Peer Effects on Violence. Experimental Evidence in El Salvador*

Lelys Dinarte†

December, 2018.

Abstract

This paper provides experimental evidence of the role of having different levels of violent peers in the context of an after-school program. By randomly assigning students to participate in the intervention with a set of similar or diverse peers in terms of violence, I measure effects of tracking on students' behavioral, neurophysiological, and academic outcomes. Participants were between 10-16 years old and enrolled in public schools in El Salvador. Results indicate that integrating students with different propensities for violence is better than segregating them, for both highly and less violent children. Particularly, the intervention can have unintended effects on misbehavior and stress if highly violent students are segregated and treated separately from their less violent peers.

Keywords: Peer effects, Tracking, Violence, After-School Programs, Education.

JEL Classification: I29, K42, Z13

*I am grateful to Claudia Martinez, Jeanne Lafortune, Rodrigo Soares, Francisco Gallego, José Tessada, Nicolás Figueroa, Micaela Sviatschi, Tomás Rau and Pablo Egana for their invaluable feedback. I appreciate the comments and suggestions of participants at the 2018 LACEA-LAMES Conference, 2017 NEUDC Conference, 2017 ESA World Meeting, 7th Annual Meeting of America Latina Crime and Policy Network AL-CAPONE, 2017 SECHI Annual Meeting, 5th Antigua Experimental Economics Conference. Additionally, I appreciate comments from participants during the seminars and job market talks at IDB, World Bank, Universidad El Rosario, Universidad de Los Andes, Universidad Javeriana, Pontificia Universidad Católica de Chile, Universidad de Chile (FEN), and Universidad Diego Portales. I am truly grateful with the incredible support of Glasswing International, as an implementer partner, specially to Stephanie Martinez and the education sector team. I am also grateful to principals, teachers, students, and instructors of the 5 public schools in El Salvador. This study was funded by CEDLAS-IDRC (Canada) in the context of "Fostering capacities in Impact Evaluation in Latin America" Project, CONICYT Chile "Soporte Operacional para Becas Nacionales", and USAID El Salvador's Office. The contents of this study are the responsibility of the author and do not necessarily reflect the views of her employer, funding agencies, or governments. This study was registered in the AEA RCT Registry with unique identifying number AEARCTR-000160

†The World Bank. Development Research Group. Email: ldinartediaz@worldbank.org

1. Introduction

Violence and crime substantially reduce productivity, increase the economic costs of health and justice services (Krug et al., 2002), and can be grave hindrances to economic growth (Soares and Naritomi, 2010). Moreover, exposure to violence in childhood and adolescence has a "snowball effect;" children and adolescents with early exposure to violence tend to be involved in other types of violence later in life (Sousa et al., 2011; Damm and Dustmann, 2014).[1]

After-school programs (ASP) are a type of intervention that can *protect* children, preventing victimization and delinquent behavior (Gottfredson et al., 2007; Mahoney et al., 2001). These programs can also act as an alternative source of *learning* and social development (Taheri and Welsh, 2016; Durlak et al., 2010; Eccles and Templeton, 2002). They are often implemented in vulnerable schools where children have a high risk of being engaged in or exposed–as victims–to criminal activities. Despite the increase in the number of programs implemented over the past years,[2] and the high incidence and economic costs of violence in the developing world,[3] the overall available non-experimental evidence of ASP's impact on social skills, crime, and violence is mixed and inconclusive (Taheri and Welsh, 2016).[4] Furthermore, experimental papers on these programs are still scarce, and all of them use data from developed countries (Goldschmidt et al., 2007; Hirsch et al., 2011; Biggart et al., 2014).[5]

Additionally, there is no evidence of how peer effects may function within an ASP setting or if such effects can help to understand the effectiveness of these programs. Many papers have explored the effects of diversity and their mechanisms but in different contexts. For example, some studies find that mixed groups are preferable when peer interactions can generate differences in the learning experience (Lafortune et al., 2016), or when the exposure to good peers improves the results of more disadvantaged individuals (Lavy et al., 2012; Rao, 2015; Griffith and Rask, 2014; Oreopoulos et al., 2017). Additional studies found that the exposure of high violent individuals to peers with different

[1]Recent papers show that this exposure can occur in all domains such as at children's households (Baker and Hoekstra, 2010), through their interaction with other peers at schools (Sousa et al., 2011; Herrenkohl et al., 2008) or in their neighborhoods (Damm and Dustmann, 2014; Chetty et al., 2016).

[2]There has also been a corresponding growth in funding for these programs. For the 2017 fiscal year, the US Congress appropriated approximately US$1.2 billion to be used for this purpose: 2% of the total Department of Education budget (U.S. Department of Education, 2017).

[3]For example, 43% of the total worldwide homicides occur among youth between 10-29 years old, and nearly all of these deaths occur in low- and middle-income countries (WHO, 2016).

[4]This article reports on the results of a systematic review and meta-analysis of the effects of ASP on delinquency. They find mixed results from 17 well-known evaluations. Additional evidence are the papers of Bellei (2009) and Berthelon et al. (2015) for Chile and Filmer and Schady (2008) for Cambodia. However, these studies are not impact assessments of ASP, but rather of other interventions oriented at supervising children.

[5]Although there is evidence of interventions that end up reducing violence and crime in developing countries, they differ from ASP. For instance, Chioda et al. (2016) find evidence of a reduction in crime due to the expansion of *Bolsa Família*, a conditional cash transfers program in Brasil. Additional evidence is from interventions in India (Banerjee et al., 2007) and in Cambodia (Filmer and Schady, 2008).

violence levels could reduce the probability of "criminal network formation" (Billings et al., 2016; Di Tella and Schargrodsky, 2013; Bayer et al., 2009). However, another strand of the literature finds that tracking individuals with similar peers can generate better results, since that segregation allows teachers to match instruction to a particular group's needs (Duflo et al., 2011), or because individuals prefer to interact with peers with whom they share particular characteristics (Carrell et al., 2013; Girard et al., 2015; Goethals, 2001).[6]

By creating an exogenous experimental variation in the propensity for violence of students' peers, this paper aims to provide experimental evidence on potential peer effects that can help study the effectiveness of the intervention.[7] The empirical design, inspired by Duflo et al. (2011) and Lafortune et al. (2016), overcomes the issues in the identification of peer effects pointed out by Angrist (2014). I find that mixing students with different levels of violence to participate in this "like-CBT" intervention is a better implementation alternative for the ASP than segregating them in more and less violent groups.

The ASP I study in this paper consists of clubs implemented after school within school facilities during the 2016 academic year, from April to mid-October. Students participated in two sessions per week, which lasted 1.5 hours each. Every session was a combination of: (i) a discussion framed in a CBT approach, which was oriented towards fostering children's conflict management, violence awareness, and social skills; and (ii) the implementation of clubs' curricula, which included activities such as scientific experiments, artistic performances, and others. The intervention was implemented by volunteers of Glasswing International, a local NGO working in Central America and Mexico. The study sample includes 1056 *enrolled* students between 10-16 years old. This age range is relevant in the Salvadorean context because that is when children and adolescents are likely to be recruited by gangs.

To measure group composition effects and to exploit that there was more demand for the program than spaces, I randomly assigned these students to a group with a heterogeneous or homogeneous combination of peers, according to their initial propensity for violence.[8] Then students in the homogeneous treatment were separated into two subgroups considering their percentile in the distribution of violence, i.e., students whose predicted violence was higher (lower) than the median were assigned to a club with peers with high (low) predicted propensity for violence. Randomization

[6]This preference for interacting with individuals of the same gender or race been extensively studied in the role model literature. Overall, this evidence has consistently shown that being assigned to mentors or supervisors of the same gender (Athey et al., 2000; Bettinger and Long, 2005; Hoffmann and Oreopoulos, 2009; Paredes, 2014) or race (Dee, 2004; Egalite et al., 2015) improves students' or workers' performance.

[7]This study was registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-"AEARCTR-0001602."

[8]This variable is a proxy of a student's vulnerability of engaging in violent acts, which was predicted using violence determinants and following the estimation strategy described by Chandler et al. (2011).

was done such that group size and club categories were balanced across both treatments.

Before the intervention, I collected self-reported data on personal and family characteristics from enrolled students. Follow-up self-reported data included questions to measure the intervention's impact on attitudes, violence and crime; exposure to risky spaces; and educational or personal expectations of enrolled children. I combined this self-reported information with neurophysiological evidence from a random subsample of the enrolled students, particularly measures of stress and emotional resilience. I used low-cost portable electroencephalograms within an in-field lab setting. Finally, I also collected administrative records on math, reading, and science grades; behavioral reports; and absenteeism data from enrolled and non-enrolled students. This data was provided by schools before and after the intervention.

Estimations indicate that, on average, the improvements in attitudes and misbehavior at school are larger when participants are in more diverse groups than in segregated ones, for both high- and low-violence children. These results are consistent with the evidence that interactions with diverse peers can generate differences in the learning experience (Lafortune et al., 2016).[9] In this sense, students in heterogeneous groups have the opportunity for exposure to both good behaviors they should follow and negative ones they should not engage in. These interactions are only weakly available for students in the homogeneous group. In the neurophysiological outcomes, I find that students treated with similar peers in terms of violence—i.e. treated in homogeneous groups—their stress levels are greater than those of students treated in a heterogeneous composition of peers. Particularly, the increase in stress is greater for children treated in HM-High groups versus the respectively comparable children treated in heterogeneous groups.

Finally, I study tracking effects on marginal students. They are defined as children located just above or below the median of the propensity distribution function within each stratum. Some of them were assigned either to high or less homogeneously violent groups. Exploiting the discontinuity around the median and using only the sample of children assigned to the homogeneous treatment, I find evidence that the marginal students are negatively affected by being assigned to the most violent group in both academic outcomes and misbehavior at school. This result contributes to the existing evidence related to how segregation by initial violence may encourage the formation of networks of violence (Billings et al., 2016; Di Tella and Schargrodsky, 2013; Bayer et al., 2009), affecting those individuals who were supposed to be the key beneficiaries from these types of intervention.

Summing up, these two pieces of evidence on peer effects indicate that having some highly violent peers can constitute a learning alternative for low violence children because they can see the type of behaviors that they should not follow. However, the jump around the median in the

[9] Alternatively, these results support the rainbow model of peer effects, whereby all individuals benefit from being exposed to a more heterogeneous set of peers (Hoxby, 2000).

tracking group also indicates that when relatively low violence children are exposed to a more significant share of bad-to-good peers, the effects are the opposite. This implies that there must be an optimal bad-to-good peers combination in the implementation of the program that allows for the maximization of the overall impact. .

This paper allows me to contribute causal evidence to the discussion of tracking versus integration as optimal strategies to allocate participants to an intervention. The greater effects on academic and non-cognitive outcomes under integration versus tracking that I present in this paper are consistent with a body of micro-level evidence, which explain that these effects are likely caused by exploiting the interaction between diverse individuals within groups.[10] My results are mainly similar to those from Rao (2015), who finds an improvement in some social preferences outcomes, such as generosity, prosocial behavior, and equity, when there is an exogenous change in wealth heterogeneity in India. The novelty of my paper is that I modify the composition regarding violence and also include analysis of peer effects on additional non-cognitive outcomes that are important in developing countries such as violence, misbehavior, and attitudes towards school and learning.

There is also a growing body of evidence that finds benefits from tracking. Theoretically, Lazear (2001) shows that – in the presence of different levels of classroom disruption – segregation by type maximizes the 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).[11] In contrast to those papers, my results indicate that the training can have unintended effects on academic and non-cognitive outcomes when it is targeted at only the most violent students.

A plausible explanation for the differences between my results and those reported in the tracking literature is the lack of specific incentives for instructors to adapt clubs' curricula to their groups' needs. In fact, my results fits into the predictions of Duflo et al. (2011)'s model under the special case in which instructors do not respond to group composition because the teacher's effort function is a constant or when the cost of effort is zero below certain target level to which teachers orient instruction. Under this assumption, tracking by violence worsens the outcomes for those above the

[10] See Sacerdote et al. (2011) for a summary of the recent literature on peer effects on student outcomes in educational settings. Specifically recent papers on random assignment of freshmen or students (Thiemann, 2013); on elite exam schools (Abdulkadiroğlu et al. (2014) and Dobbie and Fryer Jr (2014) in the United States, and Lucas and Mbiti (2014) in Kenya); and programs for gifted individuals (Bui et al., 2014) find surprisingly positive impacts of being exposed to a very different set of peers. Additional results are presented by Hoxby (2000); Zimmerman (2003); Angrist and Lang (2004); Rao (2015); Griffith and Rask (2014); Lafortune et al. (2016); Chetty et al. (2016); Oreopoulos et al. (2017)

[11] Duflo et al. (2011) find that tracking benefits both lower- and higher-ability students in Kenya. Cortes and Goodman (2014) analyze the "double-dose" algebra policy in Chicago public schools, which sorted students into algebra classes by their math ability. They find that this policy improved short- and long-term academic performance. Girard et al. (2015) study students' social networks formation and find evidence of preferences for homophila along several dimensions.

median of the original distribution of violence and increases the performance for those below the median.

The remainder of the paper is organized as follows: Section 2 describes the intervention, data collection, and study design. Specifically, this section presents details of the propensity for violence (IVV) estimation, descriptive statistics, and results of experimental design checks. Section 3 summarizes the specifications used to estimate the effects of the intervention on academic behavior, violence outcomes, and peer effects in this context. These results are presented in Section 4. Section 5 discusses the results and provides evidence of the most plausible mechanisms, and finally, the preliminary conclusions are presented in Section 6. All appendix figures and tables are at the end of this paper.

2. The Tracking-by-Violence Experiment: Intervention, Design, and Data

2.1 Intervention

After-School Clubs 2.0

This study was conducted within the context of a based-CBT after school program in El Salvador. The ASP was implemented by the NGO Glasswing International. I partnered with them to design and implement a program's impact evaluation, estimation of spillover effects, and assess how group composition in violence work within the program's set-up.

In Dinarte and Egana (2017), we present a detailed description of the intervention and estimations of its direct and spillover effects. We describe ASP curricula, instructors characteristics, beneficiary schools features, and sample distribution. I highlight here only curricula structure and enrollment process, which will help to a better understand of the peer effects I find in this study.

The ASP is implemented as part of the NGO's program *Community Schools*. According to the intervention approach, its main objective is to successfully modify children's violence and attitudes through the learning of life skills, and therefore improve their academic performance (Glasswing International, 2012). The NGO offers four categories of clubs—Leadership, Art and Culture, Sports, and Science—in the ASP by education level or (*ciclos*). Considering this intervention structure, I design the experiment by using the natural school-education level organization as the stratification variable.

Clubs meet twice a week for approximately 1.5 hours each and take place just after school ends.[12] Each session is divided into two sections: social skills development and club's curriculum. The first section is common to all participants and tries to make students aware of some behaviors, to disrupt these patterns and to promote better ones using experiential learning or role-playing. It includes topics such as conflict- and risk-management, school violence reduction, and soft skills.

