Peer Effects on Violence. Experimental Evidence in El Salvador*

Lelys Dinarte†

April, 2018.

Abstract

This paper provides experimental evidence of the overall impact of a like-CBT after-school program on students' behavioral and academic outcomes, and of the role of having different levels of violent peers in that context. Participants were between 10-16 years old and enrolled in public schools in El Salvador. I find that the program reduced bad behavior reports by 0.17 standard deviations, school absenteeism by 23%, and increased school grades by 0.11-0.13 standard deviations. Changes in highly violent students mainly drove the results. Regarding group composition, results indicate that integrating students with different propensities for violence was better than segregating them, for both highly and less violent children. Particularly, the intervention can have unintended effects if highly violent students are segregated and treated separately from their less violent peers. Finally, I find positive social spillover effects for non-enrolled children exposed to treated students.

Keywords: Peer effects, Tracking, Violence, Cognitive Behavioral Therapy (CBT), After-School Programs, Education.

JEL Classification: I29, K42, Z13

---

*I am very grateful to my advisor Claudia Martinez for her unconditional guidance, support, encouragement, and very detailed suggestions. I also thank Jeanne Lafortune, Rodrigo Soares, Francisco Gallego, José Tessada, Nicolás Figueroa, Micaela Sviatschi, Tomás Rau and Pablo Egana. I appreciate the comments and suggestions of participants at the 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 IADB, 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 also appreciate 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. All errors and omissions are our own. This study was registered in the AEA RCT Registry with unique identifying number AEARCTR-"AEARCTR-0001602"

†PhD Candidate. Economics Department. Pontifical Catholic University of Chile. email: lidinarte@uc.cl.

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. Most ASP have been implemented in developed countries[2] but more recently have been started in developing countries.[3] Despite the increase in the number of programs implemented over the past years,[4] and the high incidence and economic costs of violence in the developing world,[5] the overall available non-experimental evidence of ASP's impact on social skills, crime, and violence is mixed and inconclusive (Taheri and Welsh, 2016).[6] 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).[7]

Additionally, there is no evidence of how peer effects may function within an ASP setting. Many papers have explored the effects of diversity and their mechanisms but in different contexts. For

[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] For instance, in the US: Becoming a Man, Quantum Opportunity Program, Higher Achievement Program, Citizen Schools, Pathways, Project NAFASI, After School Matters, Safe Haven, Challenging Horizons, and others. Kremer et al. (2015) provide a more detailed review of ASP in the US

[3] For example Boys and Girls clubs in Mexico, VUELA in Colombia, Rainbow After-School Clubs in Uganda, the Amani Girls Clubs in Liberia, and Glasswing Clubs in Central America–the intervention to be evaluated in this paper.

[4] 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).

[5] 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).

[6] 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.

[7] 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).

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 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).[8]

This paper aims to fill these two gaps in the literature. First, it provides experimental evidence designed to measure the effect of an ASP – related to Cognitive Behavioral Therapy (CBT) – on participants' violence and academic outcomes in the context of a developing and highly violent country.[9] Second, by creating an exogenous experimental variation in the propensity for violence of students' peers, the same experimental design captures potential peer effects that can help study the effectiveness of the intervention.[10] 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 this "like-CBT" intervention successfully improves participants' behavior and academic performance. Moreover, I provide evidence that mixing students with different levels of violence is a better implementation alternative for the ASP than segregating them in more and less violent groups.

The field experiment was performed in public schools located in violent communities in El Salvador. This context is key for two reasons. First, it is a lower-middle-income country defined as a victim of an "epidemic of violence" since 2009 (WHO, 2011).[11] Second, its high violence levels and homicides rates have significantly affected the educational system in the last years. The country has faced a 13% reduction in its education enrollment rate (MINED, 2015),[12] with approximately 18%

[8]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.

[9]To my knowledge, this is the first experimental evaluation of a like-CBT ASP's impact from a country with these characteristics.

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

[11]Between 2009-2012 the country's average homicide rate was 69 murders per 100,000 inhabitants (PNUD, 2013). Over half those killed during this time period were 15-34 years old; approximately 80% of the victims were male; 70% were executed using firearms; and nearly 40% took place in public spaces (FUNDAUNGO, 2013). In 2015 El Salvador was the country with the highest murder rate in the world, with a murder rate of 103 per 100,000 inhabitants –As a reference, the worldwide homicide rate is 6.2 per 100,000 inhabitants–(PNUD, 2013).

[12]In 2013 the primary and secondary net enrollment rates were 93.4% and 61.6% respectively, after a relevant

of students saying that they dropped out school due to delinquency.[13] Also, in the past 5 years, more children and adolescents have been victims of homicide than in the previous two decades in the country (EPCD, 2014).[14]

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.[15] 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 the overall impact of the ASP and to exploit that there was more demand for the program than spaces, I randomly assigned these students to treatment or control groups. To study the effects of group composition, treated students were randomly allocated to a group with a heterogeneous or homogeneous combination of peers, according to their initial propensity for violence.[16] 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 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 administrative records

drop in 2015, when primary and secondary net enrollment rates were only 86.2% and 37.9% respectively (MINED, 2015).

[13]This may be a lower bound because 28.6% of students abandoned school due to change of address, which since 2010 has been highly correlated with gang threats according to testimonies elicited by local newspapers (LPG, 2016).

[14]From 2005-2013 approximately 6,300 youth were homicide victims. In 2013, 458 adolescents were charged for extortion and 321 for aggravated homicide (CSJ, 2014), which are crimes mainly related to gangs (PNC, 2014).

[15]As I explain in detail later there are two samples in this study. The first one is *enrolled* students, those who decided to participate in the ASP and then were randomly assigned to treatment or control groups (1056 students). The second sample of *non-enrolled* includes students who were not registered for the ASP but are in the same schools and classrooms as treated children (1364 children).

[16]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).

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.[17]

I find that this less intensive intervention works in the context of a developing and highly violent country, and that its short-term effects are similar – in magnitudes and signs – to those of middle intensive interventions in the U.S. (Durlak et al., 2010; Cook et al., 2010). For example, my estimations indicate that students assigned to treatment have better attitudes towards school and reduce their school absenteeism by 23%. Moreover, I find a reduction in misbehavior at school and violence in both students' and teachers' reports. A plausible explanation for these effects is that the ASP is modifying the psychological factors that give rise to those attitudes and violent behaviors such as stress or automatic responses. In Dinarte and Egana (2017), they provide evidence that the program is certainly reducing participants' overreaction to external stimuli or increasing their emotional resilience.

In line with the evidence that emotional and behavioral skills promote and indirectly influence cognitive development (Cook et al., 2011; Cunha and Heckman, 2008), I also find that the ASP successfully increases participants' academic achievement. On average, after seven months of intervention, grades were 0.11-0.13 standard deviations higher for treated students. The intervention also reduces the probability of failing any of the three core courses – a proxy of school repetition – by 2.8 points.[18]

Overall, these effects are consistent with the expected results from *learning* and *protection* services that can be delivered by a like-CBT ASP. Specifically, this intervention can provide an innovative learning structure for students, affecting their disposition towards school and learning. Additionally, the program can promote some students' skills, such as resilience, and control over automatic responses and bad behavior. Finally, the ASP could provide protection from unsafe neighborhoods, reducing the time children may spend with delinquent peers. Unfortunately, the experimental design does not allow me to disentangle these mechanisms, and I can only provide suggestive evidence that the learning channel is more likely to be driving all the effects.

I then turn to study peer effects in this context. First, the ASP also has indirect short-term effects on *non-enrolled* children. Exploiting the exogenous share of treated students within each classroom, I find positive spillovers effects from the exposure of non-enrolled students to a higher

[17] 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. These results are analyzed in a companion paper Dinarte and Egana (2017).

[18] A novel result is that the effects on academic outcomes and absenteeism are greater for the most vulnerable students, which in this setting are those with a higher propensity for violence. This result is consistent with the evidence that the probability of being engaged in criminal or violent activities after school time for these students is greater. Then, keeping them under supervision for a couple of hours and teaching life skills can generate this larger effect.

proportion of treated classmates on both academic and violence outcomes. Thus, the direct results previously described seem to be lower bounds of the total effect of the intervention. Further analysis of heterogeneous spillover effects by intensity and proximity to treated classmates indicates that: (i) the greater the exposure of non-enrolled children to their treated classmates, the higher the spillovers; and (ii) the spillover effects are greater if there is an *intermediate proximity* regarding misbehavior between treated and non-enrolled students within classrooms. This last result indicates that diversity can play an important role enhancing this positive externalities.

In the second analysis of group composition, I compare students assigned to homogeneous or heterogeneous groups using the direct variation on peers' propensity for violence in the experiment design. 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).[19] 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.

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 last 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 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 is related to a wide literature that aims to measure ASP's effects on academic

[19] 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).

outcomes and violence (Gottfredson et al., 2004; Goldschmidt et al., 2007; Hirsch et al., 2011; Taheri and Welsh, 2016). As mentioned before, even when this topic has been extensively analyzed, there are still some gaps. First, the literature has focused on the effects of these interventions in developed countries, mainly in the United States, a context that may have limited applicability for education systems in low- and middle-income countries. Thus one contribution to this literature is providing evidence of the effect of this intervention in a developing and highly violent country, where these programs can be more relevant.[20]

Second, the paper is also related to a recent and novel literature that studies the effects of CBT[21] on youths' and adults' crime and violence patterns. The seminal papers in this literature are those of Heller et al. (2017) in Chicago and Blattman et al. (2015) in Liberia. The main difference of my paper with these studies is that I am testing a hybrid structure of CBT plus ludic ASP-activities.[22] This mixed structure may be more effective in the context of Salvadorean schools for at least two reasons. First, a full CBT program may be hard to implement if the target group consists of children and adolescents, or if enrollment and participation in the program is not mandatory. Second, an only CBT intervention can have a more significant impact in contexts where there aren't any gangs or other forms of organized crime since it works better against disorganized and impulsive violence (Blattman et al., 2015).

The research design also 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.[23] 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

[20]Additionally, most of the ASP literature measures heterogeneous effects only by initial academic attainment, gender, or household income (Marshall et al., 1997; Durlak et al., 2010), without considering variables that may affect this kind of interventions, such as violence. In this sense, the novelty of my results is that the ASP in this particular context generates a differential impact according to participants' violence levels, most positively impacting the most vulnerable children's misbehavior and attitudes.

[21]CBT is a therapeutic approach that can be used to treat harmful beliefs and behaviors, making people aware of these patterns and trying to disrupt them through a "learning by doing process" (Blattman et al., 2015).

[22]The program I analyze is more similar to the third intervention in Heller et al. (2017) that included CBT approach and additional activities like sports and dancing among others.

[23]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)

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).[24] 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 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.

[24] 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.

2. Intervention, Experimental Design, and Data

2.1 Intervention

A. Glasswing's After-School Clubs (ASP)

The NGO Glasswing International implemented the ASP as part of its program *Community Schools*, which, since 2013, has taken place in 95 schools in Central America through 560 clubs, benefiting approximately 20,000 children between 8-15 years old. 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, 2012a).[25]

The NGO offers four categories of clubs in the ASP by education level (*ciclos*): Leadership, Art and Culture, Sports and Science.[26] 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. 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.[27] Each session is divided into two sections: social skills development and club's curriculum. The first section is common to all participants and includes some activities related to CBT. Specifically, it tries to make people 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. For example, if the topic is conflict management, the students participate in a role-play, where the instructor asks students to provide alternatives to get a ball from a club-mate. Some of them suggest to forcibly retrieving it either by hitting the ball or the club-mate. Then the tutor discusses other alternatives like negotiation or simply asking for the ball. The implementation of this section was uniform across schools.

The second part of the session includes the implementation of ludic activities related to each club category. Its objective is to motivate students to participate in the intervention and increase program attendance. For instance, in a science club session, if the topic is volcanos, they perform an experiment of a volcano eruption. In a Art and Culture category, children develop some artistic activities, such as dancing, painting or building handcrafts.

[25] The NGO's main activity is the provision of technical advice to private companies on social investment, and formulating and executing strategic plans for social projects.

[26] In the Science category are the discovery clubs where students do scientific experiments. In the Art and Culture category are the Glee and Art clubs. The first group includes dancing and singing and the second includes activities for developing children's fine motor skills and creativity. Finally, Leadership clubs are for those who want to develop social and leadership skills.

[27] 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.

This combination of CBT and ludic activities is another innovation from this program, compared to other full-CBT interventions, such as those evaluated by Heller et al. (2017) and Blattman et al. (2015). As explained before, this mixed approach is more appropriate in this setting given the target group's ages and the type of violence they face at their contexts.

The ASP is organized by a school coordinator who verifies the participants' attendance and drop-out rates, manages club materials, and assigns volunteers as tutors. These tutors have no formal training in social work or psychology and, unlike those from the program Becoming a Man in Chicago, they do not necessarily have similar backgrounds as the participants.[28]

To my knowledge, there are only two impact assessments (qualitative and non-experimental) reports on this ASP, showing improved primary life skills such as self-perception, self-esteem, and social skills (Glasswing International, 2012b).[29]

B. Recruitment and enrollment process

During 2016, the NGO offered and implemented the program in 5 public schools in El Salvador. Using data from the 2015 Educational Census of El Salvador, I find that they are similar to the underlying population of public educational centers in El Salvador.[30]

Out of a total of 2,420 children from the 5 schools, I recruited and enrolled 1056 students between 10-16 years of age. The age range is relevant because that's when they are more likely to be enrolled or recruited by gangs. This group of enrolled children was constituted by children interested in participating at the program and the study. Any child was allowed to self-enroll, the only requirement was to bring a signed parent's authorization.[31]

[28] There are three categories of volunteers: community volunteers are tutors living in the community who stand out for their leadership skills; corporate volunteers are part of a particular firm that has a social project with Glasswing; and independent volunteers, who are usually college students, doing social work. The NGO assessed these volunteers, and even when they did not follow a pure random allocation procedure, there is still balance in the observable characteristics of the tutors such as gender, age, and category.

[29] To estimate the effect of the intervention, this study implemented focus groups to collect student information. To sum up, authors find positive effects of the program on students' optimism and team work. The students also reported being more tolerant of others, a reduction in their interaction with bad peers, and an improvement in the overall classroom environment. Particularly, some students find that clubs reinforce their academic experience in a more fun way (Glasswing International, 2012b).

[30] Tests for differences between participant and non-participant schools are shown in the Table A1 in the appendix section. Both groups of schools are similar on schools characteristics such as location area, violence level, number of students and additional revenues. Similarly, in terms of programs, facilities and equipment, participant and non-participant schools are similar on most on these benefits, except in the share of schools with a breakfast program or access to internet: treated schools are more likely to have both benefits.

