 and persistent at 62.5% in year 2 for student-teachers. They are similarly sized (69%, and 41.2% in year 1 and year 2, respectively) for student-teacher networks. We also estimate sizable but imprecisely estimated spillover effects on those outside student-teacher networks. We note an increase in flagging in our placebo group in the first year of the program for student-teachers, but this finding does not repeat in the second academic year. In fact, we reject equality of treatment and placebo effects for all senior subgroups.

B Treatment Effect on Social Climate: 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 actions. 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 (anti-social behavior) and the cost incurred by punishing transfers (intolerance to anti-social behavior).

Figure 2 Panel A 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 B presents the distribution of incurred punishment costs (the total number of correct answers destroyed) for each transfer type, again across treatment and control groups. 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 A, transfer behavior correlates negatively with

positive indicators of social climate and socio-emotional well-being. Associations regarding the costly punishment behavior are even more pronounced (Panel B). 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 impulse control. 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.

Figure 4 presents the estimated treatment effects on the transfer decisions on the third-party punishment game. The results indicate lower transfers and higher punishment in treated classrooms relative to placebo and pure control in both years, but the estimates are much larger and more precise in the second year. Considering the full sample, pure control students transferred 1.097 points from their opponents on average and destroyed 1.497 points in cases of asymmetric transfers. Treated students transferred 0.020 and 0.131 fewer points in the first and second year, respectively, with only the second-year estimates statistically significant at the 1% level. Panel B zooms into senior subgroups. Here, we see that treated student-teachers transferred 7.2% (22.6%), seniors in student-teacher networks transferred 1.8% (18.6%), and seniors who are out of student-teacher networks transferred 4.2% (16.7%) 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, Figure 5 shows that treated students are willing to incur substantially higher costs to punish unequal transfers. For the full sample (Panel A), the effect size relative to pure control is 10.6% in year 1, significant at the 5% level, and 31.5% 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. While we do not detect a significant effect in the first year of the program, treated student-teachers were willing to sacrifice about 0.451 points to punish unequal transfers in year 2, implying an effect size of 26% relative to pure control, statistically significant at the 10% level. Again, the spillover effects are also large and statistically significant. Both seniors in and outside student-teacher networks in treatment schools sacrificed more points to punish unequal transfers, with all longer term (year 2) effects statistically significant at the 1% level. Specifically, in the second year of the program, we estimate 35.6% and 28.3% 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, relative to pure control and placebo.

C Treatment Effect on Social Climate: 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 76% (79.8%) of all friendship links and 73.7% (77.6%) 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 5 presents the estimated treatment effects on the total number of support links (sum of in-degree and out-degree ties) directed to targeted subgroups. Panel A presents links directed to student-teachers, and Panel B presents links directed 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 (0.194 and 0.175 in year 1 and year 2, respectively), as opposed to the number of ties between seniors (5.032 and 3.581 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 66.5% in year 1, and by 57% in year 2, and both estimates are statistically significant. We can rule out the pure interaction mechanism for this outcome in the first year, as the equality between placebo and treatment is rejected (see column 1 in Panel A). 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 B 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 the second year of the program.

D Treatment Effect on Social Climate: 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 behavior 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." Figures 6 and 7 present the estimated effects on these standardized indices where the pure control mean is normalized to zero. We estimate a significant improvement in behavioral norms in both years 1 and 2 in the full sample. This effect is again driven by targeted subgroups as can be seen in Panel B of Figure 6. We also estimate a positive effect on seniors' perception of adults in the first year of the program. However, this effect dissipates in the second year. For most but not all of the cases, we reject the equality of treatment and placebo effects for both behavioral norms and perception of adults.

E Treatment Effect on Academic Achievement

Taking the nationwide end-of-middle school exam for selective high schools is not compulsory. All students have a guaranteed spot in their catchment area public high school. However, all grade 8 students are strongly encouraged to take the exam, which is free. In our sample, 79.4% (65.9%) of students took the exam in 2022 (2023). While the 2022 rate is consistent with the nation's average of 83.4% in 2022, the 2023 rate is much lower than the national average for that year (82%). The latter is likely due to the devastating earthquake that hit the region on February 6, 2023. Nevertheless, we find that the program had no impact on the probability of taking the exam (p-values for 2022 and 2023 are 0.356 and 0.596, respectively).

Table 6 presents the treatment effects on the probability of admission to a selective high school within the province. In the first academic year (cohort 2022), only about 9.2% of

our student-teachers gained admission in the control group. The admission rate was even lower (5.8%) for their networks and for students outside student-teacher networks (5.6%). The impact of the program on student-teachers is remarkable. Treated student-teachers were 8.4 percentage points more likely to gain admission to one of the selective institutions compared to their pure control counterparts in 2022. Although we cannot statistically reject the equality of treatment and placebo effects in 2022, the estimated differences suggest meaningful differences between the two treatment arms. Consistent patterns are observed across both cohorts. In the second academic year (cohort 2023), the admission rate for the control was 13%, and we estimate a 7.4 percentage points treatment effect for the student-teachers, significant at the 5% level. Here, we reject the equality of treatment and placebo effects.

While our estimates for student-teacher networks are imprecise in 2022, we observe a sizable treatment effect for the 2023 cohort. A similar pattern is observed for students outside student-teacher networks, although the estimates for this group are not as consistent as those for student-teachers and their networks. Recall that while 2023 data are complete, we do not have high school admission data for six schools in 2022, four of which are treatment schools. Therefore, we perform a conservative robustness check for the 2022 cohort. Specifically, we assume that no student from these four treatment schools gained admission to selective institutions and re-estimated treatment effects. The results confirm positive treatment effects on the likelihood of admission to selective high schools for the 2022 cohort of student-teachers, even under such extreme assumptions (see Table A2 in the Appendix).

As access to the administrative data on admission to selective high schools was not initially anticipated, we did implement our own math and Turkish tests at both endlines to assess the program's effect on academic performance. Table 7 presents the estimated treatment effects for senior subgroups.[16] As can be seen, we estimate a significant treatment effect on math performance for student-teachers and student-teacher networks. The former performed 0.163 sd and the latter 0.123 sd higher than their pure control counterparts in our math test. However, this improvement is observed only in the first year of the program, and we cannot reject the equality between the treatment and placebo arms. We observe some, albeit statistically weak, improvements in Turkish scores.

It is important to contextualize our results on test scores in the second year. We imple-

[16]We do not estimate any statistically significant effect on test scores for our junior sample. Table A3 in Appendix presents the results for the full sample, junior sample, and senior sample.

mented the same achievement tests in both years based on the grade-level national curriculum. However, in the second year of the program, schools were closed for about six weeks due to the devastating earthquake that hit the region on February 6, 2023. An important factor that might influence our second-year results on test scores is the Ministry of Education's decision in March 2023 to exclude topics taught in the second semester of the 2022-2023 academic year from the end-of-middle school exam. This measure was implemented to mitigate the disadvantages experienced by students in the affected regions due to the earthquake. Another reason for caution is that our tests were implemented on students who were present during our visit. Given high absenteeism, we likely lack the statistical power to identify any cognitive improvement in 2023. Consequently, our second-year findings should be interpreted cautiously, as they may appear inconsistent with the results on admission to selective high schools in 2023. Nevertheless, our complete administrative data on 2023 selective high school admissions enabled us to circumvent the challenges we faced in the 2023 endline.

Overall, these achievement results hold a significant policy value. Given the program's objectives of empowering disadvantaged youth and helping them realize their potential to escape their neighborhood disadvantages, an increased likelihood of admission to a selective institution serves as an objective indicator of program success. The path to college from these well-resourced institutions is much smoother in Turkiye, offering socioeconomically disadvantaged adolescents a means to overcome their family and neighborhood disadvantages. The following section explores the potential mechanisms through which the program may have generated such economically and socially favorable results.

VI Potential Mechanisms

Entrusting socially influential and emotionally intelligent adolescents with the responsibility of improving their school's climate resulted in significant improvements in the relational environment of the school. This improvement was accompanied by a substantial change in these adolescents' academic trajectory for the better. The consistency of estimated treatment effects across administrative, incentivized, and survey outcomes indicates the success of our programmatic approach, with improvements primarily stemming from the intended groups: student-teachers and their networks. Note that for most of our outcomes, we were able to rule out mere interaction mechanisms generated by the program delivery through our placebo arm.