The second part of the session includes implementation of ludic activities related to each club category. Its objective is to motivate students to participate in the intervention and increase program attendance.

During 2016, the NGO offered and implemented the program in 5 public schools in El Salvador. Out of a total of 2,420 children from these schools, I recruit and enroll 1056 students between 10-16 years of age, which were interested in participating at the program and study. Any child is allow to self-enroll, the only requirement was to bring a signed parent's authorization and fill in an enrollment form during registration stage. The form collects their personal and family information and their application to participate in a club. Then, they were assigned to a group considering their preferences, parent's authorization and the aggregated demand for the club category.

The timeline of the study is shown in Figure 1.

[**Insert Figure 1 here**]

2.2 Experimental Design

This study aims to provide experimental evidence on how exposure to similar or different peers in terms of violence can modify the effectiveness of an after-school program. Therefore, the experimental design has different components: violence measure estimation per participant, random allocation of enrolled students to different treatment arms, collection of relevant data at different stages, and experimental design robustness checks. This last component includes, in addition to balance on observables across treatments before the program, other distributional criteria. In this subsection, I describe all of them.

A. Propensity for Violence Index (IVV) estimation

To assign enrolled students to each treatment group, the first requirement was to measure their propensity for violence. At the registration phase, it was not possible to directly ask about this because we could not guarantee that this personal information would be kept confidential during

[12] According to Seppanen et al. (1993), the minimal length of implementation of ASP sessions, to be cost-effective and generate impacts on violence and crime, should be between 2 to 8 hours per week.

the study.[13] Additionally, asking specific question about being an active gang member or being related to these organizations, which is highly correlated with crime and violence in El Salvador, may endanger both children and instructors.

Instead, following Chandler et al. (2011), I estimated a predictive model of violence and crime from existing data using a Two Sample Least Square strategy. First, using an existing anonymized database of youths' violence and crime from El Salvador (FUSADES, 2015),[14] I estimated the likelihood of having committed a violent act V_f as a function of a wide range of covariates:

$$V_f = \alpha_0 + \alpha_1 D_f + \epsilon_f$$

where D_f is a vector of violence determinants of student f in the FUSADES dataset.[15] This vector includes variables that indicate individuals' vulnerability to violence, such as students' characteristics (e.g. age, gender, time spent alone at home, and education level); children's household variables (e.g. residence area, mother's education, household composition); and school-level controls (e.g. school location, and commuting time to school).[16] Descriptive statistics and comparison of means (p-values) between FUSADES sample and the one from this study can be found in table A1 in the appendix section. Estimations indicate that both samples are similar in most of the determinants, except for some variables such as student's age and their report of being without adult supervision after school time.

All estimated coefficients $\hat{\alpha}_1$ have the expected sign according to the literature of violence determinants. For instance, boys are more likely to be violent than girls, adolescents are worst behaved than children (Rodrï¿œguez-Planas, 2012), and lack of parental supervision increases the probability of committing a violent act (Gottfredson et al., 2004). Statistically significant determinants are participant's age, gender, living in urban area, lack of parental supervision, and commuting time. Among all, lack of parental supervision is the most important determinant of propensity for violence in this sample. Table A2 in the appendix section summarizes the results of estimated coefficients.

Then, exploiting the availability of these variables in the registration forms of enrolled students,

[13] For example, either the local authorities or gangs organizations may force the research team or the NGO to hand them the information that completely identified each child, putting in risk not only the intervention but most importantly children's security.

[14] This database was created using the *El Salvador Youth Survey*'s instrument. It consists of a sample of 8640 students in sixth and ninth grade, enrolled in public schools in El Salvador.

[15] This database includes a great number of variables measuring crime and violence and their determinants.

[16] Some relevant papers that find evidence that these variables are determinants of crime and violence are: for gender, Bertrand and Pan (2013); Rodrï¿œguez-Planas (2012); for age, Rodrï¿œguez-Planas (2012); for area of residence, Springer et al. (2006); for maternal education, Springer et al. (2006) and Gaviria and Raphael (2001); for time spent at home, Gottfredson et al. (2004) and Aizer (2004); for commuting time to school, Springer et al. (2006); Damm and Dustmann (2014); and for household composition, Gaviria and Raphael (2001).

I predicted the measure of propensity for violence (IVV) for each child, using the vector of estimated coefficients $\hat{\alpha}_1$. There are two features of this IVV that are important to emphasize. First, since the variables included in the estimation are related to students' violence exposure at different domains—family, school and community—this measure is a more accurate proxy of students' overall propensity for violence than the reports of students' misbehavior from schools records. Second, this predicted index can be interpreted as a measure of student's *propensity* for violence, and not as an indicator of *effective* violence.

Despite the IVV is not a perfect measure of violence, I provide some evidence to argue that it is clearly the best proxy of propensity for violence under this particular context. First, according to the existing literature of violence and crime determinants for particular groups (Klassen and O'connor, 1988; Chandler et al., 2011),[17] this sort of crime and violence models estimated from existing data have a high predictive power.[18] Indeed, using misbehavior reports as the classification variable for high and low propensity for violence, estimations indicate that I would have had a similar classification in 53% of the total sample. Most importantly, there are no differences in the classification among treatments, as we can see in the last row in the appendix table A3.

An additional concern is that this index may be explaining another factor like school performance. Thus, I estimated the correlation between the predicted index and grades reported by teachers and found that it is not statistically significant. Yet, I find that the correlation between the predicted IVV and misbehavior at school is positive and statistically significant at 1%. In appendix table A4 I present these estimations, using different standardizations of academic grades and behavior reports.

Finally, the IVV predicts both intensive and extensive margins of future misbehavior. Using data from students in the control group, I find that the correlation between IVV and their bad behavior at the end of the academic year is positive and statistically significant at 5%. Estimation strategy and main results are presented in table A5 in the Appendix section.

B. Treatments

After estimating the IVV, enrolled children were randomly assigned to three groups—control (C, 25%), heterogeneous (HT, 25%), and homogeneous (HM, 50%)—within each school-by-educational level "blocks", as shown in Figure 2. Each education level consists of three years of schooling: the first is from 1st to 3rd grades, the second from 4th to 6th grades, and the third from 7th to 9th grades.

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

[18] Klassen and O'connor (1988) uses a sample of adult males at risk for violent behavior admitted as inpatients at a community mental health center. He finds that this model correctly classified 85% of the total sample.

Then, students in HM were ranked and assigned to subgroups according to their index: 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.

[**Insert Figure 2 here**]

Treatments are described below:

1. *Heterogeneous (HT):* Registered and randomly selected students are assigned to take part in a club with a heterogeneous composition of clubmates according to their IVV.

2. *Homogenous-Low (HM-Low):* Registered and randomly selected students are assigned to participate in a club with low violence peers if their IVV is lower than the median of the HM group within their respective strata.

3. *Homogenous-High (HM-High):* Registered and randomly selected students are assigned to participate in a club with high violent peers if their IVV is greater than the median the HM group within their respective strata.

4. *Control:* This group of students were not selected to participate in the clubs during the 2016 academic year.[19]

As opposed to Duflo et al. (2011) and similar to Lafortune et al. (2016), neither instructors nor participants knew details of the assignment because I wanted to capture mostly the effects of interactions between participants instead other channels, such as of teaching or curriculum adaptation. To test for changes on teaching methodologies, we collected information from a trainers' survey and present the results in upcoming sections.

2.3 Data

A. Baseline

As presented in the intervention timeline, after the NGO advertised the ASP in schools facilities, a research team returned to schools to register participants. We requested a consent form signed by a parent or tutor and asked students to fill-out a enrollment form. This instrument collected

[19]Children randomly assigned to the control group were supposed to left schools facilities after their school time. We were able to collect their information at follow up because we gave them a "participation coupon" that they could redeem next year, guaranteeing their participation in the ASP during 2017.

personal and familiar information—such as age, gender, mother's education, average commuting time, among others—which were used to estimate the IVV.

Before the intervention, I also collected schools' records of math, reading, and science grades; behavior reports,[20] and absenteeism data from all children. More details on these school reports are presented in Appendix 1.

B. Follow-Up

According to the ASP's theory of change, the intervention can directly affect behavioral and neurophysiological outcomes, such as children's violence, misbehavior at school, and emotional regulation. It may also have some indirect effects on academic performance, since existing evidence indicate that changes in non-cognitive outcomes can affect cognitive skills (Cunha and Heckman, 2008). Therefore, I collected data of these three categories of outcomes. Appendix 1 presents a detailed description of all the outcome variables used in this paper.

As we explain in Dinarte and Egana (2017), follow-up data on non-cognitive outcomes were collected from enrolled participants in school facilities at the end of October 2016, after all clubs have completely implemented their curricula. The follow-up survey included questions to measure the intervention's impact on general topics, such as students' attitudes, delinquency and violence, exposure to risky spaces, and educational, migration or personal expectations. Students filled out the questionnaire in classrooms especially set up for this purpose. Each survey took approximately 45-60 minutes. Most surveys were self-administered, with assistance from staff trained in the survey methodology.

Finally, I also collected neurophysiological evidence from a random subsample of the enrolled students, particularly measures of stress and emotional resilience. I used low-cost portable electroencephalograms within an in-field lab-setting. Data collection process for this third group of outcomes is discussed in Dinarte and Egana (2017).

However, since I do not necessarily trust self-reports, I attempted to recheck and validate these behaviors using proxies for these outcomes obtained from administrative data. In November 2016, at the end of the academic year, schools provided again math, science, and reading grades, behavior reports, and school absenteeism and drop out data, from both enrolled and non-enrolled students.

To increase statistical power to detect effects for outcomes within a family, we use indexes of variables that are expected to move in a similar direction and also to reduce the number of

[20]In El Salvador, behavior reports are reported by teachers each quarter. They are presented in the following discrete scale: Excellent (E), Very Good (MB), Good (B) and Regular (R). It can be translated in a continuous scale that is comparable to courses grades. In this paper, I used a reversed continuous scale to facilitate the interpretation and comparability to the self-reported measures of violence and crime.

hypothesis tests (Haushofer and Fehr, 2014; Heller et al., 2017).

C. Matching administrative data and attrition rates

As shown in table A6 in the Appendix section, average matching rate of administrative data of enrolled children was 94% at baseline, and 97% at follow up. All matching rates were balanced between treatments and C groups, except for the fraction of math grades at baseline between HM and C group, significant at 10%; and in absenteeism between both tracking groups, also significant at 10%. To account for this difference, I include in all specifications for the academic outcomes, the imputed grade for missing observations at baseline and a missing value indicator. Additionally, average matching rate of administrative data of non-enrolled students was 85% at baseline and 98% at follow up.

The share of initially enrolled students that filled out the follow-up survey after the intervention was 92% on average, and for the HM and HT groups, it was 91% and 94% respectively. There were no statistical differences between treatments and control groups in overall attrition rates. Therefore, results are not driven by the absence of follow-up survey data for any group.

For the neurophysiological measures, after filtering the EEG recording data, the average attrition share of 49%. In Dinarte and Egana (2017), we present several checks to verify that this attrition rate was not correlated with the intervention. Our argument is that attrition was caused mainly from the quality of the data recordings. For example, long, dense or unclean hair and/or freezing computers were the most common troubles for the Matlab toolbox to read de EEG recordings.

2.4 Summary Statistics

Descriptive statistics of the full sample and each treatment and control groups are shown in Table 1. Column 1 exhibits statistics for the total sample and columns 2-5 are for control (C), any treatment (T), and each treatment (HT and HM) groups respectively. Columns 6-7 show statistics for the two homogeneous subgroups.

Panel A presents the summary statistics of the violence determinants. Participants are on average 12 years old, 49% are male, and 73% live in an urban area. Regarding family composition, 91% of the students live with at least one parent, and 9% live with a relative or a non-related adult. On average, 62% of students' mothers have an intermediate education level (between 7-12 years), and 31% have less than six years of schooling. Regarding risk exposure, only 5% of students reported being alone at home when they are not at school. However, on average they have to travel around 18 minutes to school. Additionally, 30% of students are enrolled in the afternoon shift, increasing the probability of being without adult surveillance while their parents are at work.

Finally, the last row of Panel A shows that the average propensity for violence for any treatment and C groups is 0.038, with a standard deviation of 0.029, ranging from 0.001 to 0.215. This average propensity for violence is 14 times the mean probability that a given student will be vulnerable to violence in Chicago (Chandler et al., 2011). Even when both estimations are not completely comparable, because I use fewer violence determinants than Chandler et al. (2011), this difference sheds light on the tremendous propensity for violence of the children from this study. More descriptive statistics of the predicted propensity for violence are presented in appendix table A7.

Panel B shows academic scores and absenteeism for first quarter of the 2016 school year. In a grade scale of 0-10, requiring a minimum grade of 5 to pass each course, enrolled students have between 6.5 and 6.7 points, similar to the average grades at national level. The mean absenteeism rate in the first quarter, before the intervention, was 5.4% (2.16 out 40 days).

Finally, Panel C summarizes the clubs' characteristics: mean club size was 13 students, and community tutors ran approximately 31% of these clubs. The average take up, defined as the share of sessions attended by each student out from the total number, was 57%. Moreover, the share of enrolled students on each club category is statistically similar between treatments, except between HM-H and HM-L groups as may be expected. Finally, the mean fraction of treated students by course was 42%, statistically similar between treatments.

[**Insert Table 1 here**]

2.5 Experimental design checks

This experimental design has to meet five requirements to generate an exogenous variation that allows me to identify the causal impact of the intervention and group composition effects. First, treatments and control groups must be balanced. I find that differences between treatments and C groups are not statistically significant, except for the share of mothers with basic education and reading grades (HT vs. C), a category of household composition and reading grades (HM vs. C), and the predicted IVV (HT vs. HM, greater for the HT group). After adjusting p-values for multiple hypothesis testing of means and FWER following Jones et al. (2018), those differences are not statistically different from. These adjusted p-values are presented in appendix table A8.

Yet, to account for the difference in propensity for violence, I control by the percentile of predicted IVV in all estimations. Additionally, in academic outcomes specifications, I include the respective grades at baseline to account for the differences in academic performance before the intervention.

A second condition is that the HM-High group's IVV should be greater than that of the HM-Low group, also expressed in most of its determinants. As we can see in columns (6) and (7) in Table 1, the HM-High group has a larger proportion of male and older students than the HM-Low group. They are also more exposed to violence because face greater travel time from home to school, most of them spend time home alone, and enrolled in evening shifts.[21]

As the assignment to HM and HT was defined over the predicted violence index, the third requirement is that HT group must be more violence-diverse than any of the HM groups. Additionally, the average violence level of HT must be between the HM-Low and HM-High levels. This design fulfills these conditions, as we can see from the results in the previously presented table A7 in the Appendix. Standard deviation of the HT group's IVV is greater than those of the HM subgroups and average HT group's IVV is between those of the HM-High and HM-Low.