[31] It is important to highlight that there are two samples in this study. The first one, that I call sample of "enrolled" children, consists of the 1,056 students who applied to participate in the program, and then were assigned to treatments or control groups. The second sample of "non-enrolled" students consists of 1,364 children which were not interested in taking part in the ASP. Using available administrative data for both groups, I compare enrolled and non-enrolled children's characteristics and I find that there are no differences among the two groups. These results are presented in table A2 in the appendix section. In that sense, the individual enrollment decision is driven by other variables or preferences that are not included in the existing administrative data.

During the registration process, enrolled students fill out a registration form that 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.[32]

The timeline of the study is shown in Figure 1.

[**Insert Figure 1 here**]

2.2 Experimental Design

The experimental design allows me to simultaneously measure the impact of the intervention and study how group composition, according to a predicted violence level at baseline changes the effectiveness of the intervention.

A. Propensity for Violence Index (IVV) estimation

To assign enrolled students to each group, the first requirement was to measure their propensity for violence. However, at the registration phase was not possible to directly ask about this because we could not guarantee that this personal information would be kept confidential during the study.[33] 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),[34] I estimated the likelihood of having committed a violent act V_f as a function of a wide range of covariates:

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

[32]The original clubs number is not definitive, it depends on the number of participants interested in each option. For instance, if 30 students have chosen Discovery Club as their first preference, the NGO would open two clubs of 15 participants each. However, if only two students have ranked Glee as their most preferred club, there won't be a Glee club, and those two students are assigned to their second or third alternative. On average, and for methodological reasons, club sizes are between 13-15 participants. As will be explained later, there is balance in all club categories between both treatments.

[33]For example, either the local authorities or gangs organizations may force me 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.

[34]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.

where D_f is a vector of violence determinants of student f in the FUSADES dataset.[35] 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).[36]

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 more violent than children (Rodríguez-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.[37]

Then, exploiting the availability of these variables in the registration forms of enrolled students, 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 it is important to emphasize. First, since the variables included in the estimation are related to students' exposure to violence 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.[38] 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 can provide some evidence that it is clearly the best proxy of propensity for violence I could get given 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),[39] this sort of crime and violence models estimated from

[35] This database includes a great number of variables measuring crime and violence and their determinants. Descriptive statistics and comparison of means (p-values) between the two samples can be found in table A3 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.

[36] Some relevant papers that find evidence that these variables are determinants of crime and violence are: for gender, Bertrand and Pan (2013); Rodríguez-Planas (2012); for age, Rodríguez-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).

[37] In table A4 in the appendix section, I summarize the results of the estimated coefficients.

[38] As a robustness check, in table A5 in the appendix, I used misbehavior reports as the classification variable for high and low propensity for violence. I obtain 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 A5.

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

existing data have a high predictive power.[40] For instance, the correlation between the predicted IVV and misbehavior at school is positive and statistically significant at 1%.[41] Additionally, 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%.[42]

B. Treatments

After estimating the IVV, enrolled children were randomly assigned to three groups within each stratum: control (C, 25%), heterogeneous (HT, 25%), and homogeneous (HM, 50%) groups. Then, students in homogeneous groups 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. The randomization process is shown in Figure 2.

[**Insert Figure 2 here**]

It is important to point that as the assignment of enrolled students was done at the stratum level, the share of treated children from each course within each education level –after controlling by the share of enrolled children– was exogenous.

Treatments are described below:

1. *Heterogeneous (HT):* Registered and randomly selected students are assigned to take part in a club with a heterogeneous peer 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.

[40] 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.

[41] An additional concern is that this index is 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. In Table A6 in the appendix I present these estimations, using different standardizations of academic grades and behavior reports.

[42] The estimation strategy and main results are presented in table A7 in the appendix.

4. *Control:* This group of students were not selected to participate in the clubs during the 2016 academic year.[43]

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 the interactions between participants instead other channels such as of curriculum adaptation.

2.3 Data

Given the contents and structure of the intervention, it can directly affect non-cognitive outcomes, such as children's violence and misbehavior at school. It also may have some indirect effects on academic outcomes, since changes in violence and behavior at school could affect the learning process. Considering this, I collected data of these two categories of outcomes.[44]

During the registration phase, after the first three months of the school year and before the intervention, students provided personal and family information, as I mentioned before. I also collected schools' records of math, reading, and science grades; behavior reports,[45] and absenteeism data from both enrolled and non-enrolled children.

Follow-up data on non-cognitive outcomes were collected only from enrolled participants in school facilities at the end of October 2016, after all clubs have completely implemented their curricula.[46] Most surveys were self-administered, with assistance from staff trained in the survey methodology.

The follow-up survey included questions to measure the intervention's impact on general topics, such as students' attitudes, violence and crime, exposure to risky spaces, and educational or personal expectations. Specifically, to measure attitudes towards school and approval of a friend's criminal behavior, I used items from the Communities That Care® Youth Survey. Delinquency and violence measures were calculated using the Self-Reported Delinquency Scale (SRD). To quantify exposure to violence or crime, I used the nationwide El Salvador Youth Survey (ESYS) developed by Webb et al. (2016). It includes questions related to children's and adolescents' risk and protective factors

[43]More specifically, 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, guaranteing their participation in the ASP in 2017.

[44]Appendix 1 presents a detailed description of all the outcome variables used in this paper.

[45]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. More details on these reports are in the Appendix 1.

[46]Students took the survey in classrooms especially set up for this purpose. Each survey took approximately 45-60 minutes. Schools' teachers agreed to cover the material taught during that time with the participants.

in three domains: family, school, and community. These instruments were previously validated in at risk youth population in El Salvador by Webb et al. (2016). Finally, I included questions about educational, migration, and labor expectations. The final implemented instrument is available upon request.[47]

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.

As shown in the appendix section, the average matching rate of administrative data of enrolled children was 94% at baseline, and 97% at follow up. All the 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%.[48] To account for this difference, I include in all specifications for the academic outcomes, the imputed grade for the missing observations at baseline and a missing value indicator. Additionally, the average matching rate of administrative data of non-enrolled students was 85% at baseline and 98% at follow up.

The attrition rate was 8% on average,[49] and for the HM and HT groups, it was 9% and 6% 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.

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

[47]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. These results are analyzed in a companion paper Dinarte and Egana (2017).

[48]These results are shown in table A8 in the appendix section.

[49]I defined attrition as the absence of initially enrolled students during the implementation of the follow-up survey.

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 A9.

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.[50] In this vein, I find that differences between T and C 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). Considering the large number of hypothesis tested, these differences are acceptable. However, I account for the difference in propensity for violence controlling for the percentile of the predicted IVV in all estimations.

[50] Appendix Table A10 shows adjusted p-values for multiple hypothesis testing of means of all variables exhibited in Table 1, following Sankoh et al. (1997).

Additionally, in specifications for the academic outcomes, 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's IVV, also expressed in most of its determinants. This design meets this requirement. For example, 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.[51]

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 A5 in the Appendix. First, the standard deviation of the HT group's IVV is greater than those of the HM subgroups. Second, the average IVV of the HT group 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, the hypotheses are not rejected (p-values of 0.62 for the HT-HM comparison, 0.89 for the HT-C comparison, and 0.68 for the HM-C comparison). The similarity among distributions can be verified also in Figure 3. The second characteristic is that the distributions of the HT, HM-High, and HM-Low must differ. As Figure 4 illustrates, there are differences in the distributions of the three groups. Particularly, using the two-sample Kolmogorov-Smirnov test, I reject the hypothesis of equality of each comparison of pairs of distribution functions at 1%.

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

The last desired characteristic 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 strata, there is overlap in 67% of the sample, as shown in Figure 5. Therefore, there is still variation between IVV distribution functions of the HM subgroups at baseline that I can exploit.

[51] 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.

[**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 stratum. This design also fulfills this condition. Figure 6 shows the median of the predicted IVV of student's club mates as a function of her own IVV and the expected jump at the fiftieth percentile. Moreover, when estimating a RD-robust regression using only this homogeneous subsample, I find that students assigned to the HM-High group are enrolled with peers with a mean IVV 0.8 points greater, statistically significant at 5%.[52]

[**Insert Figure 6 here**]

3. Empirical Framework

In this section, I describe the empirical strategy used to measure ASP's effects on students' behavior, violence, and academic outcomes, and to assess the heterogeneity of the intervention by individual violence levels. Additionally, I study group composition effects and how this heterogeneity interacts with children' initial propensity for violence.

3.1 Measuring the overall ASP's impact

A. Intent-to-treat Effects of ASP Participation

To measure the ITT effects of ASP on non-cognitive and academic outcomes, I use the random variation from the experimental design and estimate the following equation:

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

where y_{ij} is the outcome of interest, measured at follow-up, of the student i in school and education level j. T_{ij} is a dummy indicating that the student was randomly offered participation in the ASP, and S_j are strata dummies. X_{ij} is a vector of control variables, including a second order polynomial of student's IVV percentile. For the academic outcomes regressions, I also included standardized grades at baseline (including imputed values) and a missing baseline grades indicator as controls. Due to the possible bias in the estimation of the IVV, standard errors are adjusted using a cluster

[52] 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.

bootstrapped at the course-school level (Treiman, 2009). In this result, θ_1 captures the short term ITT effect9 of being assigned to participate in an ASP compared to being randomly allocated to a control group.

An additional robustness check of the accuracy of the predicted IVV as a proxy for misbehavior, I estimate specification (1), but instead of controlling by a second order polynomial of students' IVV percentile, I control by a similar polynomial specification of the student's percentile in the misbehavior distribution function.

B. Heterogeneity of the Intervention by Baseline Violence

To study heterogeneous treatment effects by initial level of predicted violence level, I include in equation (1) an interaction between T_{ij} and a binary indicator IVV_high_{ij}. This dummy indicates that student i's IVV percentile at baseline is greater than the median at the group (C, HM, and HT) and stratum level. Specifically, I estimate:

$$y_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times IVV\_high_{ij} + \theta_3 IVV\_high_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{2}$$

where θ_2 indicates the marginal impact of the intervention between treated students with high and low levels of propensity for violence. The rest of variables are defined as in specification (1).

Then, exploiting the lack of correlation between IVV and baseline school grades, I also explore heterogeneous effects by initial academic attainment on the outcomes of interest. This estimation strategy is summarized in Appendix 2.

Finally, as previous studies have found (Durlak et al., 2010), it may be expected that this ASP impacts differently to boys and girls. However, since the predicted IVV includes gender as a determinant, the difference of the effects among boys and girls may be caused either by sex alone or by the combination of all determinants included in the IVV estimation. To account for this, I use an alternative specification to show that the differences in the effects I find in this section are driven mostly by students' propensity for violence. A detailed description of the equation and estimations is presented in Appendix 3.

3.2 Peer Effects

In this subsection, I estimate three measures of peer effects. First, I present the identification strategy to estimate the effects of being exposed to treated classmates on outcomes of non-enrolled children. Second, 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. Finally, 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. A comparison of the last two sets of group composition measures will clarify if the outcome is affected only by the average peer characteristics, or if there is an interaction between a student's characteristics and that of her peers.

A. Effects on non-enrolled children: Spillovers

Besides ASP direct effects, spillovers from treated students on their non-treated classmates can occur through at least two ways: First, if treated children are less disruptive during classes, this can improve the learning process for all. Second, the interaction between treated and non-treated students can allow the last group to imitate or learn some skills from the first one. If any of these situations occurs, estimations from the specification (1) may be lower-bounds of the ASP total impact due to the presence of spillovers from the program.

Recalling that (i) the assignment to treatment was done at the *ciclo*-level and (ii) each level includes three courses, then the share of enrolled children allocated to participate in the ASP at each course n –the share of treated students Sh_n– was quasi-exogenous. Considering this, I can follow Carrell et al. (2013) to measure ASP's spillover effects on non-enrolled students m.

However, a possible concern is that non-enrolled participants may have influenced the enrollment decision, thus indirectly affecting the share of classmates assigned to treatment Sh_n. To address this concern, I can include as a control the share of all enrolled students –treated and control groups– from each course, E_n. The final specification will be the following:

$$y_{mn} = \gamma_0 + \gamma_1 Sh_n + \gamma_2 X_{mn} + E_n + \epsilon_{mn} \tag{3}$$

where y_{mn} is the academic or misbehavior outcome of interest. X_{mn} is a vector of individual controls, including grades at the baseline and a missing grades indicator.[53]

Further analysis of the structure and characteristics of these spillover effects, such as optimal combination of treated with high and low violence level, intensity of exposure and proximity on misbehavior within classrooms effects are presented in Appendix 4.

[53]As I show in appendix table A2, differences on academic outcomes and bad behavior reports at baseline between enrolled and non-enrolled students are not statistically different from zero. These evidence indicate that the two groups were similar regarding academic performance and how they behave at school before the intervention, strengthening the argument that the effects on non-enrolled children are more likely caused by spillover effects.

B. 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:

$$y_{ij} = \theta_0 + \theta_1 Hom_{ij} + \theta_2 X_{ij} + S_j + \epsilon_{ij} \tag{4}$$

where y_{ij}, S_j and X_{ij} are as defined before, and Hom_{ij} is a dummy that indicates whether student i in school level j is assigned to the HM treatment. θ_1 can be interpreted as the effect on student i of receiving an offer to participate in the like-CBT ASP with a homogeneous composition of violent peers, compared to the effects of the same offer but with more diversely violent peers.

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:

$$Y_{ij} = \theta_0 + \theta_1 HomH_{ij} + \theta_2 HomL_{ij} + \theta_3 IVV\_high_{ij} + I\bar{V}V_j + \theta_4 X_{ij} + \epsilon_{ij} \tag{5}$$

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.

Specification (5) allows to compare both treatments within each half of the IVV distribution. 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 5. Finally, following Duflo et al. (2011), I present an analysis of the average group composition effects using linear-in-means and variance specifications. These equations and their identification assumptions are described in Appendix 6.

C. Effects of tracking on the marginal student

Results of equations (5) and (6) 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. Even when 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 the student 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. The assumption required for the validity of this strategy is that nothing else changes discontinuously around the point of separation between the two groups, which holds true in this design. I estimate the following equation:

$$Y_{ij} = \lambda_0 + \lambda_1 HMH_{ij} + f(IVV_{ij}) + \lambda_2 S_j + \epsilon_{ij} \tag{6}$$

where $f(\text{IVV}_{ij})$ is a flexible second order polynomial of the percentile of the individual's IVV within each stratum, and $\text{HMH}_{ij} = 1$ 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 strata.

4. Results

In this section I present reduced form estimates of the ASP's impact on students' grades, violence, bad behavior at school and positive attitudes towards school and learning. I also present heterogeneous effects of the ASP by students' initial propensity for violence. In the second section, I describe group composition effects of the ASP on the outcomes of interest. First, I show the results of spillovers on non-enrolled students. Then, I present the results of average group composition effects and the impacts of tracking on marginal students.