We argue that the content of the empowerment curriculum and and its delivery method 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 accountable for their social environment. As a result, they became more prosocial, avoided actions that triggered disciplinary flagging, extended their support to their junior schoolmates, and improved their school's climate. Their role endowed them with significant responsibility and autonomy, motivating them to take control of their outcomes and set good examples for their friends. To investigate whether our data support this conjecture, we examined various attributes that could be affected by the intervention, including various empowerment indicators, such as internal locus of control, mental well-being, self-worth, and sense of responsibility, as well as changes in perspective-taking, and impulse control, which could also be influenced by the program. Although not exhaustive, these attributes collectively cover significant aspects of adolescent socio-emotional well-being.

Figure 8 presents estimated treatment effects on the aforementioned attributes relative to pure control and placebo combined. The figure shows striking improvements in almost all these attributes among student-teachers and their networks. Consistent with our main results, although we also estimate some positive effects among juniors, the most substantial improvements were observed in the senior subgroups. For student-teachers, we estimate a 0.376 sd increase in internal locus of control, a 0.175 sd improvement in mental well-being, a 0.272 sd increase in perspective-taking ability, and a 0.261 sd increase in impulse control in the first year of the program. We also estimate a significant 0.241 sd increase in the sense of belonging to the school and a 0.286 sd increase in the sense of responsibility for world issues. The effects are similar for student-teacher networks with slightly smaller magnitudes. Furthermore, we detected improvements in the internal locus of control and the sense of responsibility even among seniors outside the student-teacher networks. Notably, these effects persist into the second year for both student-teachers and their networks.[17] These findings suggest that the program successfully empowered targeted students, changed their beliefs regarding their perceived control over their lives, and improved their overall socio-emotional well-being, all in line with the remarkable achievement results we obtained using our administrative data.

[17]We explored but did not detect a notable gender heterogeneity in treatment effects. Figures B21 and B22 in Online Appendix XI present the heterogeneity results. The estimated differential effects are not robust enough across all measures to conclude that the program had differential impacts. Additionally, we did not find any significant heterogeneity based on cognitive ability.

Although juniors did not seem to benefit as much from it, recall that the program is designed to be implemented on a rotating basis each year. This ensures that some of the juniors who were passive participants one year become student-teachers the next year. For example, in the second year of our implementation, 287 sixth-grade students who had been taught by eighth-grade student-teachers in the first year became student-teachers themselves in seventh grade and delivered the curriculum to newcomers (fifth graders). We analyzed the program effects on these 287 students and their 1114 network peers, comparing their gains as passive students in sixth grade and active student-teachers in seventh grade. As depicted in Figure A2 in the Appendix, while there were some positive effects on socio-emotional well-being in sixth grade, the impacts were substantially broader when these students became active student-teachers or were part of student-teacher networks in seventh grade. This difference highlights the role of the program's delivery method in generating impacts.

In addition to the intended positive effects, we explored the potential unintended consequences of the program. Specifically, we explored the possibility that, instead of acting responsibly, student-teachers might 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; see the individual items in the Online Appendix VI. We estimate null effects for both years (p-values 0.988 and 0.285 in years 1 and 2, respectively). Additionally, concerns were raised about the demanding nature of being a student-teacher and possible crowding out of study time. Our results on test scores and admission to selective high schools effectively rule out this possibility.

In sum, our results suggest that targeting disadvantaged adolescents in an innovative way by tapping into their desire for autonomy and social status enhanced 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, altering the targeted students' academic trajectory for the better. Importantly, these positive effects extended beyond the entrusted adolescents to their friendship networks and even to those outside their networks, amplifying the program's overall impact and enhancing its cost-effectiveness.

VII Cost-effectiveness and Potential for Scale-up

The program is highly cost-effective, with positive spillovers that significantly enhance its overall value. The first-year implementation, which included both grade 7 and 8 students, incurred $12,000 in printing costs and $3,000 in distribution costs. Based on the Turkish Ministry of Education's intern payment schedule, the total intern cost was $10,400 (20 workdays, 26 interns, 140 TRY daily wage; 1 USD = 7 TRY on average in 2021). Consequently, the total cost of the program in the first year was $25,400. With 1269 student-teachers involved, the estimated cost per student-teacher was $20.02 per year. Including the broader student-teacher networks (adding 4,897 friends of student-teachers), the cost per student decreases to $4.12 per year. However, we acknowledge that implementation costs can vary significantly across countries, especially when considering labor costs, which are a critical factor for replicating this program elsewhere. Using the OECD 2021 PPP conversion factor for Turkiye, our printing costs rise to $32,760, distribution costs to $8,190, and intern costs increase from $10,400 to $28,392 (OECD, 2021). These adjusted figures imply a cost of $54.60 per student-teacher, or $11.20 per student when including their networks.

These values indicate high cost-effectiveness compared to well-known programs such as CBT (approximately $267 per participant as reported by Blattman, Jamison and Sheridan (2017), MTO (counseling cost of $5,071 per family who utilized a voucher), and BAM (around $2,046 per participant per year). The cost-effectiveness of our program remains notable even when compared to recent school-based studies targeting student behavior. For instance, the study by Dinarte-Diaz and Egana-delSol (2023), evaluating the behavioral effects of an after-school program for disadvantaged adolescents, shows a cost of $296.50 per student per year. Similarly, the study by Sorrenti et al. (2024), evaluating a socio-emotional wellness program (PATHS), shows a cost of $67 per student per year. Our study further suggests that these school-based interventions may actually be more cost-effective than previously reported when accounting for within-school spillovers.

It is important to note that the program was implemented with near-perfect compliance in this study. Therefore, the results should be regarded as upper bounds, representing the effects expected if the program is executed as instructed with full compliance. To assess how this program would perform at scale with less monitoring, careful process testing is required. For scaling, different modalities can be considered. One approach could involve well-being teachers, familiar with the students, selecting student-teachers on a rotating basis.

Another model could involve well-being teachers delivering the program directly, potentially reducing administrative costs. A comprehensive scale-up would necessitate rigorous testing of different modalities to evaluate impacts, per-student costs, and administrative burdens.

VIII 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 first implemented in the 2021-2022 academic year and then repeated for a subset of students in the 2022-2023 academic year in Diyarbakir, Turkiye. 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.

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. The program substantially increased the targeted students' likelihood of admission to selective high schools, with some weak positive spillover effects on their friendship networks. Our results suggest that these positive effects on school climate and academic achievement may be attributed in part to a substantial improvement in the socio-emotional well-being of the targeted students.

Two caveats apply to our study. Firstly, the program was implemented in a low-income region of a large middle-income country. One might be concerned that some of the issues we highlight regarding neighborhood disadvantages and school climate issues may not be relevant to contexts outside Turkiye. However, there are two reasons why our results might be relevant beyond our setting. First, our approach to challenging adolescents is informed by 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. Secondly, while the evidence on the causal link between the relational climate in schools and students' overall well-being is robust, the impact of programs targeting school climate on critical long-term economic outcomes remains underexplored. Although our study focused on middle school students, who are too young to provide insights into long-term economic and social outcomes, the positive achievement and behavioral outcomes we observe are promising, as these outcomes are intricately linked to economic and social development. Behavioral outcomes such as locus of control, perspective-taking, impulse control, and mental well-being are pivotal in shaping individual decision-making processes, human capital investment, and long-term labor market outcomes. By fostering these outcomes in schools, we can drive better educational and labor market outcomes, ultimately contributing to economic growth and social welfare. Therefore, it is crucial for future research to explore the long-term economic and social impacts of school climate improvement programs to inform policy decisions aimed at fostering holistic student development.

References

Alan, Sule. 2024. "Transforming School Climate by Harnessing Adolescent Demand for Autonomy." AEA RCT Registry. July 31. https://doi.org/10.1257/rct.9140-1.1.

Alan, Sule, and Ipek Mumcu. 2024. "Nurturing Childhood Curiosity to Enhance Learning: Evidence from a Randomized Pedagogical Intervention." *American Economic Review*, 114(4): 1173–1210.

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

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. 2021. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking." *The Quarterly Journal of Economics*, 136(4): 2147–2194.

Alan, Sule, Gozde Corekcioglu, Mustafa Kaba, and Matthias Sutter. 2023. "DP18465 Female Leadership and Workplace Climate." https://cepr.org/publications/dp18465.

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *The Quarterly Journal of Economics*, 134(3): 1121–1162.

Andor, Mark A., James Cox, Andreas Gerster, Michael Price, Stephan Sommer, and Lukas Tomberg. 2022. "Locus of Control and Prosocial Behavior." National Bureau of Economic Research Working Paper 30359.

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

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2019. "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials." *The Review of Economic Studies*, 86(6): 2453–2490.

Baron-Cohen, Simon, Therese Jolliffe, Catherine Mortimore, and Mary Robertson. 1997. "Another Advanced Test of Theory of Mind: Evidence from Very High Func-

tioning Adults with Autism or Asperger Syndrome." *Journal of Child Psychology and Psychiatry*, 38(7): 813–822.

