PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
DOCTORADO EN ECONOMIA

TESIS DE GRADO
DOCTORADO EN ECONOMIA

Dinarte, Díaz, Lelys Ileana

Junio, 2018

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
DOCTORADO EN ECONOMIA

EDUCATION, CRIME AND VIOLENCE. EVIDENCE OF INTERVENTIONS IN DEVELOPING COUNTRIES

Lelys Ileana Dinarte Díaz

Comisión

Claudia Martínez A., Jeanne Lafortune, Rodrigo Soares, Nicolás Figueroa

Santiago, Junio de 2018

Contents

Abstract

Education, Crime and Violence.
Evidence of Interventions in Developing Countries.

This dissertation contains three essays on crime, violence, and education in the frame of development economics. The first two chapters study how specific educational interventions and the way they are implemented can determine students' cognitive and non-cognitive outcomes, including socio-emotional skills and violent behavior. Finally, the last chapter explores how providing public infrastructure in developing countries can drive some unintended short-term effects on their economic development.

Chapter 1, **"Peer Effects on Violence. Experimental Evidence in El Salvador"**, provides experimental evidence of the effect of an after-school program (ASP) on students' violence and academic outcomes; students who were enrolled in public schools in a highly violent developing country. The novelty of this intervention is that it includes some activities related to Cognitive Behavioral Therapy (CBT), which has been recently used in the economic literature to affect violent behaviors. Additionally, by creating exogenous variation in the characteristics of students' peers due to the experiment, the same design allowed to capture potential peer effects within this setting. The experimental design was inspired by Duflo et al. (2011) and Lafortune et al. (2016), adapted by myself to a different context and considering additional relevant outcomes such as violence and behavior.

This work makes two main contributions to a developing field. To start with, my research provides the first well-identified measure of how this low-intensive like-CBT intervention operates in the context of a developing and highly violent country, finding that its effects are similar (in magnitude and sign) to those of middle-intensive interventions in the U.S. (Durlak et al., 2010; Cook et al., 2015). A novel result is that the impact on academic outcomes is higher for the most vulnerable students, who in this setting are those with a higher propensity for violence. Moreover,

the intervention has indirect spillover effects on non-enrolled students.

Regarding group composition, the second main contribution is that this intervention has positive peer effects. Estimations indicate that, on average, effects on attitudes and behavior are more significant when students are treated in more diverse groups – regarding violence– than in segregated ones. Moreover, tracking has detrimental effects on the marginal student in both academic and violence outcomes. 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). Additionally, this project provides evidence that implementing this type of intervention in segregated groups of violent students can generate unintended negative results.

Then, I extended this work to measure the effect of the intervention and group composition on emotions. Recent evidence indicates that the type of emotions felt is relevant in many of a person's cognitive and behavioral outcomes such as attention, memory, and perception (Lakoff, 2008; Salzman and Fusi, 2010; Fuster, 2013). Additionally, these emotions are often determined primarily by the environment such as their communities, schools, or homes. Therefore, people exposed to highly risky conditions might face greater differences compared to their less exposed peers, creating (or widening) a gap in educational or labor market outcomes.

In the second chapter, **"How to Prevent Violence in the most Violent Contexts? Neurophysiological Evidence from El Salvador"**, Pablo Egana-Del Sol and I used the previously mentioned experimental setting, and we randomly selected a subsample of enrolled participants of the program (from both control and treated groups). Then we implemented an in-the-field lab to collect three streams of emotional and psychometric data. We relied on emotion-detection theory from affective neuroscience literature and used low-cost portable electroencephalogram recordings to obtain a proxy measure of children?s emotional state and responsiveness to stimuli.

In line with previous results, our preliminary estimations indicate that the intervention is positively affecting students' emotional regulation and socio-emotional skills. More specifically, participants' reaction towards stimuli reduces by 0.36 standard deviations and their belief that one's life can be controlled increases by 0.25 standard deviations. Moreover, the group composition effects also suggest additional gains from more violence-diverse groups structure on emotional regulation. Mainly, participants enrolled in groups with similar high-violence peers face higher levels of stress compared to other adolescents involved with diverse groups. Summing up these results, this paper sheds light on the study of the interaction between neuroscience, peer effects and like-CBT after-school programs, which has not been analyzed in the existing economic literature.