4.1 Measuring the overall ASP's impact

A. Intent-to-treat Effects of ASP Participation

Table 2 shows results of equation (1). I split them into the two sets of outcomes: positive attitudes towards school, violence, misbehavior at school (Panel A), and academic outcomes (Panel B).

First, in columns (1) - (4) in Panel A, I present the like-CBT ASP's effects on students' pro-learning attitudes from both their self-reports and from administrative data. Compared to students in the control group, ASP participants report having better attitudes towards school by 0.17 standard deviations and spending 16% more time (20.4 minutes approximately) each day doing their homework. Moreover, 7.9% report that they pay more attention during classes, compared to the control group. This improvement in attitudes is also confirmed using administrative data: treated students are absent 1.6 days fewer than students in the control group. This implies a reduction of 23% on school absenteeism. These effects shed light that the like-CBT ASP directly affects students' positive attitudes towards school as the program may allow them to be involved in a different and potentially more interesting learning approach, or to be exposed to a new category of role models - their tutors - along with their teachers.

Then I estimate the ITT effect on misbehavior and violence-related outcomes, using measures from students' and teachers' reports. As we can see in columns (5) - (9) in Panel A, after seven months of intervention, students self-report having committed fewer delinquent actions and being less violent compared to self-reports of students in the control group (in magnitudes of 0.19 and 0.14 standard deviations respectively). Similar effects are found using teachers' reports. Students randomly assigned to participate in the ASP reduced both their bad behavior at school by 0.17 standard deviations and their probability of having a misbehavior report by 6.4 percentage points.[54] Although my two sets of measures are not completely comparable,[55] results from both are consistent with an increase in participants' willingness to reduce their bad behavior and tendencies to violence.

Combining these two groups of results, the effects I find from the intervention are similar to those previously identified in the literature. For example, Durlak et al. (2010) find a reduction in criminal behavior by 0.19-0.30 standard deviations in a meta analysis of ASP implemented in

[54] Differences in number of observations in non-cognitive outcomes is because of variation in the response rate for each outcome. I estimated these results using the smallest sample (836 observations), and there are not any differences in the results.

[55] These measures are different on who made the report and on the items and domains included. For example, misbehavior outcome considers actions at school and self-report of violence includes violent actions at school, home and community.

the U.S. Similarly, Heller et al. (2017) find that the program Becoming a Man (BAM) for youth in Chicago reduced violent-crime arrests, improved school engagement and increased graduation rates.

Despite that ASP activities are not directly related to academic outcomes, there is a positive correlation between academic results and social skills. For example, as students acquire life skills and learn better behaviors, they may be less disruptive during their classes, facilitating the learning process. In this sense, it might be expected that their grades improve.

ITT results of the intervention on academic outcomes are presented in Panel B in Table 2 (columns 1 - 4). Grades have been standardized at the course-school level. At the end of the academic year, the ASP has a positive effect on math and science grades, with a magnitude of 0.11 and 0.13 standard deviations respectively (intensive margin).

Using the data on grades, I can also assess the ASP short-term effect on the extensive margin, i.e. on the probability of passing each course. Exploiting the fact that the minimum grade to pass a course in El Salvador is 5, I create a dummy that indicates if the children's score is above that value for each course. I find that the intervention increases the probability of passing reading and science courses and reduces the probability of failing any of the three courses - a proxy for grade repetition –by 2.8 percentage points (Panel B column 8). This last effect represents a reduction of 42% on course repetition compared to the control group mean.[56]

Since this is a low-intermediate intensity ASP, the effects on academic outcomes are in-between those results from highly- and low-intensive programs. Durlak et al. (2010) find that ASP in the U.S. have an average positive impact of 0.12 standard deviations on school grades. However, Shulruf (2010) concludes that extra curricular activities with a duration of three hours per session, five times per week –i.e. high-intensive programs– have an average effect of 0.30 standard deviations on math and science grades. Finally, Cook et al. (2015) find effects between 0.19 - 0.31 standard deviations on math scores in an intervention that provides individualized academic instruction.

[**Insert Table 2 here**]

B. Heterogeneity of the Intervention by baseline Violence

Table 3 summarizes the estimated effects from specification (2) for attitudes towards school and learning, violence, and bad behavior at school (Panel A), and academic outcomes (Panel B). Coef-

[56] Alternatively, I estimate the effects of the ASP on the relevant outcomes controlling by a second order polynomial of students' bad behavior at school using teachers reports. The estimated effects using this alternative specification are similar in magnitude and sign than those presented in Table 2. This result strengthens the argument that the predicted propensity for violence indeed measures students behavior. Results are presented in Appendix Table A11

ficients in row [i] in each panel show the ASP's effects on low-violence treated students compared to low-violence children in the control group, coefficients in row [ii] show the differences in effects between high-violence treated students and similar children in the control group, and coefficients in row [iii] point to the difference in effects between high- and low-violence treated students. Row [iv] indicates p-values of the test for difference in effects between high- and low-violent treated students.

Estimations from the comparison between high-violent students in treatment and control groups allow me to conclude that the ASP successfully modified behaviors and academic outcomes of students with a greater propensity for violence, as shown in Panel A row [ii]. Additionally, as we can see from row [iii] column (4), high-violence participants are two times less likely to be absent at school after the intervention than the low-violence treated students. There are no statistical differences in the rest of attitudes towards learning between both groups of treated students.

Moreover, estimations of differences in violence and misbehavior show that both groups are reducing these conducts by a similar magnitude, except in the intensive margin of bad behavior at the school –reported by teachers– where the reduction is greater for low-violence students.

On academic outcomes, as we can see in panel B, results on the intensive margin of school grades indicate that high-violent students are also driving these academic results. Row [iii] shows that differences between high- and low-violence treated students' grades are between 0.19 - 0.24 standard deviations. Although there are no statistically significant differences on the extensive margin between both groups, a notable result from row [ii] column (9) in panel B is that the total effect on the probability of failing at least one course (a proxy of course repetition) for high-violence treated students is a reduction by 4.8 points, which accounts for approximately 70% of average course repetition difference from the C group.[57]

To sum up, the second novel result from this experiment is that the most vulnerable students seem to be the main winners from this like-CBT ASP, showing higher effects on both attitudes and school grades compared to the outcomes of both highly violent students in the control group and low violent treated students.

[**Insert Table 3 here**]

As I do not find statistically significant correlation between students' school grades and their propensity for violence at baseline, it indicates that more violent students from my sample are not necessarily those with lower academic attainment. Taking advantage of this result and to contribute to the existing evidence of ASP's heterogeneous effects by initial academic performance (Marshall

[57]Further heterogeneous effects by initial level of violence are depicted in Appendix Figure A1. The graph shows the estimations of a local polynomial fit of standardized end line score grades by predicted IVV for T and C groups. There are statistical differences between both groups for students in the 55th to 95th percentiles in the IVV distribution.

et al., 1997; Durlak et al., 2010), I also estimate differences in the effects by students' school grades at baseline. I find that the ASP is not only benefiting students with a greater propensity for violence, but also those who have lower academic grades before the intervention. Particularly, low-performers treated children at baseline face a greater effect on school absenteeism and on the extensive margin of academic grades after the intervention, compared to initially high-performers treated children.[58]

4.2 Peer Effects

The second part of this paper provides evidence of peer effects in the context of an ASP. I can draw three main conclusions from this section. First, the intervention has positive spillover effects on non-enrolled children's academic and misbehavior outcomes. Second, mixing students by their initial propensity for violence generates better average effects than segregating them. Finally, tracking has detrimental effects for the marginal students.

A. Effects on non-enrolled children: Spillovers

Using the sample of non-enrolled children, I estimate specification (3) to measure how being exposed to a higher share of treated classmates affects academic and behavioral outcomes of the non-enrolled students. This model controls by the proportion of enrolled children and includes school fixed effects. Since I rely only on administrative data of non-enrolled students, spillover results are limited to school grades and behavior reports.

Table 4 shows the results of spillovers estimates. I find evidence that the interaction of students with a greater share of ASP participants generates positive effects on their reading, math and science grades, and reduces their bad behavior at school. Estimations indicate that adding 2 treated students in a classroom of 26 (almost a 1 standard deviation increase in treated students) increases academic achievement on up to 0.062 standard deviations, (for example, on math grades: $2/26 \times 0.008 = 0.062$), and reduces bad behavior reports by 0.084 standard deviations ($2/26 \times 0.011$).[59]

[58] These results are available in table A12 in the appendix section. Similarly, table A13 shows estimations of heterogeneous effects by gender. On non-cognitive outcomes, I find greater effects on absenteeism for boys compared to girls (a reduction of 2.1 days). Additionally, the effects on the extensive margin of school grades are greater for treated boys on math grades and score, compared to treated girls. However, as explained before, in table A14 I provide evidence of how these heterogeneous effects are mostly caused by differences in propensity for violence at baseline –except on absenteeism–, ruling out the only-gender heterogeneous effect.

[59] After adding individual controls, estimated coefficients are similar in magnitude and statistical significance, except for bad behavior reports which are no longer statistically significant due to the increase in the standard errors. Despite this, the sign of the effect of is negative, indicating that a higher share of treated classmates reduces the effect on bad behavior reports, providing additional evidence of reduction in the formation of violence networks or disruption during classes.

These results have similar signs to some evidence previously found in the literature. For example, Carrel and Hoekstra (2010) use the share of classmates coming from troubled families –i.e. share of children exposed to domestic violence– to measure its effect on grades and classroom misbehavior. They find that making 5% of a class troubled students –1 standard deviation– significantly decreases reading and math test scores by 0.69 percentile points, and increases misbehavior in the classroom by 0.09 more infractions.

To sum up, the spillover results shown in Table 4 give rise to two findings. First, these positive spillovers on non-enrolled students indicate that the ASP's direct effects previously described are the lower bounds of the total effect of the intervention in the context of these highly violent schools. Second, combining the results of this paper with those from Carrel and Hoekstra (2010), I can conclude that it is possible to outweigh the negative effects of misbehaving children, by incorporating students with positive behavior to their classrooms. This novel result particularly contributes to the evidence of optimal class design (Krueger, 2003; Lazear, 2001).

[**Insert Table 4 here**]

It is also noteworthy to study additional characteristics of these spillover effects. For example, there may exist a combination of high- and low-violence treated children that maximized the aggregated effect. Additionally, the intensity of these spillovers may change due to the exposure level –in terms of time length– of non-enrolled children to treated participants.[60] Finally, spillover effects may be different by misbehavior closeness of non-enrolled with treated students. Since the ASP effects are different by initial propensity for violence of treated participants, there may also exist heterogeneity in the spillover effects by initial non-enrolled students' misbehavior at school. I provide evidence addressing these additional questions in Appendix 4, and present the implications of the results in the discussion section.

Summing up the results, first I test for differences by initial propensity for violence of treated children on non-enrolled classmates' outcomes. I find that even though the differences in the effects are not statistically different from zero, due to an increase in the standard errors, estimations indicate that spillover effects on academic outcomes may be driven by the share of treated students with low level of violence. However, the reduction in misbehavior at school may be caused mainly by the share of treated students with high propensity for violence.[61]

Second, regarding intensity of exposure to treated students, I find that spillovers on non-enrolled student's academic outcomes are lead only by the share of treated students from her own classroom.

[60]For example, non-enrolled children usually spend more time with students of their own classroom compared to treated students from other classrooms.

[61]These results are summarized in table A15 in the Appendix.

Nevertheless, a novel result here is that the effect on bad behavior at school is caused by both the share of treated from their own classroom and from one course lower.[62]

Finally, in terms of closeness on misbehavior of non-enrolled children with treated students, I find that the effects are greater for students whose bad behavior at school is intermediately away –between 1 and two standard deviations– from the average misbehavior of the share of treated students within her classroom. Particularly, the effects of this medium closeness is greater on bad behavior reports. Thus, this result highlights that only certain level of similarity to treated students can have positive spillover effects.[63] This last result indicate that diversity can play an important role enhancing this positive externalities.

B. Group composition average effect

Table 5 shows estimations of group composition using specifications (4) and (5). First, from the comparison between HT and HM groups drawn from the equation (4), 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 5). They also increase their probability of having a bad behavior report at school by 5.5 percentage points (column 9, Panel A, Table 5). Finally, I do not find statistical differences between both treatments in the rest of non-cognitive and 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).

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 (5). These results are also reported in Table 5. 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

[62] I present the differences on spillovers by intensity of exposure in table A16 in the appendix section.

[63] These heterogeneous estimations by proximity to misbehavior of treated classmates are presented in table A17.

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 to be 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 0.09 percentage points, as shown in Panel A, column (9). Thus, selecting and treating together only high violence students for these programs can generate an unintended effect from the intervention. This result 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 attitudes towards school and learning and violence. Moreover, as shown in Panel B of Table 5, 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 5 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.[64]

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

[64] In appendix 5, I present details of the specifications and results. Main estimations are summarized in table A18. 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).

assumption is that after controlling for strata fixed effects, the variance and mean IVV of peer stems entirely from the random assignment.[65]

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.

C. 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 (6). 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 the effects of tracking on marginal students are summarized in Table 6. 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 intervention for the marginal participants.

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 (6) but restricting the sample to the eight students around the IVV median within each stratum. Results are also reported in Table 6. 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

[65] Details of the estimation and summary of results are presented in appendix 6 and in table A19.

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 6 here**]

5. Discussion

Despite the intensity and high costs of youth violence (WHO, 2015) and the recent increase in the number of ASP implemented in low- and middle-income countries, there is little rigorous evidence that measures the impact of these interventions on either academic or non-cognitive outcomes.

Most of the existing experimental evidence from youth interventions for developed countries supports the argument that the involvement in programs oriented to reduce participants' risky conducts, generates positive effects on both academic performance and behaviors (Heller et al., 2017; Blattman et al., 2015; Kremer et al., 2015; Durlak et al., 2010; Cook et al., 2015). A strand of this literature has focused on measuring heterogeneous effects by gender, academic attainment, and income. However, if interventions aim to reduce violent behaviors within schools and to enhance life skills, this strategy does not help to explain differential impact by violence or whether the program is indirectly affecting other children with whom the treated students interact.[66]

Furthermore, it is also important to study how ASP's group composition can improve the results. The existing evidence on this matter is mixed[67] 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).

[66] In many developing countries, 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).

[67] 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).

To my knowledge, this paper provides the first experimental evaluation of the direct impact and 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 in the program generates direct and indirect effects on academic, violence and behavioral outcomes, changes students' efforts at school, and if the group composition is relevant to affect these key results.

