CESifo WORKING PAPERS

10975
2024

February 2024

Peer Effects on Violence: Experimental Evidence from El Salvador

Lelys Dinarte-Diaz

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10975*

Peer Effects on Violence: Experimental Evidence from El Salvador

Abstract

Globally, 150 million adolescents report being victims of or engaging in peer-to-peer violence in and around school. One strategy to reduce this risk is to occupy youth in after-school programs (ASP). Yet, the question remains: how does peer group composition affect the effectiveness of an ASP? I address this question by randomly assigning youths to either a control, homogeneous, or heterogeneous peer group within an ASP implemented in El Salvador. I find that, unlike homogeneous groups, heterogeneous peer groups do help students avoid violence. These results are relevant to public policy discussions on optimal group composition for violence reduction programs.

JEL-Codes: I290, K420, Z130.

Keywords: peer effects, violence, integration, tracking, after-school programs.

Lelys Dinarte-Diaz
The World Bank, Washington DC / USA
ldinartediaz@worldbank.org

February 20, 2024
This paper has benefited greatly from the helpful feedback of Matias Busso, Jennifer Doleac, Pascaline Dupas, Francisco Gallego, Sara Heller, Kirabo Jackson, Jeanne Lafortune, David McKenzie, Claudia Martinez, Owen Ozier, Tomas Rau, Rodrigo Soares, Micaela Sviatschi, Jose Tessada, and Juan Vargas. I am indebted to Pablo Egana for his expertise in collecting and analyzing the neurophysiological data. I also appreciate the comments I received from those who participated in the conferences and seminars where I presented this paper. I would also like to express my gratitude to my research assistant Juan Duque for his support. I am truly thankful to Glasswing International for implementing the intervention and to the public school principals, teachers, students, and instructors who participated in the experiment. This study was funded by CEDLAS-IDRC (Canada) and the World Bank. The author is responsible for the content of this study, which does not necessarily reflect the views of her employer, funding agencies, or governments. This research is registered in the AEA RCT Registry under the unique identifying number AEARCTR-0001602. The Institutional Review Board (IRB) at the Pontificia Universidad Catolica de Chile reviewed and approved this study's protocol on April 2016 under ID 160314001.

1 Introduction

Evidence has shown that peer groups have powerful and lasting impacts on an individual's economic decisions. Specifically, exposure to different peer groups during childhood or adolescence can influence a person's educational decisions, (risky) behaviors, and socioemotional skills during adolescence (Zárate, 2023; Anand and Kahn, 2023; Feng et al., 2022; Alan et al., 2020; Billings et al., 2019; Billings and Hoekstra, 2019; Rao, 2019; Bursztyn and Jensen, 2015) as well as their human capital accumulation (Golsteyn et al., 2021; Carrell et al., 2018), criminal activity (Billings and Hoekstra, 2019; Damm and Dustmann, 2014; Bayer et al., 2009), and labor market outcomes (Carrell et al., 2018) throughout adulthood. In the public policy arena, peer influence is so important that most educational and violence-reduction programs targeting adolescents are implemented in group settings. Despite the abundance of evidence that exposure to peers with different levels of academic performance or income within these programs can affect their impact, and with the exception of Davis and Heller (2020), very few studies have rigorously analyzed the effects of group composition on behaviors, economic outcomes, and program effectiveness based the participants' risk levels.

This paper seeks to understand the relevance of *group composition by risk for violence* within the context of an after-school program (ASP) that aims to reduce the violent behaviors of school-aged adolescents.[1] This ASP consists of clubs implemented after school but within school facilities in El Salvador from April to mid-October 2016. The enrollees took part in two sessions per week, each of which lasted 1.5 hours. Every session combined: (i) a discussion oriented toward developing the children's social skills, promoting awareness of certain behaviors, and disrupting negative behavioral patterns to foster new ones; and (ii) club curricula that included activities such as scientific experiments, sports, and art. Volunteers with Glasswing International, a local NGO working in Central America and Mexico, implemented the intervention. The study sample includes 1,056 students

[1]The ASP's average effects on violent behaviors, attitudes toward school and learning, academic performance, and emotion regulation are presented in Dinarte-Diaz and Egana-delSol (2023).

between the ages of 10–16 years from five public schools located in areas where children are at a high risk of becoming victims of or engaging in criminal activities.

Existing evidence from the education domain does not indicate which type of group composition is preferable. On the one hand, some studies find that heterogeneous groups are more beneficial because interacting with "high-performing" peers improves outcomes for more disadvantaged individuals by enhancing their learning experience.[2] However, other strands of the literature find that grouping individuals with similar peers can generate better results because individuals prefer to interact with those who are similar to them or because homogeneous grouping is conducive to specialized instruction.[3] In light of this evidence, there were two main things to consider when designing my program. The first was who to target. Given that the program was oversubscribed, one option was to select only at-risk youths, for who the ASP is the most relevant. Targeting this population only, however, would result in the creation of relatively homogeneous peer groups. If, however, the influence of non-at-risk youth is important to help at-risk youth avoid becoming involved in violence, then selecting only at-risk students could undermine the program's effectiveness. In this case, heterogeneous groups made up of non-at-risk and at-risk students would be most beneficial. Conversely, though, non-at-risk youth's increased exposure to at-risk youth in heterogeneous groups could potentially have a negative impact on the former.

To understand the relevance and magnitude of the aforementioned peer effects by group composition, I collaborated with the NGO to randomize the study participants from the five participating schools into three groups:[4] a homogeneous peer group, a heterogeneous peer group, and a control group based on the students' initial risk level for violence.[5] Students in the homogeneous peer group were further separated into two sub-

[2]See, for example, the results found in Alan et al. (2020); Rao (2019); Oreopoulos et al. (2017); Lafortune et al. (2016); Griffith and Rask (2014); and Lavy et al. (2012).

[3]See, for example, the evidence presented in Duflo et al. (2011); Carrell et al. (2013); Girard et al. (2015); and Goethals (2001).

[4]In this paper, the terms "homogeneous" and "tracking" are interchangeable, as are "heterogeneous," "diverse," and "integration."

[5]This variable is a proxy for how vulnerable a student is to violent behaviors, which was determined using individual-level violence determinants such as gender, age, parental supervision, etc. In this sense, it

groups by their risk level: students whose risk for violence was higher (lower) than the median were assigned to a club with peers with a higher (lower) risk. Participants in the control group did not participate in the ASP; they simply returned home after school. Randomization ensured that the peer group size and club categories were balanced across both treatments.

The experimental design of this study allowed me not only to estimate the average and heterogeneous treatment effects of group composition in terms of risk for violence but also to exploit the discontinuity in the predicted risk level to estimate the treatment effects on the marginal students. I did this by considering two students with a risk level at the median, one of whom was assigned to a peer group where she was the most at-risk for violence and the other of whom was assigned to a peer group where she was the least at-risk. In addition, the random formation of groups in this study circumvents various issues that could otherwise arise concerning identification of peer effects, such as the reflection problem, the strong assumption of the separability of peer composition, and other confounding effects within groups (Manski, 1993; Angrist, 2014).

This paper focuses on the impacts of exposure to a particular composition of peers on outcomes related to violence and behaviors, attitudes towards school and learning, emotion regulation, and academic performance. To measure these outcomes, I use data from Dinarte-Diaz and Egana-delSol (2023) obtained from a self-reported survey, electroencephalograms, and administrative data. Before the intervention, we collected the enrolled students' self-reported data on personal and family characteristics and used this information to predict the measure of each participant's risk for violence. Immediately after the ASP ended, we conducted a follow-up self-reported survey that included questions to measure violent behaviors and crime as well as attitudes toward school and learning. Then, following Egana-delSol et al. (2023), we combined this self-reported information with neurophysiological data, specifically measures of stress and emotion regulation, from a random subsample of enrolled students, which we collected using a low-cost, portable electroencephalograms in a lab setting in the field. Finally, we also gathered the

can also be interpreted as a measure to identify at-risk students.

students' administrative records on grades, behavioral reports, and absenteeism from the schools both before and after the intervention took place.

Based on the data and the experimental design described above, I report three sets of results. First, assigning students with different risk levels for violence to a heterogeneous group of peers reduces negative behaviors relative to assigning them to a homogeneous group or to their not participating in the ASP at all. In addition, a heterogeneous peer group improves the participants' emotion regulation relative to student who did not participate in the ASP. These greater improvements from the heterogeneous treatment relative to the homogeneous treatment seem to be driven by reductions in the probability of having bad behavior reports for the violence and behavior outcome and by reductions in stress levels for the emotion regulation outcome.

Second, heterogeneity analyses by the student's initial risk level for violence show that a homogeneous composition of peers is detrimental relative to a heterogeneous composition, irrespective of the students' risk level at baseline. For students with a higher risk for violence, I estimate detrimental effects on violence and behaviors, attitudes toward school and learning, and emotion regulation when they are assigned to the homogeneous treatment relative to the heterogeneous treatment. Notably, the negative effect of group composition on emotion regulation is driven by increases in the students' measure of stress. Similarly, I also find that a homogeneous composition of peers can negatively affect risk levels for violence and negative behaviors among students with an initial low risk level for violence compared to students with a similarly low risk level treated in a group of more diverse peers.

Lastly, I study the treatment's effects on marginal students who fall just above or below the median of the risk for violence distribution within each stratum. By virtue of my design, very similar students around this cutoff were assigned either to a homogeneous high- or a homogeneous low-risk group. By exploiting the discontinuity around the median and using only the students assigned to the homogeneous group, I find evidence that, when assigned to the high risk group, marginal students are affected negatively on their violence and behavior, attitudes toward school and learning, and academic performance.

The three sets of results described above align with existing evidence that indicates that interacting with a diverse group of peers can result in different learning experiences (Lafortune et al., 2016). Alternatively, these results support the rainbow model of peer effects, according to which all individuals benefit from exposure to a more heterogeneous peer group (Hoxby, 2000). In other words, being in a diverse group allows at-risk students to be exposed and positively influenced by students with a low risk for violence who model appropriate social skills and good behaviors. Similarly, low-risk children benefit from witnessing the bad behaviors that they should avoid. However, the negative effects on the marginal participants in the homogeneous group also indicate that being exposed to a more significant share of at-risk peers can have the opposite effect. This implies, therefore, that there is an optimal peer combination that maximizes the program's overall impact.

I examine the robustness of the results in the following ways. First, since randomization was carried out at the individual level, the presence of spillovers in my experimental design is a potential concern. For this reason, I estimate local linear nonparametric kernel estimations and do not find any stark variance in the treatment effects between the treatments and control groups across most of the main outcomes based on the share of treated students at the grade level. Thus, I am confident that, if anything, spillovers are of the same magnitude for the control, heterogeneous, and homogeneous groups. Second, I verify the robustness of the results to the inclusion of control variables by using a double LASSO algorithm and by excluding all control variables. Third, I explore issues related to sample selection due to survey attrition and show that the results are robust to attrition. Lastly, I point out that, although some measures of violence and behaviors as well as attitudes towards school and learning are self-reported, experimenter demand effects are not of concern because the estimated effects based on these self-reports and other proxies for these measures from administrative data are in the same direction.

This paper provides causal evidence that contributes to the discussion on which strategy—tracking or integration—is optimal for assigning participants to a social intervention. To my knowledge, this paper is the first of its kind to present an experimental evaluation of risk-based peer dynamics and the impact of group composition on the effectiveness of a social violence prevention program in a developing and highly violent country. El Salvador is a poignant and relevant context to conduct this analysis because violence is pervasive in the society, including in schools. In fact, in 2019 El Salvador was one of the world's top five deadliest places for young boys to live, where 2 out of every 3 youth deaths, which affect both victims and perpetrators, are due to interpersonal violence (World Bank, 2023). Moreover, 18% of students in El Salvador reported dropping out of school because of delinquency within schools or in the surrounding neighborhoods (MINED, 2019).

The effects of integration on violence and behavior outcomes accord with the body of microlevel evidence that suggests that these effects likely stem from the interaction between diverse individuals within groups.[6] My results are similar to those of Alan et al. (2020) and Davis and Heller (2020). The former evaluate the impact of an educational program that aims to build social cohesion in ethnically mixed schools and finds that the program significantly lowers peer violence and victimization on school premises. The latter emphasize the potential gains of more flexible approaches to study treatment heterogeneity within the context of a summer job program targeted to at-risk youth. My study is novel not only in its modification of the composition regarding the participants' risk for violence but also in its analysis of peer effects on additional noncognitive outcomes such as violence, misbehavior, and attitudes toward school and learning, all of which are important in developing countries.

A growing body of evidence finds that tracking individuals by specific characteristics is beneficial. Theoretically, Lazear (2001) shows that, amid different levels of classroom disruption, tracking by type maximizes total school output. Some empirical papers also find that school tracking can improve academic results, with greater effects for low performers (Duflo et al., 2011; Cortes and Goodman, 2014; Girard et al., 2015). A plausible explanation for the differences between my results and those reported in the tracking lit-

[6]See Sacerdote et al. (2011) for a summary of the literature on peer effects on student outcomes in educational settings. Further recent evidence for peer effects on non-cognitive outcomes is provided by Anand and Kahn (2023); Feng et al. (2022); Billings and Hoekstra (2019); Billings et al. (2019); Rao (2019); Gong et al. (2019); Fletcher et al. (2019); and Fletcher and Ross (2012).

erature is the lack of incentives for instructors to adapt club curricula to their group's specific needs. In fact, my results accord with Duflo et al. (2011)'s model's predictions for the special case in which instructors do not respond to group composition because their effort function is a constant, or because the cost of their effort is zero below the target levels to which teachers orient instruction. Under this assumption, tracking by risk for violence worsens outcomes for those above the median of the risk distribution in the group to which they are assigned and improves the outcomes for those below the median.

2 Experimental Design

2.1 Intervention: After-School Clubs

I conducted this study within the context of an after-school program (ASP) that was implemented by the NGO Glasswing International in five public schools located in highly violent communities in El Salvador. I partnered with Glasswing International to design and launch the experimental evaluation in order to (i) measure the impact of this ASP on violent behaviors, attitudes toward school and learning, academic performance, and emotion regulation, and (ii) identify which group composition enhances the ASP's effectiveness. In Dinarte-Diaz and Egana-delSol (2023), we achieve the first aim.[7] In the current paper, I address the second objective by implementing an experimental design that I describe in detail in Section 2.2. In this section, I highlight only the curricula and enrollment process to explain the experimental design and its results.[8]

The intervention was implemented over a 20-week period between April and October 2016. During this time, each club met in person twice a week for approximately 1.5 hours per session within school facilities immediately after school. The intervention was

[7] Our main estimations indicate that this ASP not only improved students' attitudes toward school but also enhanced their academic performance and reduced their bad behavior at school. We also find evidence that the students' increased ability to control their emotions and automatic responses to stimuli is a potential mechanism for the ASP's effects on their behavior and academic attainment.

[8] See the appendix section in Dinarte-Diaz and Egana-delSol (2023) for a summary of deviations from the trial registered at the AEA RCT Registry.

delivered to small groups of 13 participants, which helped control costs and foster relationships.

Each session was made up of activities involving two components: social skills development and traditional club curriculum. Students first received the social skills development curriculum during the first part of each session. This component involved activities and content related to four key goals: (i) cultivating the participants' socioemotional skills, (ii) promoting awareness of certain behaviors, (iii) disrupting negative behavioral patterns, and (iv) fostering new behavioral patterns. The activities in this component followed an experiential learning or role-play approach. During the second part of the session, the students participated in the traditional club curriculum made up of extracurricular activities based specifically on the following four club categories: leadership, art and culture, sports, and science. The extracurricular club activities component aimed to encourage student participation, make the intervention fun and interactive, and increase ASP attendance.[9]

The tutors were responsible for implementing the ASP activities. They had no formal training in social work or psychology, and their backgrounds were not necessarily similar to the participants' backgrounds (Heller et al., 2017). Three categories of tutor volunteers were involved in our intervention: 1) community volunteers who lived locally; 2) corporate volunteers who were part of a firm that was involved in the project with Glasswing; and 3) independent volunteers who were college students engaged in volunteer work. Specialists from Glasswing International thoroughly trained the tutors before they started working with the students.

At the beginning of the school year, the NGO visited schools to offer the program and enroll participants. Out of a total of 2,420 students from the five beneficiary schools, 1,056 students ages 10-16 years were recruited and enrolled to participate in the ASP from April to October 2016. During the registration stage, students were asked to complete an enrollment form that collected personal and family information described below. Students

[9]This combination of components required that I exercise caution when interpreting the intervention's potential impacts. Since the program was bundled, it is impossible to isolate the effects of each of component separately. This means that the results could be driven by either component.

could self-enroll but they had to submit an authorization signed by a parent or guardian in order to participate. They were then assigned to a group in light of their preferences, parental approval, and the aggregate demand for each club category. Clubs were composed of 13 students from a single educational-level.

2.2 Experimental Design

This paper aims to provide experimental evidence to determine which group composition best increases the effectiveness of the ASP presented in Dinarte-Diaz and Egana-delSol (2023). I test whether assigning students to different groups or tracking students in the ASP groups by their baseline risk for violence levels is the best strategy within the context of this program. To this end, I created additional exogenous variation in the average baseline risk for violence level to which each treated student is exposed in the ASP groups. This design involved several different steps: (1) estimate the baseline risk for violence measure per participant (a violence risk index); (2) randomly assign enrolled students to different treatment arms; (3) collect relevant data at different stages, and (4) conduct experimental design checks. This last step included verifying other distributional criteria in addition to balancing observable characteristics across the treatments before the program began. I describe all of these steps in this section.

2.2.1 Estimating the Violence Risk Index (IVV)

To assign enrolled students to each treatment group, the NGO needed to assess each participant's risk for violence level. Given the context of the ASP, it was not possible to ask participants directly about their past behavior during the registration because there was no way to guarantee that this personal information would remain confidential. For example, the local authorities or gang organizations could potentially try to force the research team or the NGO to reveal identifying information about participants, thereby risking not only the intervention but also—and most importantly—the students' safety. In addition, asking those involved in the study specific questions about gang membership or their association with other criminal organizations in El Salvador also could have endangered both the students and their instructors.