The fourth requirement is related to three desired characteristics of the IVV distribution functions of HT, HM, and C groups, before treatment. The first one is that these distributions must be similar at the baseline. Using the two-sample Kolmogorov-Smirnov test for equality of distribution functions, hypotheses are not rejected—p-values of 0.62, 0.89 and 0.68 for HT-HM, HT-C comparison, and HM-C comparisons, respectively. The similarity among distributions can be verified also in Figure 3. The second characteristic is that the distributions of HT, HM-High, and HM-Low groups must differ. As Figure 4 illustrates, there are differences among the three groups' distributions. Particularly, using the two-sample Kolmogorov-Smirnov test, I reject the hypothesis of equality of each comparison of distribution functions pairs at 1%.

[**Insert Figures 3 and 4 here**]

The last desired feature is that the distributions of HM-High and HM-Low groups should not fully overlap in the full sample, in order to have some variability between both HM subgroups. If I had not stratified, there would not be any overlap between both groups. However, as the assignment was defined within each stratum, there is overlap in 67% of the sample, as shown in Figure 5. Therefore, there is still important variation between IVV distribution functions of the HM subgroups at baseline that I can exploit.

[**Insert Figure 5 here**]

Finally, the fifth condition is that the 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

[21] Most students in the HM-Low group have mothers with either basic or higher education. These results could be explained as follows: if their mother has basic education, it is possible that she will stay at home with her children as her potential income is low. Alternatively, if the mother has higher education, then she will probably have more financial means to pay for some sort of childcare or other presence in the home.

stratum. Figure 6 shows the predicted IVV median of student's club mates as a function of her 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 group are enrolled with peers with a mean IVV 0.8 points greater, statistically significant at 5%.[22]

[**Insert Figure 6 here**]

3. Estimating Peer Effects

In this section, I describe the empirical strategy used to study group composition effects and how this heterogeneity interacts with children' initial propensity for violence. First, I describe the specifications used to measure average effects of being treated in a particular composition of peers, exploiting the random variation generated directly from the experiment design. Second, using the discontinuity in the median of the IVV distribution function of the HM group, I evaluate the effect of tracking on the marginal participant.

3.1 Group composition average effect

Restricting the sample to treated students and using the experimental variation of this study design, I can directly test for differences in the ITT effects on the outcomes of students assigned to groups with either homogeneously or heterogeneously violent peers, using the following specification:[23]

$$y_{ij} = \theta_0 + \theta_1 Hom_{ij} + \theta_2 X_{ij} + S_j + \epsilon_{ij} \tag{1}$$

where y_{ij} is the post-intervention behavioral, neurophysiological and academic outcome for student i at school and education level j. Hom_{ij} is a dummy that indicates whether student i in school level j is assigned to the HM treatment. X_{ij} is a vector of control variables measured at or before baseline, including a second order polynomial of student's IVV percentile. For the academic outcomes regressions, we also included standardized grades at baseline (including imputed values) and a missing baseline grades indicator as controls. Finally, we also control for "randomization blocks" with school-by-education level fixed effects S_j.

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

[23] Estimations from comparing each treatment arm and the C group are presented Appendix section tables.

In this setting, θ_1 can be interpreted as the effect on student i of receiving an offer to participate in the ASP with a homogeneous composition of violent peers, compared to effects of the same offer but with more diversely violent peers.

Due to the possible bias in the estimation of the IVV, standard errors are adjusted using a cluster bootstrapped at the course-school level (Treiman, 2009). As a sensitivity analysis we also calculate p-values that account for multiple hypothesis testing.

By design, the HM group is constituted by two different subgroups (HM-High and HM-Low). In this sense, it is also interesting to explore if a particular HM subgroup is driving the results, comparing each of them with the HT group. Since the assignment variable to those subgroups was the median of the IVV distribution at each HM-stratum level, after controlling by the indicator IVV_high_{ij} and by the IVV median at the j level, $I\bar{V}V_j$, I can compare directly the results of each HM subgroup with the respective HT treatment, estimating the following specification:

$$Y_{ij} = \theta_0 + \theta_1 HomH_{ij} + \theta_2 HomL_{ij} + \theta_3 IVV\_high_{ij} + I\bar{V}V_j + \theta_4 X_{ij} + \epsilon_{ij} \tag{2}$$

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

Specification (2) allows to compare both treatments within each half of the IVV distribution, which is equivalent to include in specification (1) an interaction between Hom_{ij} and IVV_high_{ij}. In the upper half, θ_1 is an ITT estimator of assigning a child i with higher propensity for violence to a low violence-diverse group of peers, compared to allocating her to a high violence-diverse group of peers. Also, for the lower half of the IVV distribution, θ_2 is an ITT estimator of assigning a less violent children to a low violence-diverse group of peers compared to a heterogeneously violent group.

I also study nonlinear heterogeneous effects of group composition at a finer level, interacting HM and HT treatments with quartiles of the IVV distribution. Details and results of the estimation are described in Appendix 2. Finally, following Duflo et al. (2011), I present an analysis of the average group composition effects using linear-in-means and variance specifications restricting the analysis to the HT subsample. These equations and their identification assumptions are described in Appendix 3.

3.2 Effects of tracking on the marginal student

Results of equations (1) and (2) allow identification of the average effects of being treated in a particular group composition. Moreover, with this experimental design I can explore the effect of peer violence exposure on the around-the-median children in a tracking setting. I call them the *marginal participants*. This group includes a set of students just above or below the fifth percentile of the IVV distribution. Given that these just above-the-median children are similar regarding propensity for violence to those at- or below-the-median, I exploit their assignment to a group of high-IVV peers and compare with the other allocated to a low-IVV set of peers.

Studying effects on the marginal student is interesting because having high-violent peers on average also means that her is the least-violent child in her group before the intervention, and having less-violent peers implies that she is the most-violent child in her track. In this sense, the marginal participants are the most different children within their group and therefore, they may face the greater tracking impact.

To identify this impact, I use a regression discontinuity design with the median of the IVV distribution in each stratum as the discontinuity, and restrict the sample to students in the HM treatment. For the validity of this strategy, the assumption is that nothing else changed discontinuously around the point of separation between the two groups, which holds true in this design as discussed before. I estimate the following equation:

$$Y_{ij} = \lambda_0 + \lambda_1 HomH_{ij} + f(IVV_{ij}) + \lambda_2 S_j + \epsilon_{ij} \tag{3}$$

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

4. Results

In this section I present reduced form estimates of group composition and tracking effects on main outcomes of interest in the context of an ASP. I can draw two main conclusions from this section. First, mixing students by their initial propensity for violence generates better average effects than segregating them. Second, tracking has detrimental effects for marginal students with highest propensity for violence.

4.1 Group composition average effect

Table 2 shows estimations of group composition using specification (2). In Panel A we present the effects of being assigned to participate in the ASP with a homogeneous composition of peers (HM), compared to HT group behavioral outcomes. In Panel B, we present the main neurophysiological outcomes estimated through electroencephalograms recordings. Finally, Panel C we present the effects on children's academic performance.

First, from the comparison between HT and HM groups, I find that students assigned to homogeneous groups show a reduction by 0.16 standard deviations on average positive attitudes towards school, compared to students assigned to heterogeneous groups (column 1, Panel A, Table 2). They also increase their probability of having a bad behavior report at school by 5.5 percentage points (column 9, Panel A, Table 2).

As shown in Panel B, the only statistical difference between both HM and HT group compositions is in participants' stress. We are finding that when students are treated in homogeneous groups, their stress levels are greater than those of students treated in heterogeneous groups. Finally, I do not find statistical differences between both treatments in the rest of academic outcomes.

These results are consistent with the evidence that interactions with diverse peers can generate differences in the learning experience (Lafortune et al., 2016). Moreover, the rainbow peer effects model (Hoxby and Weingarth, 2005) can also explain these results. This model suggests that all students are best off when they deal with a diverse group of classmates. Additionally, these results are suggestive evidence that treating students in violence-diverse groups reduces the probability of creating networks of violent children (Billings et al., 2016).

[**Insert Table 2 here**]

Since two different subgroups regarding violence constitute the HM group, this design allows me to explore further differences in group composition comparing each HM subgroup with the HT group using specification (3). These results are reported in Table 3.

First, perhaps surprisingly, I find that HM-Low is driving the negative effect of group composition on attitudes towards school and learning. Compared with the HT group, students in the HM-Low face a reduction in their positive attitudes by 0.22 standard deviations (Panel A, column (1)) and report paying less attention in classes by 0.08 percentage points (Panel A, column (3)). This unexpected result is related to Hoxby and Weingarth (2005) invidious comparison peer effects model, that applied to this context implies that the exposure to only less violent—or well behave—students depresses the average performance of the group. An alternative explanation is that students in heterogeneous groups have the opportunity of being exposed to both good behaviors they should follow and to negative ones they should not engage in. These interactions are only weakly available for students in the homogeneous group. The second relevant result in this subsection is that the probability of having bad behavior reports is greater for high violence students when they are segregated by more than 0.09 percentage points, as shown in Panel A, column (9).

Moreover, as we can see in Panel B, the increase in stress is greater for children treated in HM-High groups when compared to respectively similar children treated in HT groups. This result explains one of Dinarte and Egana (2017)'s findings in the ASP IIT effects section. Their results indicate that, after the intervention, treated children with greater propensity for violence increase their stress levels compared to both children in the control group and to low-violence treated students.

As briefly explained before, these results may have different interpretations. First, exposure to risky environments usually increases individuals' stress level, either because they have to avoid danger or learn how to face it; defending themselves, for instance. Therefore, this can explain why children in the HM-High group are more stressed than those with a lower exposure to violence on average. Additionally, even when the coefficient of stress for HM-Low groups is not different from zero, its sign may indicate that they are also facing some level of stress, compared to low-violence children in HT groups. A plausible explanation to that result is that diversity is the social norm where these children usually perform. Thus, assigning them to similar peers may make them more stressed.

Thus, selecting and treating together only high violence students for these programs can generate an unintended effect from the intervention. This result also sheds light on that solely teaching socio-emotional skills may be not enough to reduce misbehavior or violence of highly violent students, but it seem to be also relevant that they also interact with—and probably learn good behaviors from—low violence students.

So far, results indicate that integration is better along the IVV distribution on behavioral outcomes and stress. Moreover, as shown in Panel C of Table 3, diversity regarding violence generates better results on academic outcomes for students with a high propensity for violence. The only instance where segregation seems to be better than integration is for students who are less susceptible to violence on academic outcomes. As I argue in the discussion, this last result can be driven mainly by the content of the clubs' curricula. According to the ASP structure, it may occur that more time was employed for the club's curricula in less violent HM groups, and therefore the reinforcement of "academic" content was greater here.

[**Insert Table 3 here**]

The pattern of results of heterogeneous effects of group composition at a finer level (quartiles) of student's initial propensity for violence suggests that students in both tails of the baseline IVV distribution (quartiles 1 and 4) are the most sensible to group composition, and therefore are driving the results on non-cognitive outcomes.[24]

Finally, since participants were randomly allocated to a group in the ASP, there is some variation in the group composition which stem from the fact that being assigned to HM vs HT directly affects the mean and variance of one's peers. Following Lafortune et al. (2016), the identification assumption is that after controlling for strata fixed effects, the variance and mean IVV of peer stems entirely from the random assignment. Details of the estimation and summary of results are presented in Appendix 3.

These results reinforce the previous findings using direct variation of the experiment. First, higher average clubmates' IVV negatively affects some attitudes towards school and learning and academic grades. Second, being exposed to a more violence diverse group of clubmates improves most academic outcomes, positive attitudes towards school and time employed to do homework.

4.2 Effects of tracking on the marginal student

An additional piece of evidence that can be obtained from this experiment is the effect of tracking for students in the middle of the distribution. To directly measure the effects of tracking, I can compare the two homogeneous subgroups using specification (3). This equation allows me to identify if there are differences of being assigned to a group of homogeneous peers with higher propensity towards violence.

The estimations of tracking effects on marginal students are summarized in Table 4. First, I control with a flexible second order polynomial of a student's percentile in the IVV distribution within the homogeneous group at each stratum. As shown in Panel A, I find that assigning a marginal student to a group of peers with higher propensity for violence increases her self-report of violent actions by 0.18 standard deviations. I do not find an effect on the rest of non-cognitive outcomes due to the increase in standard errors. However, despite this absence of statistical significance, the signs of coefficients of these self-reported measures of attitudes are negative and those of violence (self and teacher's reports) and absenteeism are positive, highlighting the unintended effects of the

[24] In Appendix 2, I present details of the specifications and results. Main estimations are summarized in table A11. Under integration, the reduction on misbehavior at school is greater for the most violent students (Q4) and the effects on positive attitudes towards school and learning are greater for the least violent students (Q1). Additionally, students in Q4 of the IVV distribution function are better off on academic outcomes when they are treated in violence-diverse groups. This last result is also confirmed using a more flexible estimation of differences in the group composition effect at different levels of the initial IVV distribution, as we can see in Appendix Figure A2. The differences are greater for students in the last tail of the IVV distribution (greater than 75th percentile).

intervention for the marginal participants.

On neurophysiological outcomes, I find that assigning a marginal student to a group of peers with higher propensity for violence reduces her tendency for further reflect on response by 0.77 standard deviations. This finding compares to peers with similar propensity for violence, but who were enrolled with lower-violence peers. We do not find an effect in the rest of outcomes. This result may be explained by the fact that these participants are exposed to the most violent peers of their violence distribution function. Even after the intervention, such exposure increase their misbehavior at school (**?**).

Effects of tracking on academic outcomes for marginal students are also negative. As we can see in Panel B, being assigned to a high violence group has a detrimental effect on both extensive and intensive margins on math grades (0.156 standard deviations and 0.074 percentage points, respectively) and increases the probability of failing any of the three courses by 0.048 points. As before, there is an increase in standard errors, and some coefficients are not statistically significant, but their signs suggest a negative effect.

Finally, following Duflo et al. (2011), I run specification (3) but restricting the sample to the eight students around the IVV median within each stratum. Results are also reported in the same table 4. Reducing the sample allows me to focus on the most similar students before the intervention. The downside is that it increases standard errors of the estimations, reducing statistical significance. However, the results support previous conclusions, showing that tracking generates unintended effects on marginal students, worsening their attitudes towards school and learning and increasing their bad behavior and violent actions.

In summary, the marginal student is negatively affected by being assigned to a more violent group. This is consistent with the existing evidence of endogenous formation of groups of badly behaved students when they are segregated. They seem to engage as a group member, following the group social norm of violence and negative attitudes, and indirectly impacting their academic performance.

[**Insert Table 4 here**]

5. Discussion

Despite the intensity and high costs of youth violence (WHO, 2015) and the recent increase in the number of programs oriented to reduce violence and misbehavior implemented in low- and middle-income countries, the overall evidence is still mixed and inconclusive.