Overall effects of the ASP and related interventions

The first remarkable result is that this low-intensive ASP is effective in the context of a developing and highly violent country. I find that the random assignment to the intervention successfully modified children's attitudes towards school and learning, and their misbehavior at school.[68] Additionally, the magnitude of the effects of this low-intensive intervention on non-cognitive and academic outcomes are between those found by Durlak et al. (2010); Cook et al. (2015)[69] from average ASP, and those found by Heller et al. (2017); Blattman et al. (2015) from high- and middle-intensive CBT intervention implemented in the U.S. and Liberia.

It is important to highlight that the frame and structure of some activities implemented during the ASP are closer to those from a Cognitive Behavioral Therapy (CBT) intervention.[70] For this reason the results of the ASP are lower that those found in CBT studies, but in the same direction. This recent literature on CBT includes studies of the therapy effects on youths' and adults' crime and violence patterns, such as the studies of Heller et al. (2017) in Chicago and Blattman et al. (2015) in Liberia. Overall, these papers find that CBT is a cost-effective approach to reduce criminal behavior among high-risk young men in cities across diverse contexts. Particularly, effects on BAM participants were a decrease on their arrests per students by 12% and on the number of violent

[68]The existence of such impacts from the ASP is not surprising to the extent that the neuroscience literature suggests that it is possible to affect non-cognitive skills during adolescence. Existing literature suggests that non-cognitive investments during adolescence can have a positive impact on the development of non-cognitive skills, such as behavior. In addition, studies suggests that these programs are more effective among students who are still enrolled in secondary schools (Heckman and Kautz, 2012; Cunha et al., 2010).

[69]Specifically, Durlak et al. (2010) finds an increase by 0.12 and 0.14 standard deviations on school grades and school bonding respectively in a meta-analysis of ASP in the U.S. Meanwhile, Cook et al. (2015) reports on a school-based intervention that provides disadvantaged youth with intensive individualized academic instruction, and find an increase of math grades by 0.19-0.31 standard deviations and on expected graduation rates by 46%.

[70]For example, similar to "The Fist" activity in the Becoming a Man program (BAM), the ASP included sessions in which students were asked how they would retrieve a ball from a clubmate. Some of them automatically reply that they would hit either the ball or the classmate. Then the tutor discuss with them additional ways of getting the ball, such as negotiation or just asking for it.

crime arrests by 20%. Additionally, they improved by 0.10-0.19 standard deviations on their school engagement index of enrollment, attendance and GPA, and where more likely to graduate from school.

However, as I briefly discussed before, CBT may not have full applicability in a context like public schools in El Salvador. First, it may be more effective in a setting where there are no gangs or other forms of organized crime, since it works better against disorganized and impulsive violence (Blattman et al., 2015). Second, participation in gangs in Central America starts during childhood or adolescence, around ten years of age (Rivera, 2013). Thus, a full CBT structure may be unattractive at this age. In that sense, combining it with additional activities, such as experiments, artistic performances, sports, and others, may be more attractive to guarantee children's and adolescents' attendance. Thus, my results contribute to this strand of literature providing evidence of alternative or "mixed" interventions that can work in this highly violent contexts, with greater effects on highly violent children and adolescents.

Heterogeneous effects: No children left behind!

An additional novel result is that participants with a greater propensity for violence are more likely to increase their academic achievement and reduce their school absenteeism, compared to the less violent group. These results are compatible with existing evidence that these interventions usually have a greater effect for the most disadvantaged children (Marshall et al., 1997; Durlak et al., 2010).

Despite the greater improvement on those outcomes of highly violent students, I find that although both treated groups reduced their bad behavior scores relative to the control group, the reduction on misbehavior at school was actually greater for the less violent group of treated students compared to the group of high violence.

Students' violence trends might help to explain this second heterogeneous result. First, it is possible that bad behavior is harder to modify, particularly for those used to acting in that way. From a neurophysiological perspective, Lewis et al. (1979) find that more violent individuals may have greater brain-damage, therefore reducing their tendency to violence can be harder. A second interpretation is related to Akerlof and Kranton (2002)'s ideal student theory. They state that teachers and coaches award or disapprove students according to a "school's ideal student". In this sense, teachers may have already tagged students by their initial violence level and, despite observing a reduction in their bad behavior, they report that this decrease is greater for those that already been seen as the ideal low-violence student. In any case, the take-away conclusion from heterogeneous effects estimations is that the intervention is benefiting both tails of the propensity for violence distribution function, on different sets of outcomes.

How being less violent makes me good at math?

Results of the paper also finds a positive effect on both the intensive and extensive margin of students' academic outcomes. This raises the question - how an intervention that only teaches life skills indirectly affect grades? There can be at least three channels.

First, the ASP can can modify students' classroom misconduct, reducing disruptions that affect their own or classmates' learning. For example, correlational evidence indicates that children who participate in ASP tend to exhibit better behavior in school and therefore have higher academic achievement (Scott-Little et al., 2002; Durlak et al., 2010). Moreover, Mahoney et al. (2010) and Cassel et al. (2000) posit that extracurricular involvement helps to dissuade students from becoming involved with delinquency and crime.

Second, a large body of theoretical and empirical evidence in economics and psychology (Borghans et al., 2008; Cunha and Heckman, 2008; Dodge et al., 1990; Heckman et al., 2006; Moffitt et al., 2011) shows that cognitive skills or school outcomes are defined by non-cognitive skills, such as future orientation and attitudes towards school. Finally, since there are clubs with school content in this setting, the intervention can be reinforcing academic curricula, thus improving directly students' grades. Nevertheless, as I will discuss later, this last channel operates conditionally on group composition.

Learning versus protection mechanisms

There are at least two mechanisms through this ASP may have changed behavioral outcomes. First, students may have learned social skills and conflict management directly from the clubs' curricula, through their interaction with other children, or from both. I call this the *learning mechanism*. Second, children may have reduced their violent behaviors because ASP protects them during a time when they might be left alone and exposed to external risks (Gottfredson et al., 2004; Jacob and Lefgren, 2003; Newman et al., 2000). This will be the *protection mechanism*. Although this experimental design does not allow me to perfectly disentangle between both mechanisms, I find suggestive evidence that students are indeed learning social skills, and therefore the first mechanism is more likely to be driving the effects.

First, I exploit the availability of baseline data on adult supervision after school hours to test for differences between both mechanisms.[71] The assumption is that treated students who reported being without adult supervision after school receive both effects from the intervention, and that the effects for students who are with an adult after school time are caused only by the learning

[71] As only 5% of the sample reported being without adult supervision, I face power issues. Even though, signs of the estimations provide suggestive evidence that allows me to disentangle both mechanisms.

mechanism. Then, I included in specification (1) an interaction between the treatment variable and a dummy of being alone after school hours.

Estimations are exhibited in Table A20 in the appendix. Row [i] presents the learning mechanism effects alone, row [ii] includes both effects and row [iii] shows the protection effect alone. Estimated coefficients indicate that most of the effects are mainly related to the learning mechanism, on both cognitive and non-cognitive outcomes. An interesting result drawn from row [iii] is that only protecting children may have an unintended effect compared to teaching them life skills. As we can see in columns (6) and (7), the net effect of protection alone increases violence index and approval of peer's antisocial behavior. To sum up, these results shed light on that the main mechanism of the intervention is social skills learning.[72]

As an additional attempt to study the protection mechanism, I use students' self-report of exposure to crimes, either as victims or as witnesses, and their awareness of risk within their communities or at home.[73] The assumption here is that if the protection channel is operating, they may perceive changes on their vulnerability to risky environments. I do not find statistically significant effects on most of those outcomes, except an increase on children's awareness of risk at their communities, which can be also interpreted as an skill developed through the learning channel. These results are available upon request.[74]

Better together. Group composition effects

To my knowledge, this is the first paper that provides experimental evidence of group composition regarding violence within an ASP setting. 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

[72] I also find that effects are greater when I estimate them using only the sample of students who participated in at least one session. These results are exhibit in table A21 in the appendix and shed light on how the effective participation strengthens the impact from both mechanisms.

[73] These last estimations are only an approximation, and we should be cautious in their interpretation because the question asked about crimes witnessed or experienced after school hours, which is usually from 12.30 - 2 pm. However most crimes in El Salvador occur after 5 pm.

[74] To provide further evidence to disentangle these channels, I am trying to collect information on completion of social skills curricula. The assumption here is that clubs that completed their curricula have both protection and learning channels, and for those who only partially completed the curricula, it only has a protective effect but differences in skills learning, at least from curricula.

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. (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.[75] 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

[75] 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)

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.[76]

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

[76] 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 A23 in the appendix.

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 direct effects and spillovers 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. I contribute to the literature by showing that even these low-intensive interventions have important effects on cognitive and non-cognitive outcomes, particularly for the most vulnerable students, those with a higher initial level of violence or with lower initial academic achievement. Then by exploiting three exogenous variations yielded by the experimental design, I provide evidence that the ASP's group composition has differential impact on both types of outcomes. Specifically, students assigned to more diverse groups regarding initial violence level have better results, while treating high violent students alone generates unintended, adverse effects.

In the first part of the paper, I find positive ITT effects from the intervention on most of the academic outcomes; treated participants have higher math and science grades and a greater probability of passing reading, compared to the control group. Concerning non-cognitive results, I test two groups of outcomes that could work as plausible mechanisms behind the effects on grades. First, due to the intervention, students might have better attitudes towards school and learning and therefore increase their grades. Second, participants might be less violent and have better behavior in schools. I find that treated students have better attitudes towards school, report spending more time on homework and are less likely to be absent by 1,6 days. Regarding violence, when comparing between treated and control groups, the former self-reports a greater reduction in violent and criminal activities and aversion to attitudes to antisocial behaviors. Comparing these results with teachers' behavior reports, I find similar results; treated students reduce their probability of having reports of bad behavior.

The effects of group composition are assessed in the second part of the paper. 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*.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059.

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.

Carrel, S. 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.

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.

Cassel, R. N., Chow, P., Demoulin, D. F., and Reiger, R. C. (2000). Extracurricular involvement in high school produces honesty and fair play needed to prevent delinquency and crime. *Education*, 121(2):247.

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., Dodge, K., Farkas, G., Fryer Jr, R. G., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., and Steinberg, L. (2015). Not Too Late: Improving Academic Outcomes for Disadvantaged Youth. *Institute for Policy Research Northwestern University.*

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). Emotional regulation and after-school programs in highly violent communities: Neuro-physiological evidence from el salvador.

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.

Dodge, K. A., Bates, J. E., and Pettit, G. S. (1990). Mechanisms in tie cycle of violence. *Science*, 250:1678–1683.

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 (2012a). Estudio Interno sobre impacto de programa extracurricular de seis centros escolares. Technical report, Fundación Crisálida. Glasswing International.

Glasswing International (2012b). Mi Escuela, Mi espacio. un programa educativo extracurricular para el desarrollo integral de la niñez, la adolescencia y la juventud de El Salvador. Technical report, Fundación Crisálida. 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.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482.

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

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.

Jacob, B. A. and Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. *The American Economic Review*, 93(5):1560–1577.

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.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113(485):F34–F63.

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.

Lewis, D. O., Shanok, S. S., Pincus, J. H., and Glaser, G. H. (1979). Violent juvenile delinquents: Psychiatric, neurological, psychological, and abuse factors. *Journal of the American Academy of Child Psychiatry*, 18(2):307–319.

LPG (2016). Amenazas de pandillas es la quinta causa de deserción 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., Parente, M. E., and Zigler, E. F. (2010). After-School Program participation and children's development. *Handbook of Research on Schools, Schooling, and Human Development*, pages 379–397.

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 Educación de El Salvador.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7):2693–2698.

Newman, S. A., Fox, J. A., Flynn, E. A., and Christeson, W. (2000). America's After-School Choice: The Prime Time for Juvenile Crime, or Youth Enrichment and Achievement. Technical report, Fight Crime: Invest in Kids, Washington, DC.

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*.

Rivera, J. (2013). Las maras. el fenómeno criminal del siglo xxi.

Rodríguez-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.

Sankoh, A. J., Huque, M. F., and Dubey, S. D. (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in medicine*, 16(22):2529–2542.

Scott-Little, C., Hamann, M. S., and Jurs, S. G. (2002). Evaluations of After-School Programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23(4):387–419.

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.

Shulruf, B. (2010). Do extra-curricular activities in schools improve educational outcomes? A critical review and meta-analysis of the literature. *International Review of Education*, 56(5-6):591–612.

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.

Webb, V., Nuno, L., and Katz, C. (2016). Influence of risk and protective factors on school aged youth involvement with gangs, guns, and delinquency: Findings from the El Salvador Youth Survey. Technical report, Center for Violence Prevention and Community Safety. Arizona State University.

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 ^	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	6.76^	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. OVERALL EFFECTS OF THE ASP

	(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
Any treatment	0.172***	0.336***	0.079***	-1.593***	-0.198**	-0.143***	-0.104**	-0.172***	0.064***
	(0.062)	(0.098)	(0.022)	(0.272)	(0.067)	(0.045)	(0.024)	(0.052)	(0.021)
Observations	948	935	962	836	916	956	962	1010	1010
Mean control group	-0.13	2.12	0.59	7.16	0.00	0.00	0.174	7.18	0.72
SD - control group	1.49	1.89	0.49	9.20	0.973	0.971	0.379	1.24	0.45
MDE T = C	0.108	0.109	0.108	0.173	0.108	0.108	0.131	0.135	0.123
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Any treatment	0.015	0.108***	0.131***	0.057	0.037***	0.019	0.029**	0.026	-0.028***
	(0.039)	(0.037)	(0.047)	(0.038)	(0.010)	(0.017)	(0.014)	(0.015)	(0.009)
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023
Mean control group	6.47	6.23	6.37	6.37	0.865	0.873	0.884	0.873	0.067
SD - control group	1.75	1.76	1.66	1.63	0.342	0.334	0.319	0.334	0.251
MDE T = C	0.096	0.092	0.100	0.096	0.088	0.103	0.104	0.097	0.108

***, **, * significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A is effects on non-cognitive outcomes. Positive attitudes, time spent on homework, paying attention in class, criminal and violent action indexes, and approval of antisocial behavior were estimated using collected self-reported data at follow-up. Positive attitudes towards school is an index estimated using PCA with a mean of 0 and standard deviation 1.4. Absenteeism is the number of days student missed school between April-October of the 2016 academic year. It was obtained from schools' administrative data. Criminal actions is an standardized sum of self reports of crimes. Violent actions is the standardized sum of other violent acts such as fighting at school, damage of municipal property, fighting with siblings, and others. Bad behavior reports are administrative school reports, using the control group at the school-grade level. Panel B 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. All regressions include as controls: a second order polynomial of student's IVV, and *ciclo-school* fixed effect (stratification level). Additionally, 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. Differences in number of non-cognitive outcome observations is because of variation in the response rate for each outcome.