Instead, I estimated a predictive model of violent behaviors and crime from existing data using a Two-Sample Least Squares strategy. First, drawing from an existing confidential database of youths' violence and crime in El Salvador (FUSADES, 2015),[10] I estimate each student's likelihood of having committed a violent act V_f as a function of a wide range of covariates:

$$V_f = \alpha_0 + \alpha_1 D_f + \epsilon_f \tag{1}$$

where D_f is a vector of violence determinants of student f in the FUSADES dataset. Based on existing evidence, in this vector I include some variables associated with a person's vulnerability to violence, such as student characteristics (e.g., age, gender, time spent alone at home, and education level); household variables (e.g., residence area, mother's education, and household composition); and school-level controls (e.g., school location and commute time to school).[11] Table A1 contains descriptive statistics and a comparison of means (*p*-values) between the FUSADES sample and this study sample. The estimations indicate that both samples are similar across most of the determinants except for the variables *living with only one* parent or *being alone after school*.

All estimated coefficients $\hat{\alpha}_1$ have the expected sign according to the literature on violence determinants, as shown in Table A2. For instance, boys are more likely to be violent than girls, adolescents behave worse than students (Rodriguez-Planas, 2012), and lack of parental supervision (i.e., being alone at home after school) increases the probability of committing a violent act (Gottfredson et al., 2004). The statistically significant determinants are student's age, sex, living in urban area, maternal education (i.e., intermediate

[10]This database was created using the *El Salvador Youth Survey* instrument. It consists of a sample of 6,641 students in sixth and ninth grades who are enrolled in public schools in El Salvador and for who there is information on violence determinants. This database includes a large number of variables that measure crime, violence, and their determinants.

[11]Based on existing studies, the following variables are considered determinants of crime and violence: gender (Bertrand and Pan (2013) and Rodriguez-Planas (2012)); age (Rodriguez-Planas (2012)); location of residence (Springer et al. (2006)); maternal education (Springer et al. (2006) and Gaviria and Raphael (2001)); time spent at home (Gottfredson et al. (2004) and Aizer (2004)); commute time to school (Springer et al. (2006) and Damm and Dustmann (2014)); and household composition (Gaviria and Raphael (2001)).

education), student's commute time to school, and lack of parental supervision. Overall and reassuringly, lack of parental supervision is the most important determinant of risk for violence in this sample.

I designed the registration form to collect the same vector of violence determinants that was available in the FUSADES dataset (D_f). Then, I used this data to predict the measure of risk for violence (IVV) for each child, using the vector of estimated coefficients $\hat{\alpha}_1$.[12]

Two features of this IVV are important to point out: First, since the variables included in the estimation pertain to students' exposure to violence in different domains (family, school, and community), this measure is a more accurate proxy of students' overall risk for violence than are school behavior reports. Second, this predicted index can be interpreted as a measure of students' *risk* or *propensity* for violence rather than as an indicator of *actual* violence. In this sense, it can be used to identify students who are more likely to be at risk of engaging in violent actions or behaviors, based on their individual-level characteristics.[13]

Although the IVV does not measure actual violence, I provide some evidence to document that it is the best proxy of risk for violence when, as in the context of this study, data availability and collection are restricted. First, according to the existing literature on violence and crime determinants for particular groups (Heller et al., 2022; Chandler et al., 2011; Klassen and O'Connor, 1988),[14] these types of crime and violence models estimated from existing data have an acceptable degree of predictive power.[15] As I mention below, when using school behavior reports as the classification variable in the experimen-

[12] I estimated Equation (1) and calculated the IVV in a secure room in FUSADES because extraction of data from the El Salvador Youth Survey Database is not permitted.

[13] From a program implementation perspective, an NGO and other implementing institutions and policy makers are more likely to use this IVV tool in contexts that are similar to El Salvador, where data availability and collection are restricted, because it is more feasible to do so.

[14] See Chaiken et al. (1994) for a detailed early literature review of these models and their characteristics.

[15] Heller et al. (2022) use police data to predict shootings with enough accuracy to reduce victimization by gun violence in Chicago and without distorting average risk across demographic groups. Similarly, Klassen and O'Connor (1988) base their study on a sample of adult males who are at risk for violent behavior and have been admitted as inpatients to a community mental health center. They find that this model classified 85% of the total sample correctly.

tal design, my estimations indicate that the classification would have been similar for a significant share of the total sample.

Moreover, the IVV is associated with measures of academic performance and bad behavior at school in the expected direction. For example, I estimate the correlation between the predicted index, academic score, and teachers' reports of student (negative) behavior at school, and I find that the estimated correlations between IVV and academic score are negative and the correlation between the predicted IVV and bad behavior at school before the intervention is positive and statistically significant at 1% (Table A3). In addition, the IVV predicts future misbehavior. Using data from students in the control group, I find that the correlation between the IVV and bad behavior at the end of the academic year is positive and statistically significant at 5% (Table A4).

2.2.2 The Treatments

After estimating the IVV, I randomly assigned the 1,056 enrolled students to one of two groups—either the control (C, 25%) or the treatment (T, 75%)—within each school-by-educational-level "block." In this design, I have a total of 15 blocks or strata (5 schools $\times$ 3 educational levels, each).[16] Then, in the second stage of randomization, I randomly assigned treated students to one of two treatment arms—either the heterogeneous (HT, 25%) or homogeneous (HM, 50%) group—as shown in Figure 1. Next, students in the homogeneous group were ranked and assigned to one of two subgroups based on their IVV: all students with an IVV above the median at the HM-stratum level were assigned to the High-IVV group (HM-High, 25% of the full sample) and the rest were assigned to the Low-IVV (HM-Low, 25%) group. The HM-Low and HM-High groups are defined using the median in each school-by-educational-level (stratum) block because a uniform cutoff across all randomization blocks would have generated differences in group sizes that would likely confound the effects of group composition with group size.

This design permits me to test if targeting helps improve the ASP's effects and to study

[16] Each educational level "block" consists of three years of schooling: the first is from first to third grade; the second, fourth to sixth grade; and the third, seventh to ninth grade.

the potential existence of peer effects and heterogeneity by initial risk for violence. In addition, this strategy utilizes a regression discontinuity (RD) design approach to measure the impact of varying the group composition on marginal students and compare this measure to the average impact (see Section 4 for more details). Briefly, the treatments groups are as follows:

1. *Heterogeneous (HT):* 25% of registered students were randomly assigned to take part in a club with a heterogeneous composition of clubmates according to their IVV.

2. *Homogeneous (HM):* 50% of registered students were randomly assigned to take part in a club with a homogeneous composition of clubmates according to their IVV. This group was divided into the following two subgroups:

   - *Homogeneous-Low (HM-Low):* 50% of students randomly assigned to the HM treatment participated in a club with peers with a lower risk for violence if these students' IVV was lower than the median of the HM group within their respective strata.

   - *Homogeneous-High (HM-High):* 50% of students randomly assigned to the HM treatment participated in a club with peers with a high risk for violence if the students' IVV was higher than the median of the HM group within their respective strata.

3. *Control:* This group of students was not selected to participate in the ASP clubs during the 2016 academic year. They left the school premises after school was let out.

Unlike in Duflo et al. (2011) and similar to Lafortune et al. (2016), neither the instructors nor the participants were aware of the rationale behind their assignments because I wanted to capture the effects of the participants' interactions with each other, rather than the effects of other mechanisms such as teaching or curriculum adaptation. To test for changes in teaching methodologies, I collected information from a survey for the trainers and discuss the results in upcoming sections of this paper.

3 Data and Experimental Design Checks

3.1 Data Collection Stages

In this study, I rely on data collected before the intervention (baseline) and right after the program curriculum was completed (follow-up or endline), as in Dinarte-Diaz and Egana-delSol (2023).[17] The timeline of the study is in Figure 2. Appendix A1 includes the definitions of the variables, the data used to estimate each variable, and the outcomes of this analysis.

3.1.1 Baseline Data Collection

After the NGO advertised the ASP on the school premises, a research team returned to schools to register and enroll participants in March 2016. At this stage, students were asked to complete a registration form and submit a consent form signed by a parent or guardian. This self-reported instrument collected personal and family information such as age, gender, mother's education, and average commute time, among other things. Once registered, students received a unique identification number, which enabled me to track them through all data sets. These were used to estimate the IVV as described in Section 2.2.1. In early April, the survey team also collected printed school records of academic grades, absenteeism, and behavior reports for all 1,056 enrolled students at baseline.

3.1.2 Endline Data Collection

Endline data was collected immediately after the ASP was completed, between the end of October and in November 2016. Data was collected from several sources and using different tools (e.g., self-administered survey, administrative records, and neurophysiological biomarkers) to measure four main outcomes: violence and behaviors, attitudes toward school and learning, emotion regulation, and academic performance (Dinarte-Diaz and

[17] This data is publicly available at https://academic.oup.com/jeea/advance-article/doi/10.1093/jeea/jvad068/7420182.

Egana-delSol, 2023). See Appendix A1 for a detailed description of all indicators included in each outcome and the survey instruments or tools that were used to measure them.

Most students completed the self-administered endline survey with assistance from staff trained in the survey methodology. The survey took approximately 45 minutes to complete, was conducted within school facilities, and collected measures on student violent behaviors and attitudes toward school and learning. The outcomes violence and behaviors include measures related to delinquent behaviors, violent actions, and approval of peers' antisocial behaviors. Similarly, data on positive attitudes toward school, time spent on homework, and attention paid in class were collected to measure attitudes toward school and learning.

Since experimenter demand bias might effect self-reports, in Dinarte-Diaz and Egana-delSol (2023), we attempted to recheck and validate these behaviors and attitudes using proxies for these outcomes obtained from administrative data. To this end, we complemented the self-reported measures of violent behaviors with teacher reports of students' behavior at school. Similarly, we collected data on school absenteeism as another indicator that is included in the attitudes toward school and learning index. This administrative data was collected by the end of November 2016.

To analyze the effects of group composition on emotion regulation, I used the neurophysiological recordings collected for Dinarte-Diaz and Egana-delSol (2023) from a randomly selected subsample of students.[18] Specifically, in the aforementioned study, we used electroencephalogram (EEG) recordings to measure emotional state at rest (i.e. no stimuli) and responses to positive and negative stimuli. We relied on the portable Emotiv EPOC headset, an advanced and cost-effective tool that can be used in the field.[19]

Lastly, to measure academic performance, I use data on students' academic grades for math, reading, and science that were collected by the end of November 2016 and digitized

[18]Self-reported measures of emotion regulation are an alternative way to test for these channels. However, these measures are suboptimal for estimating the impact of this intervention due to the unclear direction of self-reporting bias (Egana-delSol et al., 2023).

[19]Dinarte-Diaz and Egana-delSol (2023) provide more details on the data-collection process for this third group of outcomes.

right them immediately after, as explained in Dinarte-Diaz and Egana-delSol (2023). The indicators academic grades (the average of the three scores) and the probability of passing the course are included as indicators for the academic performance outcome.

3.2 Administrative Data Matching and Survey Completion Rates

As shown in Table A5, the control group students' average administrative data matching rate was between 68% (absenteeism) and 93% (behavior scores) at baseline and between 94% (absenteeism) and 98% (academic scores) at follow-up (Panels A and B). All matching rates were balanced across the treatment and control groups, except for absenteeism between both homogeneous groups, which was significant at 10% (Table A6). This 2 percent difference, however, is negligible relative to these two groups' high matching rate with administrative data (between 94 and 96%).

On average, 92% of the students who initially enrolled in the ASP filled out the follow-up survey after the intervention ended. There are no statistical differences between the treatment and control groups with regard to overall survey completion, except for the homogeneous and heterogeneous treatment arms, the completion rates of which were 90% and 94%, respectively (see Table A6). To address potential concerns related to differential attrition in the completion of the follow-up survey across treatment and control groups, I estimate Lee Bounds following Lee (2009).

After removing poor electroencephalogram recording data, the average attrition share was 49% for the neurophysiological measures. In Dinarte-Diaz and Egana-delSol (2023), we present several checks to verify that this attrition rate was not connected with the intervention itself; we posit that attrition was caused mainly by the quality of the data recordings. For example, we were not able to obtain good electroencephalogram recordings from students who had long, dense, or dirty hair, and in other instances the computers froze. These were some of the problems that the Matlab toolbox encountered when reading the recordings.

3.3 Summary Statistics

Descriptive statistics of the full sample and each treatment and control group are presented in Table 1. Column (1) shows statistics for the control group; Columns (2) and (3) for the heterogeneous and homogeneous treatment groups, respectively; and Columns (4) and (5) for the two homogeneous subgroups.

Panel A presents the summary statistics of the violence determinants. In this study, students in the control group are, on average, 11.9 years old, 51% are male, and 72% live in an urban area. With regard to family composition, 91% live with at least one parent, and 9% live with a relative or an unrelated adult. On average, 59% of students' mothers have an intermediate level of education (7–12 years), and 34% have fewer than six years of schooling. Regarding risk exposure, only 5% of students report being alone at home when they are not at school, and the students' commute to school is, on average, 17 minutes. In addition, the average risk for violence for the treatment and control groups is 0.040, with a standard deviation (SD) of 0.029, ranging from 0.001 to 0.215. This average risk for violence is 14 times higher than the mean probability that a given student is vulnerable to violence in Chicago (Chandler et al., 2011). Even when both estimations are not entirely comparable (because I use fewer violence determinants than Chandler et al. (2011)), this difference sheds light on the tremendous and tragic risk for violence levels that students in El Salvador face.[20]

Panel B in Table 1 shows academic scores and absenteeism for the first quarter of the 2016 academic year, before the intervention began. On a grading scale of 0–10, with a minimum grade of 5 required to pass an academic course, students from the control group received an average grade of 6.5. The mean absenteeism rate for the same quarter was 7.1%, or 2.85 out of 40 days. Lastly, Panel C in Table 1 shows that the average club included 13.4 treated students. The enrolled students attended 57% of the sessions, and an average of 74% of students within each classroom were treated. Lastly, 30% of the clubs were led by a community volunteer tutors.

[20] I present more descriptive statistics for the predicted risk for violence level in Section 3.4.

3.4 Experimental Design Checks

This experimental design must meet five requirements to generate an exogenous variation that would make it possible to identify the causal effects of ASP group composition by risk for violence on the outcomes of interest. First, the treatment and control groups must be balanced. I find just a few differences in means between the treatment arms and the comparison group. The *p*-values for all of the tests of differences in means between each treatment—the HT, HM, HM-High, HM-Low—and the control group are listed in Table A7. However, after adjusting the *p*-values for multiple hypothesis testing of means and FWER, all differences are not statistically different from zero. However, I account for all of these differences by including these variables as controls in the estimations. See more details in Section 4.

A second requirement is that the HM-High subgroup's IVV should be greater than the HM-Low subgroup's IVV. The differences between the two subgroups should also be expressed in most of the IVV determinants. As evident in Table 1, Columns (4) and (5), and which I verify by testing for differences (see the *p*-values in Table A7, Column [4]), the HM-High subgroup has a larger proportion of males and older students than the HM-Low subgroup. The students in the HM-High subgroup are also more exposed to violence because they live in an urban area, have a longer commute to school, are more likely to have mothers with intermediate education, and spend time at home alone. Finally, the average academic performance of students in the HM-High subgroup is lower than that of students in the HM-Low treatment.

Since assignment to the homogeneous and heterogeneous groups was determined based on the IVV, then the experimental design must effectively generate changes in the homogeneous and heterogeneous students' classmates' risk for violence. As I show in Table 2, consistent with the premise that non-homogeneous groups are more violence diverse than any of the homogeneous groups, the SD of the heterogeneous group was 0.007 and 0.021 points (25%–150%) higher than the same figure for the HM-High and HM-Low subgroups, respectively. Moreover, the average risk for violence level of the heterogeneous group must fall between the HM-Low and HM-High subgroup levels. As I show in Table 2, the average heterogeneous group's IVV (0.041) falls between the IVVs of the HM-High and HM-Low subgroups. which are 0.051 and 0.023, respectively.

The fourth requirement for the empirical design is that it must consider three desired characteristics of the IVV distribution functions of the heterogeneous, HM, and control (C) groups before treatment. First, these distributions must be similar at the baseline. Using the two-sample Kolmogorov-Smirnov test for equality of distribution functions, the hypotheses are not rejected, as evidence by the *p*-values of 0.619, 0.868, and 0.682 for the HT-HM comparison, the HT-C comparison, and the HM-C comparison, respectively. Figure 3 also demonstrates and affirms the similarity among distributions. Second, the distributions of the HT, HM-High, and HM-Low groups must differ. As Figure 4 illustrates, there are differences among the three groups' distributions. The two-sample Kolmogorov-Smirnov test confirms this finding and rejects the hypothesis of the equality of each comparison of distribution functions pairs at 1%. The third feature is that the distributions of the HM-High and HM-Low subgroups should not fully overlap in the full sample so that some variability exists between both HM subgroups. Without stratifying, there would be no overlap between both groups. However, since I defined the assignment within each stratum, there is overlap in 67% of the sample, as shown in Figure 5.

The fifth condition is that there must be a sharp discontinuity at the fiftieth percentile for the HM subsample, consistent with the discontinuous assignment at the median IVV within each stratum. Figure 6 shows the predicted IVV median of a student's clubmates as a function of the student's own IVV, and the expected jump at the fiftieth percentile. Moreover, a RD-robust estimation using only this homogeneous subsample indicates that students assigned to the HM-High subgroup are enrolled with peers with a mean IVV that is 0.8 points greater and statistically significant at 5%.[21]

Finally, I contend that this IVV is a good proxy for violence because even after using misbehavior reports as the classification variable for high and low risks for violence, estimations indicate a similar classification in approximately 53% of the total sample. Cru-

[21] I use a third-order local polynomial following the specification of Duflo et al. (2011). For a first- and second-order polynomial, the coefficient is 0.9 and statistically significant at 1%. This coefficient and its statistical significance are also stable when using a conventional or bias-corrected RD Method.

cially, there are no differences in the classification among treatments (see Table A8).

4 Empirical Strategy

In this section, I describe my empirical strategy to study the effects of group composition, how this variation interacts with a student's initial risk for violence level, and the effect of tracking on the marginal participant.