Finally, from the existing evidence, we know that most people in developing countries have limited access to physical infrastructures, such as highways and roads. These missing facilities have significant short- and long-term effects on individuals? decisions on education, migration, and employment in the formal and informal sectors.

In that sense, the third Chapter of this dissertation, **"Unintended Effects of Public Infrastructure: Labor, Education, and Crime Outcomes in El Salvador"**, examines the short-term consequences of highway construction (NTH) on development outcomes in an impoverished region of El Salvador. This is a joint project with Wilber Baires.

To implement our estimations, we built a unique data panel from a variety of sources at the municipal level (from administrative data from Education and Security Ministries to NASA's DMSP-OLS). We used two identification strategies –an instrumental variables approach and differences-in-differences– following Faber (2014) and Morten and Oliveira (2016).

Our estimations indicate that this highway positively impacted relevant economic activities in the region. However, it also generated unintended effects: the NTH reduced labor force participation in the formal sector –for both males and females between 15-19 years old– and increased dropout rates – especially for boys between 12-16 years of age. In this work, we propose that gang presence in these municipalities may be the mechanism driving the results. In fact, we find that districts near the NTH faced an increase in the short-term gang-related crimes growth such as homicides and extortions. Overall, understanding these effects have relevant public policy implications, emphasizing the importance of understanding local contexts, like much of my work.

Acknowledgments

I am profoundly grateful to my advisor Claudia Martínez for her encouragement, openness to share with me her experience and knowledge, and most importantly, for being a great mentor during this process. I am also grateful to the members of my committee, Jeanne Lafortune, Rodrigo Soares and Nicolas Figueroa, for their always insightful comments on my work and support to my research ideas. I also thank all the professors who provided their valuable recommendations to improve my projects.

This dissertation would not have been possible without the incredible support of Glasswing International as an implementer partner. Mainly, I thank Stephanie Martinez and the education sector team. Their work in Central America is certainly inspiring! I am also grateful to principals, teachers, students, volunteers, data collectors, and instructors of the five public schools in El Salvador, where we implemented the intervention.

With all my love I want to thank my parents Sonia and Reynaldo, my siblings Sonia, Tato and Jorge, and my dear niece Vale. They never stopped encouraging me to pursue my dreams, kept me in their prayers and, most importantly, taught me that our main purpose in this world should be to contribute to improving others' lives. You have been the best inspiration!

I also thank to all my friends in Chile and the rest of the world: My loving and affective Guate-Bolivia-CR family, the inspiring and brave Aguacatas, the fun and protective Terribles, the supportive and caring Pandilla and Pasmados, all my friends who share with me their experiences while I was here in Chile, and those in the rest of the world who have been somehow part of my life during this period. You made these five years joyful and memorable, and helped me to construct this dissertation with all your friendship and motivation.

Finally, I want to thank Faruk, who despite having come to my life almost at the end of this process, has been there cheering me and finding the way to make me happy, just when I needed him the most.

Chapter 1

Peer Effects on Violence. Experimental Evidence in El Salvador

Abstract[1]

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

[1]I am grateful to my advisor Claudia Martinez, and to Jeanne Lafortune, Rodrigo Soares, Francisco Gallego, José Tessada, Nicolás Figueroa, Micaela Sviatschi, Tomás Rau and Pablo Egana for their invaluable recommendations. I also appreciate the 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, PUC Chile, Universidad de Chile (FEN), and Universidad Diego Portales. I am very grateful for 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 participant public schools in El Salvador. All errors and omissions are our own. This study was funded by the CONICYT and CEDLAS-IDRC Canada). This project was registered in the AEA RCT Registry with unique identifying number AEARCTR-"AEARCTR-0001602"

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).[2]

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[3] but more recently have been started in developing countries.[4] Despite the increase in the number of programs implemented over the past years,[5] and the high incidence and economic costs of violence in the developing world,[6] the overall available non-experimental evidence of ASP's impact on social skills, crime, and violence is mixed and inconclusive (Taheri and Welsh, 2016).[7] 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).[8]

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

[2]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).

[3]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

[4]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.

[5]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).

[6]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).

[7]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.

[8]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).[9]

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.[10] 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.[11] 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).[12] 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),[13] with approximately 18%

[9]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.

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

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

[12]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).

[13]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.[14] 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).[15]

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.[16] 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.[17] 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).