TABLE 3. HETEROGENEOUS TREATMENT EFFECTS BY IVV

	(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
[i] Effect on low-violence students	0.097	0.428***	0.078**	-0.851	-0.229*	-0.095	-0.070**	-0.282***	-0.084***
	(0.095)	(0.153)	(0.039)	(0.678)	(0.133)	(0.077)	(0.029)	(0.081)	(0.031)
[ii] Effect on high-violence students	0.255**	0.244*	0.080**	-2.329***	-0.167**	-0.194**	-0.143***	-0.061	-0.043
[iii] Difference between high- and low-violence	0.158	-0.184	0.002	-1.478*	0.062	-0.099	-0.073	0.221**	0.041
	(0.187)	(0.225)	(0.064)	(1.050)	(0.176)	(0.139)	(0.045)	(0.098)	(0.046)
[iv] *p-value:* high-violence effect = low-violence effect	0.397	0.412	0.980	0.059	0.727	0.476	0.105	0.025	0.375
Observations	948	935	962	836	916	956	962	1010	1010
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
[i] Effect on low-violence students	-0.039	-0.009	0.031	-0.036	0.033**	-0.007	0.015	0.027	-0.009
	(0.053)	(0.055)	(0.058)	(0.050)	(0.015)	(0.023)	(0.020)	(0.023)	(0.015)
[ii] Effect on high-violence students	0.072	0.228***	0.234***	0.153***	0.040**	0.044*	0.045*	0.026	-0.048***
[iii] Difference between high- and low-violence	0.111	0.237***	0.203***	0.189***	0.008	0.050	0.030	-0.001	-0.039
	(0.073)	(0.071)	(0.075)	(0.064)	(0.026)	(0.032)	(0.038)	(0.030)	(0.026)
[iv] *p-value*: high-violence effect = low-violence effect	0.126	0.001	0.007	0.003	0.774	0.116	0.429	0.972	0.135
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023

***, **, * significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A shows effects on non-cognitive outcomes. Panel B presents results on academic outcomes. Description of outcome variables is available in Appendix 1. Row [iii] is the sum of the coefficients of the effect on low-violence treated students [i], and the coefficient of the interaction term in Row [ii]. All regressions include as controls: a second order polynomial of student's IVV, and *ciclo-school* fixed effect (stratification level). Additionally, 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. ASP SPILLOVERS. EFFECTS ON NON-ENROLLED STUDENTS.

	(1)	(2)	(3)	(4)	(5)
	Grades				Behavior reports (-)
	Reading	Math	Science	Score	
[i] Proportion of club participants within student's n classroom (coefficient)	0.007**	0.008***	0.006**	0.007***	-0.011*
	(0.003)	(0.003)	(0.003)	(0.002)	(0.006)
[ii] Spillover effect of adding 2/26 treated students (1 sd)	0.054	0.062	0.046	0.054	-0.085
Observations	1357	1358	1357	1356	1194
Mean of non-enrolled	6.78	6.47	6.54	6.60	7.63
sd of non-enrolled	1.92	1.86	1.92	1.59	1.64

***, **, * significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parenthesis. Outcome variables are standardized grades at school-grade level at follow-up. All regressions include as main control the share of enrolled students from each course. Individual controls include imputed grades in the course at baseline and a dummy indicating a missing value in the grade at baseline. Other individual controls are course and average course age. Row [i] indicates the coefficients of specification (3). Row [ii] indicates the average effect of adding 2 treated students in a classroom of 26 students (a standard deviation of treated students share) on non-enrolled academic grades and bad behavior reports. Description of outcome variables is available in Appendix 1.

TABLE 5. 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
ALL HOMOGENEOUS (Specific. 4)									
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)
BY HOMOGENEOUS SUBGROUPS (Specific. 5)									
Low Homog. 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 Homog. 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
MDE T = C	0.114	0.115	0.114	0.158	0.114	0.115	0.114	0.124	0.084
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
ALL HOMOGENEOUS (Specific. 4)									
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)
BY HOMOGENEOUS SUBGROUPS (Specific. 5)									
Low Homog. 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 Homog. 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

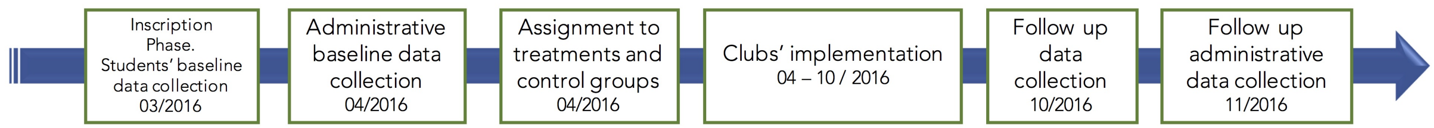
***, **, * indicates that the effect of being treated in a MH (high or low) group compared to being treated in a HT group is significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parenthesis. Panel A exhibits effects on non-cognitive outcomes. Panel 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 6. 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: 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.

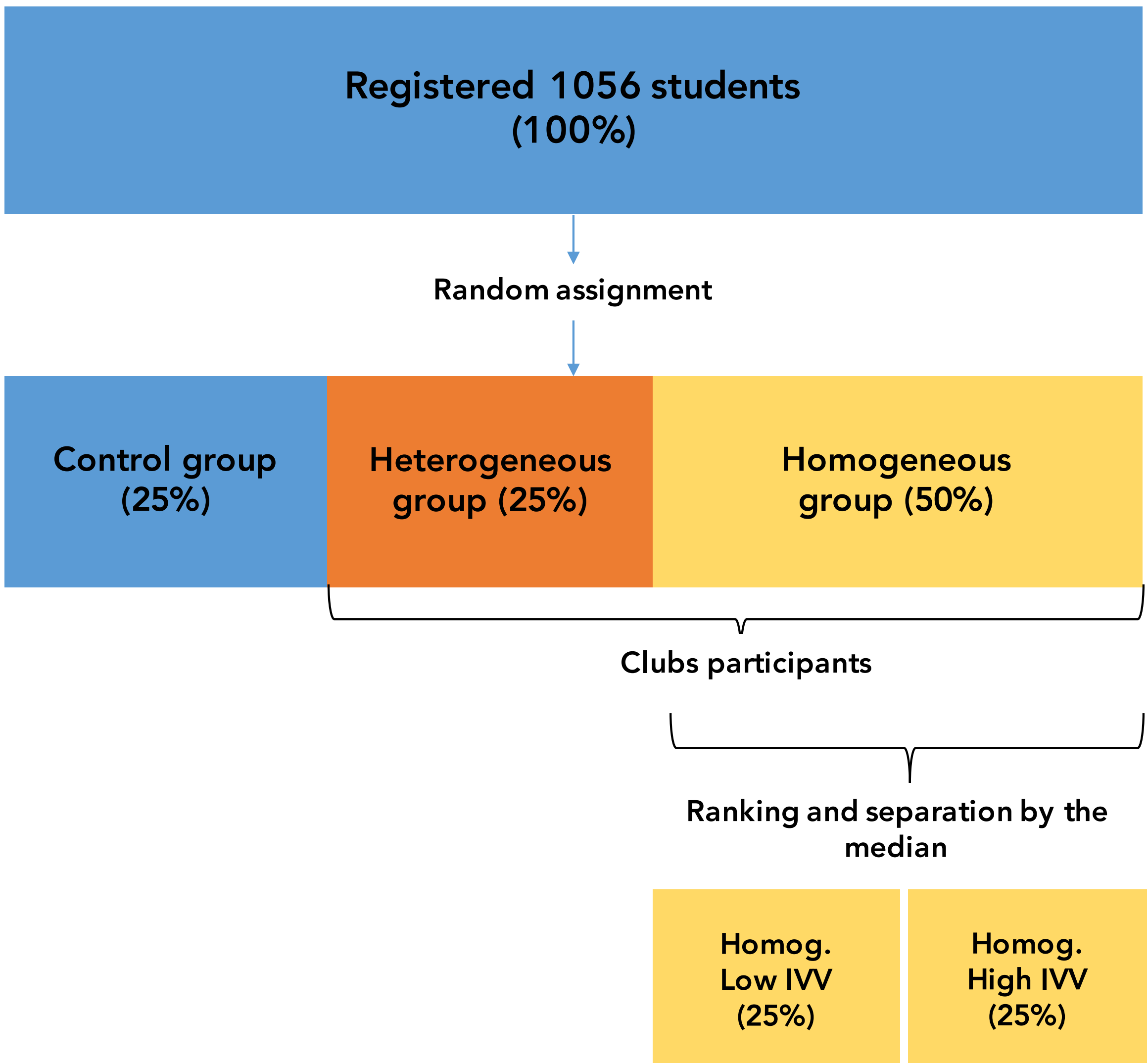
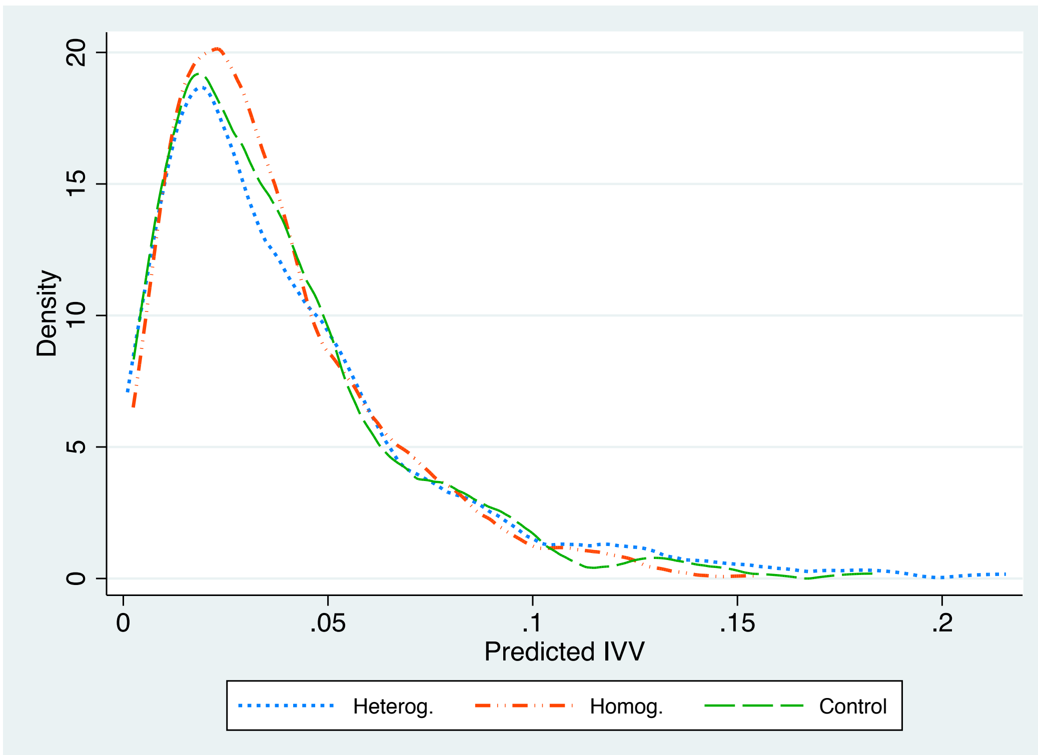
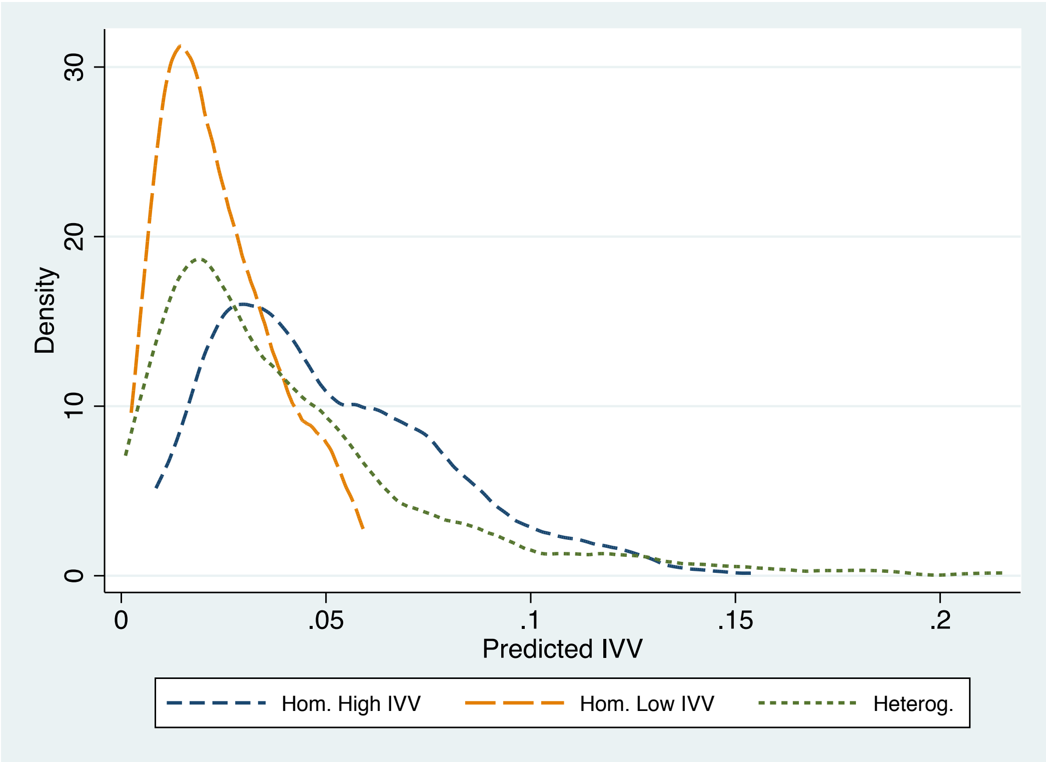


Figure 3. IVV Distribution Functions of Treatment and Control Groups.