4.1 Group Composition Average Effects

This study design creates a direct experimental variation on group composition by risk for violence. Thus, I directly test for differences in the ITT effects on the outcomes of students assigned to groups with either homogeneously or heterogeneously at-risk peers using the following specification:

$$y_{ij} = \theta_0 + \theta_1 HM_{ij} + \theta_2 HT_{ij} + X_{ij} + S_j + \epsilon_{ij} \tag{2}$$

where y_{ij} is the post-intervention outcome of student i in school and education level j. HM_{ij} and HT_{ij} are dummies that indicate whether student i in school level j is assigned to the homogeneous or heterogeneous treatment, respectively. X_{ij} is a vector of control variables measured before the intervention. To account for the differences across groups in some pre-intervention characteristics and outcomes, I include the following variables as controls: student's grade level and risk for violence score; indicators of whether the student is enrolled in the morning shift, living with both parents or with one parent; and the three outcomes at baseline (students' academic grades, behavior score, and absences). Since these outcomes at baseline include imputed values for observations with missing values, I also include missing baseline outcomes indicators as controls. Finally, I also control for "randomization blocks" with school-by-education-level fixed effects S_j.

For the inference, I follow a more agnostic approach to the structure of standard errors, allow for a potential fuzzy clustering, and estimate randomization inference standard errors (and respective p-values). Randomization inference gives me precise p-values based on the empirical distribution of all estimated treatment effects that could arise within my design and data (after randomly reassigning the treatment status 2,000 times) under the null hypothesis of no effect for any unit.

To address potential concerns regarding multiple hypothesis testing, I construct indices for each outcome category (violence and behavior, attitudes toward school and learning, neurophysiological outcomes, and academic performance) using inverse covariate weighting just like Anderson (2008). Summary indices offer three advantages: (i) they are robust to over-testing because each index represents a single test; (ii) they provide a statistical test for whether a program has a "general effect" on a category of outcomes, and (iii) they are potentially more powerful than individual-level tests by reducing random error in each outcome measure (Anderson, 2008). Each summary index is a weighted mean of several standardized outcomes.

In this setting, θ_1 (θ_2) can be interpreted as the effect on student i of receiving an offer to participate in the ASP with a homogeneous (heterogeneous) composition of at-risk for violence peers, compared to effects of the control group. To be more specific, θ_1 and θ_2 capture the effect of changing not only the mean but also both the variance and mean in the distribution of peer risk for violence, or the elements that constitute the IVV. Testing for differences between the estimated coefficients θ_1 and θ_2 indicates the net effects of group composition on the outcomes of interest.

I also exploit the variation in peer quality generated by the experiment. Since students were randomly assigned to a heterogeneous group within the ASP, they will have a random set of peers. Therefore, I restrict the sample to these groups and estimate the effect of a student's peers' mean and variance baseline IVV, as in Lafortune et al. (2016) and Duflo et al. (2011). Details of these estimations are in Appendix A2.

4.2 Group Composition Heterogeneity by Baseline Risk for Violence Level

Which students within each treatment arm benefit from the composition of their peers? An argument in favor of tracking posits that, when students are in a mixed group, the most at-risk for violence negatively influence the least at-risk. This argument, however, does not consider the fact that the least at-risk for violence could positively influence the most at-risk. In this way, a uniform group limits the potential for positive influence in the opposite direction. Since my study design includes two different subgroups in the HM group, I can further explore the differential effects of group composition for students assigned to the lower and upper section in the IVV distribution. The assignment variable to those subgroups was the median of the IVV distribution at each HM-stratum level. Therefore, after controlling by the indicator IVV_high_{ij} and by the IVV median at the j level, $I\bar{V}V_j$, I can directly compare the differential effects of group composition by student's risk for violence by estimating the following specification and restricting the sample to treated students only:

$$Y_{ij} = \theta_0 + \theta_1 HomH_{ij} + \theta_2 HomL_{ij} + \theta_3 IVV\_high_{ij} + \theta_4 I\bar{V}V_j + \theta_5 X_{ij} + \epsilon_{ij} \tag{3}$$

where $HomH_{ij}$ and $HomL_{ij}$ are dummies indicating whether student i in stratum j was assigned to the HM-High or HM-Low subgroup, respectively, with the rest of the variables defined as before.

Specification (3) allows me to compare both treatments within each half of the IVV distribution. In the upper half, θ_1 measures the average effects of assigning students with a high risk for violence to a homogeneous group of peers relative to a heterogeneous group. Also, for the lower half of the IVV distribution, θ_2 is an ITT estimator of assigning students with low IVV to a homogeneous group of peers compared to a heterogeneous group.

4.3 Effects of Tracking on the Marginal Participant

The experimental design allows me to explore the effects of exposure to peer violence on the students who fall near the median in a tracking setting. I call such students *marginal participants*. This group includes a set of students just above or below the fiftieth percentile of the IVV distribution. Given that the students just above the median have a similar risk for violence to those just at or below the median, I exploit their assignment to a group of high-IVV peers and compare them with others in a low-IVV set. Studying the effects on marginal participants is relevant since they differ the most within their group. Therefore, they may experience greater group composition impact.

To identify this impact, HM groups provide the natural prerequisites for an RD design, where the median of the IVV distribution in each stratum functions as the discontinuity (see Figure 6). In order for this strategy to be valid, the assumption is that nothing else changed discontinuously around the point of separation between the two groups, which holds true in this context. Therefore, I estimate the following equation:

$$Y_{ij} = \lambda_0 + \lambda_1 HomH_{ij} + f(IVV_{ij}) + S_j + \epsilon_{ij} \tag{4}$$

where $f(\text{IVV}_{ij})$ is a flexible second-order polynomial of an individual's IVV percentile within each stratum, and $HomH_{ij} = 1$ if the participant was in the HM-High subgroup. In this case, λ_1 is a LATE estimator that indicates the effects of tracking on the marginal participant's cognitive and noncognitive outcomes. I also estimate this specification while restricting the sample to the eight students around the cutoff within each stratum.

5 Results

5.1 Average Effect of Group Composition

Table 3 shows estimations of group composition by risk for violence on the four main outcomes of interest using Specification (2). Columns (1) and (2) in each Panel present the estimated impacts of being assigned to either a heterogeneous or homogeneous group within the ASP compared to the control group. Column (3) shows the estimated differences of the impacts between the two treatment arms.

The estimation comparisons between each type of group composition and the control group indicate that being treated within a heterogeneous group of peers reduces violence, improves behavior at school and emotion regulation, and marginally enhances attitudes toward school and learning (p-value = 0.164). I find, however, no effects of being treated in an homogeneous composition of peers in terms of risk for violence relative to being assigned to the control group.

When I compare the difference in the effects of the two group compositions, I find that participating in the ASP with more diverse peers facilitates a greater reduction (by 0.048 SD) in violence and misbehavior at school compared to the homogeneous composition of peers. Although the differences in the effects by group composition on the indices of attitudes toward school and learning and emotion regulation are not statistically significant at the conventional levels, the estimated magnitudes are economically relevant (0.05 SD and 0.09 SD, respectively), considering that they result simply from modifying how students are assigned to the groups and not from whether or not they participate in the ASP.

These greater effects of the heterogeneous relative to the homogeneous composition seem to be driven by reductions in the probability of having bad behavior reports for the violence and behavior outcome and by reductions in stress levels for the emotion regulation outcome (see Table 4). Notably, I do observe an improvement in locus of control for students treated in the HM group relative to participants treated in the heterogeneous and control groups. The marginal improvement on attitudes toward school and learning of students assigned to heterogeneous relative to those assigned to homogeneous groups is probably driven by the index on positive attitudes toward school.[22]

[22]Most of the differences in estimated effects between the heterogeneous and homogeneous groups are only marginally statistically significant, which is expected since indices are more powerful than individual-level tests because they reduce random error in each outcome measure (Anderson, 2008).

The positive impacts of a more diverse group composition are consistent with the evidence that interactions with a different peers can generate a variety of learning experiences (Lafortune et al., 2016). The rainbow peer-effects model (Hoxby and Weingarth, 2005), which claims that students do best when they have a diverse group of classmates, can also explain these results. Furthermore, the results on bad behavior at school imply that treating students in violence-diverse groups reduces the likelihood that violent students will create their own networks (Billings et al., 2019).

5.2 Group Composition Heterogeneity by Baseline IVV

The results of the differential impacts of group composition by initial risk for violence obtained from Specification (3) are presented in Table 5. Column (1) lists the estimated group composition differences for students in the lower half of the IVV distribution. Column (2) shows the same differences but for students in the upper half of the IVV distribution.

Based on the results, I find that the effects of being treated in an homogeneous composition of peers are particularly detrimental for students with higher IVV compared to peers with similar levels of IVV but treated in heterogeneous groups. These negative effects extend to my estimates for violence and behaviors, attitudes toward school and learning, and emotion regulation. Second, I also observe that being treated in a homogeneous peer group can negatively affect violence and behavior for students with an initially low IVV compared to other students with an initially low IVV who were treated in a more diverse peer group. Notably, regardless of the students' risk for violence at baseline, increases in stress drive the more negative effect on emotion regulation among the students treated in the homogeneous groups relative to those treated in the heterogeneous groups (Table A9).

In sum, the implementation of this ASP in homogeneous groups instead of heterogeneous groups can negatively impact the violence and behavior, attitudes toward school and learning, and emotion regulation of participants with a greater risk for violence. These results can be interpreted in different ways. First, peers with diverse IVV levels can be beneficial because they enable highly violent students to be exposed to less violent students and to learn social skills and good behaviors from them. Similarly, less violent students benefit from witnessing the more violent students' misbehaviors and then choose to avoid behaving similarly. Students in a homogeneous group, on the other hand, miss out on the opportunity to learn and imitate positive behaviors and witness negative behaviors that they should avoid. An argument in favor of heterogeneous peer groups is that diversity is normative in the real-life environments in which students function. Thus, assigning students to a similar peer group may, in fact, stress them more. For example, assigning highly violent students to the same group would simply reinforce and exacerbate their negative behaviors and, thereby, produce unintended effects.

Finally, since students were allocated to a group in the ASP randomly, some variation in group composition stems from changes in the mean and variance of one's peers' IVV. Following Lafortune et al. (2016) and Duflo et al. (2011), my identification assumption is that after controlling for strata fixed effects, the variance and mean IVV of peers arise entirely from the random assignment. I include details of the estimation and a summary of results in Appendix A2 and Table A10. In sum, these results reinforce my previous findings obtained by using a direct variation of the experiment. First, higher group average IVV increases violent behaviors and worsens emotion regulation and academic performance, whereas exposure to a more IVV-diverse group of peers reduces students' violent behaviors and improves their attitudes towards school and learning and academic performance.

5.3 Effects of Tracking on Marginal Participants

An additional feature of this experiment is the evidence it provides on the effects of tracking on marginal students. Consider two students who fall near the median of the IVV distribution. The assumption is that these two students are very similar in all their observables. One student, however, is assigned to an ASP group with other students that have, on average, a higher IVV (subgroup HM-High), while the other student is assigned to a group with clubmates with a lower IVV (subgroup HM-Low). In this scenario, the first student is the least at-risk for violence within a highly at-risk group, and the second student is the most at-risk for violence within a far at-risk group. To directly measure the effects of these assignments, I can compare the two homogeneous subgroups using specification (4), which allows me to identify the differences of being assigned to a homogeneous peer group with higher risk for violence.

The estimations of the tracking effects on the marginal participants' outcomes are summarized in Table 6. Column (1) presents the estimated coefficients using all students assigned to the homogeneous treatment. Following Duflo et al. (2011), I run Specification (4) but restrict the sample to the eight students around the IVV median within each stratum; I report these results in Column (3). This sample restriction allows me to focus on the students who were most similar before the intervention began. The downside of this approach is that it increases standard errors of the estimations, thereby reducing statistical significance.

Overall, the results indicate that tracking can have some unintended effects on students least at-risk for violence. I find that assigning a marginal participant to a group of peers with a higher risk for violence increases this student's violence and misbehavior by 0.048 SD. Although not statistically significant at the conventional level, I also find that tracking has a 0.06 SD negative effect on attitudes toward school and learning and a 0.10 SD negative effect on academic performance (p-values = 0.157 and 0.160, respectively). Moreover, when restricting the sample to the 8 students around the cut-off, I also find large negative effects on attitudes toward school and learning (0.36 SD).

In brief, being the least at-risk for violence member of a highly at-risk (and less diverse) group negatively affects the former's violence and attitudes toward school and learning-related outcomes and academic performance. Such students seem to follow the group's social norms regarding violence and negative attitudes, which indirectly impacts their academic performance.

Moreover, combined with my previous results that indicate that the presence of a few at-risk peers within a diverse group can have positive average effects, the results for the marginal participant indicate that, when the share of high-to-low-risk peers is too high, the overall effects are detrimental. In this sense, there is an optimal risk-level ratio that can maximize the program's overall beneficial impact.

6 Robustness Checks

In this section, I address potential concerns regarding the results presented in Section 5, such as spillovers, sensitivity arising from the selection of control variables, and self-reports on outcomes measured using survey data.

6.1 Spillovers

Since the random assignment of students occurred at the individual level, the presence of spillovers from treated students on their untreated classmates likely affected the experiment. In Dinarte-Diaz and Egana-delSol (2023), we exploit quasi-exogenous variation in our design on the share of treated students at the classroom level to test for the presence of spillovers. We followed Baird et al. (2018) to compare students in school-grade level clusters with a high and low share of treated classmates. Overall, we do find positive spillover effects, which are of the same magnitude for both the treated and untreated students' outcomes. Therefore, when comparing groups of treated and control students, these spillovers cancel each other out, indicating that the ASP's estimated effects are as close as possible to the intervention's causal effect.

As explained in Dinarte-Diaz and Egana-delSol (2023), we can detect only large spillover effects (up to 0.34 SD on the treated students and up to 0.40 SD on the control group). In the current design, I am even less powered to detect spillovers when separating the treatment group into homogeneous and heterogeneous groups. For this reason, I follow a different approach and estimate local linear nonparametric kernel estimations to assess whether the treatment effect varies according to the saturation at each school-grade level cluster. As I show in Figure 7, I do not find stark differences in terms of treatment effect variance across most of the outcomes according to the share of treated students at the grade level.[23] Thus, I am confident that, if anything, spillovers are of the same magnitude and direction for the control, heterogeneous, and homogeneous groups.

6.2 Assessing Sensitivity from the Selection of Control Variables

As I discuss in Section 4, I include in the vector of controls the baseline characteristics and outcomes for which I find statistically significant differences across groups, including missing baseline outcome indicators for missing observations. Given this approach, one potential concern is that the results are driven by the inclusion of these variables. First, I show that the exclusion of these covariates does not change the results. As presented in Appendix Tables A12 to A14, all estimation results remain similar in magnitude and statistical significance after excluding the control variables.[24] Second, I also use a double LASSO approach to identify the variables that should be included in the estimations as control variables and test for the stability of the estimated coefficients after including these variables for each of my main outcomes. As I show in Tables A15 to A18, the estimated coefficients and their statistical significance do not change after including these additional control variables selected by LASSO.[25]

6.3 Assessing Potential Bias Due to Differential Survey Attrition

As presented in Table A5, students in the heterogeneous group were 4 percentage points more likely to complete the follow-up survey relative to students in the homogeneous group. To address potential differential attrition in the follow-up survey, I estimate Lee bounds to account for sample selection (Lee, 2009) and present these results in Table A23. This procedure is a conservative estimate of the treatment effect, as it corresponds to ex-

[23] I confirm these results by formally testing for the equality of the distributions presented in Figure 7 using a Kolmogorov-Smirnov test. *P*-values from this test are presented in Table A11. The only statistically significant difference occurs between the control group and the heterogeneous treatment (p-value = 0.021).

[24] If anything, some of the estimated coefficients are larger and statistically significant. This is the case for academic performance. Yet, I prefer to use the most conservative estimated coefficients as my main results.

[25] LASSO selects variables that can be used as controls for each outcome, but these variables differ across outcomes and models. For instance, LASSO selects age and gender for some outcomes, and education level and household composition for others. See Tables A19 to A22 for further details.

treme assumptions about the missing information. I find that for the heterogeneous treatment, all upper and lower bounds for the violence and behavior as well as the emotion regulation outcomes differ significantly from zero, except the upper bound of the emotion regulation outcome. Moreover, the upper and lower bounds for the difference between heterogeneous and homogeneous groups in the effects on violence and behavior differ from zero. All in all, these results suggest that my main results are mostly robust to differential attrition.

6.4 Differences in ASP Attendance by Treatment Assignment

Another potential concern is that ASP attendance varied by type of assignment. In this sense, if students assigned to the heterogeneous treatment were more likely to attend the sessions, then this would explain why this group experienced greater positive effects on the different outcomes. Using administrative data on attendance provided by the ASP tutors, I estimate the differences in attendance within the treatment group using specifications (2) and (3). As I show in Table A24, there are no differences in ASP attendance across the heterogeneous and homogeneous treatment groups (Column [1]) or between the two homogeneous subgroups (Columns [2] and [3]) relative to the heterogeneous groups or between each other.

6.5 Assessing Potential Bias Due to Self-Reports of Survey-based Outcomes

When assessing the ASP's effects on violence and behaviors as well as attitudes towards school and learning, self-reports for these measures can be problematic because participants might be influenced by experimenter demand effects. To address this potential concern, and as I explained in Section 3, I attempted to recheck and validate these behaviors and attitudes by using proxies for these outcomes that I obtained from administrative data. In this sense, I complement the self-reported measures of violent behaviors with recorded reports of students' behavior at school. Similarly, I collect data on school absenteeism as another proxy for attitudes toward school and learning. As I show in Table 4, the estimated effects based on self-reported measures are in the same direction as those based on more objective measures of violent behaviors or attitudes toward school and learning. In this sense, I do not think that the experimenter demand effect is relevant in this context.

7 Discussion of Potential Mechanisms

Why does integration typically generate better outcomes? In this section, I discuss and provide suggestive evidence that points to four potential mechanisms that may underlie these results. First, one potential mechanism could be the rainbow model of peer effects, whereby all individuals benefit from exposure to a more heterogeneous peer set (Hoxby, 2000). On the one hand, students in heterogeneous groups benefit from being exposed to both good behaviors they should adopt and bad behaviors that they should avoid. Specifically, a heterogeneous group composition would allow students at a high risk for violence to be exposed to students with a lower risk and learn social skills and good behaviors from them. Similarly, lower-risk students in heterogeneous groups benefit from witnessing bad behaviors that they should avoid. On the other hand, students in a homogeneous group miss out on the opportunity to learn good behaviors from the students on the opposite end of the risk for violence distribution. However, the jump observed around the median of the tracking group also indicates that being exposed to a more significant share of peers who behave badly can have the opposite effect. This implies there is an optimal "bad-to-good" peer ratio within a group that can maximize the program's overall impact.