Bierhoff, Hans Werner, Renate Klein, and Peter Kramp. 1991. "Evidence for the Altruistic Personality from Data on Accident Research." *Journal of Personality*, 59(2): 263–280.

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

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review*, 107(4): 1165–1206.

Brandts, Jordi, and Gary Charness. 2011. "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics*, 14(3): 375–398.

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

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

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 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, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *American Economic Review*, 108(10): 3028–3056.

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

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

Dahl, Ronald 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, Ronald E., Nicholas B. Allen, Linda Wilbrecht, and Ahna Ballonoff Suleiman. 2018. "Importance of Investing in Adolescence from a Developmental Science Perspective." *Nature*, 554(7693): 441–450.

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

Dinarte-Diaz, Lelys, and Pablo Egana-delSol. 2023. "Preventing Violence in the Most Violent Contexts: Behavioral and Neurophysiological Evidence from El Salvador." *Journal of the European Economic Association*, jvad068.

Dustmann, Christian, Mikkel Mertz, and Anna Okatenko. 2023. "Neighbourhood Gangs, Crime Spillovers and Teenage Motherhood." *The Economic Journal*, 133(653): 1901–1936.

Fehr, Ernst, and Urs Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior*, 25(2): 63–87.

Findley, Maureen J., and Harris M. Cooper. 1983. "Locus of Control and Academic Achievement: A Literature Review." *Journal of Personality and Social Psychology*, 44(2): 419–427.

Galeotti, Andrea, and Sanjeev Goyal. 2009. "Influencing the Influencers: A Theory of Strategic Diffusion." *The RAND Journal of Economics*, 40(3): 509–532.

Galeotti, Andrea, Benjamin Golub, and Sanjeev Goyal. 2020. "Targeting Interventions in Networks." *Econometrica*, 88(6): 2445–2471.

Guryan, Jonathan, Sandra Christenson, Ashley Cureton, Ijun Lai, Jens Ludwig, Catherine Schwarz, Emma Shirey, and Mary Clair Turner. 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.

Hadsell, Lester. 2010. "Achievement Goals, Locus of Control, and Academic Success in Economics." *American Economic Review*, 100(2): 272–276.

Heller, Sara B. 2014. "Summer Jobs Reduce Violence among Disadvantaged Youth." *Science*, 346(6214): 1219–1223.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *The Quarterly Journal of Economics*, 132(1): 1–54.

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

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

Kempe, David, Jon Kleinberg, and Éva Tardos. 2003. "Maximizing the Spread of Influence through a Social Network." *KDD '03*, 137–146. New York, NY, USA:Association for Computing Machinery.

Kesavayuth, Dusanee, Dai Binh Tran, and Vasileios Zikos. 2022. "Locus of Control and Subjective Well-Being: Panel Evidence from Australia." *PLoS ONE*, 17(8): e0272714.

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

Kipchumba, Elijah Kipkech, Catherine Porter, Danila Serra, and Munshi Sulaiman. 2021. "Infuencing Youths' Aspirations and Gender Attitudes through Role Models: Evidence from Somali Schools." Working Papers 20210224-002, Texas A&M University, Department of Economics.

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

Luthar, Suniya S., and Thomas J. McMahon. 1996. "Peer Reputation among Inner-City Adolescents: Structure and Correlates." *Journal of Research on Adolescence*, 6(4): 581–603.

Midlarsky, Elizabeth, and Manus Midlarsky. 1973. "Some Determinants of Aiding under Experimentally-Induced Stress." *Journal of Personality*, 41(2): 305–327.

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

Nguyen, Trang. 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." MIT Working Paper.

OECD. 2021. "Purchasing Power Parities (PPP)." https://www.oecd.org/en/data/indicators/purchasing-power-parities-ppp.html.

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

Oreopoulos, Philip, Robert S. Brown, and Adam M. Lavecchia. 2017. "Pathways to Education: An Integrated Approach to Helping At-Risk High School Students." *Journal of Political Economy*.

Paluck, Elizabeth Levy, Hana Shepherd, and Peter M. Aronow. 2016. "Changing Climates of Conflict: A Social Network Experiment in 56 Schools." *Proceedings of the National Academy of Sciences*, 113(3): 566–571.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics*, 12(3): 226–254.

Raven, John C, and John Hugh Court. 1998. *Raven's Progressive Matrices and Vocabulary Scales.* Vol. 759, Oxford pyschologists Press Oxford.

Resnjanskij, Sven, Jens Ruhose, Simon Wiederhold, Ludger Woessmann, and Katharina Wedel. 2024. "Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor Market Prospects." *Journal of Political Economy*, 132(3): 1013–1062.

Richardson, Deborah R., Laura R. Green, and Tania Lago. 1998. "The Relationship between Perspective-Taking and Nonaggressive Responding in the Face of an Attack." *Journal of Personality*, 66(2): 235–256.

Rotter, J. B. 1966. "Generalized Expectancies for Internal versus External Control of Reinforcement." *Psychological Monographs*, 80(1): 1–28.

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

Sacerdote, Bruce. 2011. "Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*. Vol. 3, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 249–277. Elsevier.

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

Sharkey, Patrick. 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.

Shinde, Sachin, Helen A. Weiss, Beena Varghese, Prachi Khandeparkar, Bernadette Pereira, Amit Sharma, Rajesh Gupta, David A. Ross, George Patton, and Vikram Patel. 2018. "Promoting School Climate and Health Outcomes with the SEHER Multi-Component Secondary School Intervention in Bihar, India: A Cluster-Randomised Controlled Trial." *Lancet (London, England)*, 392(10163): 2465–2477.

Shinde, Sachin, Helen A. Weiss, Prachi Khandeparkar, Bernadette Pereira, Amit Sharma, Rajesh Gupta, David A. Ross, George Patton, and Vikram Patel. 2020. "A Multicomponent Secondary School Health Promotion Intervention and Adolescent Health: An Extension of the SEHER Cluster Randomised Controlled Trial in Bihar, India." *PLoS Medicine*, 17(2): e1003021.

Sorrenti, Giuseppe, Ulf Zölitz, Denis Ribeaud, and Manuel Eisner. 2024. "The Causal Impact of Socio-Emotional Skills Training on Educational Success." *The Review of Economic Studies*, rdae018.

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

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

Wolff, Steven B, Anthony T Pescosolido, and Vanessa Urch Druskat. 2002. "Emotional Intelligence as the Basis of Leadership Emergence in Self-Managing Teams." *The Leadership Quarterly*, 13(5): 505–522.

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

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

Zimmerman, David 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		Student-Teachers		
	N	Mean	N	Mean	p-value [NST = ST]
Student Demographics:					
Proportion of Males	14957	0.511	1269	0.516	0.515
Age (months)	14700	153.441	1269	152.880	0.020
No. Siblings	11588	3.619	1130	3.818	0.696
Computer at Home	11589	0.307	1130	0.279	0.263
Internet at Home	11589	0.603	1130	0.588	0.065
Social Climate:					
Behavioral Norms	11578	-0.003	1129	0.035	0.215
Perceived Adult Behavior	11583	-0.007	1130	0.070	0.069
Experienced Anti-social Behavior	11581	0.014	1130	-0.139	0.000
Having a Friend	14803	0.835	1269	1.000	0.000
Friendship Ties (in-degree)	14803	2.962	1269	7.593	0.000
Popularity (in-degree)	14799	1.906	1269	8.352	0.000
Socio-Emotional Well-being:					
Locus of Control	11565	-0.014	1129	0.142	0.000
Mental Well-being	11540	-0.004	1129	0.045	0.109
Perspective Taking	11542	-0.024	1129	0.240	0.000
Impulse Control	11536	0.003	1129	-0.027	0.414
Sense of Belonging	11556	-0.028	1129	0.282	0.000
Sense of Responsibility	11579	-0.008	1130	0.080	0.001
Cognitive Skills:					
Math Score	11589	-0.015	1130	0.154	0.000
Turkish Score	11589	-0.021	1130	0.217	0.000
Fluid IQ (Raven)	11589	-0.019	1130	0.199	0.000
Emotional Intelligence (RMET)	11553	-0.037	1269	0.338	0.000

Notes: Reported statistics use the baseline data. Cognitive test scores and survey measures of social climate (except for network ties) and socio-emotional well-being are standardized to have a zero mean and a unit standard deviation. P-values of the equality test between student-teachers (ST) and non-student-teachers (NST) are obtained by controlling for grade and district fixed effects and clustering standard errors at the school level (unit of randomization).