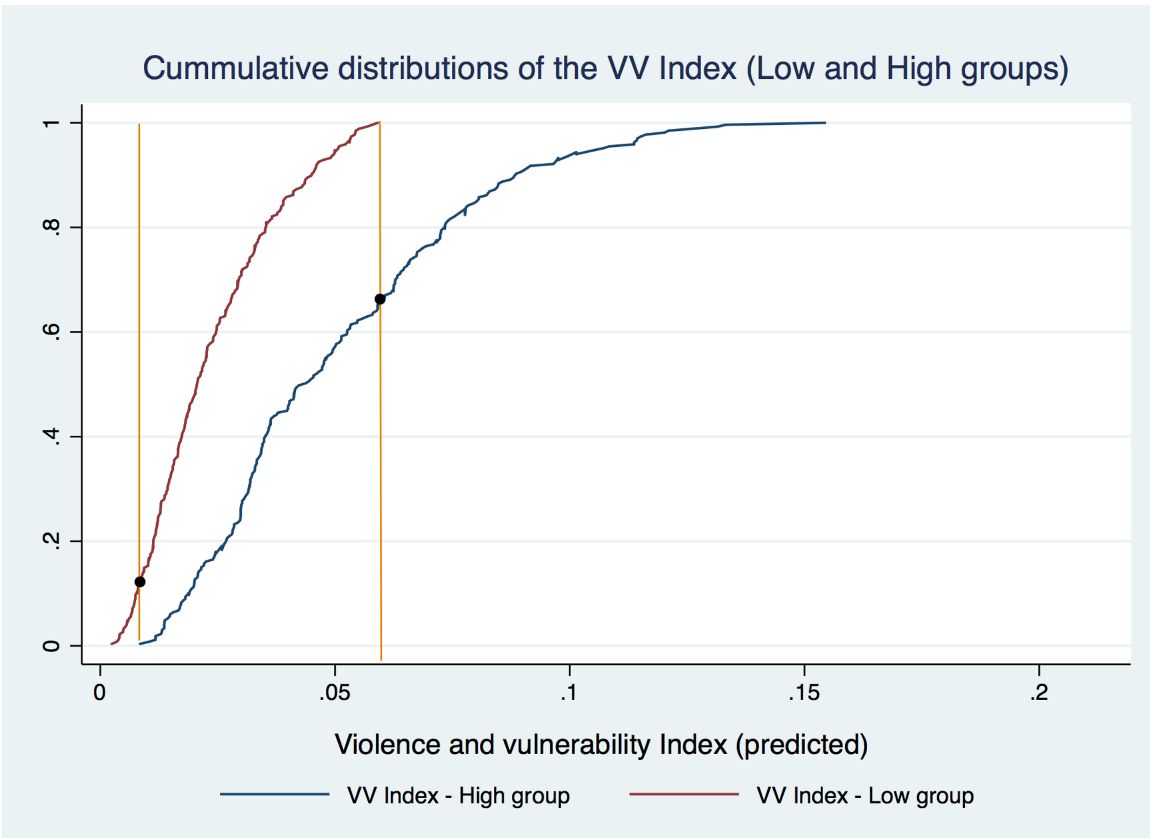
Predicted IVV distribution functions for the control and any treatment (homogeneous and heterogeneous) groups prior to treatment, for the whole study sample.

Figure 4. IVV Distribution Functions of Treated Groups.



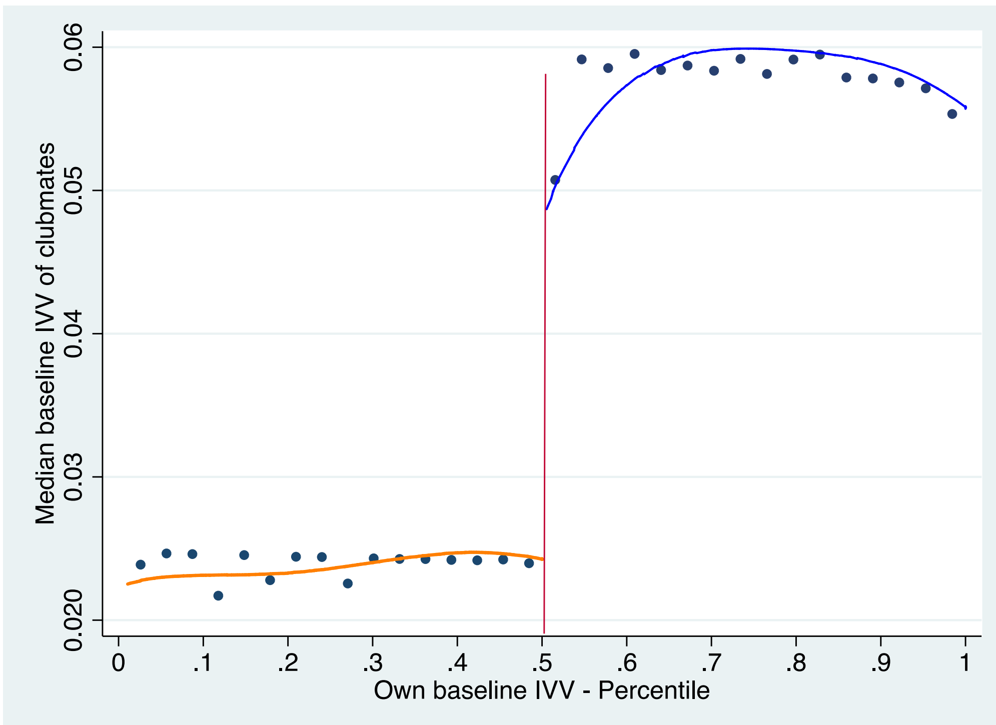
Predicted IVV distribution functions generated by the experimental design for the heterogeneous treatment group and each of the homogeneous subgroups (High and Low IVV) in the whole study sample

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.

Here is a discussion of the construction of the outcome variables used in the paper:

1. Positive attitudes towards school and learning is an index estimated using PCA with mean 0 and standard deviation 1.4. I used 5 items from the self-reported follow-up survey.

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

3. Pay attention in class was a self report from students. The question was: *During the last 3 months, did you pay attention during classes?*

4. Delinquent actions index is an standardized sum of self report crimes such as theft, mugging someone, etc.

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.

6. Approval of peers' antisocial behavior is a binary indicator that takes the value of 1 if students approve some peer behavior such as alcohol and drugs consumption, fighting, etc.

7. Absenteeism is the number of days the student was not at school between April-October of the 2016 academic year. Administrative data was provided by schools.

8. Drop-out is a binary indicator taking the value of 1 if the student has followed the formal school process to abandon school. The Ministry of Education in El Salvador requires students and their parents to show up to school and ask for student's documents to declare that she is no longer enrolled in that school.

9. 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, I used a reversed continuous scale to facilitate the interpretation and comparability to the self-reported measures of violence and crime.

Appendix 2. Heterogeneity of the ASP by baseline grades.

Since the intervention provides life skills training and promotes positive attitudes towards school and learning, according to the NGO's theory of change it may also improve children's academic attainment. As previous papers have shown (Durlak et al., 2010), it is plausible that the ASP may be affecting differently those students with low academic performance compared to the rest of their class.

The main concern in the estimation of heterogeneous effects by baseline academic performance under this experiment design is that the differences can be caused mostly by children's propensity for violence than by their initial academic attainment. However, this is addressed since the predicted IVV is not correlated with grades at the baseline (see Appendix Table A6).

Exploiting this lack of correlation between grades and estimated IVV in this sample, I can assess the heterogeneous effects by initial academic achievement. I include a dummy variable A_{ij}, which indicates whether child i was in the bottom half of the baseline score[1] distribution in her course, and an interaction between it and the treatment dummy. The resulting equation used to identify differential effects of the program by academic performance at baseline is the following:

$$y_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times A_{ij} + \theta_3 A_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{1}$$

The rest of variables are defined as before. Results are shown in table A10. As before, Panel A shows violence and attitudes outcomes and Panel B shows academic performance results. Row [i] in both panels shows the results for students with low academic performance before the intervention and row [ii] shows the results for students with a score higher than the median within her course.

I find that students with higher initial academic achievement reduce their absenteeism by 1.9 days more than students with high academic performance. There are no differences in the effects on the rest of behavioral outcomes for either group. Regarding academic outcomes, results indicate that the effects on the extensive margin are higher for those students in the bottom of the grade distribution, including a reduction in the probability to failing any of the three main courses.

Combining these results with the heterogeneous effects results by initial IVV presented before, I can conclude that the ASP is benefiting the most vulnerable children, which are those with either higher propensity for violence or lower academic performance.

[**Insert Table A10 here**]

[1] This score is an average of the grades achieved by the student in her three main courses: math, reading and science during the first quarter of the 2016 academic year, i.e. before the intervention.

Appendix 3. Gender vs. propensity for violence heterogeneous effects.

Previous studies have found that after-school programs usually impact differently to boys and girls (Durlak et al., 2010). They regularly identify this difference by incorporating an interaction between gender and the treatment dummy. However, in this study, it can not be done in that way since the estimation of the IVV includes sex as a determinant. Thus, the difference in the effects among boys and girls may be caused either by gender alone or by the combination of it and the rest of determinants included in the IVV estimation.

Under this naive approach, I would estimate the following equation:

$$y_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times G_{ij} + \theta_3 G_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{2}$$

where G_{ij} is a dummy that takes the value of 1 if the child is a boy. The coefficient of the interaction term would indicate the difference in the effects of the ASP between boys and girls. Results of this naive approach are presented in table A11 in the appendix. I find higher effects on absenteeism for boys compared to girls (a reduction of 2.1 days of absenteeism). Additionally, the impact on the extensive margin of school grades is more significant for treated boys on math and score, compared to treated girls.

[**Insert Table A11 here**]

As we can see from the previous results, most of the differences by gender are found on the same outcomes as the differences by initial propensity for violence. To verify which of the measures are generating the differences, I use the following alternative specification:

$$y_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times G_{ij} + \theta_3 T_{ij} \times IVV_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{3}$$

where θ_2 indicates the difference of the ASP effects by gender (boys versus girls) and θ_3 shows the difference of the impact by the propensity for violence (highly versus low violent children). In the control variables vector, I include gender, high-IVV dummy and a second order polynomial of students' percentile of initial IVV.

Appendix table A12 shows the results, separated in the two main panels. Rows [i] and [ii] show the estimations of θ_2 and θ_3 respectively. Results reinforce the previous conclusion that the heterogeneous effects on academic and non-cognitive outcomes reported in Table 3 are in fact driven by students' initial propensity for violence, except for absenteeism. Gender heterogeneous effects are found only on attitudes towards school and learning outcomes.

[**Insert Table A12 here**]

Appendix 4. Further analysis and evidence of spillovers.

In this Appendix, I present further evidence of spillovers' characteristics in the context of this ASP. First, in the primary analysis of the intervention impact, I find that students with a higher propensity for violence benefit more from the program. However, the results of group composition effects indicate that these gains of the high violent students are driven mainly because they are exposed to a diversity of peers regarding violence. Therefore, treating both groups of students maximies the overall results.

To test if this also holds on the spillovers estimations, I divide the share of treated students in groups of high- and low-propensity for violence. The estimation equation is the following:

$$y_{mn} = \gamma_0 + \gamma_1 ShH_n + \gamma_2 ShL_n + \gamma_3 X_{mn} + E_n + \epsilon_{mn} \tag{4}$$

where ShH_n and ShL_n are the share of treated students with high and low IVV at the classroom level, respectively and the rest of variables are defined as in specification (4).

Results are shown in table A13. I find that even though the differences in the effects after comparing shares of treated students with low and high level of violence are not statistically different from zero. However, from their signs we may think that that spillover effects on academic outcomes can be driven by the share of treated students with low level of violence. However, the reduction in misbehavior at school is caused mainly by the share of treated students with high propensity for violence.

[**Insert Table A13 here**]

The second analysis I implemented was to test if the intensity of these spillovers may change due to the level of exposure –in terms of time length– of non-enrolled children to treated participants. To measure intensity of exposure, I exploit the fact that non-enrolled children usually spend more time with students of their own classroom compared to treated students from other classrooms. To study this between-classrooms closeness, I estimate the following equation:

$$y_{mn} = \gamma_0 + \gamma_1 Sh_n + \gamma_2 Sh_{n-1} + \gamma_3 Sh_{n+1} + \gamma_4 X_{mn} + E_n + \epsilon_{mn} \tag{5}$$

where Sh_n is the share of treated children at own student's classroom n, and Sh_{n-1} and Sh_{n+1} are the share of treated students in the previous and next course, respectively. The rest of variables are defined as in specification (4).

As we can see in table A14, spillovers on non-enrolled students' academic outcomes are lead only by the share of treated students from her own classroom. Nevertheless, a novel result here is that the effect on bad behavior at school is caused by both the percentage of treated from their classroom and one course below. To understand better this last result is necessary a further analysis on the social interactions within schools, using sociograms, for example. However, from the results I can infer that most of the interaction seem to come from treated children with whom non-enrolled students spend relatively more time.

[**Insert Table A14 here**]

Finally, spillover effects may be different by misbehavior closeness of non-enrolled with treated students within the same classroom. Since the ASP effects are modified by the initial propensity for violence of treated participants, there may also exist heterogeneity in spillover effects by non-enrolled students' misbehavior at school before the intervention.

Since I rely only on administrative data of non-enrolled students –i.e. I do not have an IVV measure for them–, to test this within-classroom closeness I use misbehavior reports at school for all children. Then I created dummies indicating if each non-enrolled student is less than i standard deviations away from the average of her group. Finally I estimate the following specification:

$$y_{mn} = \gamma_0 + \gamma_1 Sh_n + \gamma_2 Sh_n \times C1_{mn} + \gamma_3 Sh_n \times C2_{mn} + \gamma_4 X_{mn} + E_n + \epsilon_{mn} \tag{6}$$

where Ci_{mn} are dummies indicating whether student m has a bad behavior level that is less than i standard deviations from the average behavior of treated children at her classroom m, with $i \in \{1, 2, +2\}$. The rest of variables are defined as before.

Results are presented in table A15. I find that the effects are more significant for students whose lousy behavior at school is between 1 and two standard deviations away from the mean of misbehavior of the share of treated students from her classroom. Notably, the effects of this intermediate closeness are more significant on bad behavior reports. Thus, this result highlights that only certain level of similarity to treated students can have positive spillover effects.