A second channel that might explain the results is that diversity in terms of risk for violence is the social norm in the settings where students typically function (i.e., school, home, etc). Being in a diverse peer group is familiar and, therefore, more comfortable. Although my data makes it impossible to formally test this mechanism, I use two sets of results to document it. First, from Appendix A2 and Table A10, I find evidence that being exposed to a more IVV-diverse group in the ASP improves violence and behavior, attitudes toward school and learning, and academic performance. Second, during focus group discussions with participants conducted in mid-2017, students assigned to the heterogeneous treatment reported being more satisfied with the intervention and with the peers to which they were assigned.[26] Moreover, based on in-depth interviews with the tutors, I document that the tutors for heterogeneous groups were more likely than tutors for the homogeneous groups to report that implementing the activities went more smoothly and their the students were more engaged and collaborative. In fact, tutors in the homogeneous groups reported more negative incidents and less collaboration between and interactions among students.[27]

A possible third mechanism underlying the group composition results is that homogeneous peer groups can increase the potential for students to establish violent networks amongst themselves, an issue that has been discussed in the literature (Billings et al., 2019; Bayer et al., 2009). The implementation of interventions for groups comprised only of students with high or low risk for violence can generate unintended effects on both groups, particularly for the most at risk. In line with this existing evidence, and based on my analysis of heterogeneity of group composition by student's risk for violence, I find that the homogeneous treatment can have negative effects on violence and behaviors, attitudes toward school and learning, and emotion regulation.

Lastly, the fourth mechanism behind the negative effect of tracking students aligns with the results of Duflo et al. (2011) and is related to the club tutors' lack of incentive to target the club materials to the particular needs of their participants. From my extensive interviews with the ASP tutors, I find that only 32% of instructors reported adjusting their methods or club activities, with no differences between instructors of heterogeneous or homogeneous groups. The reasons why instructors made any adjustments at all, however, was not based on the participants' needs but on the fact that they lacked materials or they

[26]In May 2017, I conducted 6 focus group discussions with a total of 37 treated students—2 discussion per each treatment group (HT, HM-High, and HM-Low). On average, Between 5-6 students per group, participated in each discussion. The goal of these discussions was to gather feedback regarding student satisfaction with the program and to explore this potential mechanism.

[27]In May 2017, I also interviewed a total of 15 tutors, 5 tutors per each treatment group. The objective of these interviews was twofold. First, I sought tutors' feedback on students' behaviors during the implementation of the ASP. Second, I tried to understand if the tutors were able to identify the characteristics (i.e., composition) of the groups that they led and whether they modified how they implemented the ASP activities based on the composition of the group.

did not have an adequate amount of space to implement the activities.

8 Concluding Remarks

Although my experiment in schools located in violent communities of El Salvador suggests that involving youth in an after-school program can successfully reduce their risk of engaging in violence (e.g., joining gangs), how this program is implemented is key. For example, targeting at-risk youths and then implementing activities for at-risk youths within the same group would be detrimental to them because being in a homogeneous group only reinforces their exposure to risk. It is much better to recruit all youth, irrespective of their risk levels, and group them together since interacting with a diverse group of peers helps at-risk youth to witness an alternative path. Importantly, this approach does not entail the risk of increasing violence among those with a low baseline risk, as long as there are enough other low-risk students in the group.

The results presented in this paper have implications for public policy discussions regarding interventions oriented to reducing violence within schools and improving students' self-regulation. More specifically, this paper takes the first step toward understanding the relevance of group composition in an ASP. It shows that, within this context, peer effects are an important mechanism that can improve relevant outcomes, especially when interventions are implemented in heterogeneous groups. It is also likely that these results are relevant to other social and educational programs that aim to modify participants' behaviors and attitudes by allowing them to interact and learn from each other.

Will these results persist over time? Because the NGO's donors required that the research team allow students in the control group to participate in the intervention the following year, I cannot measure the ASP's long-term effects. Nonetheless, this experimental design can be potentially helpful in other contexts where an implementing practitioner or policy maker would like to evaluate the usefulness of targeting while maintaining coverage of all initial beneficiaries.

References

Aizer, A. (2004). Home alone: Supervision after school and child behavior. *Journal of Public Economics*, 88(9):1835–1848.

Alan, S., Baysan, C., Gumren, M., and Kubilay, E. (2020). Building social cohesion in ethnically mixed schools: An intervention on perspective taking.

Anand, P. and Kahn, L. B. (2023). The effect of a peer's teen pregnancy on sexual behavior. Technical report, National Bureau of Economic Research.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Baird, S., Bohren, J. A., McIntosh, C., and Özler, B. (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844–860.

Bayer, P., Hjalmarsson, R., and Pozen, D. (2009). Building criminal capital behind bars: Peer effects in juvenile corrections. *Quarterly Journal of Economics*, 124(1):105–147.

Bertrand, M. and Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, 5(1):32–64.

Billings, S. B., Deming, D. J., and Ross, S. L. (2019). Partners in crime. *American Economic Journal: Applied Economics*, 11(1):126–50.

Billings, S. B. and Hoekstra, M. (2019). Schools, neighborhoods, and the long-run effect of crime-prone peers. Technical report, National Bureau of Economic Research.

Boozer, M. and Cacciola, S. E. (2001). Inside the "Black Box" of Project STAR: Estimation of peer effects using experimental data. Technical report, Economic Growth Center Discussion Paper Series.

Bursztyn, L. and Jensen, R. (2015). How does peer pressure affect educational investments? *The quarterly journal of economics*, 130(3):1329–1367.

Carrell, S. E., Hoekstra, M., and Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108(11):3377–3415.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81(3):855–882.

Chaiken, J., Chaiken, M., and Rhodes, W. (1994). Predicting violent behavior and classifying violent offenders. *Understanding and preventing violence*, 4:217–295.

Chandler, D., Levitt, S. D., and List, J. A. (2011). Predicting and preventing shootings among at-risk youth. *American Economic Review*, 101(3):288–292.

Cortes, K. E. and Goodman, J. S. (2014). Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement. *American Economic Review*, 104(5):400–405.

Damm, A. P. and Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, 104(6):1806–1832.

Davis, J. M. and Heller, S. B. (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, 102(4):664–677.

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

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking. *American Economic Review*, 101(5 (August 2011)).

Egana-delSol, P., Sun, X., and Sajda, P. (2023). Neurophysiological markers of emotion regulation predict efficacy of entrepreneurship education. *Scientific Reports*, 13(7260):1–15.

Feng, S., Kim, J. H., and Yang, Z. (2022). Effects of childhood peers on personality skills.

Fletcher, J., Kim, J., Nobles, J., Ross, S., and Shaorshadze, I. (2019). The effects of foreign-born peers in us high schools and middle schools. Technical report, National Bureau of Economic Research.

Fletcher, J. M. and Ross, S. L. (2012). Estimating the effects of friendship networks on health behaviors of adolescents. Technical report, National Bureau of Economic Research.

FUSADES (2015). El Salvador Youth Survey Database.

Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2):257–268.

Girard, Y., Hett, F., and Schunk, D. (2015). How individual characteristics shape the structure of social networks. *Journal of Economic Behavior & Organization*, 115:197–216.

Goethals, G. R. (2001). Peer Effects, Gender, and Intellectual Performance among Students at a Highly Selective College: A Social Comparison of Abilities Analysis. WPEHE discussion paper series, Williams College, Williams Project on the Economics of Higher Education.

Golsteyn, B. H., Non, A., and Zölitz, U. (2021). The impact of peer personality on academic achievement. *Journal of Political Economy*, 129(4):1052–1099.

Gong, J., Lu, Y., and Song, H. (2019). Gender peer effects on students' academic and noncognitive outcomes: Evidence and mechanisms. *Journal of Human Resources*, pages 0918–9736R2.

Gottfredson, D. C., Gerstenblith, S. A., Soulé, D. A., Womer, S. C., and Lu, S. (2004). Do After-School Programs reduce delinquency? *Prevention Science*, 5(4):253–266.

Griffith, A. L. and Rask, K. N. (2014). Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review*, 39:65–77.

Heller, S. B., Jakubowski, B., Jelveh, Z., and Kapustin, M. (2022). Machine learning can predict shooting victimization well enough to help prevent it. Technical report, National Bureau of Economic Research.

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

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

Hoxby, C. M. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Working paper.

Klassen, D. and O'Connor, W. A. (1988). A prospective study of predictors of violence in adult male mental health admissions. *Law and Human Behavior*, 12(2):143.

Lafortune, J., Perticará, M., and Tessada, J. (2016). The benefits of diversity: Peer effects in an adult training program in Chile.

Lavy, V., Paserman, M. D., and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *Economic Journal*, 122(559):208–237.

Lazear, E. P. (2001). Educational Production. *Quarterly Journal of Economics*, 116(3):777–803.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Lyle, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics*, 89(2):289–299.

Manski, C. F. (1993). Identification of endogenous social effects:The reflection problem. *Review of Economic Studies*, 60(3):531–542.

MINED (2019). Estadisticas del Ministerio de Educacin de El Salvador.

Oreopoulos, P., Brown, R. S., and Lavecchia, A. M. (2017). Pathways to education: An integrated approach to helping at-risk high school students. *Journal of Political Economy*, 125(4):947–984.

Rao, G. (2019). Familiarity Does not Breed Contempt: Diversity, Discrimination and Generosity in Delhi schools. *American Economic Review*, 109(3):774–809.

Rodriguez-Planas, N. (2012). Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States. *American Economic Journal: Applied Economics*, 4(4):121–39.

Sacerdote, B. et al. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far. *Handbook of the Economics of Education*, 3(3):249–277.

Springer, A. E., Selwyn, B., and Kelder, S. H. (2006). A descriptive study of youth risk behavior in urban and rural secondary school students in El Salvador. *BMC International Health and Human Rights*, 6(1):3.

World Bank, T. (2023). Central america human capital review. phase 2: El salvador, honduras, and nicaragua.

Zárate, R. A. (2023). Uncovering peer effects in social and academic skills. *American Economic Journal: Applied Economics*, 15(3):35–79.

Tables and Figures

Table 1: **SUMMARY STATISTICS: MEANS OF VARIABLES BY TREATMENT GROUP AT BASELINE**

	(1)	(2)	(3)	(4)	(5)
		Treatments		Tracking Groups	
	Control Group (C)	Hetero. Group (HT)	Homo. Group (HM)	Homo.-High (HM-H)	Homo.-Low (HM-L)
Panel A: Individual and Household Characteristics					
Student is male	0.51	0.48	0.49	0.76	0.22
Student's age (years)	11.86	12.04	11.93	12.43	11.44
Student lives in an urban area	0.72	0.73	0.74	0.78	0.70
Student's household composition					
Student lives with both parents	0.47	0.53	0.56	0.53	0.59
Student lives with one parent	0.37	0.34	0.30	0.33	0.26
Student lives with a parent and stepparent	0.07	0.06	0.07	0.06	0.07
Student lives with other relative/adult	0.09	0.07	0.08	0.08	0.07
Mother's education level					
Basic education (1-6 years)	0.34	0.27	0.31	0.22	0.40
Intermediate education (7-12 years)	0.59	0.65	0.62	0.72	0.52
University or higher (13 and +)	0.07	0.08	0.07	0.06	0.08
Student's commute time from home to school (minutes)	16.98	17.84	17.86	19.58	16.13
Student is alone at home after school	0.05	0.07	0.04	0.08	0.01
Student's grade level	5.67	5.81	5.76	6.03	5.49
Student is enrolled in the morning shift	0.71	0.68	0.73	0.73	0.74
Student's risk level for violence (index)	0.04	0.04	0.04	0.05	0.02
Panel B: Main Outcomes					
Academic grades (score)	6.49	6.68	6.67	6.55	6.80
Pr. of passing course	0.83	0.87	0.90	0.86	0.93
Behavior at school (score)	7.25	7.41	7.43	7.36	7.51
Pr. of having bad behavior reports	0.29	0.26	0.24	0.27	0.21
Absences (days)	2.85	2.03	1.77	2.13	1.43
Panel C: Club Characteristics					
Average club size (N)		13.43	13.38	13.13	13.63
Average club take-up (%)		0.57	0.57	0.56	0.59
Community tutors		0.29	0.32	0.35	0.29
Treated students by course (%)	72.77	77.17	76.13	76.23	76.04

Notes: Table 1 shows descriptive statistics of the available variables at baseline. Panel A summarizes information obtained from the enrollment form that was used as determinants in the IVV estimation. Panel B presents consenting students' administrative and academic data for 2016 academic year, before the clubs were implemented. The grading system for academic courses and behavior reports are based on a 0–10 point scale in El Salvador. All variables are dummies except when noted. The definitions of the variables are listed in Appendix A1. In Columns (1), (2), and (3), I present the control (C), heterogeneous (HT), and homogeneous (HM) groups' average characteristics, respectively. The last two columns show the average for the two homogeneous subgroups, HM-H and HM-L. Unadjusted p-values are presented in Table A7 in the Appendix. Statistical differences occur between HM and C in two categories of household composition and in average absenteeism, between HT and C in the average course and absenteeism, and between HT and HM in the IVV score. Most of the differences between the HM-H and HM-L subgroups are statistically relevant, which is desirable in this particular experimental design.

Table 2: **DESCRIPTIVE STATISTICS OF THE RISK FOR VIOLENCE INDEX MEASURE BY TREATMENT GROUP**

	(1)	(2)	(3)	(4)	(5)
		Treatments		Tracking Groups	
	Control Group (C)	Hetero. Group (HT)	Homo. Group (HM)	Homo.-High (HM-H)	Homo.-Low (HM-L)
Mean	0.038	0.041	0.037	0.051	0.023
SD	0.029	0.035	0.026	0.028	0.014
Min.	0.003	0.001	0.002	0.012	0.002
Median	0.029	0.030	0.031	0.044	0.020
Max.	0.183	0.216	0.154	0.154	0.059
Obs.	258	263	535	267	268

Notes: This table provides summary statistics for the Risk for Violence Index (IVV) predicted using FUSADES (2015) dataset and variables available during the ASP enrollment phase. According to this experimental design, I expect that the HT group's average IVV should fall between the HM-H and HM-L groups' IVV. In addition, the IVV variation in the HT group should be greater than the respective values for both groups.

Table 3: **EFFECTS OF ASP GROUP COMPOSITION ON THE PRIMARY OUTCOMES**

Full Sample. Results from Specification (2)

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Violence and Behavior	-0.071***	-0.023	-0.048**	0.000	1,014
	(0.023)	(0.025)	(0.019)		
	[0.003]	[0.236]	[0.015]		
Attitudes Toward School and Learning	0.063	0.010	0.053	-0.017	1,004
	(0.047)	(0.041)	(0.042)		
	[0.164]	[0.808]	[0.175]		
Emotion Regulation (-)	-0.170*	-0.081	-0.089	-0.001	308
	(0.097)	(0.078)	(0.069)		
	[0.070]	[0.292]	[0.269]		
Academic Performance	0.057	0.052	0.005	0.000	1,023
	(0.059)	(0.048)	(0.045)		
	[0.306]	[0.275]	[0.921]		

Notes: This table presents the effects of tracking and integration on the main outcomes under analysis: violence and behavior, attitudes toward school and learning, emotion regulation, and academic performance. These outcomes were estimated using indices following Anderson (2008). See Appendix A1 for a description of the variables included in each index. Negative estimated coefficients for Emotion Regulation should be interpreted as improvements in the outcome. All regressions include the following controls: grade level, student living with both parents, student living with one parent, morning shift, risk for violence, academic grades (score), behavior, and absences at the baseline as well as three dummy variables indicating the missing values of academic grades, behavior, and absences at baseline. All regressions are estimated using the model of Specification (2). The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, and * indicate that the estimated effect is statistically significant at 1%, 5%, and 10%, respectively.

Table 4: **THE ASP GROUP COMPOSITION'S EFFECTS ON COMPONENTS OF THE MAIN OUTCOMES**

Full sample. Results from Specification (2)

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Panel A: Violence and Behavior					
Delinquency (index)	-0.206**	-0.184**	-0.022	0.000	916
	(0.123)	(0.088)	(0.085)		
	[0.042]	[0.025]	[0.795]		
Violent actions (index)	-0.090	-0.138*	0.047	0.000	956
	(0.083)	(0.065)	(0.066)		
	[0.344]	[0.077]	[0.548]		
Approval of antisocial behavior	-0.100***	-0.101***	0.001	0.174	956
	(0.032)	(0.031)	(0.018)		
	[0.000]	[0.000]	[0.969]		
Behavior at school	0.128*	0.152**	-0.025	0.000	1,010
	(0.082)	(0.058)	(0.066)		
	[0.093]	[0.016]	[0.726]		
Pr. of Having bad behavior reports	-0.085***	-0.039	-0.046+	0.331	1,010
	(0.037)	(0.025)	(0.026)		
	[0.007]	[0.157]	[0.144]		
Panel B: Attitudes Toward School and Learning					
Positive attitude toward school (index)	0.182***	0.068	0.113+	0.000	948
	(0.064)	(0.067)	(0.070)		
	[0.022]	[0.312]	[0.113]		
Time spent on homework (hours)	0.339+	0.256	0.084	2.123	935
	(0.195)	(0.162)	(0.196)		
	[0.105]	[0.159]	[0.627]		
Attention paid in class	0.101**	0.065*	0.036	0.591	962
	(0.048)	(0.030)	(0.040)		
	[0.016]	[0.058]	[0.329]		
Absences (days)	-1.639**	-1.329**	-0.311	7.156	843
	(0.526)	(0.461)	(0.554)		
	[0.014]	[0.013]	[0.544]		

	Heter. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Panel C: Emotion Regulation					
Arousal (stress)	-0.133	0.170+	-0.303***	0.000	308
	(0.116)	(0.121)	(0.086)		
	[0.303]	[0.116]	[0.007]		
Valence	-0.304	-0.382*	0.078	0.000	308
	(0.189)	(0.188)	(0.219)		
	[0.266]	[0.087]	[0.750]		
Positive Valence difference	-0.313	-0.337*	0.024	0.000	308
	(0.144)	(0.152)	(0.177)		
	[0.156]	[0.053]	[0.876]		
Negative Valence difference	-0.206	-0.306	0.100	0.000	308
	(0.308)	(0.189)	(0.291)		
	[0.560]	[0.274]	[0.725]		
(External) Locus of control	-0.129	-0.366**	0.237*	0.000	295
	(0.149)	(0.118)	(0.117)		
	[0.443]	[0.013]	[0.088]		
Panel D: Academic Performance					
Academic grades	0.006	0.028	-0.022	0.000	1,023
	(0.064)	(0.048)	(0.053)		
	[0.915]	[0.544]	[0.645]		
Pr. of passing course	0.027	0.021	0.006	0.873	1,023
	(0.023)	(0.020)	(0.018)		
	[0.233]	[0.281]	[0.756]		

Notes: This table presents the effects of tracking and integration on the components of each outcome of interest: violence and behavior (Panel A), attitudes toward school and learning (Panel B), emotion regulation (Panel C), and academic performance (Panel D). Descriptions of the outcome variables are available in Appendix A1. The control variables included are: grade level, student living with both parents, student living with one parent, morning shift, risk for violence, academic grades (score), behavior, and absences at the baseline as well as three dummy variables indicating a missing value of academic grades, behavior, and absences at baseline. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, *, and + indicate that the estimated coefficient is statistically significant at 1%, 5%, 10%, and 15%, respectively.