This issue is of foremost relevance in many developing countries, wherein violent children are more likely to drop out of school to enroll in an outside option like the formal or informal job market, migration, or criminal organizations. This is certainly the case in El Salvador where, despite the implementation of some macro measures to reduce crime and violence nationally, there is no rigorous evidence of programs providing protection or surveillance to students who usually engage in criminal organizations such as gangs (MINED, 2015).

This mixed results can be explained not by the intervention's curricula or mentors, but by how ASP's group composition in terms of violence. The existing evidence on this matter is mixed[25] and mostly related to other contexts, such as educational settings (Duflo et al., 2011), female labor training (Lafortune et al., 2016) and first-year students at the United States Air Force Academy (Carrell et al., 2013).

To my knowledge, this paper provides the first experimental evaluation of the group composition effects of a like-CBT ASP implemented in a developing and highly violent country. My research experimentally manipulates the participation of 1056 students in an ASP implemented in five public schools in El Salvador. I additionally manipulated whether students participated in the program in homogeneous or heterogeneous groups according to their initial predicted propensity for violence. My analysis focuses on studying whether the participation with a particular composition of peers have some effects on academic, neurophysiological, and behavioral outcomes, changes students' efforts at school.

Better together. Group composition effects

Using the direct source of variation yielded by this experimental design, I find evidence that an average student is better off in a more diverse ASP group than in a segregated one. Specifically, mixing is better for non-cognitive outcomes regardless of the student's initial violence level. However, regarding academic grades, mixing is still better for the high-violence group, but segregation generates greater effects for the less violent children.

These results are consistent with a body of micro-level evidence, such as papers on random assignment of freshmen or students (Thiemann, 2013); on elite exam schools (Abdulkadiroğlu et al., 2014; Dobbie and Fryer Jr, 2014; Lucas and Mbiti, 2014) and programs for gifted individuals (Bui et al., 2014). Additional evidence on academic and labor contexts is presented by Hoxby (2000); Zimmerman (2003); Angrist and Lang (2004); Rao (2015); Griffith and Rask (2014); Lafortune et al.

[25] Some papers find that participating in groups with more similar peers generates greater effects due to homophile preferences or curriculum adaptation (Girard et al., 2015; Goethals, 2001; Duflo et al., 2011). However, most of the evidence finds that being involved in diverse groups generates greater impact due to positive peer effects (Zimmerman, 2003; Angrist and Lang, 2004; Lafortune et al., 2016; Griffith and Rask, 2014; Rao, 2015; Oreopoulos et al., 2017; Dobbie and Fryer Jr, 2014).

(2016); Chetty et al. (2016); Oreopoulos et al. (2017). Overall, these papers find positive impacts of being exposed to a very different set of peers. They argue that the integration effects occur due to the interaction between different individuals within groups, supporting the rainbow model of peer effects (Hoxby and Weingarth, 2005).

Particularly, as I briefly explained before, my results are mostly related to those from Rao (2015), who provides the first evidence of how changes on peers composition at school can shape a student's social preferences, through an improvement on her generosity, prosocial behavior and equity. My paper contributes to these results providing additional experimental evidence that is particularly relevant for the developing world. I test how the exposure to diversity regarding violence impacts positively additional non-cognitive outcomes, such as violence, approval of peers' antisocial behavior, misbehavior and attitudes towards school and learning. An additional outstanding characteristic in Rao (2015) is that he uses well constructed measures of social preferences. In my paper, I collected measures of non-cognitive outcomes from students' self-reports and administrative data provided by schools. These two sources of information allow me to contrast and validate the results.

Additional evidence that can be drawn from my experimental design are the tracking effects for marginal individuals. For example, an individual at the median in the violence distribution who is assigned to a high-violence group can be either contaminated by her peers and increase her violence level; or, according to the invidious comparison model, she can become less violent because she does not want to be like her fellow group members (Hoxby and Weingarth, 2005). Restricting the analysis to the homogeneous group, I find that students with the same level of violence at baseline seem to be "contaminated" by the predominant level of violence of the group to which they have been assigned.

In contrast to some theoretical and empirical pro-tracking papers (Lazear, 2001; Duflo et al., 2011; Cortes and Goodman, 2014; Girard et al., 2015), my results indicate that the training can have unintended effects on academic and non-cognitive outcomes when it is targeted at only the most violent students. This result reinforces the main conclusion of the paper of the benefits of diversity regarding violence, since it allows high violence students to be exposed to less violent children and learn social skills and good behaviors from them.

Why does integration generate better results?

In this subsection, I provide suggestive evidence to understand how these group composition impacts on average and marginal students may have operated. I start exploring peer effects in *social skills learning.* Students in heterogeneous groups are benefiting from being exposed to both "good behaviors" that they should follow and "misbehaviors" that they must avoid, as predicted by the rainbow peer effects model (Hoxby and Weingarth, 2005). However, students in a homogeneous group are losing the opportunity to learn from behaviors of the other tail of the violence distribution function.

A second channel that could explain the results is that *diversity is the social norm* in the scenarios -particularly at public schools- where students usually perform, making them feel more comfortable as it is the setting with which they are familiar. In this sense, one can assume that students in heterogeneous groups may have attended more sessions than those in homogeneous groups. I test for differences in attendance to the ASP between each HM group compared to the HT group and present the results in table A22 in the Appendix. Due to an increase in the standard errors, I find a small but not significant reduction on clubs attendance by both HM groups. Despite this lack of statistical significance, this result sheds light on preferences for diversity.

To provide further evidence to support the preference for diversity mechanism, I use data from spillovers and find different effects regarding proximity to misbehavior between non-enrolled and treated students. The results are higher for students whose bad behavior at school is in between 1 and two standard deviations from the average misconduct of treated students from her classroom. Notably, the effects of this intermediate proximity are more significant on bad behavior reports.

Further evidence to support the preference for diversity mechanism is the intensity of treatment by exposure. The assumption here is that if children have preferences for diversity, then the effects of the intervention should be lower when they are exposed to a higher share of clubmates who are also their classmates. I interact the treatment with the share of clubmates that are also classmates and could not find differential effects on non-cognitive outcomes. These results are presented in table A9 in the Appendix section.

The last mechanism that may drive the group composition results is that tracking can strengthen the possibility of *creating violence networks*, which has been previously analyzed in the literature (Billings et al., 2016; Bayer et al., 2009). Implementing interventions while keeping high or low violent students together can generate unintended effects on both groups, particularly for the most violent children. These results also match those of Pekkarinen et al. (2009), who find benefits of ending school tracking in Finland on the performance of students from lower ability backgrounds.

Explaining the puzzle from the less violent children's outcomes

It is puzzling that the effects on academic outcomes for low-violence students are greater under tracking even when mixing improves their attitudes towards school and learning. One explanation is that the time dedicated on each part of the session was conditional on the group composition. For instance, tutors in Low-HM clubs may have had to use less time on social skills training than on the particular club's curriculum, compared to the High-HM or HT groups. Thus, it may be expected that Low-HM clubs with academic curricula are driving the improved academic results compared to the HT clubs. I test this channel by including in the specification (3) an interaction between each HM treatment and a dummy for academic clubs on academic outcomes. I find that in the comparison of Low-HM and HT groups, the effects on academic outcomes are driven by students enrolled clubs focusing on academic topics. Results are shown in table A24 in the appendix.

6. Conclusions

This paper provides experimental evidence of group composition effects of an ASP on participants' academic outcomes, behavior, and violence level. The intervention was implemented in schools located in highly violent communities in a developing country, El Salvador.

First, by exploiting the direct variation from the experimental design, I find that - regarding academic outcomes - tracking benefits only low violence students and worsens these results for the high violence students when both are compared to the heterogeneous group. Additionally, concerning behavior and violence, tracking generates adverse effects for low violence students and increases the probability of bad behavior reports for ex-ante high violent students. These results are confirmed using the exogenous variation in the peer's composition. I find that there are positive academic and non-cognitive effects of being treated in more diverse groups concerning levels of violence than in less diverse ones. Additionally, for those students with an initial violence level around the median, being assigned to clubs with similarly high violent peers generates negative effects on both groups of outcomes.

These results have implications for public policy discussions on interventions oriented to improve academic outcomes and reduce violence within schools. First, participating in an ASP, where students learn about life skills and conflict management, has benefits both regarding academic and non-cognitive outcomes, mainly benefiting the most vulnerable students. Additionally, increasing adult supervision of students for some hours during the week reduces their exposure to risk and, particularly for boys at this age, may reduce their probability of being recruited by gangs (Cruz, 2007; Aguilar and Carranza, 2008; Aguilar, 2006). Furthermore, this paper provides a first step in understanding the relevance of group composition in an ASP, showing that within this context, peer effects are an important mechanism that can improve the relevant outcomes, motivating special attention to the implementation of these interventions in heterogeneous groups.

Since the intervention keeps students away from potential risk contexts for some hours and under supervision, and since during this time they also learn some life skills, the positive effects can be caused either because they are learning these skills in the program or because they are less involved with bad peers outside of school. I provide suggestive evidence that the life skills learning mechanism is driving the results. However, further rigorous research on these two channels is still necessary and would have significant implications for the design of this programs.

Another question for further research is if these results will persist over time. Due to this NGO's donors, a requirement for financing the impact evaluation was that students in the control group must be allowed to participate in the intervention the following year. This will make difficult to measure the ASP's long term effect.

Finally, in the literature of interventions aimed at reducing crime and violence, one important aspect of these programs is the developing of new and more healthy social ties, fostering a sense of belonging for participants that positive influences identity (Heller et al., 2017). In this aspect, there is still lack of evidence of how this intervention can be improved if students participate in the program within their closer network, exploiting their preferences for similar peers.

References

Abdulkadiroğlu, A., Angrist, J., and Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, 82(1):137–196.

Aguilar, J. (2006). Pandillas juveniles transnacionales en Centroamérica, México y Estados Unidos: Diagnóstico de El Salvador. *San Salvador: IUDOP*.

Aguilar, J. and Carranza, M. (2008). Las maras y pandillas como actores ilegales de la región. *Ponencia presentada en el marco del Informe Estado de la Región en Desarrollo Humano Sostenible, San Salvador*.

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

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

Angrist, J. D. and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco Program. *The American Economic Review*, 94(5):1613–1634.

Athey, S., Avery, C., and Zemsky, P. (2000). Mentoring and diversity. *American Economic Review*, 90(4):765–786.

Baker, A. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1):211–228.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122(3):1235–1264.

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

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review*, 28(5):629–640.

Berthelon, M., Kruger, D. I., and Oyarzun, M. A. (2015). The Effects of Longer School Days on Mothers' Labor Force Participation.

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.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review*, 95(2):152–157.

Biggart, A., Sloan, S., and O?Hare, L. (2014). A Longitudinal Follow Up Study of the Doodle Den After School Programme. Technical report, Childhood Development Initiative.

Billings, S. B., Deming, D. J., and Ross, S. L. (2016). Partners in crime: Schools, neighborhoods and the formation of criminal networks.

Blattman, C., Jamison, J. C., and Sheridan, M. (2015). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *National Bureau of Economic Research*.

Bui, S. A., Craig, S. G., and Imberman, S. A. (2014). Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy*, 6(3):30–62.

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. *The American Economic Review*, 101(3):288–292.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *The American Economic Review*, 106(4):855–902.

Chioda, L., De Mello, J. M., and Soares, R. R. (2016). Spillovers from conditional cash transfer programs: Bolsa Família and crime in urban Brazil. *Economics of Education Review*, 54:306–320.

Cook, P. J., Gottfredson, D. C., and Na, C. (2010). School crime control and prevention. *Crime and Justice*, 39(1):313–440.

Cook, P. J., Ludwig, J., and McCrary, J., editors (2011). *Controlling Crime*. University of Chicago Press.

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

Cruz, J. M. (2007). *Street Gangs in Central America*. UCA Editores.

CSJ (2014). Estadisticas de la Corte Suprema de Justicia de El Salvador.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and non-cognitive skill formation. *Journal of human resources*, 43(4):738–782.

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

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1):195–210.

Di Tella, R. and Schargrodsky, E. (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy*, 121(1):28–73.

Dinarte, L. and Egana, P. (2017). Reducing violence in the most violent contexts: Behavioral and neurophysiological evidence.

Dobbie, W. and Fryer Jr, R. G. (2014). The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools. *American Economic Journal: Applied Economics*, 6(3):58–75.

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

Durlak, J. A., Weissberg, R. P., and Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American journal of community psychology*, 45(3-4):294–309.

Eccles, J. S. and Templeton, J. (2002). Chapter 4: Extracurricular and other after-school activities for youth. *Review of research in education*, 26(1):113–180.

Egalite, A. J., Kisida, B., and Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, 45:44–52.

Filmer, D. and Schady, N. (2008). Getting girls into school: Evidence from a scholarship program in Cambodia. *Economic development and cultural change*, 56(3):581–617.

FUNDAUNGO (2013). Atlas de la violencia en el salvador (2009-2012). *San Salvador*.

FUSADES (2015). El Salvador Youth Survey Database.

Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behavior. *The 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.

Glasswing International (2012). Estudio Interno sobre impacto de programa extracurricular de seis centros escolares. Technical report, Fundacii¿œn Crisi¿œlida. Glasswing International.

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.

Goldschmidt, P., Huang, D., and Chinen, M. (2007). The long-term effects of after-school programming on educational adjustment and juvenile crime: A study of the LA's BEST After-School Program. *Los Angeles: UCLA/CRESST. Retrieved September*, 8:2008.

Gottfredson, D. C., Cross, A., and Soulé, D. A. (2007). Distinguishing characteristics of effective and ineffective after-school programs to prevent delinquency and victimization. *Criminology & Public Policy*, 6(2):289–318.

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.

Haushofer, J. and Fehr, E. (2014). On the psychology of poverty. *Science*, 344(6186):862–867.

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

Herrenkohl, T. I., Sousa, C., Tajima, E. A., Herrenkohl, R. C., and Moylan, C. A. (2008). Intersection of child abuse and children's exposure to domestic violence. *Trauma, Violence, & Abuse*, 9(2):84–99.

Hirsch, B. J., Hedges, L. V., Stawicki, J., and Mekinda, M. A. (2011). After-School Programs for high school students: An evaluation of After School Matters. *Evanston, IL: Northwestern University.*

Hoffmann, F. and Oreopoulos, P. (2009). A professor like me the influence of instructor gender on college achievement. *Journal of Human Resources*, 44(2):479–494.

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.

Jones, D., Molitor, D., and Reif, J. (2018). What do workplace wellness programs do? evidence from the illinois workplace wellness study. Technical report, National Bureau of Economic Research.

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.

Kremer, K. P., Maynard, B. R., Polanin, J. R., Vaughn, M. G., and Sarteschi, C. M. (2015). Effects of after-school programs with at-risk youth on attendance and externalizing behaviors: a systematic review and meta-analysis. *Journal of youth and adolescence*, 44(3):616–636.