Table 2: Balance at Baseline - Full Sample

	N	Control Mean	Placebo Mean	Treatment Mean	p-value [T = C]	p-value [T = P]	p-value [C = P]
Student Demographics:							
Proportion of Males	22875	0.500	0.511	0.514	0.172	0.715	0.393
Age (months)	22348	142.900	142.572	142.258	0.595	0.120	0.346
No. Siblings	17867	3.849	3.869	3.744	0.835	0.701	0.846
Computer at Home	17869	0.266	0.252	0.288	0.764	0.901	0.919
Internet at Home	17869	0.550	0.511	0.572	0.758	0.677	0.869
Social Climate:							
Behavioral Norms	17852	0.022	0.036	0.067	0.419	0.637	0.871
Perceived Adult Behavior	17862	0.011	0.042	-0.006	0.494	0.328	0.574
Experienced Anti-social Behavior	17858	1.517	1.539	1.503	0.609	0.572	0.816
Having a Friend	22462	0.811	0.780	0.807	0.824	0.125	0.162
Friendship Ties (in-degree)	22462	2.805	2.649	2.817	0.906	0.138	0.302
Socio-Emotional Well-being:							
Locus of Control	17827	-0.027	-0.030	-0.007	0.822	0.948	0.809
Mental Well-being	17797	-0.003	-0.001	0.031	0.409	0.642	0.978
Perspective Taking	17798	-0.036	-0.084	-0.024	0.915	0.565	0.648
Impulse Control	17790	0.006	0.009	0.030	0.488	0.459	0.847
Sense of Belonging	17817	-0.008	-0.011	0.028	0.461	0.523	0.998
Sense of Responsibility	17854	-0.016	-0.069	-0.074	0.147	0.593	0.503
Cognitive Skills:							
Math Score	17870	0.017	0.034	0.117	0.429	0.637	0.921
Turkish Score	17870	0.025	0.032	0.118	0.460	0.659	0.916
Fluid IQ (Raven)	17870	-0.072	-0.105	-0.028	0.664	0.821	0.919
Emotional Intelligence (RMET)	18055	-0.055	-0.096	-0.038	0.799	0.847	0.998
Joint test p-value: [T vs. C]	0.291						
Joint test p-value: [P vs. C]	0.177						

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

Table 3: Treatment Effects on Disciplinary Flagging

Panel A: Academic Year 1

	Full Sample	Juniors (Grades 5 and 6)	Seniors (Grades 7 and 8)
Treatment	-0.010**	-0.009	-0.010*
	(0.005)	(0.007)	(0.005)
Placebo	0.002	-0.004	0.008
	(0.006)	(0.007)	(0.007)
Control Mean	0.018	0.021	0.016
p-value [TR = P]	0.029	0.374	0.004
WB p-value [TR]	0.065	0.265	0.050
WB p-value [P]	0.722	0.632	0.311
Observations	27028	10802	16226

Panel B: Academic Year 2

	Full Sample	Juniors (Grades 5 and 6)	Seniors (Grades 7 and 8)
Treatment	-0.011**	-0.011	-0.012**
	(0.005)	(0.007)	(0.006)
Placebo	0.015	0.014	0.015
	(0.012)	(0.014)	(0.012)
Control Mean	0.026	0.028	0.025
p-value [TR = P]	0.019	0.060	0.022
WB p-value [TR]	0.048	0.101	0.100
WB p-value [P]	0.287	0.432	0.302
Observations	27841	10818	17023

Notes: The table presents the estimated treatment effects (via OLS) on the probability of disciplinary flagging for the full, junior, and senior samples. The binary dependent variable equals one for students flagged as having extreme behavioral issues and zero otherwise. Regressions control for gender, age in months, baseline cognitive scores, school type fixed effects, and district fixed effects. P-value [TR=P] shows the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 4: Treatment Effects on Disciplinary Flagging - Senior Subgroups

Panel A: Academic Year 1

	Student-Teachers (ST)	Seniors in ST Networks	Seniors outside ST Networks
Treatment	-0.024**	-0.020**	-0.006
	(0.011)	(0.009)	(0.005)
Placebo	0.039**	-0.003	0.005
	(0.017)	(0.011)	(0.006)
Control Mean	0.035	0.029	0.014
p-value [TR = P]	0.000	0.041	0.061
WB p-value [TR]	0.062	0.043	0.269
WB p-value [P]	0.038	0.839	0.473
Observations	1269	5260	4896

Panel B: Academic Year 2

	Student-Teachers (ST)	Seniors in ST Networks	Seniors outside ST Networks
Treatment	-0.035**	-0.014*	-0.012
	(0.016)	(0.008)	(0.008)
Placebo	-0.003	0.005	0.006
	(0.021)	(0.012)	(0.012)
Control Mean	0.056	0.034	0.032
p-value [TR = P]	0.095	0.064	0.102
WB p-value [TR]	0.057	0.085	0.214
WB p-value [P]	0.868	0.700	0.660
Observations	565	2381	4510

Notes: The table presents the estimated treatment effects (via OLS) on the probability of disciplinary flagging for student-teachers and student in and outside of student-teacher networks. The binary dependent variable equals one for students flagged as having extreme behavioral issues and zero otherwise. Regressions control for gender, age in months, baseline cognitive scores, school type fixed effects, and district fixed effects. P-value [TR=P] shows the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 5: Treatment Effects on Social Support Networks

Panel A: Support Ties directed to Student-Teachers

	Academic Year 1		Academic Year 2	
	From Juniors	From Seniors	From Juniors	From Seniors
Treatment	0.129***	0.476*	0.134**	0.346
	(0.044)	(0.278)	(0.067)	(0.390)
Placebo	0.039	0.556*	0.020	1.502***
	(0.053)	(0.295)	(0.064)	(0.525)
Control Mean	0.194	5.032	0.235	3.469
p-value [TR = P]	0.070	0.784	0.127	0.018
WB p-value [TR]	0.010	0.155	0.075	0.457
WB p-value [P]	0.567	0.129	0.782	0.030
Observations	1269	1269	565	565

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

	Academic Year 1		Academic Year 2	
	From Juniors	From Seniors	From Juniors	From Seniors
Treatment	-0.028	0.098	0.062**	0.007
	(0.021)	(0.165)	(0.028)	(0.306)
Placebo	-0.009	0.222	-0.001	0.413
	(0.025)	(0.175)	(0.037)	(0.250)
Control Mean	0.175	3.581	0.123	2.608
p-value [TR = P]	0.457	0.516	0.138	0.118
WB p-value [TR]	0.263	0.615	0.064	0.994
WB p-value [P]	0.761	0.279	0.970	0.192
Observations	5247	5247	2381	2381

Notes: The table presents the estimated treatment effects on the number of support ties formed within the school. Panel A presents ties directed to student-teachers, and Panel B 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 B replicates Panel A for student-teacher networks. Reported estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for respective baseline outcomes, gender, age in months, baseline cognitive scores, class size, share of boys in class, session and school type fixed effects, school size, and district fixed effects. P-value [TR=P] shows the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 6: Treatment Effects on Admission to Selective High Schools

Panel A: Academic Year 1 (Cohort 2022)

	Student-Teachers (ST)	Seniors in ST Networks	Seniors outside ST Networks
Treatment	0.084***	0.019	-0.009
	(0.028)	(0.013)	(0.012)
Placebo	0.042	0.007	-0.024**
	(0.037)	(0.013)	(0.010)
Control Mean	0.092	0.058	0.056
p-value [TR = P]	0.249	0.327	0.251
WB p-value [TR]	0.007	0.241	0.515
WB p-value [P]	0.320	0.651	0.036
Observations	573	2270	2048

Panel B: Academic Year 2 (Cohort 2023)

	Student-Teachers (ST)	Seniors in ST Networks	Seniors outside ST Networks
Treatment	0.074**	0.027*	0.020*
	(0.028)	(0.014)	(0.011)
Placebo	0.010	0.008	0.009
	(0.027)	(0.015)	(0.018)
Control Mean	0.130	0.057	0.043
p-value [TR = P]	0.019	0.136	0.522
WB p-value [TR]	0.015	0.089	0.113
WB p-value [P]	0.711	0.571	0.632
Observations	633	2685	2624

Notes: The table presents the estimated treatment effects on the probability of admission to selective high schools for senior subgroups who graduated in 2022 (Panel A) and 2023 (Panel B). The dependent variable is a binary variable, which equals 1 if the student is admitted to a selective high school and zero otherwise. Reported estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for gender, age in months, baseline cognitive scores, class size, share of boys in class, school type fixed effects, and district fixed effects. P-value [TR=P] shows the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 7: Treatment Effects on Academic Outcomes - Senior Subgroups