Table 5: **HETEROGENEOUS EFFECTS OF GROUP COMPOSITION**
Treated Subsample Only. Results from Specification (3)

	HM-Low (1)	HM-High (2)	Observations (3)
Violence and Behavior	0.036*	0.064***	766
	(0.027)	(0.027)	
	[0.060]	[0.001]	
Attitudes Toward School and Learning	-0.045	-0.056+	763
	(0.056)	(0.061)	
	[0.260]	[0.177]	
Emotion Regulation (-)	0.032	0.118+	238
	(0.096)	(0.111)	
	[0.689]	[0.142]	
Academic Performance	0.041	-0.046	771
	(0.061)	(0.049)	
	[0.357]	[0.313]	

Notes: This table shows the differential effects of group composition by initial levels of risk for violence. Following Anderson (2008), indices were constructed as an inverse covariance index. See Appendix A1 for descriptions of the variables included in each index. These estimated coefficients were obtained from estimating Equation (3). All estimations include the following variables as controls: median IVV by stratum, dummy if student has an IVV higher than the median, grade level, student living with both parents, student living with one parent, morning shift, risk for violence, academic grades (score), behavior, and absences at baseline as well as three dummy variables that indicate the missing values of academic grades, behavior and absences at baseline. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, *, and + indicate that the estimated coefficient is statistically significant at 1%, 5%, 10%, and 15%, respectively.

Table 6: **ASP GROUP COMPOSITION EFFECTS ON MARGINAL PARTICIPANTS**

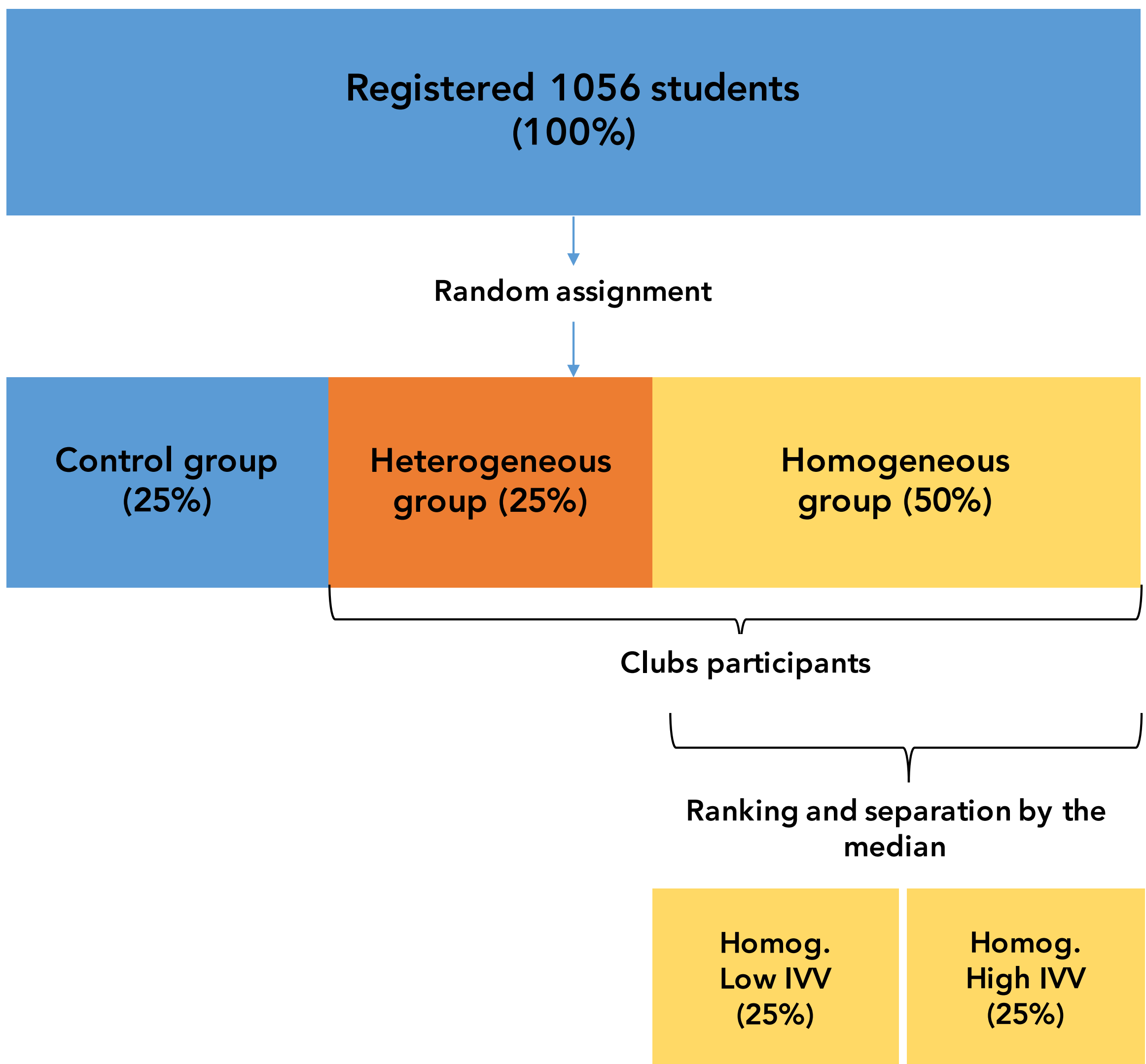
Tracking Groups Only. Results from Specification (4)

	All HM Groups		Only Around Cutoff	
	HM-High	Obs.	HM-High	Obs.
	(1)	(2)	(3)	(4)
Indices				
Violence and Behavior	0.048**	512	0.056	114
	(0.041)		(0.063)	
	[0.045]		[0.305]	
Attitudes Toward School and Learning	-0.058	510	-0.358***	115
	(0.090)		(0.169)	
	[0.157]		[0.000]	
Emotion Regulation	-0.015	151	0.176	14
	(0.171)		(0.387)	
	[0.817]		[0.493]	
Academic Performance	-0.097	516	-0.045	115
	(0.114)		(0.147)	
	[0.160]		[0.795]	

Notes: This table shows the effects of group composition on the main outcomes of interest for marginal students. See Appendix A1 for descriptions of the variables included in each index. I restrict the sample to students assigned to the homogeneous treatment and use Equation (4) to estimate the effects. Column (1) presents the estimated effects on the marginal students using the full sample of students assigned to the HM treatment. Column (3) shows the estimated coefficients when restricting the sample to the eight students around the IVV median within each stratum. Following Anderson (2008), indices were constructed as an inverse covariance index. All estimations include a second-order polynomial of the cumulative IVV distribution among students. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, and * indicate that the estimated coefficient is statistically significant at 1%, 5% and 10%, respectively.

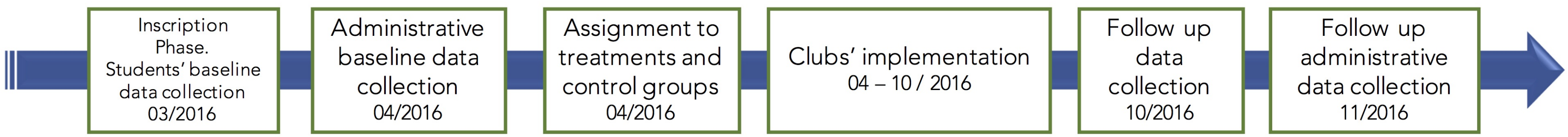
9 Figures

Figure 1: **EXPERIMENTAL DESIGN**



Notes: This figure shows the sample composition and randomization procedure applied in this design. From the total number of enrolled students within each educational level $\{1,2,3\} \in$ school A, I randomly assigned 25% to the C group, 25% to the HT group, and 50% to the HM groups. I followed the same approach for each of the remaining schools.

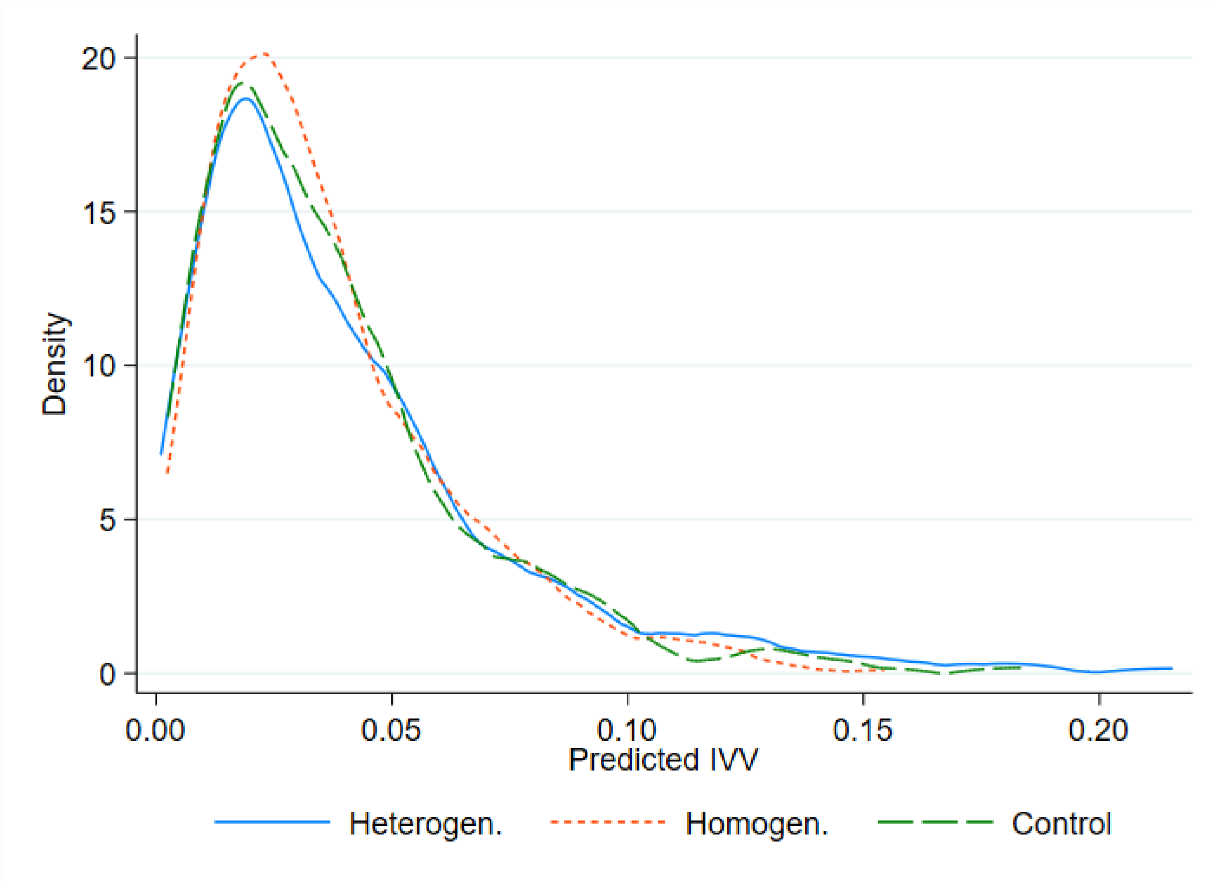
Figure 2: **PROJECT TIMELINE**



Timeline of the intervention and data collection.

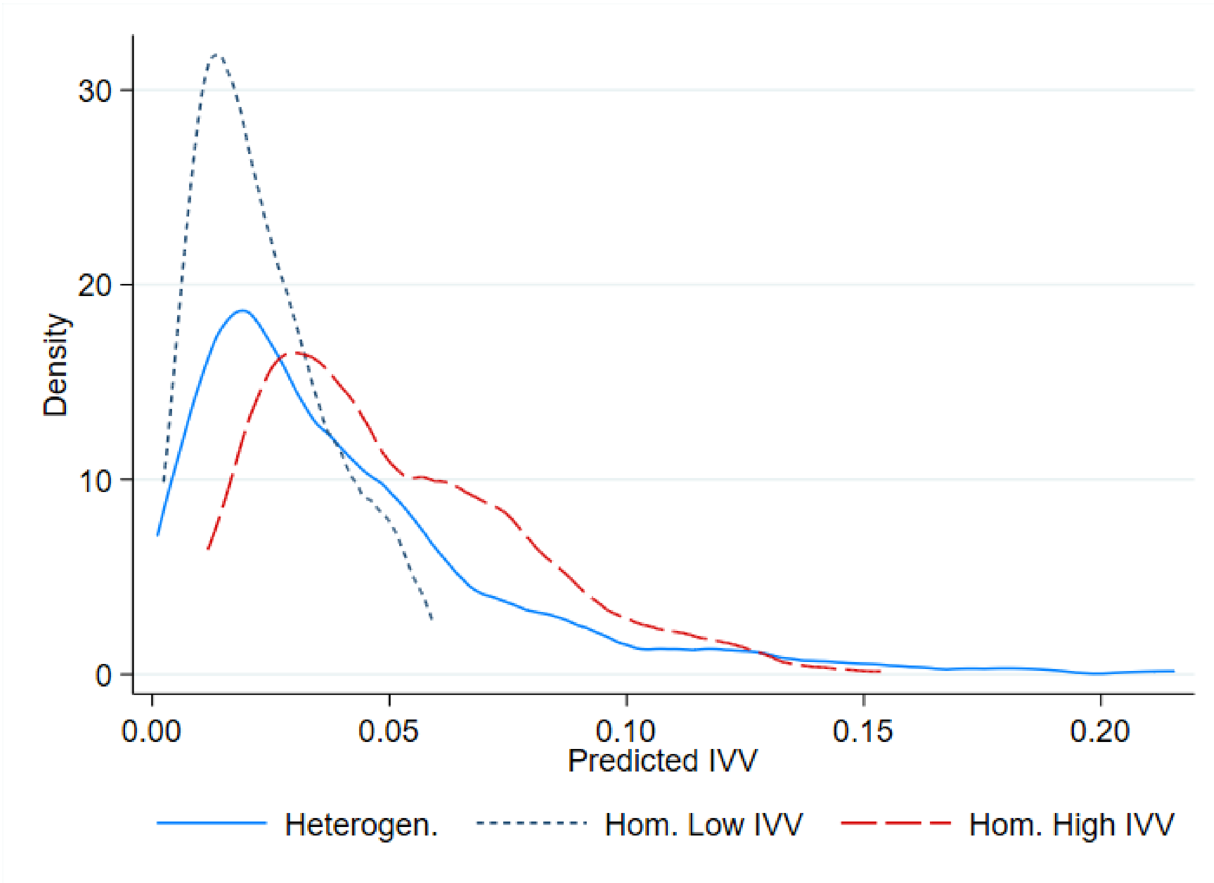
Notes: This figure shows when each of the stages of the project took place, from student enrollment to the collection of follow-up data.