Krug, E. G., Mercy, J. A., Dahlberg, L. L., and Zwi, A. B. (2002). The World Report on Violence and Health. *The Lancet*, 360(9339):1083–1088.

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. *The Economic Journal*, 122(559):208–237.

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

LPG (2016). Amenazas de pandillas es la quinta causa de deserciï¿œn escolar.

Lucas, A. M. and Mbiti, I. M. (2014). Effects of school quality on student achievement: Discontinuity evidence from Kenya. *American Economic Journal: Applied Economics*, 6(3):234–263.

Mahoney, J. L., Stattin, H., and Magnusson, D. (2001). Youth recreation centre participation and criminal offending: A 20-year longitudinal study of swedish boys. *International journal of behavioral development*, 25(6):509–520.

Marshall, N. L., Coll, C. G., Marx, F., McCartney, K., Keefe, N., and Ruh, J. (1997). After-school time and children's behavioral adjustment. *Merrill-Palmer Quarterly (1982-)*, pages 497–514.

MINED (2015). Estadisticas del Ministerio de Educaciï¿œn 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.

Paredes, V. (2014). A teacher like me or a student like me? role model versus teacher bias effect. *Economics of Education Review*, 39:38–49.

Pekkarinen, T., Uusitalo, R., and Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93(7):965–973.

PNC (2014). Estadisticas de la Policia Nacional Civil de El Salvador.

PNUD (2013). Informe Regional de Desarrollo Humano 2013-2014. Seguridad ciudadana con rostro humano: diagnóstico y propuestas para América Latina. Technical report, Programa de Naciones Unidas para el Desarrollo.

Rao, G. (2015). Familiarity does not breed contempt: Diversity, discrimination and generosity in Delhi schools. *Job Market Paper*.

Rodrï¿œguez-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.

Seppanen, P. S. et al. (1993). National Study of Before-and After-School Programs. Final Report. Technical report, Wellesley Coll., MA.; Mathematica Policy Research, Princeton, NJ.; RMC Research Corp., Portsmouth, NH.

Soares, R. R. and Naritomi, J. (2010). Understanding high crime rates in Latin America: The role of social and policy factors. In *The Economics of Crime: Lessons for and from Latin America*, pages 19–55. University of Chicago Press.

Sousa, C., Herrenkohl, T. I., Moylan, C. A., Tajima, E. A., Klika, J. B., Herrenkohl, R. C., and Russo, M. J. (2011). Longitudinal study on the effects of child abuse and children?s exposure to domestic violence, parent-child attachments, and antisocial behavior in adolescence. *Journal of interpersonal violence*, 26(1):111–136.

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.

Taheri, S. A. and Welsh, B. C. (2016). After-school programs for delinquency prevention: A systematic review and meta-analysis. *Youth violence and juvenile justice*, 14(3):272–290.

Thiemann, P. (2013). Social Planning with Spillovers: The Persistent Effects of Short-Term Peer Groups.

U.S. Department of Education (2017). U.S. Department of Education Budget 2017.

WHO (2011). World Health Organization Statistics.

WHO (2015). *Preventing youth violence: An overview of the evidence*. World Health Organization.

WHO (2016). *World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals*. World Health Organization.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and statistics*, 85(1):9–23.

TABLE 1. SUMMARY STATISTICS: MEANS OF VARIABLES BY TREATMENT GROUP PRIOR TO TREATMENT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Treatments		Tracking groups	
	Full Sample	Control Group (C)	Any Treatment (T)	Heterogen. group (HT)	Homogen. group (HM)	Homog. High (HM-H)	Homog. Low (HM-L)
PANEL A: IVV DETERMINANTS							
Student is male	0.49	0.51	0.49	0.48	0.49	0.76	0.22***
Student's age	11.94	11.86	11.96	12.04	11.93	12.41	11.4*
Student lives in urban area	0.73	0.72	0.74	0.73	0.74	0.78	0.70
Student's household composition							
Student living with both parents	0.53	0.49	0.55	0.53	0.56"	0.53	0.59
Student living only with one parent	0.32	0.37	0.31	0.34	0.30"	0.33	0.26
Student living with one parent and step-parent	0.06	0.07	0.06	0.06	0.07	0.06	0.07
Student living with other relative /adults	0.09	0.10	0.08	0.07	0.09	0.09	0.08***
Student's mother level of education:							
Basic education (1-6 years)	0.31	0.34	0.30	0.27 ˆ	0.31	0.22	0.40***
Intermediate education (7-12 years)	0.62	0.59	0.63	0.65	0.62	0.72	0.52***
University or higher (13 and +)	0.07	0.07	0.07	0.08	0.07	0.06	0.08
Student's travel time from house to school (min.)	17.64	16.98	17.85	17.84	17.86	19.58	16.13**
Student is alone at home after school	0.05	0.05	0.05	0.07	0.04	0.08	0.01***
Student's school year	5.75	5.69	5.77	5.81	5.76	6.02	5.49***
Student enrolled in the morning shift	0.704	0.704	0.704	0.69	0.71	0.69	0.74**
Student's violence index	0.04	0.04	0.04	0.04	0.04	0.05	0.02***
PANEL B: ACADEMIC OUTCOMES							
Academic scores Q1 2016 (Baseline)							
Reading scores	6.67	6.46	6.73	6.76ˆ	6.71""	6.54	6.88
Math scores	6.48	6.41	6.51	6.46	6.49	6.52	6.44
Science scores	6.62	6.46	6.67	6.62	6.54	6.63	6.55
Behaviour scores	7.18	7.15	7.16	7.21	7.16	7.28	7.12
Absenteeism Q1 2016	2.16	2.78	1.81	1.91	1.76	2.09	1.44
PANEL C: CLUBS' CARACTHERISTICS							
Average club size	-	-	13.4	13.43	13.38	13.13	13.63
Average club take up	-	-	0.57	0.57	0.57	0.56	0.59
Community tutors	-	-	0.31	0.29	0.32	0.35	0.29
Club category							
Leadership	-	-	0.29	0.14	0.16	0.18	0.13
Art and Culture	-	-	0.16	0.28	0.30	0.18	0.44***
Sports	-	-	0.26	0.25	0.27	0.32	0.21**
Science	-	-	0.29	0.33	0.27	0.32	0.22**
Share of treated by course	-	-	0.42	0.42	0.42	0.43	0.42
Retention rate (1 - attrition)	0.92	0.92	0.92	0.94	0.91	0.90	0.91

Table 1 shows descriptive statistics of the available variables at baseline for the full sample. Panel A summarizes information obtained from the enrollment form that was used as determinants in the IVV estimation. Panel B presents administrative data provided by schools only from students who had consented. This data is from the first quarter of academic year 2016 before the clubs were implemented. The scale of grades in El Salvador is 0-10 points. Panel C presents clubs characteristics. Take up is estimated as number of hours attended by student i / max hours attended by any student in each club. p-values are presented in Table A4 in the Appendix. indicates differences statistically significant at 10 % between HT and Control;" and "" indicate differences statistically significant at 5 and 10 % respectively between HM and Control; and *, **, *** indicate differences statistically significant at 1 %, 5 % and 10 % respectively between HM-Low and HM-High groups. In the comparison between HT = HM groups, there are statistically significant differences in the predicted violence index.

TABLE 2. EFFECTS OF ASP GROUP COMPOSITION

Only Treated Subsample. Results from specification (2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: BEHAVIORAL OUTCOMES									
	Attitudes towards school and learning				**Violence and Behavior**				
	Positive attitudes towards school	Time to do homework (hours)	Pay attention in class	Absenteeism (days)	Delinquency (Index)	Violent actions (Index)	Approval of antisocial behavior	Behavior reports (-)	Probability of bad behavior report
Homog. Group	-0.158*	-0.099	-0.030	0.249	0.020	-0.043	-0.002	-0.034	0.055***
	(0.082)	(0.151)	(0.032)	(0.401)	(0.070)	(0.054)	(0.017)	(0.056)	(0.017)
Observations	716	707	727	631	691	722	720	762	762
PANEL B: NEUROPHYSIOLOGICAL OUTCOMES									
	Arousal (stress)	Valence	Positive Valence Difference	Negative Valence Difference	Locus of control	CRT	Raven		
Homog. Group	0.257***	-0.048	-0.014	-0.084	-0.178	0.042	-0.114		
	(0.071)	(0.257)	(0.283)	(0.343)	(0.128)	(0.114)	(0.129)		
Observations	238	238	238	238	227	227	227		
PANEL C: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Homog. Group	0.041	0.075	0.029	0.040	0.018	0.003	0.004	-0.018	-0.010
	(0.037)	(0.053)	(0.041)	(0.034)	(0.011)	(0.014)	(0.011)	(0.014)	(0.009)
Observations	771	771	771	771	771	771	771	771	771

***, **, * indicates that the effect of being treated in a MH (high or low) group compared to being treated in a HT group is significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parenthesis. Panel A exhibits effects on non-cognitive outcomes. Panel B presents results on academic outcomes. Description of outcome variables is available in Appendix 1. All regressions are estimated using only treated group and models of specifications (4) - (5). All regressions include as controls: a second order polynomial of student's IVV, and ciclo-school fixed effect (stratification level), except those from specification (5). In estimations for academic outcomes, absenteeism and bad behavior reports, I also include the corresponding imputed outcome at the baseline and a dummy indicating a missing value at baseline.

TABLE 3. HETEROGENEOUS EFFECTS OF GROUP COMPOSITION

Only Treated Subsample. Results from specification (3)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: BEHAVIORAL OUTCOMES									
	Attitudes towards school and learning				**Violence and Behavior**				
	Positive attitudes towards school	Time to do homework (hours)	Pay attention in class	Absenteeism (days)	Delinquency (Index)	Violent actions (Index)	Approval of antisocial behavior	Behavior reports (-)	Probability of bad behavior report
Low Hom. Group	-0.219*	0.100	-0.084**	0.288	0.083	-0.049	0.004	-0.022	0.021
	(0.125)	(0.295)	(0.043)	(0.752)	(0.112)	(0.079)	(0.030)	(0.072)	(0.023)
High Hom. Group	-0.120	-0.253	0.028	0.321	0.009	-0.040	-0.007	-0.046	0.092***
	(0.141)	(0.266)	(0.052)	(0.855)	(0.130)	(0.091)	(0.023)	(0.090)	(0.032)
Observations	716	707	727	631	691	722	720	762	762
PANEL B: NEUROPHYSIOLOGICAL OUTCOMES									
	Arousal (stress)	Valence	Positive Valence Difference	Negative Valence Difference	Locus of control	CRT	Raven		
Low Hom. Group	0.142	-0.263	-0.184	-0.286	-0.182	0.197	0.070		
	(0.094)	(0.374)	(0.445)	(0.416)	(0.160)	(0.130)	(0.152)		
High Hom. Group	0.336***	0.049	0.025	-0.010	-0.181	-0.024	-0.245		
	(0.082)	(0.347)	(0.436)	(0.450)	(0.161)	(0.160)	(0.168)		
Observations	238	238	238	238	227	227	227		
PANEL C: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Low Hom. Group	0.118	0.180	0.100	0.143*	0.030**	0.033	-0.008	-0.014	-0.014
	(0.076)	(0.106)	(0.085)	(0.078)	(0.021)	(0.027)	(0.019)	(0.027)	(0.014)
High Hom. Group	-0.061	-0.050	-0.065	-0.082*	0.005	-0.026*	0.015	-0.023	-0.007
	(0.059)	(0.058)	(0.067)	(0.049)	(0.028)	(0.027)	(0.020)	(0.024)	(0.017)
Observations	771	771	771	771	771	771	771	771	771
MDE T = C	0.081	0.091	0.100	0.085	0.109	0.110	0.111	0.112	0.156

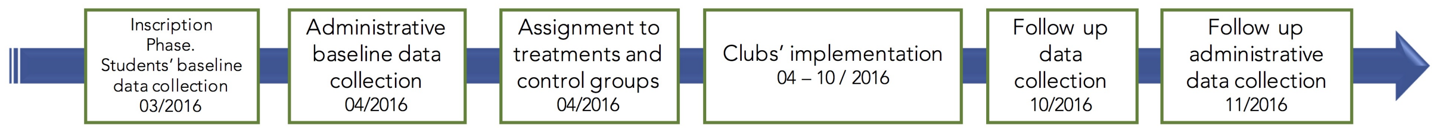
***, **, * indicates that the effect of being treated in a MH (high or low) group compared to being treated in a HT group is significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parenthesis. Panel A exhibits effects on non-cognitive outcomes. Panel B presents results on academic outcomes. Description of outcome variables is available in Appendix 1. All regressions are estimated using only treated group and models of specifications (4) - (5). All regressions include as controls: a second order polynomial of student's IVV, and ciclo-school fixed effect (stratification level), except those from specification (5). In estimations for academic outcomes, absenteeism and bad behavior reports, I also include the corresponding imputed outcome at the baseline and a dummy indicating a missing value at baseline.