[14]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).

[15]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).

[16]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).

[17]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.[18]

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.[19]

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

[18] 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).

[19] 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).[20] 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

[20] 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.[21]

Second, the paper is also related to a recent and novel literature that studies the effects of CBT[22] 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.[23] 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.[24] 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

[21]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.

[22]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).

[23]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.

[24]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).[25] 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.

[25] 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).[26]

The NGO offers four categories of clubs in the ASP by education level (*ciclos*): Leadership, Art and Culture, Sports and Science.[27] 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.[28] 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.

[26] 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.

[27] 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.

[28] 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.[29]

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).[30]

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.[31]

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.[32]

[29] 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.

[30] 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).

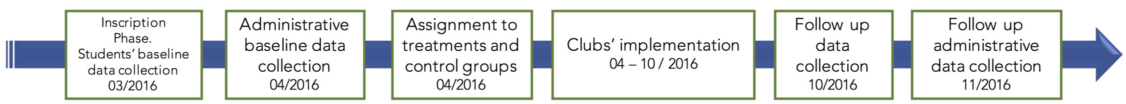
[31] Tests for differences between participant and non-participant schools are shown in the Table A1 in the appendix section. Both groups of schools are similar on schools characteristics such as location area, violence level, number of students and additional revenues. Similarly, in terms of programs, 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.

[32] 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.[33]

The timeline of the study is shown in Figure 1.

Figure 1.1: Intervention and data collection timeline



Timeline of the intervention and data collection.

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.[34] 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

[33] 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.

[34] 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.

database of youths' violence and crime from El Salvador (FUSADES, 2015),[35] I estimated the likelihood of having committed a violent act V_f as a function of a wide range of covariates:

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

where D_f is a vector of violence determinants of student f in the FUSADES dataset.[36] 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).[37]

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.[38]

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.[39] Second, this predicted index can be interpreted as a measure of student's *propensity* for violence, and not as an indicator

[35] 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.

[36] 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.

[37] 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).

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

[39] 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.

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),[40] this sort of crime and violence models estimated from existing data have a high predictive power.[41] For instance, the correlation between the predicted IVV and misbehavior at school is positive and statistically significant at 1%.[42] 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%.[43]

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.

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.

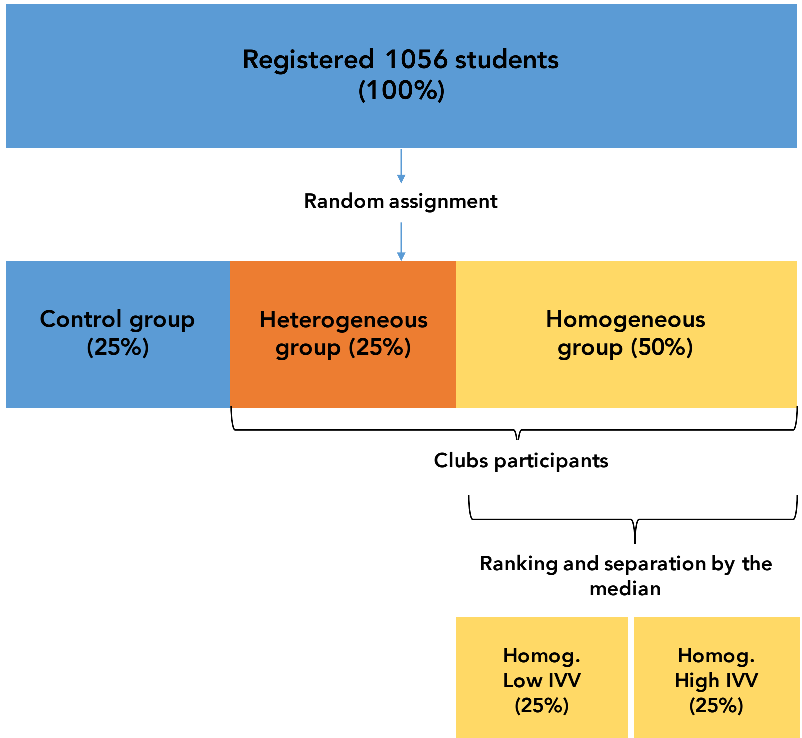
[40] See Chaiken et al. (1994) for a detailed early literature review of these models and their characteristics.

[41] 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.

[42] 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.