Figure 3: **RISK FOR VIOLENCE DISTRIBUTION FUNCTIONS OF THE TREATMENT AND CONTROL GROUPS**



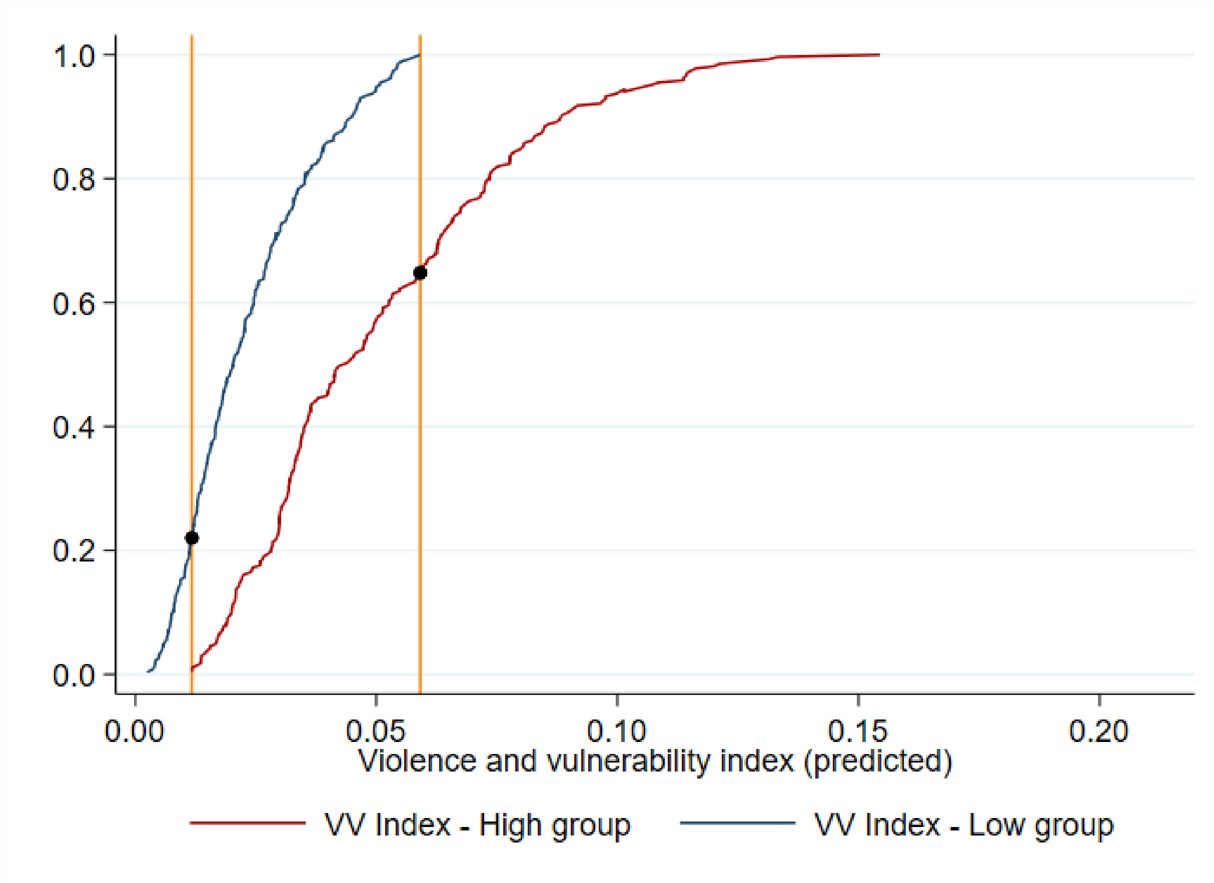
Notes: This figure shows the predicted Risk for Violence Index (IVV) distribution functions for the entire study sample for the control and treatment (homogeneous and heterogeneous) groups prior to treatment. A Two-sample Kolmogorov–Smirnov test for distribution equality was performed while maintaining the null hypothesis of equality of distributions for each of the following comparisons: C vs.HT (p-value = 0.868), C vs. HM (p-value = 0.682), and HT vs. HM (p-value = 0.619).

Figure 4: **RISK FOR VIOLENCE DISTRIBUTION FUNCTIONS OF THE TREATED GROUPS**



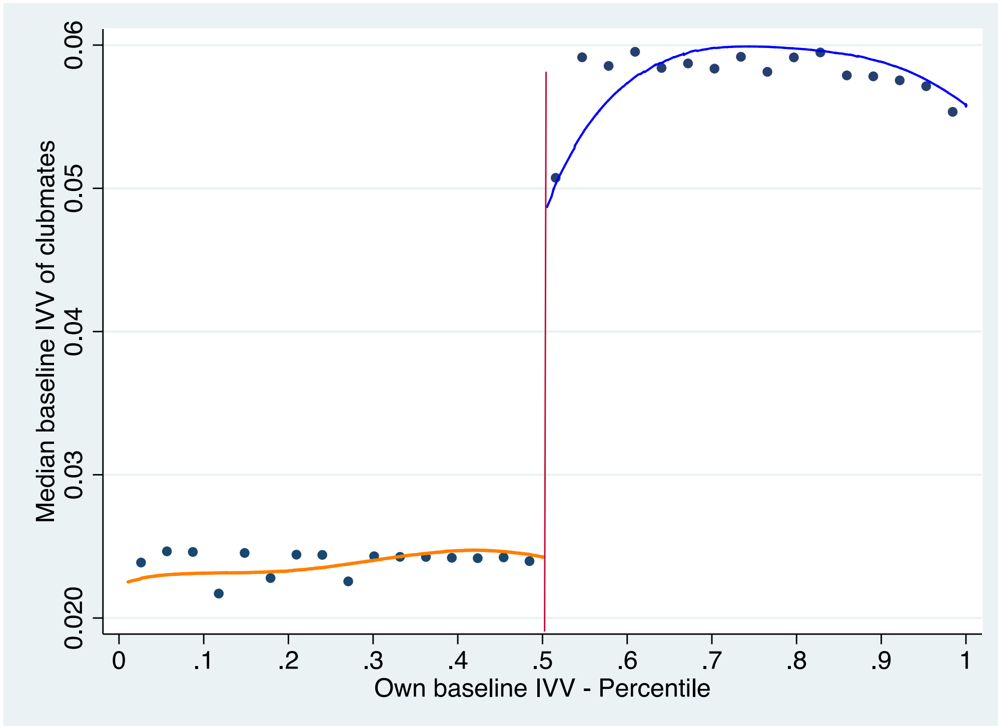
Notes: This figure shows the predicted Risk for Violence Index (IVV) distribution functions generated by the experimental design for the heterogeneous treatment group and each of the homogeneous subgroups (High and Low IVV), meaning the entire study sample. Two-sample Kolmogorov–Smirnov test for equality of distribution was carried out. We reject the null hypothesis of equality of distributions in all comparisons across the different groups: HT vs. HM-Low (p-value = 0.000), HT vs. HM-High (p-value = 0.000), and HM-Low vs. HM-High (p-value = 0.000).

Figure 5: **RISK FOR VIOLENCE CUMULATIVE DISTRIBUTION FUNCTIONS OF THE HOMOGENEOUS SUBGROUPS**



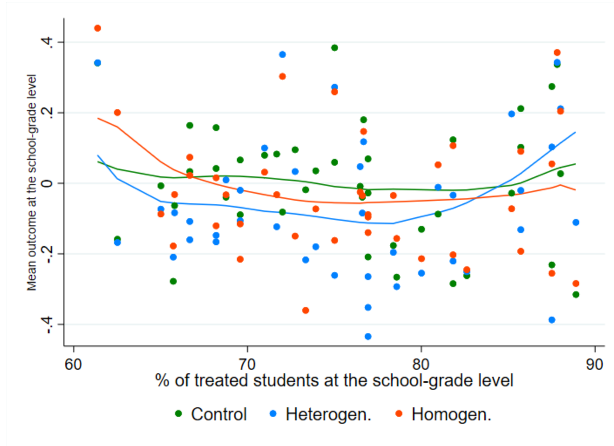
Notes: This figure presents the cumulative distribution function for each of the high- and low-homogeneous treatment groups' predicted risk for violence. The vertical yellow lines define the overlap limits for both distribution functions. This overlap in the risk for violence level occurs because the assignments took place at the stratum level, and the median level was different within each stratum.

Figure 6: **EXPERIMENTAL VARIATION IN IVV PEER COMPOSITION PRIOR TO TREATMENT**

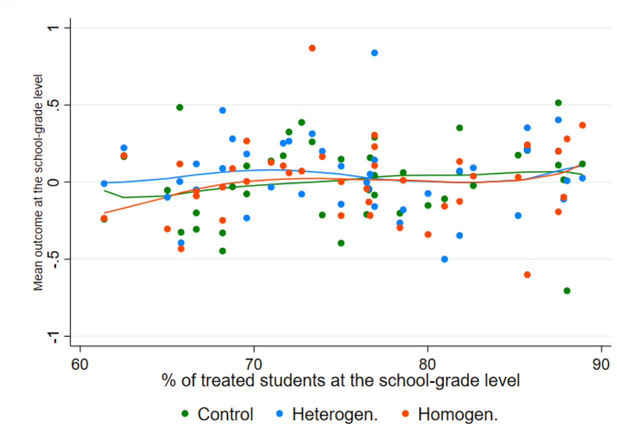


Notes: This figure shows the median predicted IVV of student's peers as a function of the student's own baseline IVV in the high- and low-homogeneous treatment groups. Consistent with the discontinuous assignment at the median IVV, there is a sharp discontinuity at the fiftieth percentile for the entire subsample.

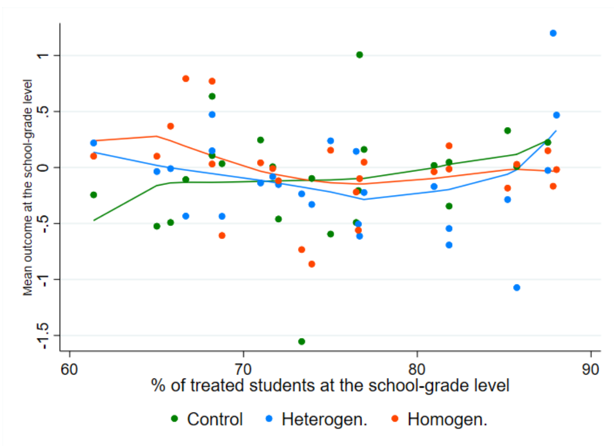
Figure 7: **AVERAGE OUTCOME AND SHARE OF TREATED STUDENTS AT THE SCHOOL-GRADE LEVEL BY TREATMENT STATUS**



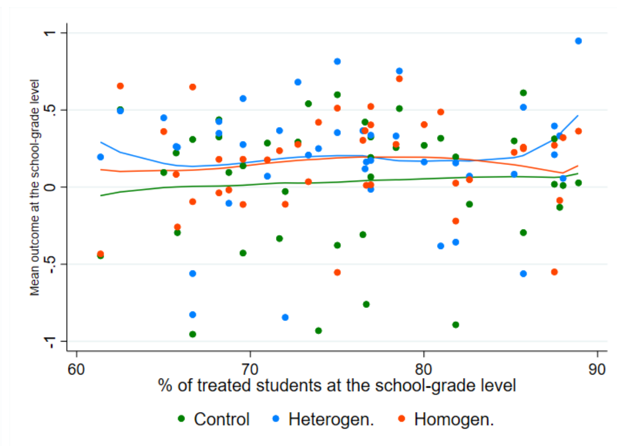
(a) Violence and Behavior



(b) Attitudes Toward School and Learning



(c) Emotion Regulation



(d) Academic Performance

Notes: These figures show the mean outcome and share of treated students (treatment saturation) at the grade level for all outcomes under analysis. The lines correspond to the nonparametric estimations of the average outcome by treatment status (the heterogeneous [HT], homogeneous [HM], or control [C] groups). I exclude the clusters at the tails of the saturation distribution (i.e., those with a saturation of <60% or =100%) from the graphs. *P*-values from Kolmogorov-Smirnov test of the equality of the distributions across groups for each outcome are presented in Table A11.

Appendix

For Online Publication Only

A1 Description of Outcome Variables.

In this appendix, I present the definitions of all the variables included in the construction of each of the four main outcomes (violence and behavior, attitudes toward school and learning, emotion regulation, and academic performance).

A. Violence and Behavior This index is made up of five variables: three measures of violent behaviors estimated using survey data (delinquency index, violent actions index, and approval of antisocial behaviors) and two measures of misbehavior at school (behavior at school and probability of having a bad behavior report) estimated using teachers' reports.

1. *Delinquency index*: Corresponds to a standardized sum of self-reported committal of delinquent actions such as theft, mugging, etc. The question was: *In the last 3 months, have you ...?* This variable is the sum of delinquent actions that I then standardized relative to the control group at the school-by-education level.

2. *Violent actions index*: Consists of the standardized sum of other violent acts such as fighting at school, damaging municipal property, fighting with siblings, etc. The question was: *In the last 3 months, have you ...?* This index variable can be interpreted as the sum of violent actions that I standardized relative to the control group at the school-by-education level.

3. *Approval of peers' antisocial behavior:* A binary indicator that equals 1 if a student approves of any peer behaviors related to alcohol and drug consumption, fighting, etc. The question was: *What would you think if one of your closest friends ...?*

4. *Behavior reports*: In El Salvador, teachers prepare student behavior reports based on the following discrete scale: Excellent (E), Very Good (VG), Good (G), and Fair (F). This scale can be translated into a numerical scale that is comparable to academic

grades. In this paper, I used a reversed continuous scale to facilitate the interpretation and comparison of the behavior reports to the self-reported measures of violence and crime. For the estimations, I standardized these values relative to the control group at the school-by-education level.

5. *Probability of having a bad behavior report:* A dummy variable that equals 1 if the student's behavior report was Fair (F) (or >5.5 points), or 0 otherwise.

B. Attitudes Toward School and Learning. This index is made up of four variables: three measures of attitudes toward school estimated using survey data (positive attitude toward school, time spent on homework after school, and attention paid in class) and one measure of school absenteeism (number of absences) estimated using teachers' reports.

1. *Positive attitude towards school:* A standardized index that includes the following 5 items from the self-reported follow-up survey: (i) an indicator of how the student values learning, (ii) an indicator of how the student likes school, (iii) an indicator of the student's willingness to study as a means to a better future, (iv) an indicator of whether the student thinks hard work at school pays off, and (v) an indicator of whether the student thinks that what he/she learns at school is relevant for the future. Each item was ranked on a sale of 1 to 4, where 1 equals very important, and 4 equals not important. For the estimations, I have standardized these values relative to the control group at the school-by-education level.

2. *Time spent on homework*: A student self-reported variable measured in hours. The question was: *In the last 3 months, how much time did you spend doing your homework outside of school?*

3. *Attention paid in class*: A self-reported outcome. The student responded to the statement: *I pay attention when I am in class* on a scale of 1 to 4, where 1 equals always, and 4 equals never. Based on this scale, I created a dummy that equals 1 if the student reported paying attention most of the time (i.e., scale value: 1 or 2), or 0 if the student reported barely paying attention during class (i.e., scale value: 3 or 4).

4. *Absences*: Corresponds to the number of days the student did not attended school. Absences at baseline were measured as the number of days the student was not present at school between January and March of the 2016. Absences at endline were measured as the number of days a student was absent between April and October of the 2016 academic year. This information was part of the administrative data provided by schools.

C. Emotion Regulation. I use the data on neurophysiological markers via electroencephalograms (EGG), which was collected in Dinarte-Diaz and Egana-delSol (2023). Using the recordings, I followed Egana-delSol et al. (2023) and estimated the outcomes I describe below. I standardized the values of the estimations relative to the control group at the school-grade level.

1. *Arousal*: A pre-test measure of alertness at rest obtained directly from the student's brain activity measured using EEG while the student looked at a black cross in the center of a gray screen for 30 seconds.
2. *Valence*: A pre-test state measure of valence at rest estimated directly from the student's brain activity using EEG while the student looked at a black cross in the center of a gray screen for 30 seconds. This variable can be interpreted as either a positive or negative mood as well as either an attraction to or withdrawal from a stimulus (Harmon-Jones et al, 2018; Kassam et al, 2013).
3. *Positive stimuli valence*: The difference between the response intensity measure after exposure to positive stimuli and the valence index described above.
4. *Negative stimuli valence*: The difference between the response intensity measure after exposure to negative stimuli and the valence-at-rest index described above. Both differences can be interpreted as the student's hyperreactivity level. In other words, the student has become calmer and moves towards physical and emotional withdrawal.
5. *Locus of control*: A psychometric test developed by Rotter (1966) that indicates the degree to which a person feels he has agency in his life.

D. Academic Performance. I collected this information at baseline (before the intervention commenced) and at the end of the 2016 academic year (after the program finished) using administrative data (teachers' reports).

1. *Academic grades*: A numeric variable that reports the average of the student's combined scores for math, science, and reading. I use the average of the three subjects since there is no indication that the intervention would affect any of the three subjects differently. Grades are given on a 0 to 10 scale, where 0 is the worst academic performance and 10 is the best. For the analysis, I replaced missing values with the school-by-education level mean and subsequently standardized the latter against the control group at the school-by-education level.

2. *Probability of passing the course*: A dummy variable that equals 1 if a student passed the course (i.e., if the student's average school grade is equal to or greater than 5), and 0 otherwise.

A2 Exploiting the random allocation of peers

In addition to the main group composition effects obtained from the direct variation generated by this experiment, I also exploit an additional variation in peer quality generated by my tracking-by-risk-for-violence design. Since participants in the HT subsample were assigned randomly to a group in the ASP, they will have a random set of peers. I can restrict the sample to these groups and estimate the effect of a student's peers' mean and variance baseline IVV by OLS using the following equation:

$$y_{ij} = \gamma_0 + \gamma_1 \bar{x}_{-ij} + \gamma_2 \bar{X}_{-ij} + \gamma_3 var(x_{-ij}) + \gamma_3 Var(x_{-ij}) + \gamma_4 X_{ij} + \epsilon_{ij} \tag{A1}$$

where $\bar{x}_{-ij}$ and $var(x_{-ij})$ are the club's mean and variance to which student i was assigned, excluding the student's personal IVV score. The latter allows me to address the reflection problem. In addition, $\bar{X}_{-ij}$ and $Var(x_{-ij})$ are the IVV mean and variance of all treated students, also excluding student i's IVV. The vector of control variables X_{ij} includes student's own baseline IVV. The rest of the variables are defined as before. The estimated coefficients of interest are γ_1 and γ_3, which reflect the causal effect of student i's peers' mean and variance in risk for violence on the student's own results.

In addition, since the intervention participants were randomly allocated to a group in the ASP, there is some variation in group composition which stems from the fact that being assigned to the homogeneous treatment or the heterogenenous treatment directly affects the mean and variance of one's peers. As in Lafortune et al. (2016), once I control for a strata fixed effect, the variance and mean IVV for peers stems entirely from the participant's random assignment. Carrell et al. (2013); Duflo et al. (2011); Boozer and Cacciola (2001); Lyle (2007) have all used a similar approach. The estimating equation for the sample of students selected to participate in the ASP is:

$$Y_{ij} = \gamma_0 + \gamma_1 \bar{x}_{-ij} + \gamma_2 var(x_{-ij}) + \gamma_3 S_j + \gamma_4 X_{ij} + \epsilon_{ij} \tag{A2}$$

All variables are defined as in Specification A1. With this equation I can directly provide evidence of how student i's non-cognitive and/or academic outcomes are affected by his/her peers' average or variance in risk for violence at baseline.