Panel A: Academic Year 1

	Student-Teachers (ST)		Seniors in ST Networks		Seniors outside ST Networks	
	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.163**	0.096	0.123***	0.020	0.107**	0.012
	(0.075)	(0.070)	(0.041)	(0.032)	(0.053)	(0.041)
Placebo	0.091	0.103	0.040	-0.017	0.116*	0.020
	(0.097)	(0.068)	(0.064)	(0.050)	(0.065)	(0.047)
p-value [TR = P]	0.409	0.904	0.225	0.413	0.888	0.873
WB p-value [TR]	0.048	0.235	0.006	0.571	0.077	0.772
WB p-value [P]	0.412	0.158	0.573	0.770	0.116	0.693
Observations	991	991	3975	3975	2636	2636

Panel B: Academic Year 2

	Student-Teachers (ST)		Seniors in ST Networks		Seniors outside ST Networks	
	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	-0.014	0.038	-0.062	0.134**	0.035	0.075*
	(0.111)	(0.119)	(0.058)	(0.055)	(0.058)	(0.042)
Placebo	-0.134	0.034	0.036	0.083	-0.080	0.041
	(0.136)	(0.130)	(0.075)	(0.066)	(0.066)	(0.052)
p-value [TR = P]	0.377	0.973	0.160	0.432	0.060	0.468
WB p-value [TR]	0.893	0.753	0.330	0.036	0.589	0.096
WB p-value [P]	0.385	0.788	0.654	0.273	0.293	0.441
Observations	356	356	1363	1363	2298	2298

Notes: The table presents estimated treatment effects on academic test scores for senior subgroups. The dependent variables are standardized math and Turkish test scores. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline scores, gender, age in months, baseline cognitive scores, class size, share of boys in class, school type fixed effects, and district fixed effects. P-value [TR=P] shows the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figures

Figure 1: Study Timeline

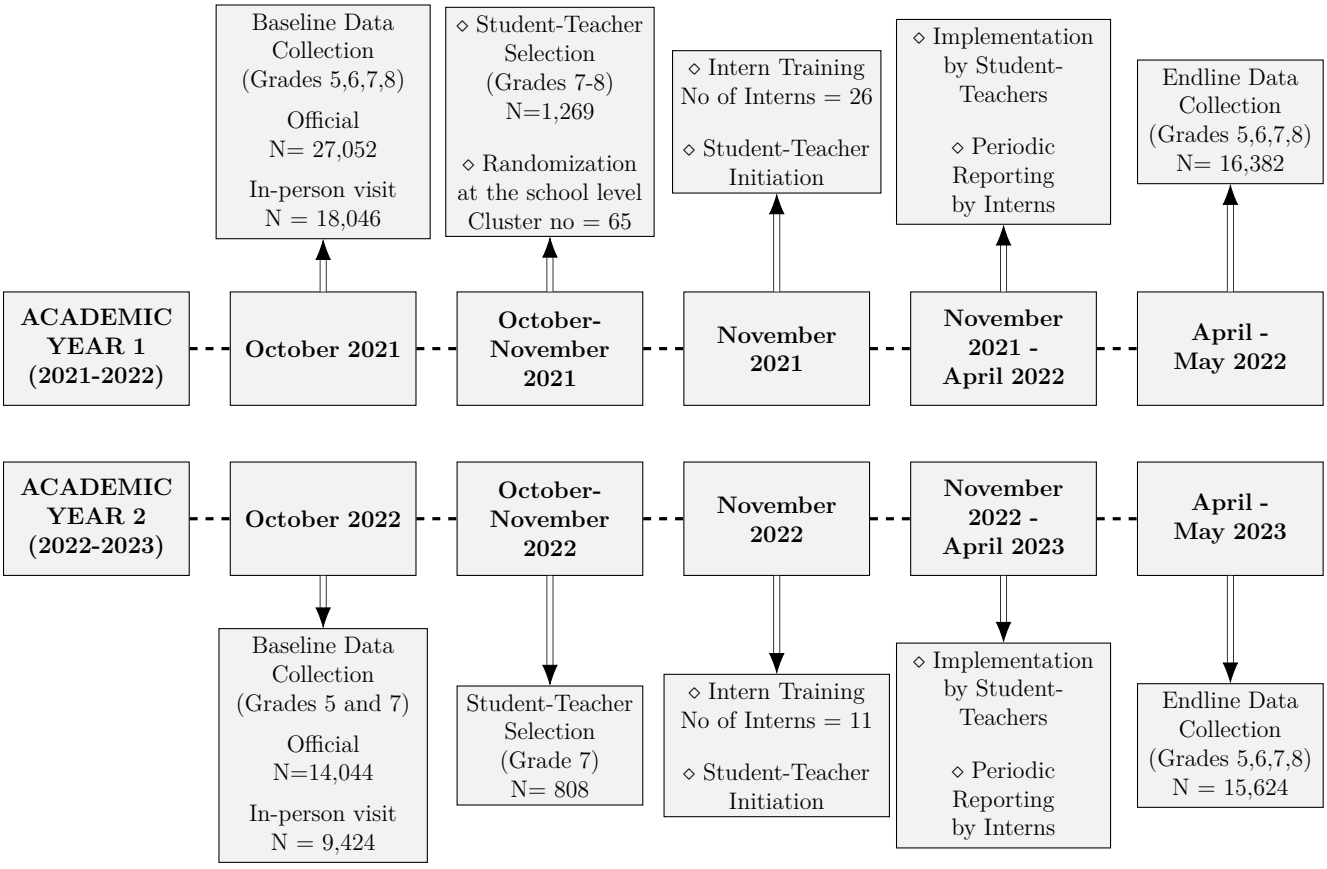
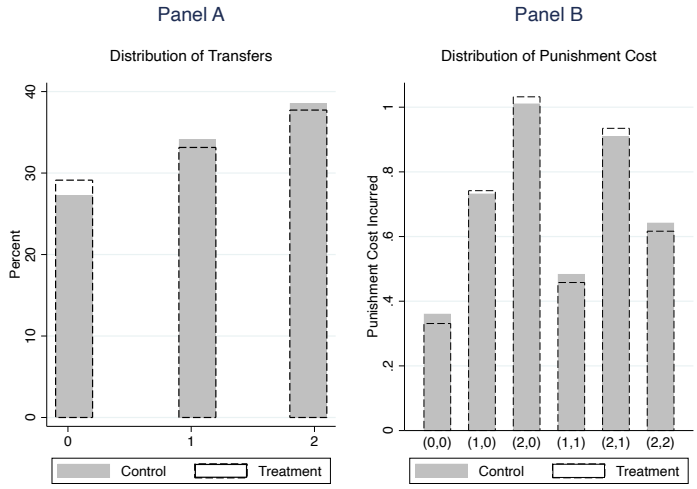
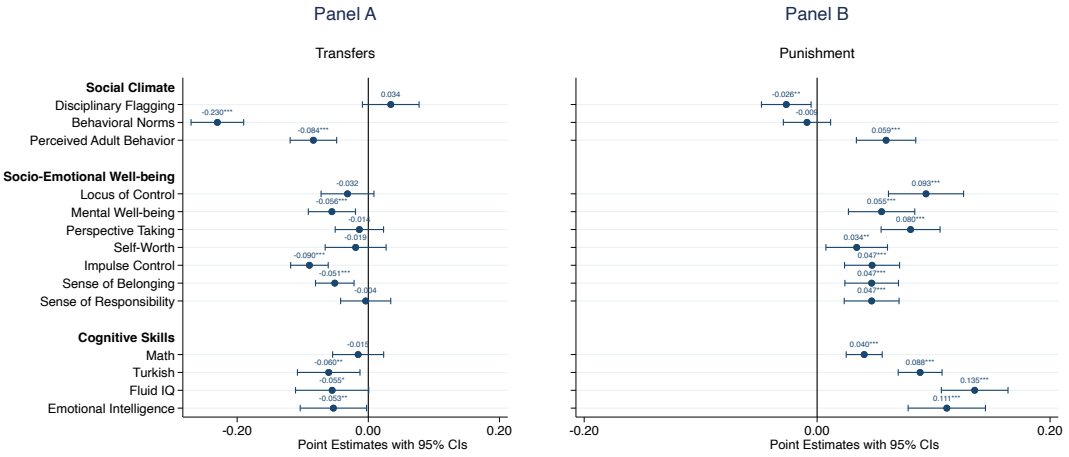