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

Figure 1.2: Experimental Design and Randomization Process



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

4. *Control:* This group of students were not selected to participate in the clubs during the 2016 academic year.[44]

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.

[44]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, guaranteeing their participation in the ASP in 2017.

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.[45]

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,[46] 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.[47] 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 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.[48]

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

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

[46] 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.

[47] 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.

[48] 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).

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%.[49] 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,[50] 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 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 com-

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

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

parable, 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.

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.[51] 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. 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,

[51] 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).

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.

and enrolled in evening shifts.[52]

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%.

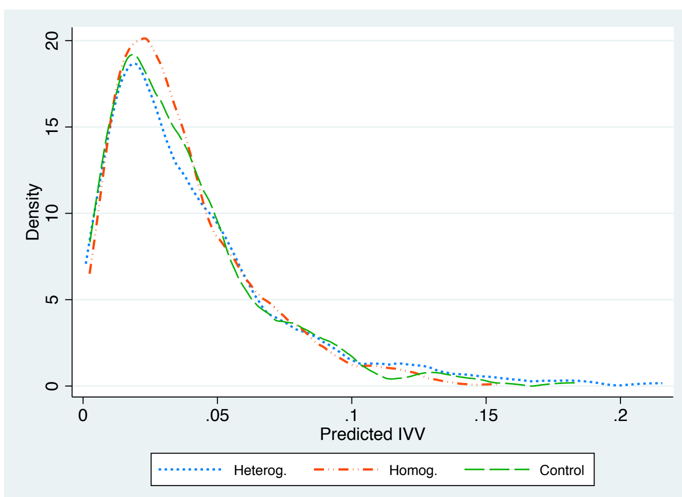


Figure 1.3: IVV Distribution Functions of Treatment and Control Groups.

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,

[52]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.

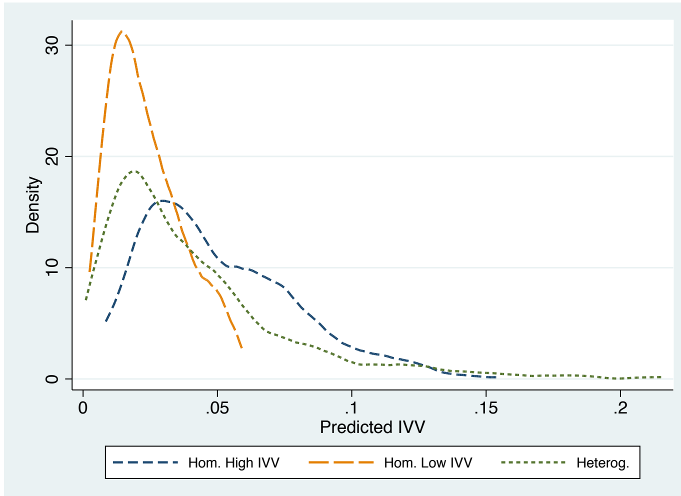


Figure 1.4: IVV Distribution Functions of Treated Groups.

there is still variation between IVV distribution functions of the HM subgroups at baseline that I can exploit.

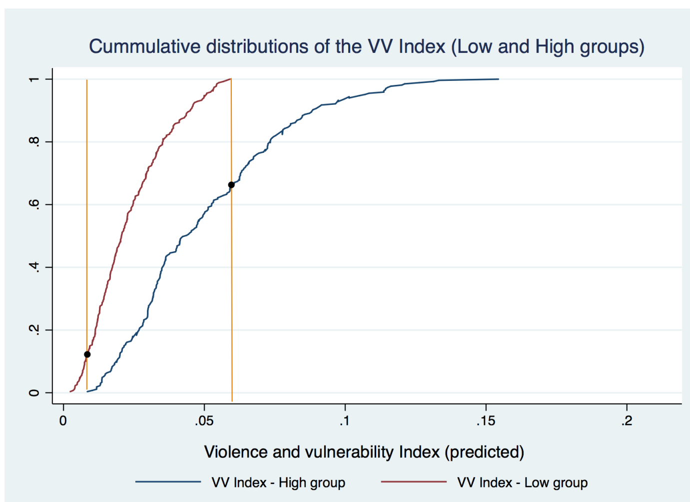


Figure 1.5: IVV Cumulative Distribution Functions of Homogeneous Sub-groups.