I present the results from both specifications in table A10 in the Appendix Tables Section. Panel A shows the effects on behaviors and attitudes towards school and learning. Panel B presents the estimated effects of each group's composition on emotion regulation, stress, and psychometric outcomes. Finally, Panel C shows the estimated effects of both of the specifications on academic performance. Columns (1) to (3) present the results using Model 1, while Columns (4) to (6) show similar estimations obtained from Model 2. Using these alternative estimation approaches, I obtain results similar to those when using direct variation on group composition generated by the experiment. Panel A shows that a higher average peer IVV reduces the student's self-reported amount of time spent on homework, whereas being in a more diverse group increases both self-reported time spent on homework and reduces absenteeism. In terms of violence, I do not find an effect from either the mean or standard deviation of the peers' IVV. Despite the lack of power from neurophysiological outcomes and psychometric tests, I find that greater diversity reduces the participants' stress levels. Finally, the effects on academic performance indicate that a higher level of peer risk for violence can improve a student's academic performance, which is in line with Hoxby (2000)'s invidious comparison model. However, greater diversity can also improve the extensive margin of academic grades.

Appendix Tables

Table A1: **COMPARISON OF THE STUDY AND FUSADES (2015) SAMPLES**

	(1)	(2)	(3)	(4)	(5)
	Study Sample		FUSADES (2015) Sample		
	Mean	SD	Mean	SD	*P*-value
Student is male	0.49	0.50	0.47	0.50	0.23
Student lives in an urban area	0.73	0.44	0.66	0.47	0.10
Household composition					
Student living with both parents	0.53	0.49	0.54	0.50	0.55
Student living with only one parent	0.32	0.47	0.30	0.46	0.19
Student living with one stepparent	0.06	0.25	0.08	0.27	0.02
Student living with other relative/adult	0.08	0.27	0.07	0.26	0.25
Student's commute time from home to school (minutes)	17.64	14.37	17.25	12.98	0.37
Student's mother's education level (basic, 1-6 years)	0.34	0.46	0.4	0.49	0.40
Student is alone at home after school	0.05	0.22	0.11	0.31	0.00
Student's age	11.95	2.95	13.87	1.67	0.10
Student's grade level	5.75	2.71	5.5	2.52	0.29
N	1056		6641		

Notes: This table shows the means and standard deviations of the main baseline characteristics of students that coincide in this study and in the FUSADES (2015) samples. These variables were used to estimate the IVV for each student in the study sample. Column (5) shows the *p*-value of the comparison of means between both samples. ***, **, and * denote difference significance at the 1%, 5%, and 10% level, respectively, when comparing the means.

Table A2: **IVV ESTIMATION RESULTS AND DETERMINANTS USING THE FUSADES (2015) SAMPLE**

	Violence
Student is male	0.258***
	(0.054)
Student's age	0.092***
	(0.017)
Student lives in an urban area	0.195***
	(0.066)
Student's household composition	
Student living with only one parent	0.033
	(0.062)
Student living with other relative	0.042
	(0.112)
Student living with other nonrelative adult	0.723
	(0.466)
Student not living with any adults	0.362
	(0.290)
Student's mother's level of education:	
Intermediate education (7–12 years)	0.113*
	(0.061)
University or higher (13 and +)	0.057
	(0.079)
Student's commute time from home to school (minutes)	0.005**
	(0.002)
Student is alone at home after school	0.391***
	(0.070)
Student's school year	0.067
	(0.089)
Student is enrolled in the morning shift	-0.002
	(0.087)

Notes: I estimated Specification (1) denoted by $V_f = \alpha_0 + \alpha_1 D_f + \epsilon_f$. The FUSADES (2015) survey defined V_f as a violence dummy indicating that a child or adolescent has done one of the following: *Have you ever: (i) carried a gun, (ii) attacked someone with the intention to hurt him, (iii) shot someone with a gun, (iv) used a gun or violence to get money or things from someone?*. D_f is a vector of violence determinants, including gender, age, mother's education, etc.
***, **, * indicate whether the estimated coefficients α_1 are statistically different from zero. Standard errors are in parentheses. The o mitted category from mother's education is: mother has a basic education (1st–6th grade). The omitted category from household composition is: children living with both parents.

Table A3: **RELATIONSHIP BETWEEN IVV, ACADEMIC GRADES, AND MISBEHAVIOR REPORTS AT BASELINE**

	Academic Score (1)	Misbehavior at School (2)
Panel A: Standardized and Imputed Grades		
IVV	-3.611**	5.735***
	(1.514)	(1.551)
Observations	1,056	1,056
Panel B: Standardized Grades at the Course Level		
IVV	-3.992**	5.644***
	(1.642)	(1.551)
Observations	970	1,000
Panel C: Nonstandardized Grades		
IVV	-5.528**	7.485***
	(2.330)	(2.040)
Observations	970	1,000

Notes: This table uses administrative data (teachers' reports) to present the correlations between the IVV prediction for all students and their academic score and misbehavior reports before the intervention began. The estimations are based on the following specification: $y_{ij} = \beta_1 IVV_{ij} + \alpha_i + \epsilon_{ij}$, where y_{ij} is the academic score or misbehavior report for student i in school j, IVV_{ij} is the estimated risk for violence, and α_i are strata fixed effects. Panel A includes a dummy variable that indicates the missing value of academic score (for Column [1]) or misbehavior at school (Column [2]) as a control in the estimations. ***, **, and * indicate that coefficients are significant at 1%, 5%, and 10%, respectively. Clustered standard errors at the course level are in parentheses.

Table A4: **IVV PREDICTION POWER OF MISBEHAVIOR AT SCHOOL USING DATA FROM THE CONTROL GROUP**

	Misbehavior at School	
	(1)	(2)
IVV	9.287***	8.479**
	(3.330)	(3.282)
Observations	258	258
Controls	No	Yes

Notes: This table presents the correlation between the IVV prediction at baseline and misbehavior reports at the end of the 2016 academic year using administrative data only for the control group, meaning those who did not receive treatment. The following was the estimated specification: $y_{ijt} = \alpha_0 + \alpha_1 IVV_{ijt-1} + \epsilon_{ijt}$, where y_{ijt} is the misbehavior report for student i in school j in the period t (one year after), and IVV_{ijt-1} is the estimated risk for violence one year before. ***, **, and * indicate that coefficients are significant at 1%, 5%, and 10%, respectively. Robust standard errors at course-school level are in parentheses.

Table A5: **MATCHING RATE WITH ADMINISTRATIVE DATA AND ATTRITION RATE**

		Treatments		Tracking Groups	
	Control Group (C)	Hetero. Group (HT)	Homo. Group (HM)	Homo.-High (HM-H)	Homo.-Low (HM-L)
Panel A. Share of Students with Matched Administrative Data, Baseline (Q1 2016)					
Academic Scores	0.90	0.93	0.92	0.92	0.93
Behavior Scores	0.93	0.95	0.95	0.94	0.96
Absenteeism	0.68	0.68	0.67	0.65	0.69
Panel B. Share of Students with Matched Administrative Data, Endline (Q4 2016)					
Academic Scores	0.98	0.97	0.96	0.97	0.96
Behavior Scores	0.96	0.95	0.96	0.96	0.95
Absenteeism	0.94	0.96	0.95	0.96	0.94
Panel C. Number of Students at Baseline and Follow-up					
Number of Students Present at Baseline	258	263	535	267	268
Number of Students Present at Follow-up	237	248	483	239	244
Retention Rate (1-attrition)	0.92	0.94	0.90	0.90	0.91

Notes: This table provides the enrolled students match rates with administrative data at baseline and follow-up. These rates are calculated as the share of enrolled students who could be matched with administrative data from schools at baseline (Panel A) or at endline (Panel B). Panel C presents the retention rate—that is, the share of enrolled students for who we collected self-reported data during the follow-up survey.

Table A6: ***P*-VALUES FOR THE DIFFERENCES IN MATCHING RATE WITH ADMINISTRATIVE DATA AND ATTRITION RATE**

	P-values			
	Treatments			Tracking Groups
	C = HT (1)	C = HM (2)	HT = HM (3)	HM-High = HM-Low (4)
Share of Students with Matched Administrative Data, Q1 2016				
Academic Scores	0.214	0.103	0.794	0.884
Behavior Scores	0.339	0.444	0.796	0.482
Absenteeism	0.682	0.413	0.129	0.874
Share of Students with Matched Administrative Data, Q4 2016				
Academic Scores	0.425	0.200	0.709	0.529
Behavior Scores	0.583	0.616	0.970	0.126
Absenteeism	0.320	0.506	0.530	0.072
Number of Students at Baseline and Follow-up				
Retention Rate (1-attrition)	0.220	0.385	0.045	0.380

Notes: This table shows unadjusted *p*-values of tests for differences in matching rates with baseline (Panel A) and endline (Panel B) administrative data. Panel C presents the *p*-values of tests of differences in completion rates of the self-reported endline survey. In Columns (1) and (2), I present the *p*-values of the balance tests between the control group and treatment arms. Column (3) presents similar values for the comparison between the treatment arms, and Columns (4) and (5) show the results of the test between the two homogeneous subtreatments, HM-High and HM-Low.

Table A7: ***P*-values OF DIFFERENCES BETWEEN THE TREATMENT AND CONTROL GROUPS**

	(1)	(2)	(3)	(4)
	Differences Across Treatments			Differences Across Tracking Groups
	C = HT	C = HM	HT = HM	HM-High = HM-Low
Panel A: Individual and Household Characteristics				
Student is male	0.543	0.740	0.616	0.000
Student's age	0.163	0.405	0.216	0.000
Student lives in an urban area	0.949	0.504	0.563	0.102
Student's household composition				
Student lives with both parents	0.228	0.050	0.300	0.137
Student lives with one parent	0.590	0.056	0.248	0.340
Student lives with a parent and stepparent	0.548	0.793	0.672	0.907
Student lives with other relative/adults	0.516	0.632	0.780	0.844
Mother's education level				
Basic education (1-6 years)	0.105	0.481	0.147	0.002
Intermediate education (7-12 years)	0.131	0.566	0.369	0.001
University or higher (13 and +)	0.432	0.997	0.625	0.630
Student's commute time from home to school (minutes)	0.539	0.501	0.991	0.036
Student is alone at home after school	0.189	0.835	0.143	0.001
Student's grade level	0.086	0.120	0.368	0.012
Student is enrolled in the morning shift	0.283	0.349	0.051	0.462
Student's risk for violence (index)	0.232	0.716	0.051	0.000
Panel B: Main Outcomes				
Academic grades	0.280	0.206	0.846	0.043
Pr. of passing course (%)	0.165	0.017	0.305	0.002
Behavior at school	0.137	0.095	0.837	0.053
Pr. of having bad behavior reports (%)	0.339	0.197	0.549	0.165
Absences (days)	0.074	0.013	0.404	0.069
Panel C: Club Characteristics				
Average club size	0.926	0.926	0.926	0.385
Average club take-up (%)	0.910	0.910	0.910	0.286
Community tutors	0.354	0.354	0.354	0.352

Notes: This table shows the unadjusted *p*-values of the balance tests for all variables available at baseline. Panel A presents the *p*-values of the tests for variables obtained from the enrollment form. Panel B shows similar values for variables from the administrative data for consenting students. In Columns (1) and (2), I present the *p*-values of the balance tests between the control group and each treatment arm. Column (3) presents similar values for the comparison between the treatment arms. Column (4) shows the results of the test between the two homogeneous subtreatments, HM-High and HM-Low.

Table A8: **CLASSIFICATION USING MISBEHAVIOR REPORTS OR ESTIMATED RISK FOR VIOLENCE (IVV)**

	(1)	(2)	(3)	(4)	(5)
	Full Sample	Treated (T)	Control (C)	Hetero. (HT)	Homo. (HM)
Similar Classification	0.527	0.528	0.527	0.513	0.535
Observations	1,056	798	258	263	535
		Two-sample T-Test for Differences			
		T = C	C = HT	C = HM	HT = HM
P-value of Differences		0.990	0.753	0.844	0.572

Notes: The variable = 1 if a student would have been classified as having a high risk for violence based on where the student's falls within the IVV and misbehavior reports distribution functions, at the stratum-treatment arm (C, T, Hetero., Homo.) level. Tests include strata fixed effects. Robust standard errors at the course-school level are in parentheses.

Table A9: **HETEROGENEOUS EFFECTS OF GROUP COMPOSITION**

Treated Subsample Only. Results from Specification (3)

	HM-Low (1)	HM-High (2)	Observations (3)		HM-Low (1)	HM-High (2)	Observations (3)
Panel A: Violence and Behavior				**Panel C: Emotion Regulation**			
Delinquency (index)	0.048	0.000	691	Arousal (stress)	0.197*	0.324***	238
	(0.103)	(0.130)			(0.132)	(0.083)	
	[0.595]	[0.997]			[0.062]	[0.001]	
Violent actions (index)	-0.009	-0.094	722	Valence	-0.177	0.069	238
	(0.078)	(0.100)			(0.382)	(0.369)	
	[0.914]	[0.237]			[0.530]	[0.776]	
Approval of antisocial behavior	-0.002	-0.004	720	Positive Valence difference	-0.026	0.058	238
	(0.029)	(0.022)			(0.294)	(0.353)	
	[0.907]	[0.836]			[0.908]	[0.753]	
Behavior at school	0.046	0.023	762	Negative Valence difference	-0.022	-0.114	238
	(0.082)	(0.077)			(0.399)	(0.486)	
	[0.479]	[0.719]			[0.946]	[0.722]	
Pr. of having bad behavior reports	0.018	0.071**	762	(External) Locus of control	-0.237	-0.295**	227
	(0.035)	(0.036)			(0.139)	(0.154)	
	[0.534]	[0.013]			[0.123]	[0.046]	
Panel B: Attitudes Toward School and Learning				**Panel D: Academic Performance**			
Positive attitude toward school (index)	-0.122*	-0.107	716	Academic grades	0.081*	-0.026	771
	(0.080)	(0.094)			(0.067)	(0.056)	
	[0.065]	[0.111]			[0.089]	[0.573]	
Time spent on homework (hours)	0.091	-0.239	707	Pr. of passing course	0.007	-0.018	771
	(0.307)	(0.260)			(0.025)	(0.020)	
	[0.624]	[0.198]			[0.717]	[0.319]	
Attention paid in class	-0.078**	0.013	727				
	(0.044)	(0.055)					
	[0.030]	[0.724]					
Absences (days)	0.327	0.106	638				
	(0.640)	(0.847)					
	[0.501]	[0.815]					

Notes: For all outcomes, the same controls were used, including: median IVV by strata, dummy if student has an IVV higher than median, grade level, student living with both parents, student living with one parent, morning shift, risk for violence, academic grades (score), behavior, and absences at baseline as well as three dummy variables indicating the missing values of academic grades, behavior, and absences at baseline. Outcomes are from data collected during the follow-up survey, at the end of the 2016 academic year. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses. Randomization inference *p-values* are in brackets. These are the results of 2000 randomizations. ***, **, and * indicate that the effect of being treated in a HM-High or -Low group compared to being treated in a HT group is significant at 1%, 5%, and 10%, respectively. Panel A presents the effects on behaviors and attitudes toward school and learning. Panel B shows the results for emotion regulation outcomes and psychometric tests. Panel C lists the estimated effects on academic performance. All regressions are estimated using models of Specification (3).

Table A10: **FURTHER MEASURES OF GROUP COMPOSITION EFFECTS FOR INDICES**

	Nontracking Group Only. Results from Specification (A1)			*All Treated Participants. Results from Specification (A2)*		
	Mean(IVV) of Peers	SD (IVV) of Peers	N	Mean (IVV) of Peers	SD (IVV) of Peers	N
	(1)	(2)	(3)	(4)	(5)	(6)
Indices						
Violence and Behavior	5.645*	-5.226*	254	2.574*	-4.149***	766
	(2.945)	(2.886)		(1.353)	(1.358)	
Attitudes Toward School and Learning	10.410	-6.925	253	-1.603	4.178**	763
	(6.644)	(9.108)		(1.399)	(1.874)	
Emotion Regulation (-)	0.374	11.283	87	7.423**	-4.246	236
	(7.697)	(9.943)		(3.453)	(3.398)	
Academic Performance	-1.772	6.469	255	-5.126**	5.310*	771
	(7.362)	(9.091)		(2.510)	(2.746)	

Notes: This table presents the effects of changes in the mean and variance of risk for violence (IVV) of one's peers on the main outcomes under analysis. As in Lafortune et al. (2016) and Duflo et al. (2011), the identification assumption is that the variance and mean IVV of peers arise entirely from the random assignment after controlling for strata fixed effects. Details of the estimation and a summary of the results are included in Appendix A2. Negative estimated coefficients for emotion regulation should be interpreted as improvements in the outcome. All regressions are estimated using a model of Specification (2). The sample size for each specification varies according to the amount of data available for each subsample and output. Clustered standard errors are in parentheses. ***, **, and * indicate that the estimated effect is statistically significant at 1%, 5%, and 10%, respectively.

Table A11: ***P*-VALUES FOR THE KOLMOGOROV-SMIRNOV TEST FOR EQUALITY OF DISTRIBUTION FUNCTIONS PRESENTED IN FIGURE 7**

	C = HT (1)	C = HM (2)	HT = HM (3)
Violence and Behavior	0.021	0.212	0.472
Attitudes Toward School and Learning	0.933	0.998	0.969
Emotion Regulation	0.502	0.799	0.303
Academic Performance	0.446	0.411	0.735

Notes: This table presents the exact *p*-values of the Kolmogorov-Smirnov equality of distributions test for figure 7.

Table A12: **EFFECTS OF ASP GROUP COMPOSITION ON THE PRIMARY OUTCOMES, EXCLUDING CONTROLS**

Full Sample. Results from Specification (2)

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Indices					
Violence and Behavior	-0.069***	-0.018	-0.050**	-0.000	1,014
	(0.023)	(0.025)	(0.019)		
	[0.004]	[0.358]	[0.011]		
Attitudes Toward School and Learning	0.046	-0.010	0.056	-0.017	1,004
	(0.046)	(0.040)	(0.042)		
	[0.306]	[0.796]	[0.152]		
Emotion Regulation (-)	-0.147+	-0.064	-0.084	-0.001	308
	(0.102)	(0.077)	(0.066)		
	[0.109]	[0.391]	[0.294]		
Academic Performance	0.164**	0.152**	0.012	-0.000	1,023
	(0.088)	(0.061)	(0.064)		
	[0.022]	[0.017]	[0.865]		

Notes: This table presents the effects of tracking and integration on the main outcomes under analysis: violence and behavior, attitudes toward school and learning, emotion regulation, and academic performance. These outcomes were estimated using indices following Anderson (2008). Negative estimated coefficients for emotion regulation should be interpreted as improvements in the outcome. All regressions are estimated using the model of Specification (2), excluding the vector of control variables. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, *, and + indicate that the estimated effect is statistically significant at 1%, 5%, 10%, and 11%, respectively.