TABLE 4. EFFECTS OF ASSIGNMENT TO HIGH VIOLENCE HOMOGENEOUS GROUP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: VIOLENCE AND ATTITUDES									
	Attitudes towards school and learning				**Violence and Behavior**				
	Positive attitudes towards school	Time to do homework (hours)	Pay attention in class	Absenteeism (days)	Delinquency (Index)	Violent actions (Index)	Approval of antisocial behavior	Behavior reports (-)	Probability of bad behavior report
Second Order Polinomial Specification									
High-Homog group	-0.079	-0.420	-0.024	0.027	0.093	0.180*	0.021	0.071	0.026
	(0.183)	(0.296)	(0.048)	(0.889)	(0.131)	(0.107)	(0.041)	(0.178)	(0.019)
Observations	472	468	480	423	455	476	474	511	511
Restricting the sample to 8 students around the cut-off									
High-Homog group	-0.645**	-1.596***	-0.244**	0.294	0.579***	0.250*	-0.018	0.369**	0.132*
	(0.287)	(0.383)	(0.112)	(1.408)	(0.221)	(0.143)	(0.041)	(0.169)	(0.080)
Observations	106	106	108	92	92	102	108	114	114
PANEL B: NEUROPHYSIOLOGICAL OUTCOMES									
	Arousal (stress)	Valence	Positive Valence Difference	Negative Valence Difference	Locus of control	CRT	Raven		
Second Order Polinomial Specification									
High-Homog group	0.107	0.656	0.478	0.666	-0.279	-0.773**	-0.008		
	(0.157)	(0.830)	(0.990)	(0.800)	(0.208)	(0.340)	(0.251)		
Observations	238	238	238	238	227	227	227		
PANEL C: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Second Order Polinomial Specification									
High-Homog group	-0.034	-0.156*	0.004	-0.054	-0.038	-0.074**	-0.033	-0.051*	0.048**
	(0.083)	(0.092)	(0.100)	(0.081)	(0.028)	(0.032)	(0.023)	(0.029)	(0.020)
Observations	516	516	516	516	516	516	516	516	516
Restricting the sample to 8 students around the cut-off									
High-Homog group	-0.090	-0.151	0.085	0.026	-0.045	-0.095	0.002	0.007	0.031***
	(0.201)	(0.161)	(0.181)	(0.119)	(0.041)	(0.061)	(0.037)	(0.042)	(0.012)
Observations	115	115	115	115	115	115	115	115	115

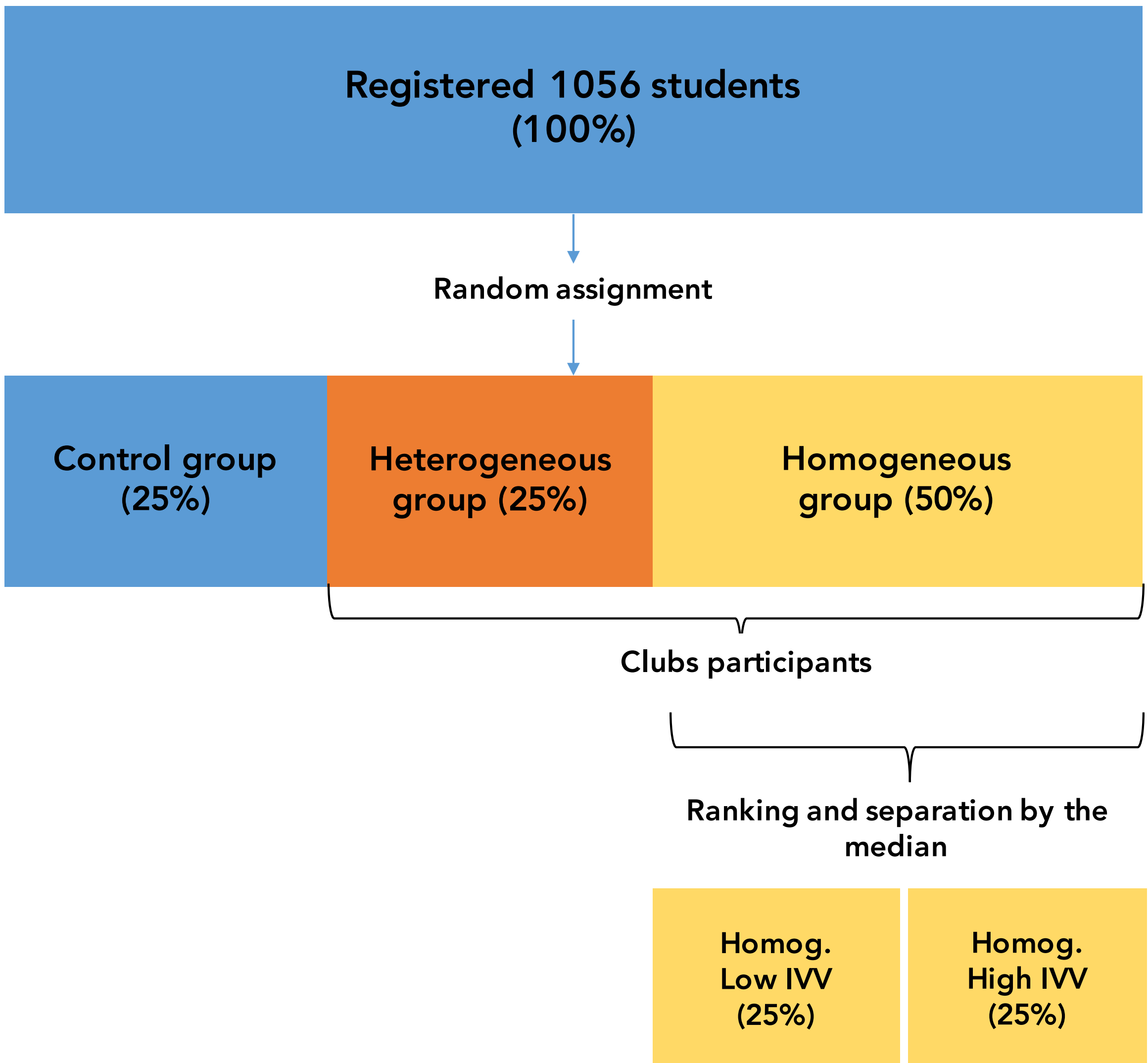
***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A present results on academic outcomes. Reading, math, and science grades are standardized values from control groups at the school-grade level at follow-up. Score is an average of the three courses. Panel B shows effects on non-cognitive outcomes. All regressions include the following controls: second order polynomial IVV, grades in the respective course before treatment, a dummy indicating a missing value in the grade before treatment, and ciclo-school fixed effect (stratification level). Estimations first use the homogeneous groups subsample and then the 8 students around the cut-off. These estimations correspond to the model from specification (7).

Figure 1. Timeline of the Intervention.



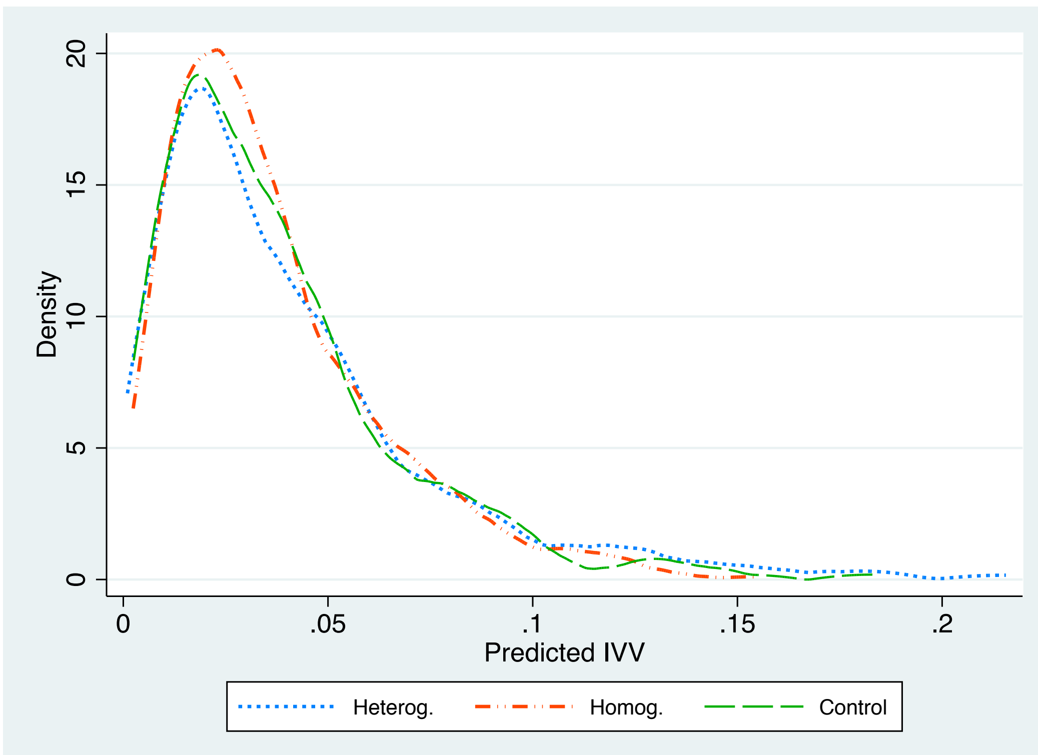
Timeline of the intervention and data collection.

Figure 2. Experimental Design.



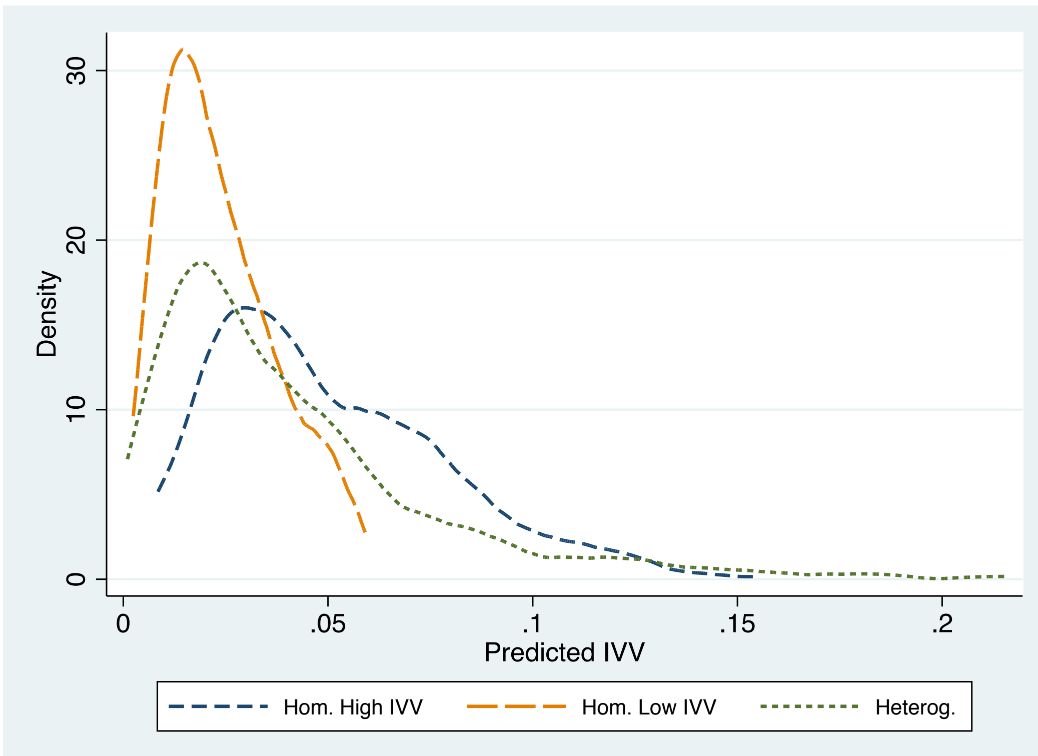
This figure shows the samples composition and randomization procedure applied in this design. From the total of enrolled children in each educational level $\{1,2,3\} \in$ school A, we randomly assigned 25% to C and 25% to HT treatment arm and 50% to HM groups. The same procedure was implemented in the remaining schools.

Figure 3. IVV Distribution Functions of Treatment and Control Groups.



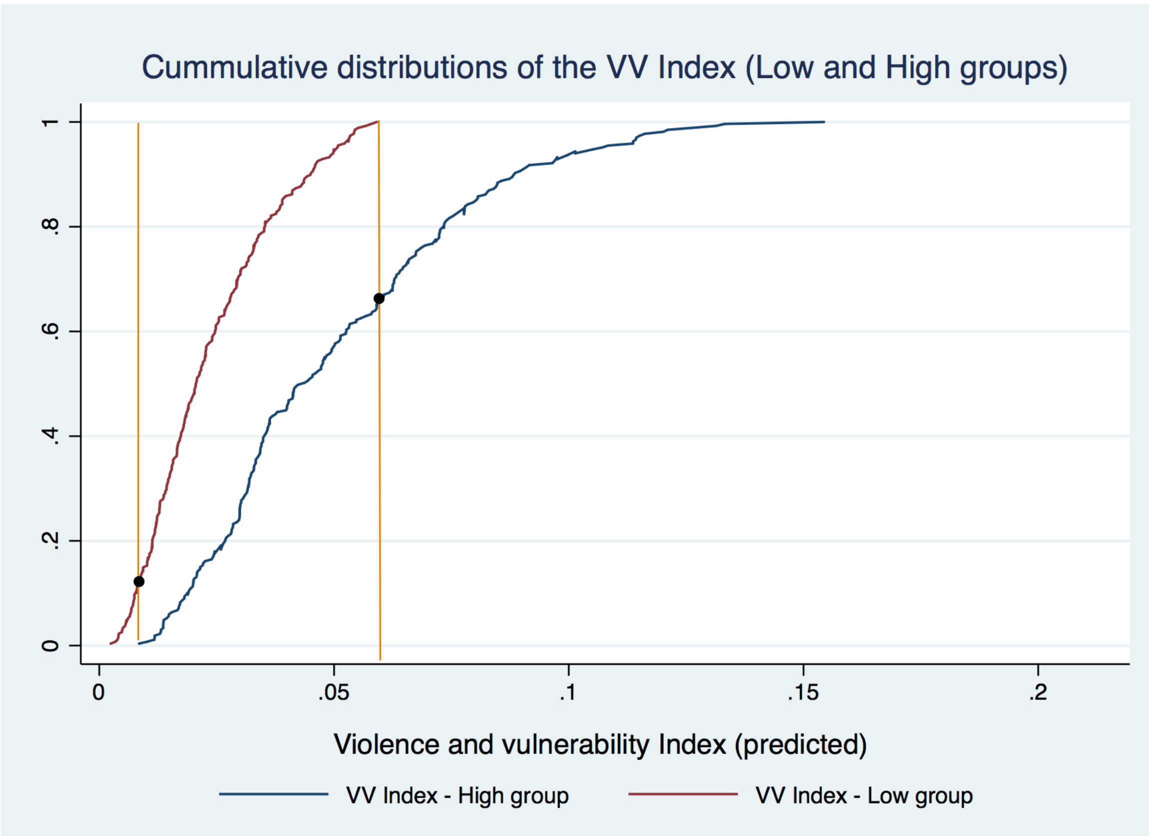
Predicted IVV distribution functions for the control and any treatment (homogeneous and heterogeneous) groups prior to treatment, for the whole study sample.

Figure 4. IVV Distribution Functions of Treated Groups.



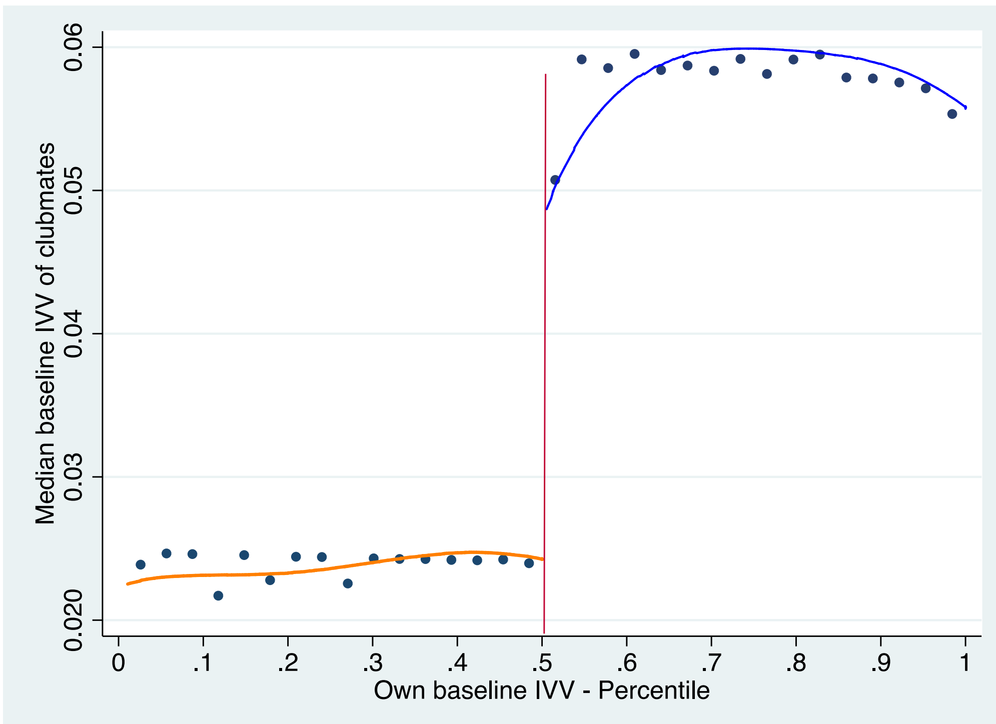
Predicted IVV distribution functions generated by the experimental design for the heterogeneous treatment group and each of the homogeneous subgroups (High and Low IVV) in the whole study sample

Figure 5. IVV Cumulative Distribution Functions of Homogeneous Sub-groups.