[**Insert Table A15 here**]

Appendix 5. 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:

$$Y_{ij} = \alpha_0 + \alpha_1 HT_{ij} + \alpha_2 HM_{ij} + \alpha_3 \sum_{k=1}^{4} HT_{ij} \times Qk_{ij} + \alpha_4 \sum_{k=1}^{4} HM_{ij} \times Qk_{ij} + \alpha_5 X_{ij} + S_j + \epsilon_{ij} \quad (7)$$

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 A16 here**]

Appendix 6. 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:

$$Y_{ij} = \gamma_0 + \gamma_1 \bar{x}_{-ij} + \gamma_2 var(x_{-ij}) + \gamma_3 S_j + \gamma_4 X_{ij} + \epsilon_{ij} \tag{8}$$

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 A17 here**]

TABLE A1. TESTS FOR DIFFERENCES BETWEEN PARTICIPANT AND NON-PARTICIPANT SCHOOLS

	(1)	(2)	(3)	(4)	(5)
Panel A. Schools characteristics					
	School is located in urban area	School is in a top ten most violent municip.	Enrollment (Number of students)	Share of Indigenous students	Additional school revenues
Participant school	0.125	0.000	130.41	-0.003	575.973
	(0.182)	(0.000)	(107.25)	(0.003)	(2,109.54)
Mean non-participant schools	0.245***	0.093	256.14***	0.041***	1,798.6***
	(0.000)	(0.000)	(0.104)	(0.000)	(2.054)
Panel B. Schools programs					
Does school has a	EITP Program	School kits program	*Vaso de leche* Program	Food Program	Psychological professional
Participant school	-0.109	-0.184	0.191***	0.034	0.148
	(0.070)	(0.134)	(0.043)	(0.024)	(0.115)
Mean non-participant schools	0.149***	0.979***	0.572***	0.983***	0.035***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel C. Schools facilities or equipment					
Does school has access to	Computer	Water	Electricity	Sanitation	Internet
Participant school	22.462	-0.072	0.003	0.172	0.416**
	(13.949)	(0.183)	(0.002)	(0.199)	(0.205)
Mean non-participant schools	9.024***	0.774***	0.976***	0.031***	0.217***
	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	5,134	5,134	5,134	5,134	5,134

Data source: El Salvador Educational Census (2015). For these estimations, I restricted the sample to public schools only and estimated the following specification: $y_{ij} = \alpha_0 + \alpha_1 P_{ij} + F_j + \epsilon_{ij}$, where y_{ij} is the characteristic of interest of school i in department j –geographic division–, α_0 is the mean of non-participant schools, P_{ij} is an indicator for participant schools, and F_{ij} are departments fixed effects. *Vaso de leche* corresponds to a breakfast program, and EITP is an acronym for *Escuela Inclusiva a Tiempo Pleno*

***, **, * indicates that coefficients are significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parentheses.

TABLE A2. BALANCE BETWEEN ENROLLED AND NON-ENROLLED STUDENTS

	(1)	(2)	(3)	(4)	(5)
	Grades				Behavior reports (-)
	Reading	Math	Science	Score	
Enrolled students	-0.106	-0.041	-0.051	-0.051	0.040
	(0.101)	(0.147)	(0.163)	(0.111)	(0.107)
Mean non-enrolled students	6.82	6.56	6.67	6.69	7.54
	(0.131)	(0.130)	(0.174)	(0.110)	(0.088)
Observations	2,415	2,415	2,415	2,415	2,334

The sample includes a total of 2,420 students from the 5 public participant schools. The estimated specification was the following: $y_{ij} = \alpha_0 + \alpha_1 E_{ij} + F_j + \epsilon_{ij}$, where y_{ij} is the non-standardized grades or misbehavior report of student i in the school-course j, α_0 is the mean of non-enrolled children, E_{ij} is an indicator of student's decision to participate in the ASP at baseline, i.e. if they and their parents signed a consent form, and F_{ij} are school-courses fixed effects. Outcomes include imputed missing data at baseline and a missing data indicator. This data was obtained from administrative schools' records. ***, **, * indicate that the estimation is significant at 1 %, 5 %, and 10 % respectively. Clustered standard errors at the course-school level are in parentheses.

TABLE A3. 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 A4. 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 A5. 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 A6. 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 A7. 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 A8. 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 A9: 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 A10. 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 A11. OVERALL EFFECTS OF THE ASP CONTROLING BY STUDENTS' BEHAVIOR AT SCHOOL

	(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**
Any treatment	0.174***	0.343***	0.080***	-1.622***	-0.205***	-0.147***	-0.104***	-0.239***	-0.086***
	(0.067)	(0.101)	(0.023)	(0.250)	(0.065)	(0.047)	(0.024)	(0.043)	(0.018)
Observations	1010	935	962	836	916	956	962	1010	1010
Mean control group	-0.13	2.12	0.59	7.16	0.00	0.00	0.174	7.18	0.72
SD - control group	1.49	1.89	0.49	9.20	0.973	0.971	0.379	1.24	0.45
MDE T = C	0.108	0.109	0.108	0.173	0.108	0.108	0.131	0.135	0.123
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				**Failing at least one course**
	Reading	**Math**	**Science**	**Score**	**Reading**	**Math**	**Science**	**Score**	
Any treatment	0.039	0.126***	0.156***	0.080**	0.041***	0.022	0.029**	0.031**	-0.030***
	(0.040)	(0.037)	(0.047)	(0.040)	(0.009)	(0.015)	(0.013)	(0.014)	(0.008)
Observations	1010	1010	1010	1010	1010	1010	1010	1010	1010
Mean control group	6.47	6.23	6.37	6.37	0.865	0.873	0.884	0.873	0.067
SD - control group	1.75	1.76	1.66	1.63	0.342	0.334	0.319	0.334	0.251
MDE T = C	0.096	0.092	0.100	0.096	0.088	0.103	0.104	0.097	0.108

****, **, * significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A is effects on non-cognitive outcomes. Panel B presents results on academic outcomes. All regressions include as controls: a second order polynomial of student's bad behavior at school using teachers reports -before the intervention- and *ciclo-school* fixed effect (stratification level). Additionally, 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. Differences in number of non-cognitive outcome observations is caused by the differences in the response rate for each outcome.

TABLE A12. HETEROGENEOUS TREATMENT EFFECTS BY INITIAL ACADEMIC ACHIEVEMENT

	(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
[*i*] Effect on low performance students	0.222	0.270	0.072*	-2.706***	-0.126	-0.137	-0.116***	-0.187***	-0.097***
	(0.151)	(0.167)	(0.040)	(0.616)	(0.103)	(0.094)	(0.030)	(0.046)	(0.027)
[*ii*] Effect on high performance students	0.157**	0.439***	0.091***	-0.898**	-0.285***	-0.166**	-0.095***	-0.235***	-0.059*
[*iii*] Difference between high and low performers	-0.065	0.170	0.020	1.808**	-0.160	-0.029	0.021	-0.048	0.038
	(0.199)	(0.206)	(0.058)	(0.910)	(0.145)	(0.146)	(0.042)	(0.099)	(0.046)
[*iv*] p-value: effect on Top half = effect on bottom half	0.746	0.410	0.737	0.047	0.270	0.842	0.622	0.627	0.412
Observations	948	935	962	833	916	956	962	1010	1010
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
[*i*] Effect on low performance students	0.154***	0.183***	0.235***	0.211***	0.098***	0.061***	0.089***	0.103***	-0.073***
	(0.043)	(0.037)	(0.052)	(0.037)	(0.014)	(0.021)	(0.026)	(0.029)	(0.018)
[*ii*] Effect on high performance students	0.076**	0.121***	0.184***	0.128***	0.002	-0.002	-0.003	-0.005	0.001
[*iii*] Difference between high and low performers	-0.077	-0.062	-0.050	-0.083	-0.096***	-0.064***	-0.092***	-0.108***	0.073***
	(0.058)	(0.055)	(0.066)	(0.051)	(0.019)	(0.024)	(0.028)	(0.034)	(0.019)
[*iv*] p-value: effect on Top half = effect on bottom half	0.182	0.259	0.449	0.106	0.000	0.009	0.001	0.001	0.000
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023

***, **, * 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. Panel B presents effects on non-cognitive outcomes. Description of the outcome variables is available in Appendix 1. Row Total effects on High Scores is the sum of the coefficients of any treatment dummy and the coefficient of the interaction term. All regressions include as controls: a second order polynomial of student's IVV, and ciclo-school fixed effect (stratification level). Additionally, 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 A13. HETEROGENEOUS TREATMENT EFFECTS BY GENDER.

	(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
[*i*] Any treatment	0.064	0.300**	0.031	-0.604	-0.255**	-0.101	-0.100***	-0.241***	-0.085***
	(0.094)	(0.138)	(0.034)	(0.576)	(0.112)	(0.064)	(0.029)	(0.087)	(0.029)
[*ii*] Boy x Any T	0.232*	0.076	0.096	-2.088**	0.122	-0.067	-0.006	0.141	0.044
	(0.128)	(0.229)	(0.062)	(0.985)	(0.168)	(0.105)	(0.038)	(0.132)	(0.061)
[*iii*] Total effects on Boys	0.296***	0.376**	0.127**	-2.692***	-0.133	-0.168**	-0.106***	-0.100	-0.041
Observations	948	935	962	836	916	956	962	1010	1010
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				**Failing at least**
	Reading	Math	Science	Score	Reading	Math	Science	Score	one course
[*i*] Any treatment	-0.031	0.010	0.097	-0.007	0.010	0.002	0.022	0.017	-0.010
	(0.055)	(0.061)	(0.064)	(0.055)	(0.015)	(0.023)	(0.023)	(0.020)	(0.016)
[*ii*] Boy x Any T	0.088	0.191**	0.066	0.126*	0.053**	0.032	0.015	0.016	-0.035
	(0.078)	(0.083)	(0.090)	(0.077)	(0.021)	(0.032)	(0.037)	(0.031)	(0.023)
[*iii*] Total effects on Boys	0.057	0.201***	0.163**	0.119**	0.063***	0.034	0.037	0.033	-0.045***
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023

***, **, * 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. Panel B presents effects on non-cognitive outcomes. Description of outcome variables is available in Appendix 1. Row Total effects on Boys is the sum of the coefficients of any treatment dummy and the coefficient of the interaction term. All regressions include as controls: a second order polynomial of student's IVV, and education level fixed effect (stratification level). Additionally, 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 the baseline.

TABLE A14. HETEROGENEOUS TREATMENT EFFECTS BY GENDER AND VIOLENCE.

	(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
[*i*] T x Boy	0.202	0.272	0.140**	-1.831**	0.128	-0.014	0.053	-0.024	-0.032
	(0.201)	(0.312)	(0.066)	(0.854)	(0.201)	(0.123)	(0.035)	(0.151)	(0.071)
[*ii*] T x High IVV	0.048	-0.339	-0.077	-0.449	-0.012	-0.092	-0.103**	-0.205*	-0.021
	(0.261)	(0.302)	(0.068)	(0.951)	(0.210)	(0.156)	(0.046)	(0.109)	(0.052)
Observations	948	935	962	836	916	956	956	1010	1010
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				**Failing at least**
	Reading	Math	Science	Score	Reading	Math	Science	Score	one course
[*i*] T x Boy	0.031	0.076	-0.078	0.024	0.071**	0.005	-0.004	0.024	-0.018
	(0.123)	(0.118)	(0.139)	(0.121)	(0.035)	(0.047)	(0.056)	(0.052)	(0.027)
[*ii*] T x High IVV	0.094	0.193*	0.245*	0.175	-0.032	0.047	0.032	-0.014	-0.029
	(0.118)	(0.104)	(0.126)	(0.108)	(0.039)	(0.047)	(0.056)	(0.051)	(0.031)
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023

***, **, * 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. Panel B presents effects on non-cognitive outcomes. Description of outcome variables is available in Appendix 1. Row Total effects on Boys is the sum of the coefficients of any treatment dummy and the coefficient of the interaction term. All regressions include as controls: a second order polynomial of student's IVV, and ciclo-school fixed effect (stratification level). Additionally, 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 the baseline.

TABLE A15. SPILLOVERS BY STUDENTS PROPENSITY FOR VIOLENCE

	(1)	(2)	(3)	(4)	(5)
	Grades				Bad behavior
	Reading	Math	Science	Score	reports (-)
[*i*] Proportion of treated students with high propensity for violence	0.004	0.007*	0.006	0.005	-0.014**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
[*ii*] Proportion of treated students with low propensity for violence	0.009***	0.008**	0.005*	0.008***	-0.010
	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)
[*iii*] p-value [i] = [ii]	0.219	0.980	0.948	0.594	0.667
Observations	1357	1357	1357	1357	1194

***, **, * significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parenthesis. Outcome variables are standarized grades at school-grade level at follow-up. All regressions include as main control the share of enrolled students from each course. Individual controls include imputed grades in the course at baseline and a dummy indicating a missing value in the grade at baseline. Row [i] indicates the effect of the share of treated students with high propensity for violence withing each classroom. Similarly, row [ii] indicates the effect of the proportion of treated students with lower propensity for violence. Row [iii] is the p-value of the hypothesis that the difference between both coefficients is statistically different from 0.

TABLE A16. RELATIVE SPILLOVERS EFFECTS

	(1)	(2)	(3)	(4)	(5)
	Grades				Bad behavior
	Reading	Math	Science	Score	reports (-)
[i] Proportion of treated students at classroom m (own classroom)	0.007**	0.007***	0.006**	0.007***	-0.009*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
[ii] Proportion of treated students at classroom $m-1$	-0.001	0.001	0.001	0.001	-0.005*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
[iii] Proportion of treated students at classroom $m+1$	-0.001	0.000	-0.002	-0.001	0.005
	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)
p-value [i] = [ii]	0.0349	0.0342	0.0785	0.0386	0.4485
p-value [i] = [iii]	0.0485	0.0254	0.0164	0.0253	0.0392
p-value [ii] = [iii]	0.9835	0.8009	0.2131	0.5130	0.0352
Observations	1357	1.327	1.326	1.356	1135

***, **, * significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parenthesis. Outcome variables are standarized grades at school-grade level at follow-up. All regressions include as main control the share of enrolled students from each course. Individual controls include imputed grades in the course at baseline and a dummy indicating a missing value in the grade at baseline. Row [i] indicates the affect of the share of treated students within own student's classroom (m). Row [ii] indicates the effect of the proportion of treated students within one course lower ($m-1$) than student's own classroom. And row [iii] is similar to the previous row but related to the share of treated students one course greater ($m+1$). p-values are related to the null hypothesis that the difference between each pair of coefficients is different from 0.

TABLE A17. RELATIVE SPILLOVERS HETEROGENEOUS EFFECTS

	(1)	(2)	(3)	(4)	(5)
	Grades				Bad behavior
	Reading	Math	Science	Score	reports (-)
behavior report is within 1sd from treated students	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
[*ii*] Spillovers on non-enrolled whose bad behavior report is at most 2sd away from treated students	0.006**	0.007***	0.008***	0.006**	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)
[*iii*] Spillovers on non-enrolled whose bad behavior report is more than 3sd away from treated students	0.001	0.008**	-0.001	0.004	0.002
	(0.004)	(0.004)	(0.005)	(0.004)	(0.007)
p-value $[i] = [ii]$	0.867	0.858	0.417	0.979	0.076
p-value $[i] = [iii]$	0.036	0.623	0.168	0.366	0.121
p-value $[ii] = [iii]$	0.286	0.700	0.127	0.578	0.018
Observations	1.357	1.327	1.326	1.356	1.135

***, **, * significant at 1 %, 5 % and 10 % respectively. Robust standard errors at course-school level are in parenthesis. Outcome variables are standardized grades at school-grade level at follow-up. All regressions include as main control the share of enrolled students from each course. Individual controls include imputed grades in the course at baseline and a dummy indicating a missing value in the grade at baseline. Row [i] shows spillover effect on outcomes for non-enrolled students with a 1 sd- bad behavior level away from her treated classmates (at baseline). Row [ii] shows the spillover effect on those non-enrolled which were 2 sd - bad behavior level away for the average of her treated classmates. And row [iii] exhibits the spillovers for non-enrolled students with a bad behavior level at baseline that was three or more sd away from her treated classmates.