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%.[53]

[53] I use a third order local polynomial in order following the specification of Duflo et al. (2011). For a first and sec-

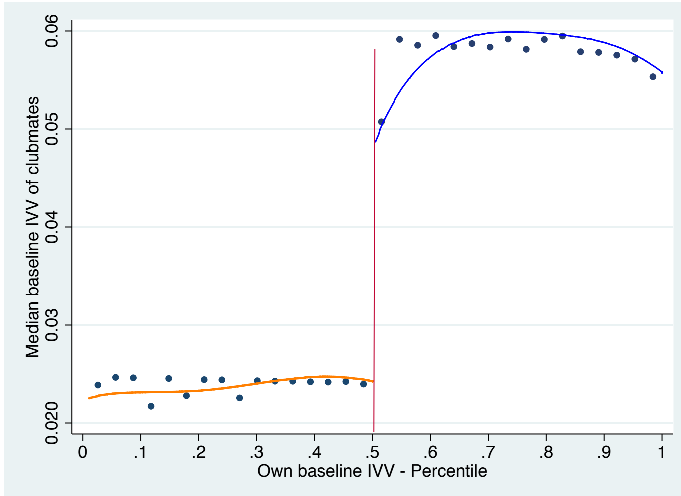


Figure 1.6: Experimental Variation in IVV Peer Composition, prior to treatment.

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:

$$y_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 X_{ij} + S_j + \epsilon_{ij} \tag{1.1}$$

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 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.

ond 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.

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:

$$y_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times IVV\_high_{ij} + \theta_3 IVV\_high_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{1.2}$$

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:

$$y_{mn} = \gamma_0 + \gamma_1 Sh_n + \gamma_2 X_{mn} + E_n + \epsilon_{mn} \tag{1.3}$$

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.[54]

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.

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:

$$y_{ij} = \theta_0 + \theta_1 Hom_{ij} + \theta_2 X_{ij} + S_j + \epsilon_{ij} \tag{1.4}$$

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.

[54] 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.

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:

$$Y_{ij} = \theta_0 + \theta_1 HomH_{ij} + \theta_2 HomL_{ij} + \theta_3 IVV\_high_{ij} + I\bar{V}V_j + \theta_4 X_{ij} + \epsilon_{ij} \tag{1.5}$$

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:

$$Y_{ij} = \lambda_0 + \lambda_1 HMH_{ij} + f(IVV_{ij}) + \lambda_2 S_j + \epsilon_{ij} \tag{1.6}$$

where $f(\mathrm{IVV}_{ij})$ is a flexible second order polynomial of the percentile of the individual's IVV within each stratum, and $\mathrm{HMH}_{ij} = 1$ 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 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.[55] Although my two sets of measures are not completely comparable,[56] 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 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.[57]

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.

[55] 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.

[56] 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.

[57] 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

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.

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). Coefficients 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.[58]

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.

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 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,

[58]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.

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.

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.[59]

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 ($2/26 \times 0.011$).[60]

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

[59] 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.

[60] 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.

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).

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.

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.[61] 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

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

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.[62]

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. 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.[63]

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.[64] 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

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

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

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

Weingarth (2005) invidious comparison peer effects model, that applied to this context implies that the exposure to only less violent –or well behave– students depresses the average performance of the group. An alternative explanation is that students in heterogeneous groups have the opportunity 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.

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.[65]

Finally, since participants were randomly allocated to a group in the ASP, there is some variation in the group composition which stem from the fact that being assigned to HM vs HT directly affects the mean and variance of one's peers. Following Lafortune et al. (2016), the identification assumption is that after controlling for strata fixed effects, the variance and mean IVV of peer stems entirely from the random assignment.[66]

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.

[65]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).

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

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.

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 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.

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).

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.[67]

Furthermore, it is also important to study how ASP's group composition can improve the results. The existing evidence on this matter is mixed[68] and mostly related to other contexts, such as educational settings (Duflo et al., 2011), female labor training (Lafortune et al., 2016) and first-year students at the United States Air Force Academy (Carrell et al., 2013).

To my knowledge, this paper provides the first experimental evaluation of the 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.[69] Additionally, the

[67]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).

[68]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).

[69]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 sec-

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)[70] 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.[71] 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 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

ondary schools (Heckman and Kautz, 2012; Cunha et al., 2010).