Cumulative distribution function for high- and low-homogeneous treatment groups' predicted propensity for violence. Vertical yellow lines define the limits of overlap between both distribution functions. This overlap in the violence level occurs because assignment was at the strata level, and the median level was different within each strata.

Figure 6. Experimental Variation in IVV Peer Composition, prior to treatment



Median predicted IVV of student's clubmates as a function of the student's own baseline IVV in homogeneous high and low groups. Consistent with the discontinuous assignment at the median IVV, there is a sharp discontinuity at the fiftieth percentile for the entire subsample.

Appendix

Appendix 1. Description of Outcome Variables.

In our follow-up survey, we have multiple variables that measure some behavioral outcomes, such as attitudes, delinquency, and violent behavior. In order to have a single continuous measure that can be compared to previous evidence in the literature, we have built for some of them a standardized index that is an average of the multiple variables measured in the survey.

In the following section, we provide details of each outcome variable. For the index outcomes construction, we provide information of the main items included.

A. Behavior and Academic Outcomes

1. *Positive attitudes towards school and learning:* standardized index estimated using principal components analysis. We used 5 items from the self-reported follow-up survey, each of them was on a 1-4 scale, where 1 = very important, and 4 = no important. The items were the following: (i) Indicator of how important is to learn for them, (ii) indicator of how they like their school, (iii) indicator of studying is an opportunity for having a better future, (iv) indicator of whether hard work at schools pays off, and (v) indicator of whether what they learn at school is relevant for their future.

2. *Time spent on homework*: This variable was a self report from students in time units (minutes). The question was: *During the last 3 months, how much time did you spend doing your homework aside from the time you were at school or in classes?*

3. *Paying attention in class*: This outcome was a self report from students. The question was: *When I am in class, I pay attention* The answers were on a 1-4 scale, in which 1 = always and 4 = never. Using this scale, we created a dummy that takes the value of 1 if student reports that she pay attention most of the time (1 or 2) and 0 if she barely pays attention during classes (3 or 4).

4. *Delinquency Index*: corresponds to a standardized sum of self reported delinquent actions, such as theft, mugging someone, etc. The question was: *During the last 3 months, did you ...?* This variable can be interpreted as an aggregation of delinquency actions.

5. *Violent Actions Index*: is the standardized sum of other violent acts such as fighting at school, damage of municipal property, fight with siblings, etc. The question was: *During the last 3 months, did you ...?* As before, this index variable can be interpreted as an aggregation of violent actions.

6. *Approval of peers' antisocial behavior* is a binary indicator that takes the value of 1 if students approve any peer behavior related to alcohol and drugs consumption, fighting, etc. The question was: *What do you think if someone of your closest friends ...?*

7. *Absenteeism*: corresponds to the number of days the student did not attended school between April-October of the 2016 academic year. This was administrative data provided by schools at follow-up.

8. *Bad behavior reports*: In El Salvador, these are reported by teachers each quarter. They are presented on the following discrete scale: Excellent (E), Very Good (MB), Good (B), and Regular (R). It can be translated in a continuous scale that is comparable to course grades. In this paper, we used a reversed continuous scale to facilitate the interpretation and comparability to the self-reported measures of violence and crime.

9. *Reading, math and science grades*: Variables that indicate performance on each course. It is a 0-10 scale, where 0 is the worst performance and 10 is the best. We have standardized these values from control groups at the school-grade level. Score is an average of the three courses.

10. *Passing course*: Is a dummy variable that takes the value of 1 if student has been promoted to the following course and 0 otherwise.

B. Neurophysiological Outcomes

1. *Arousal*: pre-test resting measure of individual's stress, estimated directly from her brain activity using EEG recordings measured while children were watching a black cross in the center of a gray screen for a period of 30 seconds.

2. *Valence*: pre-test resting state measure estimated directly from participants' brain activity using EEG recordings. As the arousal measure, this recordings were estimated while children were watching a black cross in the center of a gray screen for a period of 30 seconds. This variable can be interpreted as a positive or negative mood, as well as an attitude of either approach or withdrawal towards/from a stimulus (**??**).

3. *Locus of control*: Psychometric test developed by (**?**). This indicates that children think that they are not able to control what happens in their lives.

4. *Cognitive Reflection Test (CRT)*: Is a test designed to measure if an individual tends to automatically choose an initially incorrect response and then engage in a deeper reasoning to find a correct answer.

5. *Raven*: Is a measure of abstract reasoning and a non-verbal estimate of intelligence. It is implemented as a set of matrices in progressive order.

6. *Positive Valence Difference*: Corresponds to the difference between the response intensity measure after exposure to positive stimuli and the valence-at-resting-state index described before.

7. *Negative Valence Difference*: Is a measure of the variation in the valence index recorded when the stimulus was negative net of the individual's baseline resting state valence index. Both differences can be interpreted as a lower level of overreaction of participants –they become more phlegmatic or cold headed– or that individuals move towards a more withdrawal behavior or attitude.

Appendix 2. Group composition heterogeneous effects

I also explore non-linear heterogeneous effects of group composition by initial propensity for violence in a finer level. Thus, I interact HM and HT treatments with dummies of quartiles of the IVV distribution, using the following specification:

$$Y_{ij} = \alpha_0 + \alpha_1 HT_{ij} + \alpha_2 HM_{ij} + \alpha_3 \sum_{k=1}^{4} HT_{ij} \times Qk_{ij} + \alpha_4 \sum_{k=1}^{4} HM_{ij} \times Qk_{ij} + \alpha_5 X_{ij} + S_j + \epsilon_{ij} \quad (1)$$

which is equivalent to:

$$Y_{ij} = \alpha_0 + \alpha_1 HT_{ij} + \alpha_2 HM_{ij} + \alpha_3 \sum_{m=1}^{4} HT_{ij} \times Qs_{ij}$$

$$+\alpha_{4a} \sum_{m=1}^{2} HomL_{ij} \times Qs_{ij} + \alpha_{4b} \sum_{m=3}^{4} HomH_{ij} \times Qs_{ij} + \alpha_5 X_{ij} + S_j + \epsilon_{ij}$$

where $Qs_{ij} = 1$ if student i is in quartile $s \in \{1, 2, 3, 4\}$ of the IVV distribution function at the stratum j level. The omitted category is Q1 and the interaction between it and the treatment dummy. Results are shown in Appendix Table A16. At each panel, I present the total effect of each treatment by quartile and then the p-values of the test of differences among the effects of each treatment by quartile.

On outcomes related to attitudes towards school and learning, I find that least and most violent students (Q1 and Q4 respectively) are more responsive to group composition. For example, Q1 students improve their positive attitudes and pay more attention during classes when are treated in heterogeneous groups compared to students treated in homogeneous group from the same quartile.

Moreover, in terms of violence-related outcomes, students in Q4 face a reduction in the probability of having a misbehavior report when they are treated in heterogeneous group compared to those in heterogeneous groups. These results do not seem to be at expense of students in Q1, because even though the reduction on misbehavior is greater when they are treated in homogeneous groups, they actually reduce their bad behavior at school under both treatments. In the rest of outcomes, differences between HT and HM treatments for students in similar quartiles are not statistically different from zero.

On academic outcomes, the most violent students (Q4) are more sensitive to group composition. According to the results, they have greater academic outcomes when treated in heterogeneous groups. These results also seem not to be at the expense of low violent children. For example, I do not find statistical differences between the effects of assigning students of the rest of quartiles to homogeneous or heterogeneous groups on academic outcomes, except on the extensive margin of reading grades.

Similarly, I estimate a local polynomial fit of standardized end line score grades by predicted violence index, and find that the children in the least violent quartile ($Q1$) and in the most violent quartile ($Q4$) are more sensitive to their group composition as shown in Appendix Figure A2.

This pattern of results suggests that students driving most of the impact estimates are those in both tails of the baseline IVV distribution, that is the students for whom the exposure to certain level of violence from their peers is usually greater than the exposure than those located closer to the middle of the violence distribution. One of these groups is constituted by the students expected to benefit the most from the ASP.

[**Insert Table A10 here**]

[**Insert Table A11 here**]

Appendix 3. Exploiting the random allocation of peers

Since participants were randomly allocated to a group in the ASP, there is some variation in the group composition which stem from the fact that being assigned to HM vs HT directly affects the mean and variance of one's peers. As in Lafortune et al., 2016, after controlling for a strata fixed effect, the variance and mean IVV of peer stems entirely from the random assignment. Similar approaches have been used by Carrell et al., 2013; Duflo, Dupas and Kremer (2011), and Lyle et al (2007). The estimating equation for the sample of students selected to participate in the ASP is:

$$Y_{ij} = \gamma_0 + \gamma_1 \bar{x}_{-ij} + \gamma_2 var(x_{-ij}) + \gamma_3 S_j + \gamma_4 X_{ij} + \epsilon_{ij} \tag{2}$$

where $\bar{x}_{-ij}$ and $var(x_{-ij})$ are the club's mean and variance to which student i was assigned, excluding her personal IVV - this allows me to address the reflection problem. The rest of variables are defined as before. With this specification I can directly provide evidence of how student's i non-cognitives and/or her academic outcomes are affected by the average baseline or variance in the violence of her peers.

Using this and restricting the sample to treated students, I find terms of non-cognitive outcomes. Panel A shows that a higher average clubmates' IVV reduces the self reported time spent doing homework but being in a more diverse group increases both positive attitudes towards school and learning and self reported time spent doing homework. In terms of violence, I do not find an effect from either the mean or average of clubmates' IVV.

I also find that on average, students exposed to a group of peers with higher mean of propensity for violence reduce their math and reading scores, showing a negative peer effect of violence on grades. However, being exposed to a more diverse group of clubmates increases math grades and reduces the probability of grade repetition.

[**Insert Table A12 here**]

TABLE A1. COMPARISON OF THE STUDY AND FUSADES (2015) SAMPLES

	(1)	(2)	(3)	(4)	(5)
	Study Sample		FUSADES(2015) Sample		p-value
	Mean	Std. Dev.	Mean	Std. Dev.	
Student is male	0.49	0.50	0.47	0.50	0.23
Student lives in 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 only with one of his/her parents	0.32	0.47	0.30	0.46	0.19
Student living with one parent	0.06	0.25	0.08	0.27	0.02
Student living with other relative	0.08	0.27	0.07	0.26	0.25
Student's travel time from house to school (minutes)	17.64	14.37	17.25	12.98	0.37
Student's mother's level of education	0.31	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	0.09
Student's course	5.75	2.71	5.5	2.52	0.29
N	1056		6641		

The table provides means and standard deviations of the main variables from this study and 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 * denotes difference significant at the 1 %, 5 % and 10 % level respectively when comparing the means.

TABLE A2. IVV ESTIMATION RESULTS AND DETERMINANTS. FUSADES (2015) SAMPLE

	Violence
Student is male	0.258***
	(0.054)
Student's age	0.092***
	(0.017)
Student lives in urban area	0.195***
	(0.066)
Student's household composition	
Student living only with one of his/her parents	0.033
	(0.062)
Student living with other relative	0.042
	(0.112)
Student living with other non-relative adult	0.723
	(0.466)
Student living with no adults	0.362
	(0.290)
Student's mother level of education:	
Intermediate education (7-12 years)	0.113*
	(0.061)
University or higher (13 and +)	0.057
	(0.079)
Student's travel time from house to school (min.)	0.005**
	(0.002)
Student is alone at home after school	0.391***
	(0.070)
Student's school year	0.067
	(0.089)
Student enrolled on morning shift	-0.002
	(0.087)

I estimated the following specification $V_f = \alpha_0 + \alpha_1 D_f + \epsilon_f$. In FUSADES (2015) survey, they defined V_f as a violence dummy indicating that a child or adolescent has committed at least one of the following actions: *Have you ever: (i) bring a gun, (ii) attacked someone with the intention to hurt him, (iii) attacked someone with a gun, (iv) used a gun or a violent attitude to get money or things from someone?*. D_f is a vector of violence determinants, including gender, age, mothers' education, etc.

***, **, * indicate if estimated coefficients α_1 are statistically different form zero. Standard error in parentheses. Mother's education omitted category: mother has basic education (1-6th grades). Household composition omitted category: children living with both parents.

TABLE A3. CLASSIFICATION USING MISBEHAVIOR REPORTS OR ESTIMATED PROPENSITY FOR VIOLENCE (IVV)

	(1)	(2)	(3)	(4)	(5)
	Full Sample	Treated [T]	Control [C]	Heterog. [Het]	Homog. [Hom]
Similar classification	0.527	0.528	0.527	0.513	0.534
Observations	1056	798	258	263	535
Test for differences		T = C	C = Het	C = Hom	Het = Hom
		0.998	0.773	0.871	0.560

The variable "similar classification" = 1 if a student would have been classified as high violence child using their position in the IVV and misbehavior reports distribution functions, at the stratum-treatment arm (C, T, Het, Hom) level. Tests include strata fixed effects. Robust standard errors at course-school level are in parentheses.

TABLE A4. CORRELATION BETWEEN IVV, ACADEMIC GRADES AND MISBEHAVIOR REPORTS AT BASELINE

	(1)	(2)	(3)	(4)	(5)
	GRADES				Behaviour
	Reading	Math	Science	Score	
Panel A. Standardized and imputed grades					
IVV	-0.013	0.021	-0.021	-0.011	0.056***
	(0.017)	(0.039)	(0.020)	(0.020)	(0.021)
Constant	0.176*	-0.048	0.179*	0.143	0.304***
	(0.096)	(0.150)	(0.104)	(0.087)	(0.104)
Observations	1,056	1,056	1,056	1,056	1,056
Panel B. Standardized grades at the course level					
IVV	-0.015	-0.007	-0.021	-0.018	0.050**
	(0.019)	(0.028)	(0.018)	(0.021)	(0.020)
Constant	0.059	0.025	0.078	0.067	0.190*
	(0.103)	(0.104)	(0.097)	(0.090)	(0.101)
Observations	1,034	984	1,007	970	1,000
Panel C. Non-standardized grades					
IVV	-0.029	-0.005	-0.031	-0.024	0.066**
	(0.031)	(0.042)	(0.026)	(0.027)	(0.026)
Constant	6.772***	6.499***	6.740***	6.723***	7.202***
	(0.161)	(0.164)	(0.143)	(0.118)	(0.130)
Observations	1,034	984	1,007	970	1,000

I estimated the correlation between the IVV prediction with academic grades and misbehavior reports before the intervention using administrative data.The estimated specification was the following: $y_{ij} = \alpha_0 + \alpha_1 IVV_{ij} + \epsilon_{ij}$, where y_{ij} is the academic grade or misbehavior report for student i in school j, IVV_{ij} is the estimated propensity for violence. ***, **, * indicates that coefficients are significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parentheses.