Figure 2: Distribution of Decisions in the Third-Party Punishment Game



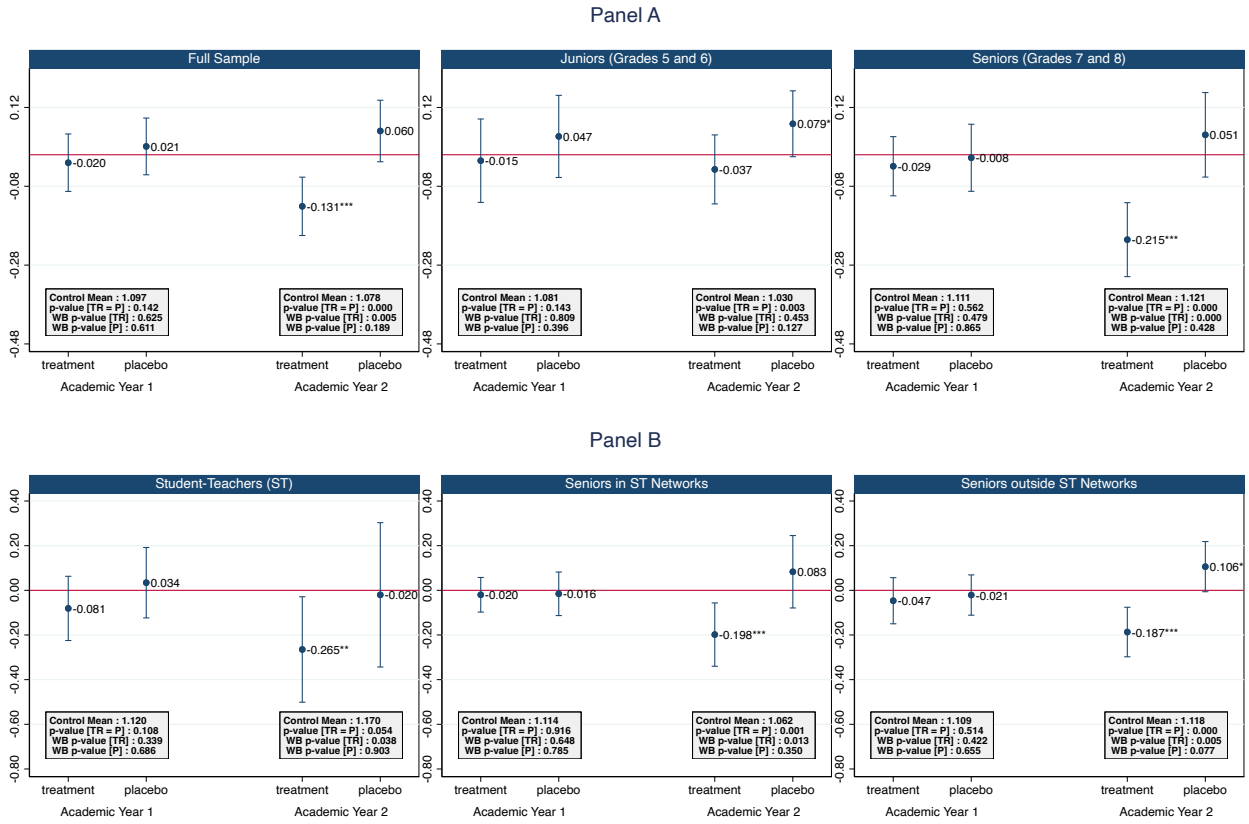
Notes: The figure depicts the distribution of transfers (0, 1, and 2) and punishment cost incurred for each transfer scenario (0,0), (1,0), (2,0), (1,1), (2,1), and (2,2) separately for control and treatment. The analysis uses the data collected in the first year of the program.

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



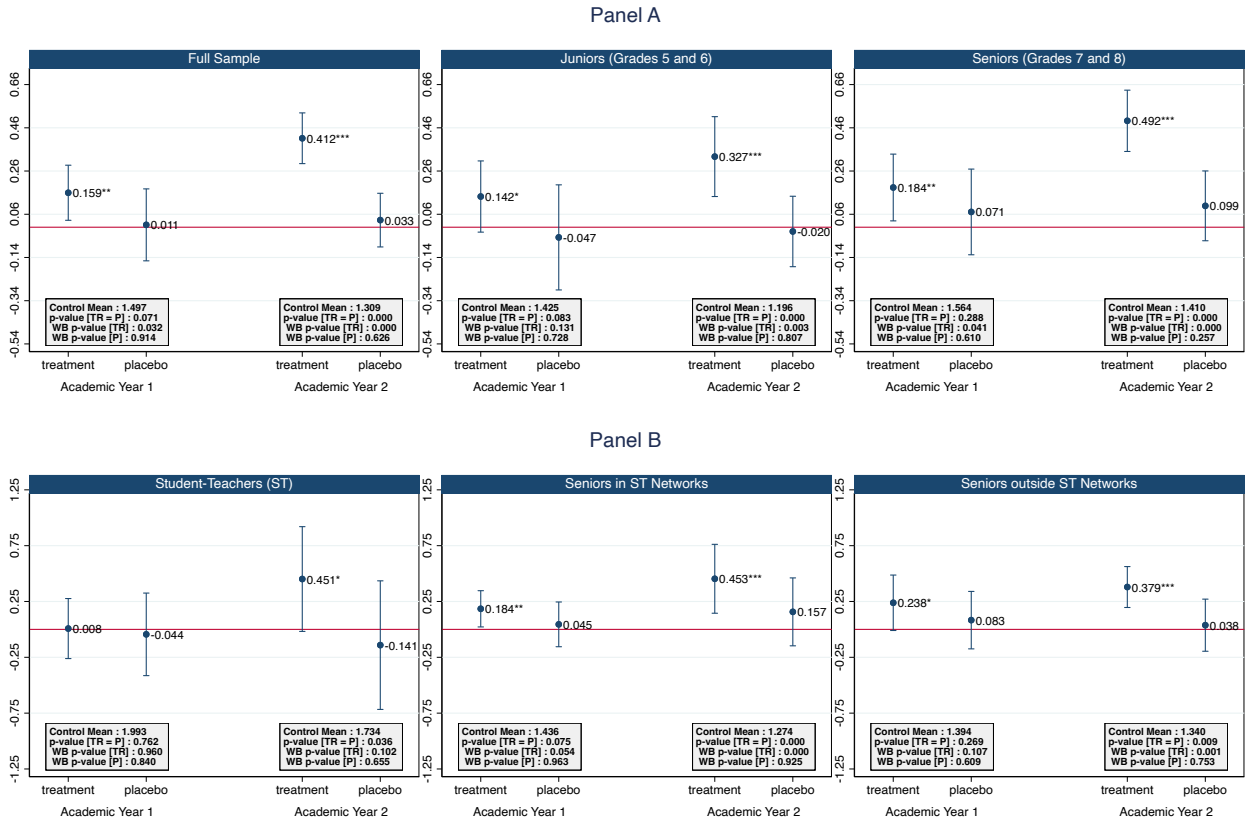
Notes: The figure depicts the predictive validity of the decisions made in the third-party punishment game. Panel A presents the OLS coefficients from the regressions of the number of transfers on the indicators of perceived social environment, socio-emotional well-being, and cognitive skills. Panel B depicts the OLS coefficients from the regressions of incurred punishment cost on the same indicators. The analysis combines both years of the data and uses only the control group. Standard errors are clustered at the school level. Regressions control district fixed effects. Asterisks indicate that the coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 4: Treatment Effects on Anti-Social behavior