[70]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%.

[71]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.

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.[72] 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 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.[73]

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.[74] 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.[75]

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

[72] 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.

[73] 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.

[74] 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.

[75] 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.

group, but segregation generates greater effects for the less violent children.

These results are consistent with a body of micro-level evidence, such as papers on random assignment of freshmen or students (Thiemann, 2013); on elite exam schools (Abdulkadiroğlu et al., 2014; Dobbie and Fryer Jr, 2014; Lucas and Mbiti, 2014) and programs for gifted individuals (Bui et al., 2014). Additional evidence on academic and labor contexts is presented by Hoxby (2000); Zimmerman (2003); Angrist and Lang (2004); Rao (2015); Griffith and Rask (2014); Lafortune et al. (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.[76] Restricting the analysis to the homogeneous group, I find that students with the same level of violence at baseline seem to be "contaminated" by the predominant level of violence of the group to which they have been assigned.

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

Why does integration generate better results?

In this subsection, I provide suggestive evidence to understand how these group composition impacts on average and marginal students may have operated. I start exploring peer effects in *social skills learning*. Students in heterogeneous groups are benefiting from being exposed to both "good behaviors" that they

[76] 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)

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.[77]

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

Explaining the puzzle from the less violent children's outcomes

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

[77]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.

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.

Chapter 2

How to Prevent Violence in the Most Violent Contexts? Neurophysiological Evidence from El Salvador

Abstract[1]

We use a randomized controlled trial to study the impact of an After-School Program (ASP) and its group composition on emotional regulation and socio-emotional skills among at-risk children. We randomly assign participants at the school-education level, and within the treatment group we randomize participation in mixed or segregated groups by students' initial propensity for violence. By conducting lab-in-the-field experiments, we measured emotional regulation using physiological recordings of the widely-studied arousal and valence dimensions of emotions. We find that the program has a significant impact on emotional regulation and socio-emotional skills: participants' reaction towards stimuli reduces by 0.36 standard deviations and their belief that one's life can be controlled increases by 0.25 standard deviations. Regarding group composition, we found that the impact on our proxy for stress is greater when students are enrolled in homogeneous groups compared to heterogeneous groups. Together, these results suggest that the program implemented with integrated groups have significant impacts on psychological well-being.

Keywords: Emotional Regulation, After-School Programs, Violence

JEL Classification: I25, D87

[1]We are very grateful for the comments of Claudia Martínez and participants at the 2017 Phd Meeting at the SECHI, 10th Maastricht Behavioral and Experimental Economics Symposium, 3rd Maastricht Behavioral Economic Policy Symposium, and 7th Annual Interdisciplinary Symposium on Decision Neuroscience. We also appreciate the support of Glasswing International as an implementer partner, and 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"

1. Introduction

Violence and crime cause critical welfare losses in the developing world. They do not only affect the availability of labor force in the market – individuals can change their behavior to avoid crime or engage in criminal activities–, but they also force countries to spend substantial amounts of public and private resources on reducing their adverse effects. For example, 43% of 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). Specifically, in Latin America and the Caribbean–the most violent region in the world–the homicide rate is approximately four times the global average and crimes cost the region between 3-6% of GDP (Jaitman et al., 2017).[2] This represents an average cost of around US$300 per capita for each country.[3]

Despite the high incidence and economic costs of crime and violence, recent years have shown an increase in programs oriented to reducing violent behaviors. The available papers indicate that these interventions efficiently improve participants' behavioral and cognitive outcomes (Heller et al., 2017; Blattman et al., 2015; Carrell et al., 2013; Heckman and Kautz, 2012; Cunha and Heckman, 2008). However, except for Heller et al. (2017), there is no hard evidence of how these programs change youth behavior, considering the psychology of emotions, automaticity and impulse control (Kahneman, 2011; Fudenberg and Levine, 2006). These outcomes are relevant to the crime economics perspective since an individual's criminal actions may reflect how they manage their automatic responses and self-control when facing different events or levels of violence exposure. Additionally, there is evidence that the type of emotions a person faces is relevant to many of their cognitive and behavioral outcomes, such as attention, memory, and perception (Damasio, 1994; Lakoff, 2008; Salzman and Fusi, 2010a; Fuster, 2013).[4] In that sense, people exposed to highly risky environments might suffer more substantial differences compared to their less exposed peers when learning and developing cognitive and socio-emotional skills, which in turn may create or widen a gap in educational or labor market outcomes.