Table A13: **EFFECTS OF ASP GROUP COMPOSITION ON COMPONENTS OF THE MAIN OUTCOMES, EXCLUDING CONTROLS**

Full Sample. Results from Specification (2)

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Panel A: Violence and Behavior					
Delinquency (index)	-0.213**	-0.190**	-0.024	0.000	916
	(0.122)	(0.088)	(0.082)		
	[0.035]	[0.020]	[0.783]		
Violent actions (index)	-0.115	-0.160**	0.044	0.000	956
	(0.079)	(0.064)	(0.066)		
	[0.223]	[0.034]	[0.566]		
Approval of antisocial behavior	-0.104***	-0.105***	0.001	0.174	956
	(0.033)	(0.029)	(0.018)		
	[0.000]	[0.000]	[0.966]		
Behavior at school	0.216**	0.242***	-0.025	0.000	1010
	(0.096)	(0.081)	(0.084)		
	[0.029]	[0.002]	[0.767]		
Pr. of having bad behavior reports	-0.116***	-0.070**	-0.046+	0.331	1010
	(0.041)	(0.029)	(0.033)		
	[0.002]	[0.027]	[0.149]		
Panel B: Attitudes Toward School and Learning					
Positive attitude toward school (index)	0.183**	0.080	0.102	0.000	948
	(0.066)	(0.068)	(0.074)		
	[0.023]	[0.247]	[0.152]		
Time spent on homework (hours)	0.404**	0.303+	0.101	2.123	935
	(0.198)	(0.157)	(0.197)		
	[0.060]	[0.106]	[0.553]		
Attention paid in class	0.098**	0.069**	0.029	0.591	962
	(0.049)	(0.030)	(0.041)		
	[0.023]	[0.049]	[0.459]		
Absences (days)	-2.062***	-1.954***	-0.108	7.156	843
	(0.665)	(0.639)	(0.571)		
	[0.002]	[0.000]	[0.845]		

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Panel C: Emotion Regulation					
Arousal (stress)	-0.102	0.170+	-0.272**	0.000	308
	(0.125)	(0.121)	(0.070)		
	[0.424]	[0.125]	[0.015]		
Valence	-0.319	-0.369+	0.050	0.000	308
	(0.185)	(0.186)	(0.240)		
	[0.257]	[0.108]	[0.855]		
Positive Valence difference	-0.295	-0.312*	0.017	0.000	308
	(0.153)	(0.159)	(0.184)		
	[0.164]	[0.064]	[0.943]		
Negative Valence difference	-0.186	-0.270	0.084	0.000	308
	(0.282)	(0.177)	(0.312)		
	[0.613]	[0.331]	[0.743]		
(External) Locus of control	-0.129	-0.310**	0.181	0.000	295
	(0.163)	(0.112)	(0.137)		
	[0.465]	[0.038]	[0.196]		
Panel D: Academic Performance					
Academic grades	0.150*	0.170**	-0.020	0.000	1023
	(0.109)	(0.077)	(0.085)		
	[0.088]	[0.018]	[0.775]		
Pr. of passing course	0.057**	0.048**	0.009	0.873	1023
	(0.029)	(0.022)	(0.022)		
	[0.021]	[0.036]	[0.701]		

Notes: This table presents the effects of tracking and integration on the outcomes of interest, which were estimated using Specification (2) without any controls variables. Panel A exhibits the effects on behaviors and attitudes toward school and learning. Panel B presents the results for the emotion regulation outcomes and psychometric tests. Panel C shows the estimated effects of each group composition on academic performance. Descriptions of the outcome variables are available in Appendix 1. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses. Randomization inference *p-values* are in brackets. These are the results of 2000 randomizations. ***, **, and * indicate that the effect of being treated in a specific group (HT or HM) compared to the control (C) group is significant at 1%, 5%, and 10%, respectively.

Table A14: **HETEROGENEOUS EFFECTS OF SUBGROUP COMPOSITION, EXCLUDING CONTROLS**

Treated Subsample Only. Results from Specification (3)

	HM-Low (1)	HM-High (2)	Observations (3)
Indices			
Violence and Behavior	0.036*	0.064***	766
	(0.027)	(0.027)	
	[0.058]	[0.001]	
Attitudes Toward School and Learning	-0.058	-0.055	763
	(0.055)	(0.061)	
	[0.153]	[0.194]	
Emotion Regulation (-)	0.038	0.122+	238
	(0.093)	(0.100)	
	[0.618]	[0.119]	
Academic Performance	0.026	-0.046	771
	(0.082)	(0.071)	
	[0.670]	[0.410]	

Notes: This table shows the differential effects of group composition by initial levels of risk for violence. Following Anderson (2008), the indices were constructed as inverse covariance indices. See Appendix A1 for descriptions of the variables included in each index. These estimated coefficients were obtained from estimating Equation (3) when excluding the vector of control variables. The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference *p*-values are in brackets. ***, **, *, and + indicate that the estimated coefficient is statistically significant at 1%, 5%, 10%, and 12%, respectively.

Table A15: **EFFECTS OF ASP GROUP COMPOSITION ON THE PRIMARY OUTCOMES USING THE CONTROL VARIABLES SELECTED BY LASSO**

Full Sample. Results from Specification (2).

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Indices					
Violence and Behavior	-0.066***	-0.017	-0.049**	0.000	1,014
	(0.023)	(0.026)	(0.019)		
	[0.005]	[0.387]	[0.014]		
Attitudes Toward School and Learning	0.053	-0.006	0.059+	-0.017	1,004
	(0.047)	(0.040)	(0.042)		
	[0.234]	[0.881]	[0.135]		
Emotion Regulation (-)	-0.154*	-0.078	-0.076	-0.001	308
	(0.100)	(0.078)	(0.064)		
	[0.094]	[0.308]	[0.335]		
Academic Performance	0.097*	0.089*	0.009	0.000	1,023
	(0.062)	(0.049)	(0.049)		
	[0.084]	[0.079]	[0.865]		

Notes: This table presents the effects of tracking and integration on the main outcomes under analysis. These outcomes were estimated using indices following Anderson (2008). See Appendix A1 for descriptions of the variables included in each index. Negative estimated coefficients for emotion regulation should be interpreted as improvements in the outcome. All regressions are estimated using the model of Specification (2) and include the vector of control variables selected by LASSO for each outcome (See Table A19). The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference *p*-values are in brackets. ***, **, *, and + indicate that the estimated coefficient is statistically significant at 1%, 5%, 10%, and 15%, respectively.

Table A16: **EFFECTS OF ASP GROUP COMPOSITION ON COMPONENTS OF THE MAIN OUTCOMES USING CONTROL VARIABLES SELECTED BY LASSO**

Full Sample. Results from Specification (2)

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Panel A: Violence and Behavior					
Delinquency (index)	-0.213**	-0.190**	-0.024	0.000	916
	(0.122)	(0.088)	(0.082)		
	[0.030]	[0.024]	[0.786]		
Violent actions (index)	-0.105	-0.152**	0.048	0.000	956
	(0.079)	(0.062)	(0.066)		
	[0.255]	[0.043]	[0.557]		
Approval of antisocial behavior	-0.104***	-0.105***	0.001	0.174	956
	(0.033)	(0.029)	(0.018)		
	[0.000]	[0.000]	[0.975]		
Behavior at school	0.124+	0.163***	-0.039	0.000	1010
	(0.080)	(0.056)	(0.064)		
	[0.104]	[0.009]	[0.557]		
Pr. of having bad behavior reports	-0.095***	-0.043+	-0.053**	0.332	967
	(0.034)	(0.024)	(0.024)		
	[0.003]	[0.130]	[0.068]		
Panel B: Attitudes Toward School and Learning					
Positive attitude toward school (index)	0.210**	0.099+	0.110+	0.000	948
	(0.066)	(0.069)	(0.071)		
	[0.010]	[0.148]	[0.118]		
Time spent on homework (hours)	0.399**	0.303*	0.096	2.123	935
	(0.197)	(0.157)	(0.197)		
	[0.051]	[0.093]	[0.555]		
Attention paid in class	0.100**	0.068*	0.032	0.591	962
	(0.048)	(0.030)	(0.041)		
	[0.012]	[0.052]	[0.393]		
Absences (days)	-1.639***	-1.380**	-0.259	7.156	843
	(0.577)	(0.501)	(0.575)		
	[0.009]	[0.014]	[0.622]		

	Hetero. Group (1)	Homo. Group (2)	Difference Hetero.-Homo. (3)	Mean C Group (4)	Observations (5)
Panel C: Emotion Regulation					
Arousal (stress)	-0.102	0.170+	-0.272**	0.000	308
	(0.125)	(0.121)	(0.070)		
	[0.442]	[0.126]	[0.018]		
Valence	-0.319	-0.369+	0.050	0.000	308
	(0.185)	(0.186)	(0.240)		
	[0.247]	[0.119]	[0.837]		
Positive Valence difference	-0.295	-0.312**	0.017	0.000	308
	(0.153)	(0.159)	(0.184)		
	[0.162]	[0.080]	[0.927]		
Negative Valence difference	-0.186	-0.270	0.084	0.000	308
	(0.282)	(0.177)	(0.312)		
	[0.599]	[0.355]	[0.757]		
(External) Locus of control	-0.129	-0.310**	0.181	0.000	295
	(0.163)	(0.112)	(0.137)		
	[0.463]	[0.035]	[0.220]		
Panel D: Academic Performance					
Academic grades	0.043	0.069	-0.025	0.000	1023
	(0.067)	(0.051)	(0.057)		
	[0.472]	[0.175]	[0.604]		
Pr. of passing course	0.034+	0.024	0.010	0.878	947
	(0.020)	(0.019)	(0.019)		
	[0.130]	[0.203]	[0.638]		

Notes: This table presents the effects of tracking and integration on the components of each outcome of interest: violence and behavior (Panel A), attitudes toward school and learning (Panel B), emotion regulation (Panel C), and academic performance (Panel D). Descriptions of the outcome variables are available in Appendix A1. All coefficients are estimated using the model of Specification (2) and include vector of control variables selected by LASSO for each outcome (see Table A20). The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, *, and + indicate that the estimated coefficient is statistically significant at 1%, 5%, 10%, and 15%, respectively.

Table A17: **HETEROGENEOUS EFFECTS OF SUBGROUP COMPOSITION USING VARIABLES SELECTED BY LASSO**

Treated Subsample Only. Results from Specification (3)

	HM-Low (1)	HM-High (2)	Observations (3)
Indices			
Violence and Behavior	0.036*	0.058***	766
	(0.028)	(0.027)	
	[0.072]	[0.001]	
Attitudes Toward School and Learning	-0.051+	-0.056	763
	(0.054)	(0.062)	
	[0.138]	[0.168]	
Emotion Regulation (-)	0.044	0.127+	238
	(0.093)	(0.102)	
	[0.618]	[0.119]	
Academic Performance	0.059	-0.049	771
	(0.067)	(0.052)	
	[0.581]	[0.354]	

Notes: This table shows the differential effects of group composition by initial levels of risk for violence. Following Anderson (2008), indices were constructed as an inverse covariance indices. See Appendix A1 for descriptions of the variables included in each index. All coefficients are estimated using the model of Specification (3) and include vector of control variables selected by LASSO for each outcome (see Table A21). The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, *, and + indicate that the estimated coefficient is statistically significant at 1%, 5%, 10%, and 12%, respectively.

Table A18: **ASP GROUP COMPOSITION EFFECTS ON MARGINAL PARTICIPANTS, INCLUDING CONTROL VARIABLES SELECTED BY LASSO**

Tracking Groups Only. Results from Specification (4)

	All HM Groups		Only Around Cutoff	
	HM-High (1)	N (Students) (2)	HM-High (3)	N (Students) (4)
Indices				
Violence and Behavior	0.043**	512	0.053	114
	(0.041)		(0.062)	
	[0.035]		[0.244]	
Attitudes Toward School and Learning	-0.082	510	-0.424***	115
	(0.093)		(0.157)	
	[0.175]		[0.000]	
Emotion Regulation	0.005	151	0.176	14
	(0.173)		(0.422)	
	[0.838]		[0.492]	
Academic Performance	-0.159*	516	-0.043	115
	(0.094)		(0.171)	
	[0.061]		[0.722]	

Notes: This table shows the effects of group composition on the main outcomes of interest for the marginal students. See Appendix A1 for descriptions of the variables included in each index. Column (1) presents the estimated effects on the marginal students based on the full sample of students assigned to the HM treatment. Column (3) shows the estimated coefficients when restricting the sample to the eight students who fall near the IVV median within each stratum. All coefficients are estimated using the model of Specification (4) and include a vector of control variables selected by LASSO for each outcome (see Table A22). The sample size for each specification varies according to the amount of data available for each output. Clustered standard errors are in parentheses, and randomization inference p-values are in brackets. ***, **, and * indicate that the estimated coefficient is statistically significant at 1%, 5%, and 10%, respectively.

Table A19: **CONTROL VARIABLES SELECTED BY LASSO FOR THE MODELS ESTIMATED IN TABLE A15**

	Student's Grade Level	Academic Grades
Violence and Behavior	X	
Attitudes Toward School and Learning	X	
Emotion Regulation	X	
Academic Performance		X

Notes: This table presents the variables selected by the LASSO in the model that estimates each of the outcomes, Specification (2).

Table A20: **CONTROL VARIABLES SELECTED BY LASSO FOR THE MODELS ESTIMATED IN TABLE A16**

	Student is male	Student's age	Student's grade level	Student's risk for violence (index)	Academic grades	Behavior at school	Absences (days)	Pr. of passing course (%)	Pr. of having bad behavior reports (%)
Panel A: Violence and Behavior									
Delinquency (index)									
Violent actions (index)	X								
Approval of antisocial behavior									
Behavior at school	X				X	X			
Pr. of having bad behavior reports					X	X			X
Panel B: Attitudes Toward School and Learning									
Positive attitude toward school (index)		X	X						
Time spent on homework (hours)				X					
Attention paid in class		X			X				
Absences (days)						X	X		
Panel C: Emotion Regulation									
Arousal (stress)									
Valence									
Positive Valence difference									
Negative Valence difference									
Locus of control									
Panel D: Academic Performance									
Academic grades					X				
Pr. of passing course					X			X	

Notes: This table presents the variables selected by LASSO in the model that estimates each of the outcomes, Specification (3).

Table A21: **CONTROL VARIABLES SELECTED BY LASSO FOR THE MODELS ESTIMATED IN TABLE A17**

	Student is male	Student's age	Student lives in an urban area	Basic education (1-6 years)	Student's travel time from house to school (minutes)	Student is alone at home after school	Student's grade level	Student's risk for violence (index)	Academic grades
Indices									
Violence and Behavior	X	X		X		X	X	X	
Attitudes Toward School and Learning	X	X		X		X	X	X	
Emotion Regulation	X	X	X					X	
Academic Performance	X	X			X	X	X	X	X

Notes: This table presents the variables selected by LASSO in the model that estimates each of the outcomes. I implemented a double-LASSO selection.

Table A22: **CONTROL VARIABLES SELECTED BY LASSO FOR THE MODELS ESTIMATED IN TABLE A18**

	Student is male	Basic education (1-6 years)	Intermediate education (7-12 years)	Student's travel time from house to school (minutes)	Student is alone at home after school	Student's grade level	Student's risk for violence (index)	Academic grades
Indices								
Violence and Behavior	X	X	X	X	X	X	X	
Attitudes Toward School and Learning	X	X	X	X	X	X	X	
Emotion Regulation	X				X		X	
Academic Performance	X	X	X	X	X		X	X

Notes: This table presents the variables selected by LASSO in the model that estimates each of the outcomes. I implemented a double-LASSO selection.

Table A23: **LEE BOUNDS FOR ATTRITION ANALYSIS**

	Violence and Behavior	Attitudes toward School and Learning	Emotion Regulation (-)	Academic Performance
	(1)	(2)	(3)	(4)
Heterogeneous Group Treatment				
Lower	-0.084***	-0.008	-0.155**	0.021
	(0.023)	(0.040)	(0.076)	(0.053)
Upper	-0.042*	0.082**	0.003	0.156**
	(0.023)	(0.041)	(0.077)	(0.064)
Homogeneous Group Treatment				
Lower	-0.002	-0.060	-0.101	-0.032
	(0.020)	(0.039)	(0.074)	(0.060)
Upper	0.033	0.021	0.084	0.076
	(0.021)	(0.039)	(0.073)	(0.048)
Difference Heterogeneous - Homogenenous				
Lower	-0.045**	0.008	-0.143*	0.113
	(0.022)	(0.041)	(0.082)	(0.073)
Upper	-0.042*	0.023	-0.066	0.124**
	(0.022)	(0.041)	(0.087)	(0.060)
Observations	1056	1054	363	1053

Notes: This table presents the Lee bounds associated with the estimates for the treatment effects, following (Lee, 2009). No control variables are included in these estimations. The sample size for each specification varies according to the number of observations available for each outcome. Clustered standard errors at the course-school level are in parentheses. ***, **, and * indicate that the estimated bound coefficient is statistically significant at 1%, 5%, and 10%, respectively.

Table A24: **TREATED STUDENTS' ASP ATTENDANCE**

	Homo. Group (1)	HM-Low Group (2)	HM-High Group (3)	Observations (4)
Sessions Attended	-0.289	0.223	-0.737	798
	(0.992)	(1.313)	(1.261)	

Notes: This table shows the students' average ASP attendance by treatment assignment. The outcomes are based on attendance collected by tutors. The results in Column (1) are obtained from the estimation of Specification (2). The results in Columns (2) and (3) are from estimations of Equation (3). All estimations include the following control variables: grade level, student living with both parents, student living with one parent, morning shift, risk for violence, academic grades (score), behavior, and absences at baseline as well as three dummy variables indicating the missing values of academic grades, behavior, and absences at baseline. For these estimations, we restrict the sample to students assigned to any treatment arm. The outcome mean for the heterogeneous treatment group is 25.6 sessions attended (57% take-up rate). *P*-values for the difference between the HM-Low Group and the HM-High Group is 0.499. Clustered standard errors are in parentheses.