TABLE A5. IVV PREDICTION POWER OF MISBEHAVIOR AT SCHOOL

Using only the control group

	(1)	(2)	(3)	(4)
	Intensive margin		*Extensive margin*	
IVV	0.227***	0.129**	0.101***	0.061**
	(0.074)	(0.064)	(0.034)	(0.031)
Observations	248	248	248	248
Controls	No	Yes	No	Yes

Results of the correlation between IVV prediction and misbehavior reports one year after the estimation. I used administrative data only for the control group (those who where not directly treated). The estimated specification was the following: $y_{ijt} = \alpha_0 + \alpha_1 IVV_{ijt-1} + \epsilon_{ijt}$, 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 propensity for violence one year before. ***, **, * indicates that coefficients are significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parentheses.

TABLE A6. MATCHING RATE WITH ADMINISTRATIVE DATA AND ATTRITION RATE.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Treatments		Tracking groups	
	Full Sample	Control Group (C)	Any Treatment (T)	Heterogen. group (HT)	Homogen. group (HM)	Homog. High (HM-High)	Homog. Low (HM-Low)
Fraction of students with matched administrative data, Q1 2016							
Reading scores	0.98	0.97	0.98	0.98	0.98	0.98	0.98
Math scores	0.91	0.89	0.92*	0.90	0.92+	0.92	0.93
Science scores	0.95	0.94	0.96	0.96	0.96	0.96	0.96
Behaviour scores	0.93	0.91	0.94	0.94	0.94	0.94	0.94
Abseenteism	0.68	0.68	0.67	0.68	0.67	0.65*	0.69
Fraction of students with matched administrative data, Q4 2016							
Reading scores	0.97	0.98	0.97	0.97	0.96	0.96	0.97
Math scores	0.97	0.98	0.97	0.97	0.96	0.96	0.97
Science scores	0.97	0.98	0.97	0.97	0.96	0.96	0.97
Behaviour scores	0.96	0.96	0.97	0.95	0.95	0.95	0.96
Abseenteism	0.80	0.79	0.80	0.80	0.80	0.76	0.83”
Number of students at baseline and follow up							
Number of students present at baseline	1056	258	798	263	535	267	268
Number of students present at follow-up	968	237	731	248	483	239	244
Retention rate (1-attrition)	0.92	0.92	0.92	0.94	0.91	0.90	0.91

The table provides the match rate with administrative data, calculated as the fraction of students present at the survey at the baseline whom could be matched with administrative data from schools. In comparing T and C, * denotes difference significant at the 10 % level. A similar notation is used to indicate statistically significant differences between HM and C (+) and between HM-High and HM-Low (”).

TABLE A7: DESCRIPTIVE STATISTICS OF THE IVV BY TREATMENT GROUP.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full Sample	Control Group (C)	Any Treatment (T)	Treatments		Tracking groups	
				Heterogen. group (HT)	Homogen. group (HM)	Homog. High (HM-H)	Homog. Low (HM-L)
Mean	0.038	0.038	0.038	0.041	0.037	0.051	0.023
Std. Dev	0.029	0.029	0.029	0.035	0.026	0.028	0.014
Median	0.030	0.029	0.030	0.001	0.031	0.044	0.021
Min	0.001	0.003	0.001	0.001	0.002	0.009	0.002
Max	0.216	0.183	0.216	0.216	0.154	0.154	0.059
N	1056	258	798	263	535	267	268

The table provides summary statistics for the Vulnerability and Violence Index (IVV) predicted using FUSADES (2015) dataset and variables available at during the clubs' enrollment phase.

TABLE A8. p-values OF DIFFERENCES BETWEEN TREATMENT AND CONTROL GROUPS.

	(1)	(2)	(3)	(4)	(5)
	Adjusted *p-values*				
	Control = Tratado	Control = Heterog.	Control = Homog.	Heterog. = Homog.	Homog. High = Homog. Low
PANEL A: IVV Determinants					
Student is male	0.652	0.511	0.723	0.627	0.000
Student's age	0.227	0.151	0.391	0.192	0.081
Student lives in urban area	0.491	0.901	0.509	0.548	0.115
Student's household composition					
Student living with both parents	0.161	0.414	0.082	0.279	0.323
Student living with only one parent	0.103	0.741	0.071	0.228	0.905
Student living with a parent and a step-parent	0.652	0.639	0.987	0.668	0.841
Student living with other relative /adult	0.541	0.653	0.757	0.728	0.000
Student's mother's level of education:					
Basic education (1-6 years)	0.265	0.084	0.463	0.112	0.000
Intermediate education (7-12 years)	0.364	0.117	0.549	0.326	0.000
University or higher (13 and +)	0.771	0.428	0.993	0.629	0.622
Student's travel time from house to school (min.)	0.446	0.533	0.507	0.976	0.021
Student is alone at home after school	0.801	0.184	0.822	0.110	0.000
Student's school year	0.173	0.140	0.294	0.346	0.004
Student enrolled on morning shift	0.859	0.286	0.897	0.319	0.055
Student's violence index	0.786	0.221	0.705	0.031	0.000
PANEL B: Academic outcomes					
Academic scores Q1 2016					
Reading scores	0.136	0.073	0.046	0.377	0.165
Math scores	0.690	0.260	0.927	0.215	0.259
Science scores	0.105	0.278	0.083	0.546	0.114
Behaviour scores	0.115	0.111	0.150	0.971	0.149
Absenteeism Q1 2016	0.646	0.747	0.650	0.889	0.172
PANEL C: Sample composition and response rate					
Average club size at baseline	-	-	-	0.926	0.385
Take up	-	-	-	0.910	0.286
Retention rate (1-attrition)	0.398	0.202	0.390	0.051	0.383
Communitary tutor	-	-	-	0.139	0.113

TABLE A9. ASP ATTENDANCE OF TREATED STUDENTS

	(1) Sessions attended	(2) Days attended
Low Homog. group	-0.258	-0.184
	(1.502)	(1.195)
High Homog. group	-0.580	-1.653
	(1.485)	(1.191)
Observations	798	798

***, **, * indicates that the club attendance from the HM (high or low) group compared to being treated in a HT group is significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at course-school level are in parenthesis. Two measures of attendance are number of sessions and days. Regressions are estimated using only treated group and models of specifications (5).

TABLE A10. HETEROGENEOUS EFFECTS OF GROUP COMPOSITION BY IVV.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: NON-COGNITIVE OUTCOMES									
	Attitudes towards school and learning				**Violence and Behavior**				
	Positive attitudes towards school	Time to do homework (hours)	Pay attention in class	Absenteeism (days)	Delinquency (Index)	Violent actions (Index)	Approval of antisocial behavior	Behavior reports (-)	Probability of bad behavior report
(1) Total Hom effect on Q4	0.090	0.000	0.127**	-2.608	-0.283**	-0.322**	-0.104***	0.017	0.039
(2) Total Hom effect on Q3	0.232	0.389*	0.054	-1.732	-0.122	-0.123	-0.168***	-0.178	-0.081
(3) Total Hom effect on Q2	0.016	0.576***	0.032	-0.670	-0.184	-0.302**	0.018	-0.221**	-0.025
(4) Total Hom effect on Q1	0.141	0.268	0.055	-0.776	-0.172	0.152**	-0.169***	-0.383***	-0.131***
(5) Total Het effect on Q4	0.181	0.545*	0.078	-3.376	-0.064	0.027	-0.092***	-0.059	-0.090*
(6) Total Het effect on Q3	0.398*	0.387*	0.079	-1.107	-0.245*	-0.258**	-0.132**	-0.124	-0.150**
(7) Total Het effect on Q2	0.111	0.645*	0.063	-0.025	-0.165	-0.351***	-0.016	-0.282**	-0.086
(8) Total Het effect on Q1	0.453***	-0.019	0.193***	-2.633**	-0.419*	0.149	-0.187***	-0.120	-0.064
Observations	948	935	962	833	916	956	962	1010	1010
p-value test HomQ4 = HetQ4 [row (1) = row (5)]	0.4432	0.1006	0.4451	0.3124	0.1188	0.0212	0.6121	0.5145	0.0010
p-value test HomQ3 = HetQ3 [row (2) = row (6)]	0.3755	0.9933	0.6977	0.5968	0.3788	0.2761	0.2084	0.5440	0.0813
p-value test HomQ2 = HetQ2 [row (3) = row (7)]	0.5826	0.8548	0.6372	0.5061	0.8835	0.5670	0.2497	0.5790	0.2099
p-value test HomQ1 = HetQ1 [row (4) = row (8)]	0.0465	0.4030	0.0027	0.1598	0.1150	0.9841	0.6820	0.0166	0.1523
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
(1) Total Hom effect on Q4	-0.019	0.151**	0.134	0.056	0.052*	0.026	0.053	0.014	-0.080
(2) Total Hom effect on Q3	0.042	0.295***	0.229**	0.149**	0.036	0.075**	0.052	0.046	-0.024
(3) Total Hom effect on Q2	0.147**	0.100	0.120	0.119	0.101***	-0.004	0.029	0.048*	-0.027
(4) Total Hom effect on Q1	-0.063	-0.044	0.061	-0.059	-0.026	-0.025	-0.017	-0.033	0.011
(5) Total Het effect on Q4	0.131	0.237**	0.299***	0.183	0.100***	0.082**	0.083*	0.061*	-0.080**
(6) Total Het effect on Q3	-0.022	0.191**	0.149	0.136*	-0.016	0.078**	0.001	0.050	-0.003
(7) Total Het effect on Q2	0.006	0.044	0.105	0.032**	0.051**	-0.059	0.009	0.058	-0.018
(8) Total Het effect on Q1	-0.202*	-0.310	-0.148	-0.281*	-0.053	-0.051	0.009	-0.024	0.028
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023
p-value test HomQ4 = HetQ4 [row (1) = row (5)]	0.0344	0.2606	0.0671	0.0330	0.0300	0.0112	0.2443	0.0241	0.9464
p-value test HomQ3 = HetQ3 [row (2) = row (6)]	0.4795	0.2616	0.2807	0.8347	0.1100	0.8969	0.0369	0.8726	0.3897
p-value test HomQ2 = HetQ2 [row (3) = row (7)]	0.0387	0.6141	0.8182	0.1630	0.0025	0.1346	0.5790	0.6949	0.5964
p-value test HomQ1 = HetQ1 [row (4) = row (8)]	0.2078	0.1697	0.1050	0.1126	0.3237	0.4021	0.3429	0.7934	0.4856

***, **, * indicates that the effect for a student in quartile Qi of being treated in a HM or HT group compared to the control group is significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors in parentheses at course-school level. All regressions are estimated using only treated sample. Panel A present results on academic outcomes. Panel B exhibit effects on non-cognitive outcomes. Description of outcome variables is available in Appendix 1. All regressions include as control variables grades in the respective course at baseline, dummy indicating a missing value in the grade at baseline, and ciclo-school fixed effect (stratification level).

TABLE A11. HETEROGENEOUS EFFECTS OF GROUP COMPOSITION BY ACADEMIC COURSES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least
	Reading	Math	Science	Score	Reading	Math	Science	Score	one course
[*i*] HM-Low enrolled in a	0.013	0.105	0.053	0.057	-0.009	0.018	-0.013	-0.050*	-0.005
non-academic course	(0.085)	(0.075)	(0.080)	(0.064)	(0.020)	(0.024)	(0.019)	(0.026)	(0.013)
[*ii*] HM-Low enrolled in a	0.389*	0.363	0.257	0.363*	0.105**	0.052	-0.000	0.054	-0.032
academic course	(0.201)	(0.297)	(0.184)	(0.207)	(0.055)	(0.057)	(0.039)	(0.047)	(0.034)

***, **, * indicate that the comparison between HM-Low vs HT at their respective category of course is significant at 1 %, 5 %, and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Sample is restricted to N = 771 treated students. All regressions include as controls: a second order polynomial of student's IVV, IVV median at the group-stratum level, a binary indicator of high violence, imputed outcome at the baseline, and a dummy indicating a missing value at the baseline.

TABLE 12. EFFECTS OF ASP GROUP COMPOSITION (Only Treated Subsample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: VIOLENCE AND ATTITUDES									
	Attitudes towards school and learning				**Violence and Behavior**				
	Positive attitudes towards school	Time to do homework (hours)	Pay attention in class	Absenteeism (days)	Delinquency (Index)	Violent actions (Index)	Approval of antisocial behavior	Behavior reports (-)	Probability of bad behavior report
Clubmates' IVV Mean	-0.012	-0.178**	-0.016	0.046	0.012	-0.006	0.009	-0.004	0.012
	(0.046)	(0.073)	(0.010)	(0.227)	(0.039)	(0.036)	(0.010)	(0.040)	(0.016)
Clubmates' IVV Variance	0.032**	0.060***	0.005	-0.071	-0.014	0.002	-0.001	0.003	-0.006
	(0.014)	(0.019)	(0.006)	(0.046)	(0.013)	(0.013)	(0.003)	(0.017)	(0.006)
Observations	716	707	727	631	691	722	720	762	762
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Clubmates' IVV Mean	-0.034	-0.059***	-0.014	-0.039*	-0.011*	-0.023**	0.009	-0.016	0.002
	(0.021)	(0.022)	(0.030)	(0.022)	(0.006)	(0.010)	(0.011)	(0.010)	(0.005)
Clubmates' IVV Variance	0.009	0.012***	0.006	0.010	0.001	0.004*	0.002	0.005*	-0.000
	(0.007)	(0.004)	(0.006)	(0.007)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)
Observations	771	771	771	771	771	771	771	771	771

***, **, * indicates that the effect of being treated in a MH (high or low) group compared to being treated in a HT group is significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors in parentheses at course-school level. Panel A present results on academic outcomes. Panel B exhibit effects on non-cognitive outcomes. Description of outcome variables is available in Appendix 1. All regressions include as controls: a second order polynomial of student's IVV and ciclo-school fixed effect (stratification level). In estimations for academic outcomes, absenteeism and bad behavior reports, I also include the corresponding imputed outcome at baseline and a dummy indicating a missing value at baseline.