Our paper aims to contribute to the current literature in three dimensions. First, we measure the impact of a violence prevention program on emotional regulation and socio-emotional skills among at-risk children. By conducting a randomized controlled trial (RCT) in El Salvador, we are able to estimate the impact of the ASP on emotional regulation and socio-emotional skills outcomes. Second, these variables are measured using physiological recordings of the arousal –a proxy of stress– and valence –a proxy for intrinsic attractiveness or aversiveness to an event– dimensions of emotions, which provided unbiased estimations. Third, we estimate whether the composition of the groups in which students participate, in terms of violence, is a determinant in their emotional resilience level.

[2]Crime costs heterogeneity is large among Latin American and Caribbean countries. The lower bound is 2.41% and the upper bound is 3.55%. However, costs of crime in Central America are double the regional average. Paradoxically, this percentage is approximately the same GDP share assigned to the sum of education and health budgets in El Salvador.

[3]These costs are broken down into 42% in public spending (mostly police services), 37% in private spending, and 21% in social costs of crime, mainly victimization (Jaitman et al., 2017).

[4]According to DellaVigna (2009a), even slight manipulations of the individual's mood have a substantial impact on their behavior, both in the short and medium term. These emotions are often defined by the environment in which individuals are involved, such as their communities, schools or homes.

"Emotional regulation" can be defined as a mixture of cognitive and emotional processes that affect a disposition to act (Salzman and Fusi, 2010b). That is to say, it involves developing the ability to consciously affect one's own emotional and physical responses to given stimuli. Thus self-control, internal locus of control, grit, or emotional intelligence are different from emotional regulation.[5] Emotional regulation matters because emotions influence decision making and economic behavior (DellaVigna, 2009b; Loewenstein, 2000; Haushofer and Fehr, 2014). In the context of poverty and violence, stress and emotional instability can even generate vicious cycles or psychological poverty traps, which could be as bad as other well-studied financial traps.

We use a random subsample from the experimental design developed by Dinarte (2017). Participants were between 10-16 years old and enrolled in public schools in El Salvador. In this setting, students were allowed to participate in an After-School Program (ASP) from April to mid-October of the 2016 academic year, and were sorted by random assignment into treatment and control groups, stratified by school and academic level. Among the treated students, Dinarte (2018) defined two categories of treatments. First, some students were randomly granted participation in the ASP with a particular group of heterogeneous peers according to their predicted propensity for violence (IVV). Second, other students were randomly assigned to participate in the ASP with a set of homogeneous peers. Within this second treatment, students were separated into two groups: those with an IVV higher than the median were assigned to a club constituted by peers with high predicted propensity for violence (HM-High). The others were assigned to a club of peers with low predicted propensity for violence (HM-Low).

All randomly selected participants attended two sessions per week which lasted 1.5 hours each, took place just after school hours, and were implemented by volunteers. Every session was a combination of: (i) a discussion framed in a CBT approach, and oriented towards fostering children's conflict management, violence awareness, and social skills; and (ii) the implementation of club curricula that included activities such as scientific experiments, artistic performances and others. In the first part, instructors discuss concrete methods for regulating participants' violent behavior using experiential learning or role-playing. For instance, problem-solving therapy is a part of CBT intervention that uses cognitive and behavioral interventions to help students work directly on their life's challenges. It allows participants to take action in their lives, face their difficulties, and learn how to be proactive in solving their own problems. To apply this therapy in the ASP, the tutor implemented an experiential learning activity in which she mentioned some day-to-day difficulties that children usually face in low income contexts. Then, she asked them to propose alternatives to solve those problems by themselves and, after brainstorming, evaluate the solutions.

To measure the overall impact of program participation on emotional regulation, we compare the subsample of students randomly assigned to attend the program (with any kind of group composition)

[5] It is worth noting that emotional regulation is a different concept altogether from emotional intelligence. On one hand, emotional regulation is a mixture of cognitive and emotional processes that shape a mental state, and thus can be thought of as a disposition to act (Salzman and Fusi, 2010b). On the other hand, emotional intelligence traditionally defined is a synthesis of four capabilities or competencies: self-awareness, self-management, social awareness and social skills (Goleman, 2010).

and the control subsample. To measure the impact of group composition, we compare the two sub treatments: program participation in heterogeneous and homogeneous groups. Additionally, to analyze the effect of tracking, we follow Duflo et al. (2011) and restrict the sample to treated students in the homogeneous group, and estimate a model that included flexible polynomials of the students' percentile in the IVV distribution function.