TABLE A18. 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 A19. 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.

TABLE A20. LEARNING AND PROTECTION MECHANISMS

	(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
[*i*] Effect on "protected" students	0.192***	0.323***	0.077***	-1.603***	-0.199***	-0.163***	-0.109***	-0.178***	-0.062***
	(0.060)	(0.108)	(0.023)	(0.291)	(0.075)	(0.047)	(0.026)	(0.057)	(0.023)
[*ii*] Effect on "non-protected" students	-0.255	0.534	0.103	-1.190	-0.185	0.363***	0.010	-0.035	-0.082
	(0.380)	(0.604)	(0.109)	(1.110)	(0.403)	(0.134)	(0.025)	(0.202)	(0.099)
[*iii*] Net protection effect	-0.447	0.212	0.026	0.410	0.013	0.527***	0.120***	0.143	-0.020
	(0.375)	(0.646)	(0.112)	(1.204)	(0.436)	(0.141)	(0.034)	(0.222)	(0.106)
Observations	948	935	962	836	916	956	956	1,010	1,010
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
[*i*] Effect on "protected" students	0.011	0.089**	0.118**	0.045	0.034***	0.017	0.027**	0.028*	-0.027***
	(0.041)	(0.038)	(0.047)	(0.039)	(0.010)	(0.017)	(0.013)	(0.015)	(0.009)
[*ii*] Effect on "non-protected" students	0.082	0.449*	0.376	0.308	0.089	0.039	0.077	-0.023	-0.049
	(0.223)	(0.272)	(0.341)	(0.280)	(0.098)	(0.092)	(0.102)	(0.101)	(0.081)
[*iii*] Net protection effect	0.072	0.360	0.258	0.264	0.055	0.022	0.050	-0.051	-0.023
	(0.233)	(0.279)	(0.348)	(0.286)	(0.101)	(0.091)	(0.101)	(0.102)	(0.081)
Observations	1,023	1,023	1,023	1,023	1,023	1,023	1,023	1,023	1,023

***, **, * significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A presents effects on non-cognitive outcomes. Panel B present results on academic outcomes. Row [i] indicates the effect on students reporting being with adult supervision after school hours (i.e. the learning effect). Row [ii] shows results of the ASP on "non-protected" students, or those without adult supervision after school hours (i.e. both learning and protection mechanism). And row [iii] shows the net protection effect, i.e. the difference between non-protected and protected students. All regressions include as controls: a second order polynomial of student's IVV, and ciclo-school fixed effect (stratification level). Additionally, 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 the baseline.

TABLE A21. OVERALL EFFECTS OF THE ASP - LEARNING MECHANISM

	(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
Any treatment	0.154**	0.360***	0.064***	-1.474***	-0.209***	-0.161***	-0.107***	0.187***	0.071***
	(0.061)	(0.092)	(0.024)	(0.252)	(0.071)	(0.046)	(0.025)	(0.056)	(0.024)
Observations	845	830	857	741	814	850	850	897	897
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Any treatment	0.037	0.124***	0.146***	0.076**	0.048***	0.025	0.034**	0.032**	-0.031***
	(0.041)	(0.041)	(0.047)	(0.037)	(0.008)	(0.016)	(0.014)	(0.015)	(0.009)
Observations	907	907	907	907	907	907	907	907	907

***, **, * significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A presents effects on non-cognitive outcomes. Panel B present results on academic outcomes. Estimations are restricted to sample that attended at least one session of the ASP. All regressions include as controls: a second order polynomial of student's IVV, and ciclo-school fixed effect (stratification level). Additionally, 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 the baseline. Differences in number of observations of non-cognitive outcomes is because of variation in the response rate for each outcome.

TABLE A22. 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 A23. 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 one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
[*i*] HM-Low enrolled in a non-academic course	0.013	0.105	0.053	0.057	-0.009	0.018	-0.013	-0.050*	-0.005
	(0.085)	(0.075)	(0.080)	(0.064)	(0.020)	(0.024)	(0.019)	(0.026)	(0.013)
[*ii*] HM-Low enrolled in a academic course	0.389*	0.363	0.257	0.363*	0.105**	0.052	-0.000	0.054	-0.032
	(0.201)	(0.297)	(0.184)	(0.207)	(0.055)	(0.057)	(0.039)	(0.047)	(0.034)

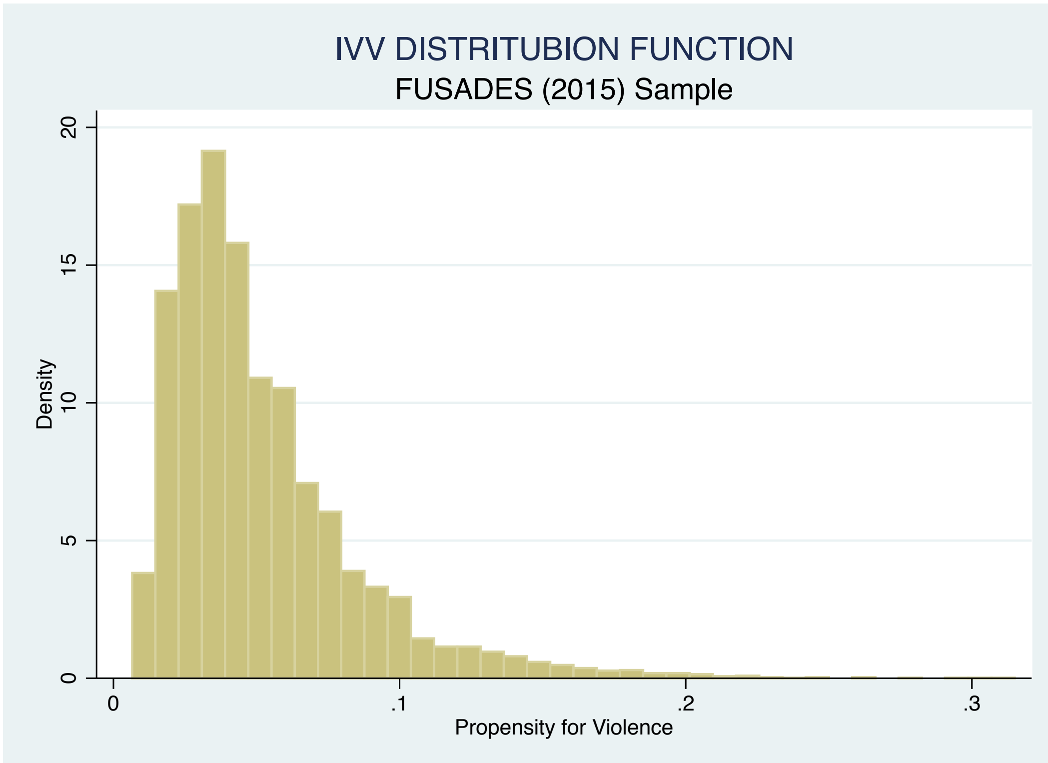
***, **, * indicate that the comparison between HM-Low vs HT at their respective category of course is significant at 1 %, 5 %, and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Sample is restricted to N = 771 treated students. All regressions include 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 A24. INTENSITY OF TREATMENT BY EXPOSURE

	(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
Any treatment (T)	0.151	0.178	0.043	-1.359*	-0.168	-0.151*	-0.108***	-0.090	-0.051
	(0.122)	(0.174)	(0.036)	(0.803)	(0.109)	(0.081)	(0.031)	(0.092)	(0.043)
T x Share of clubmates	0.054	0.417	0.093	-0.616	-0.080	0.021	0.009	-0.214	-0.034
and classmates	(0.288)	(0.389)	(0.088)	(1.841)	(0.199)	(0.188)	(0.058)	(0.145)	(0.071)
Observations	948	935	962	836	916	956	956	1,010	1,010
PANEL B: ACADEMIC OUTCOMES									
	Grades				**Probability of passing**				Failing at least one course
	Reading	Math	Science	Score	Reading	Math	Science	Score	
Any treatment (T)	-0.079	-0.068	0.032	-0.079	0.002	-0.023	0.017	0.025	-0.001
	(0.064)	(0.088)	(0.080)	(0.067)	(0.015)	(0.033)	(0.022)	(0.022)	(0.014)
T x Share of clubmates	0.247*	0.462**	0.259	0.358**	0.090**	0.108	0.032	0.002	-0.070**
and classmates	(0.139)	(0.211)	(0.160)	(0.144)	(0.035)	(0.069)	(0.047)	(0.041)	(0.034)
Observations	1023	1023	1023	1023	1023	1023	1023	1023	1023

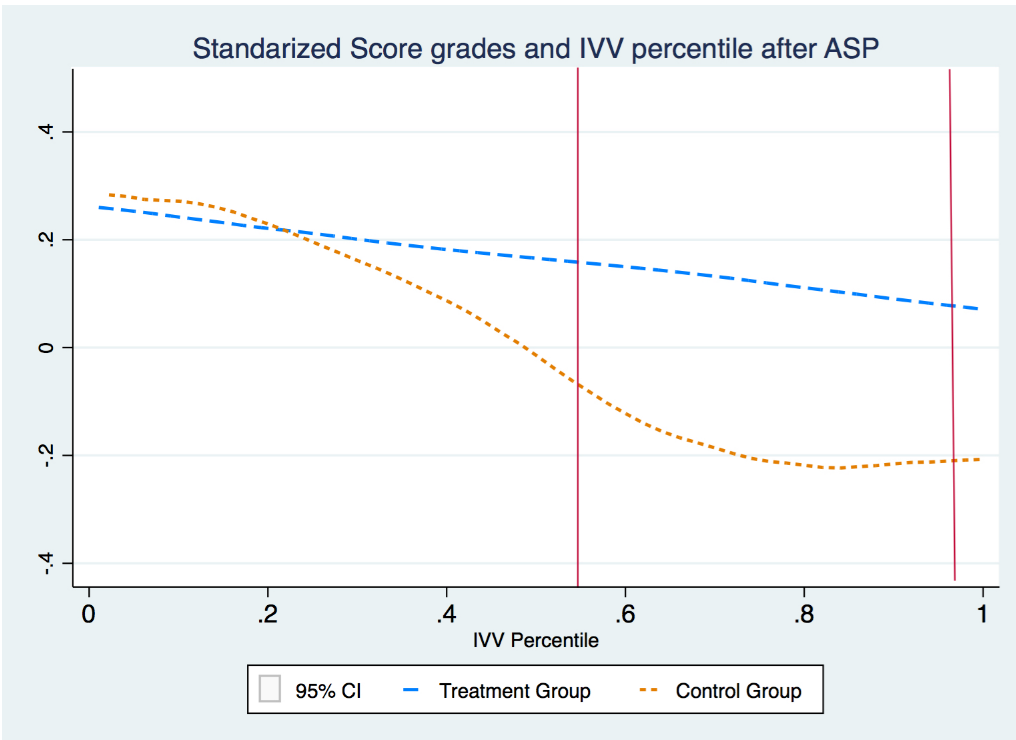
***, **, * significant at 1 %, 5 % and 10 % respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A is effects on non-cognitive outcomes. Panel B present results on academic outcomes. All regressions include an interaction between treatment and the share of clubmates and classmates at the classroom level. I also include as controls: a second order polynomial of student's IVV, and education level fixed effect (stratification level). Additionally, 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.

Figure A1. Propensity for Violence distribution function



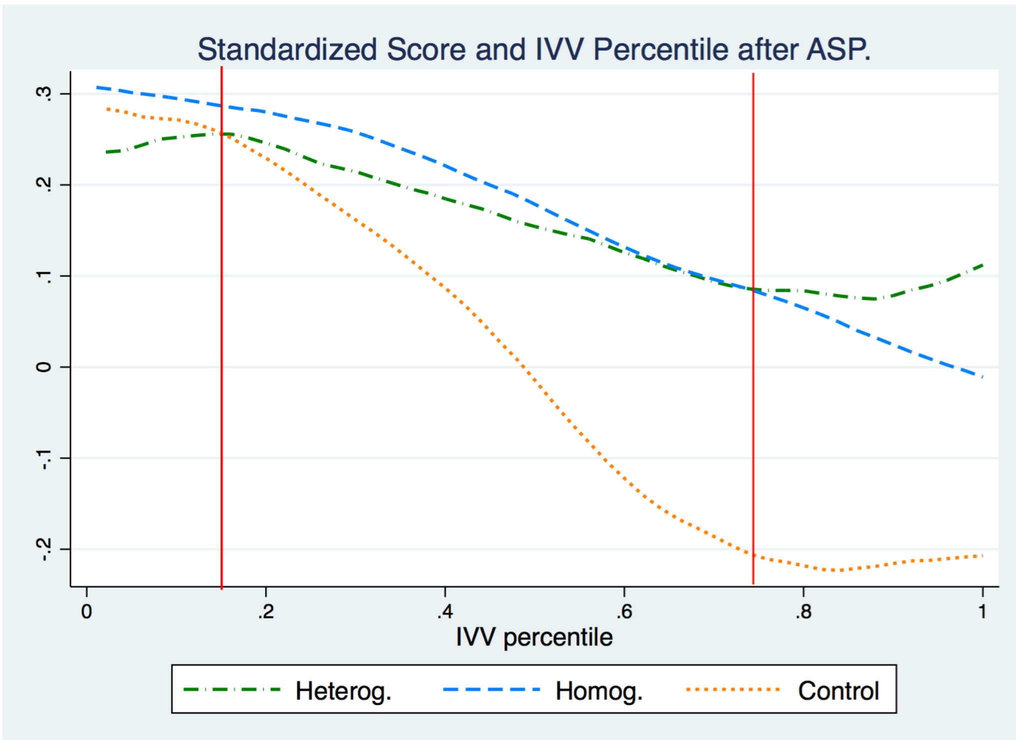
Distribution function of the estimated propensity for violence using the available determinants in the FUSADES (2015) dataset.

Figure A2. Non-linear ASP effects on endline score grades.



Local polynomial fit of standardized endline score grades by percentiles of predicted IVV. There are statistical differences between treated and control groups for students in the 55% to 95% violence percentiles.

Figure A3. Non-linear group composition effects on endline score grades.



Local polynomial fit of standardized end line score grades by predicted IVV. Children in the least violent quartile ($Q1$) and in the most violent quartile ($Q4$) are more sensible to their group composition.