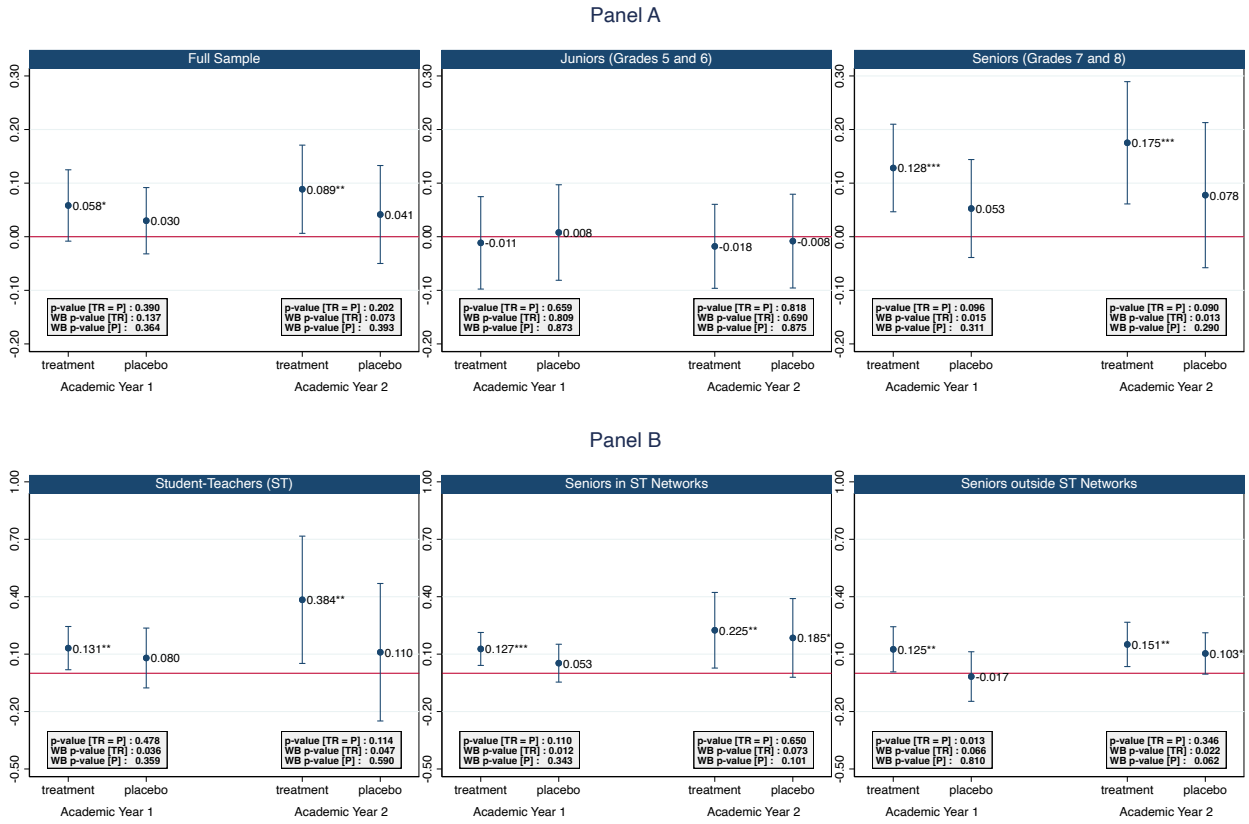
Notes: The figure plots the estimated treatment effects on the transfer decisions in the third-party punishment game. The dependent variable is the number of tokens transferred from the opponent. Reported estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for gender, age in months, baseline cognitive scores, class size, share of boys in class, school type fixed effects, and district fixed effects. P-value [TR=P] shows the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization), and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 5: Treatment Effects on Punishment of Unequal Transfers



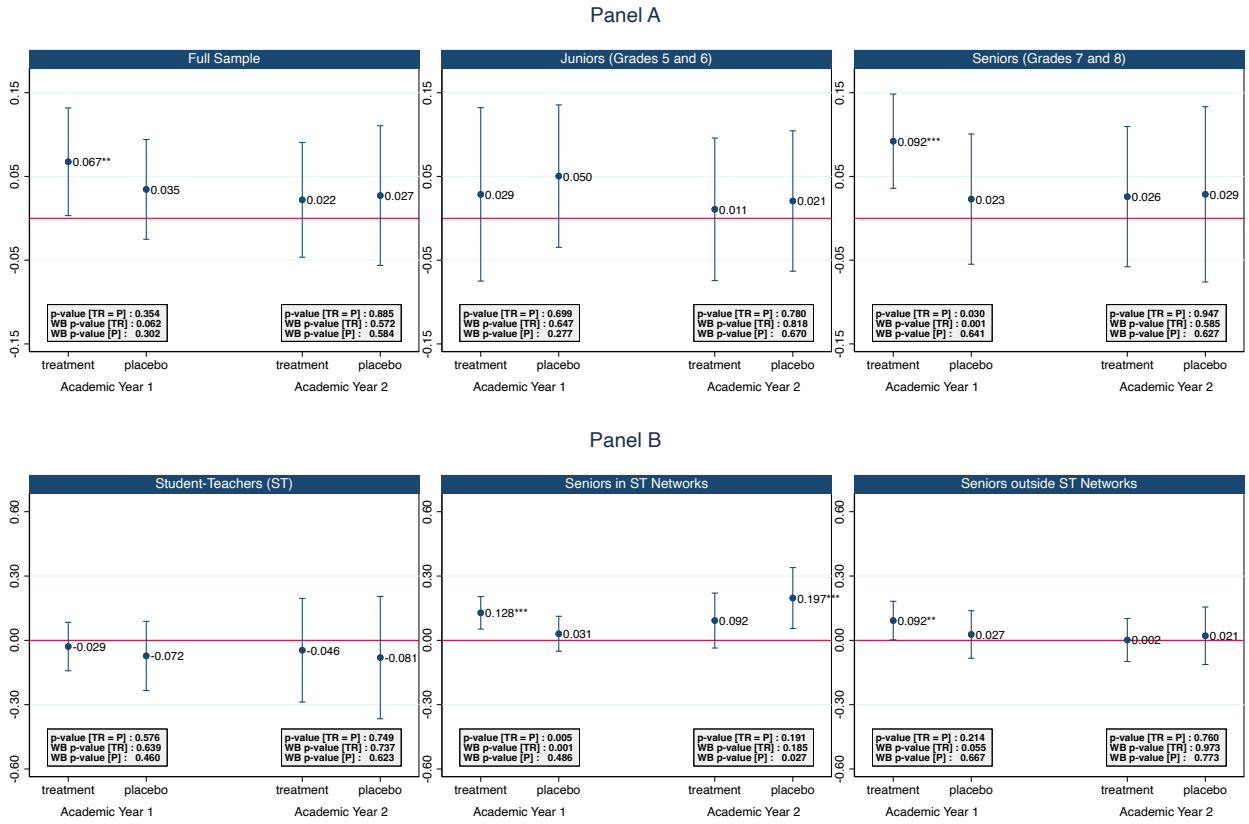
Notes: The figure plots the estimated treatment effects on punishment decisions in the third-party punishment game. The dependent variable is the number of tokens forgone to punish unequal transfers. Reported estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for gender, age in months, baseline cognitive scores, class size, share of boys in class, school type fixed effects, and district fixed effects. P-value [TR=P] presents the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization), and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 6: Treatment Effects on Behavioral Norms



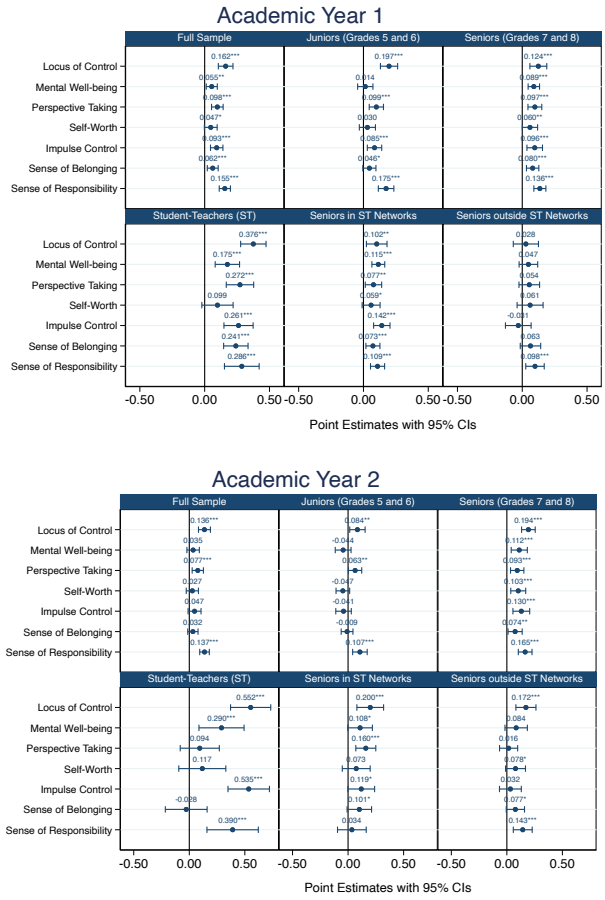
Notes: The figure plots the estimated treatment effects on the perceived behavioral norms. The dependent variable is in standard deviation units, and the estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for baseline outcome, gender, age in months, baseline cognitive scores, school type fixed effects, school size, and district fixed effects. P-value [TR=P] presents the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization), and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 7: Treatment Effects on Perceived Adult Behavior



Notes: The figure plots the estimated treatment effects on the perceived adult behavior. The dependent variable is in standard deviation units, and the estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for baseline outcome, gender, age in months, baseline cognitive scores, school type fixed effects, school size, and district fixed effects. P-value [TR=P] presents the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization), and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 8: Treatment Effects on Socio-emotional Well-being



Notes: The figure depicts the estimated treatment effects on socio-emotional well-being for the first and the second year of the program, separately. 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 standard deviation units. Regressions control for baseline values of the corresponding outcome when available, gender, age in months, baseline cognitive scores, school type fixed effects, and district fixed effects.

Appendix

Tables

Table A1: Correction for Multiple Hypothesis Testing

	Original	Romano Wolf
Full Sample:		
Social Climate	0.003	0.044
Socio-emotional Outcomes	0.000	0.014
Achievement	0.078	0.134
Juniors:		
Social Climate	0.009	0.092
Socio-emotional Outcomes	0.010	0.092
Achievement	0.359	0.351
Seniors:		
Social Climate	0.013	0.098
Socio-emotional Outcomes	0.000	0.006
Achievement	0.029	0.110
Student-Teachers:		
Social Climate	0.001	0.012
Socio-emotional Outcomes	0.000	0.010
Achievement	0.012	0.046
Seniors in Student-Teacher Networks:		
Social Climate	0.001	0.016
Socio-emotional Outcomes	0.000	0.002
Achievement	0.007	0.028
Seniors outside Student-Teacher Networks:		
Social Climate	0.078	0.234
Socio-emotional Outcomes	0.055	0.234
Achievement	0.052	0.234

Notes: The table provides original p-values and p-values corrected for multiple hypothesis testing using Romano-Wolf algorithm. The analysis combines both years of the data. Outcome indices are constructed for each outcome group by taking the average of the following outcomes: Social climate: disciplinary flagging, transfers and cost of punishment in the third-party punishment game, behavioral norms and perceived adult behavior, total number of support ties with juniors and total number of support ties with seniors (the latter two only included for student-teachers and their networks), (ii) socio-emotional outcomes: Locus of control, mental well-being, perspective-taking, self-worth, impulse control, sense of belonging, sense of responsibility, (iii) achievement: standardized math and Turkish test scores and admission status to selective high schools (the latter is only for final year students in 2022 and 2023). The number of replications is set to 500.

Table A2: Treatment Effects on Admission to Selective High Schools: Robustness for 2022

Panel A: Academic Year 1 (Cohort 2022)

	Student-Teachers (ST)	Seniors in ST Networks	Seniors outside ST Networks
Treatment	0.060**	0.011	-0.011
	(0.029)	(0.014)	(0.012)
Placebo	0.038	0.009	-0.022**
	(0.039)	(0.013)	(0.011)
Control Mean	0.092	0.058	0.056
p-value [TR = P]	0.570	0.863	0.352
WB p-value [TR]	0.061	0.541	0.406
WB p-value [P]	0.453	0.548	0.052
Observations	611	2415	2164

Panel B: Academic Year 2 (Cohort 2023)

	Student-Teachers (ST)	Seniors in ST Networks	Seniors outside ST Networks
Treatment	0.074**	0.027*	0.020*
	(0.028)	(0.014)	(0.011)
Placebo	0.010	0.008	0.009
	(0.027)	(0.015)	(0.018)
Control Mean	0.130	0.057	0.043
p-value [TR = P]	0.019	0.136	0.522
WB p-value [TR]	0.015	0.089	0.113
WB p-value [P]	0.711	0.571	0.632
Observations	633	2685	2624

Notes: The table presents the estimated treatment effects on the probability of admission to selective high schools for senior subgroups who graduated in 2022 (Panel A) and 2023 (Panel B). The dependent variable is a binary variable, which equals 1 if the student is admitted to a selective high school and zero otherwise. Reported estimates are obtained from ordinary least squares (OLS) regressions. Regressions control for gender, age in months, baseline cognitive scores, school type fixed effects, and district fixed effects. The 2022 analysis assumes that no student in the four missing treatment schools gained admission to a selective high school; therefore, the analysis was conducted using 63 schools, with only one placebo and one control school missing. P-value [TR=P] presents the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table A3: Treatment Effects on Academic Outcomes

Panel A: Academic Year 1

	Full Sample		Juniors (Grades 5 and 6)		Seniors (Grades 7 and 8)	
	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	0.075**	0.028	0.045	0.038	0.112***	0.022
	(0.035)	(0.041)	(0.052)	(0.061)	(0.038)	(0.030)
Placebo	0.014	-0.003	-0.028	-0.028	0.056	0.028
	(0.047)	(0.050)	(0.062)	(0.072)	(0.052)	(0.040)
p-value [TR = P]	0.158	0.443	0.084	0.204	0.284	0.885
WB p-value [TR]	0.062	0.546	0.466	0.613	0.008	0.478
WB p-value [P]	0.793	0.962	0.706	0.744	0.349	0.551
Observations	16402	16402	8143	8143	8259	8259

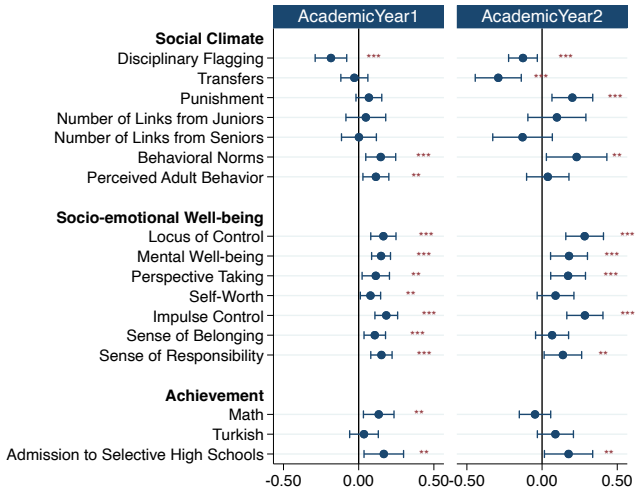
Panel B: Academic Year 2

	Full Sample		Juniors (Grades 5 and 6)		Seniors (Grades 7 and 8)	
	Math	Turkish	Math	Turkish	Math	Turkish
Treatment	-0.008	0.059	0.009	0.055	-0.014	0.062
	(0.045)	(0.050)	(0.061)	(0.061)	(0.040)	(0.046)
Placebo	-0.015	0.043	0.019	0.049	-0.052	0.022
	(0.055)	(0.061)	(0.073)	(0.083)	(0.049)	(0.047)
p-value [TR = P]	0.870	0.715	0.860	0.925	0.333	0.212
WB p-value [TR]	0.895	0.296	0.891	0.432	0.780	0.230
WB p-value [P]	0.834	0.556	0.837	0.623	0.355	0.684
Observations	15749	15749	7791	7791	7958	7958

Notes: The table presents estimated treatment effects on academic test scores for full, junior, and senior samples. The dependent variables are standardized math and Turkish test scores. Reported estimates are obtained from ordinary least squares (OLS) regressions. The regressions control for respective baseline scores, gender, age in months, baseline cognitive scores, class size, share of boys in class, school type fixed effects, and district fixed effects. P-value [TR=P] presents the p-value from the test of equality of treatment and placebo effects. "WB p-value [TR]" and "WB p-value [P]" stand for wild bootstrapped p-value for the estimated treatment and placebo effects, respectively. Standard errors are clustered at the school level, and all equality tests use clustered-robust inference. Asterisks indicate that the coefficient is statistically significant at the 1% , 5% , and 10% levels.

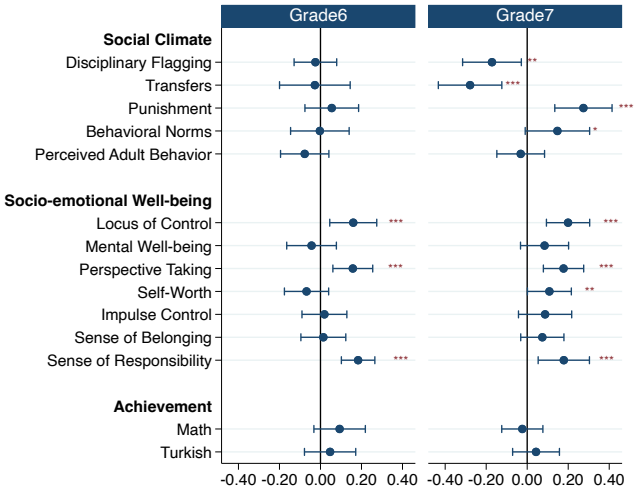
Figures

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



Notes: 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: Treatment Effects on Student-Teachers and Seniors in Student-Teacher Networks: Grades 6 and 7



Notes: The table presents the estimated treatment effects for the sample of students who were exposed to the program's first year as 6th graders (juniors) and then selected as student-teachers (or as student-teacher networks)in the second year of the program. The first column shows the estimated effects when the students were in grade 6 (2021-2022 academic year), the second column is when the same students were in grade 7 (2022-2023 academic year). The 95% confidence intervals are based on standard errors clustered at the school level (unit of randomization). Regressions control for baseline values of the corresponding outcome when available, gender, age in months, baseline cognitive scores, school type fixed effects, and district fixed effects. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.