We present three main results. First, for the estimations of the overall intention to treat (ITT) effects of program participation, we used neurophysiological recordings and found an effect on emotional regulation in the clubs' participants as well as an impact on responsiveness to positive emotionally-laden stimuli, compared to the control group. In particular, the program reduces their valence outcome by 0.36 standard deviations, indicating that participants become more phlegmatic and move more towards a withdrawal attitude or behavior, relative to the control group. Additionally, treated children report a reduction in their internal locus of control test by 0.25 standard deviations, compared to what children in the control group reported. Thus, treated students are perceiving that they can manage or control what happens in their lives at a greater magnitude than non-treated children.

Second, comparing low-violence students assigned to treatment versus those assigned to a control group, we find that the effects on treated children are driving both the reductions in their withdrawal behavior and in their perception that they can control their circumstances by 0.46 and 0.49 standard deviations respectively. Something surprising but expected, per previous results from Dinarte (2018), is that we also find highly violent treated children to have higher stress levels compared to both similarly violent peers in the control group and to treated children with low propensity for violence.

To sum up, these heterogeneous effects indicate that the ASP is benefiting both groups of students, but the net effect on the highly violent group is not clear. First, even when the most violent treated children experience a great reduction in positive valence difference, their stress levels also increase. In that sense, we can create unintended neurophysiological effects on the highly violent children, who were supposed to benefit more from the ASP.

What are the effects of group composition? Our estimations indicate two main results. First, for students treated with similar peers in terms of violence are driving effects on withdrawal behavior, locus of control and reaction to positive stimuli, compared to students assigned to the control group. Similarly, comparing students treated in violence-diverse peer groups to those in the control group, we find a reduction in their reaction to positive stimuli. Second, comparing the effects between the two sub treatments, we find that when students are treated in homogeneous groups, their stress levels are greater than those of students treated in a heterogeneous composition of peers. Particularly, the increase in stress is greater for children treated in HM-High groups versus the respectively comparable children treated in heterogeneous groups.

When combined, these two pieces of evidence indicate that both group compositions can have positive effects on different emotional regulation and socio-emotional skill measures when compared to ASP nonparticipants. However, similar to Dinarte's (2018) findings, the second result particularly indicates that heterogeneity is a superior group composition, since tracking participants can increase their stress level and create unintended effects from the intervention.

This last result may have different interpretations. For instance, exposure to risky environments usually increases individuals' stress level, either because they have to avoid danger or learn how to face risks –e.g. defending themselves. Therefore, it can explain why children in HM-High groups are more stressed on average than those with less exposure to violence.

Finally, in the analysis of tracking effects on the marginal student, the Cognitive Reflection Test (Frederick, 2005) shows that when assigned to a more violent group, they either increase their automatic responses or reduce their tendency to override incorrect responses in order to analyze the correct ones. Additionally, we find that marginal students assigned to HM-High groups perceive that they can control their circumstances more than those assigned to HM-Low. However, this result may be explained by participants' exposure to the most violent peers of their violence distribution function, which increases their misbehavior at school even after the intervention (Dinarte, 2017). Thus, this context may be forcing them to learn to defend themselves and, therefore, feel that they can handle their own destiny.

The rest of the paper is organized as follows: In section 2 we describe our sample, the data collection procedure and the results of balance tests between treatment and control groups. In section 3 we present the specifications to estimate the outcomes of interest. Finally, in Section 4 we present the results. All Appendix tables are at the end of this paper.

2. Intervention, Experimental Design and Data

2.1 Intervention: After-School Program to Reduce Violent Behaviors

The ASP was implemented by the NGO Glasswing International as part of its program *Community Schools*. Since 2013, Community Schools has taken place in 95 schools across Central America and benefited approximately 20,000 children between 8-15 years old through 560 clubs. According to the NGO's theory of change, the program's main objective is to successfully modify children's violence and attitudes through the learning of life skills, which consequently improves their academic performance (Glasswing International, 2012a).

The NGO offers four categories of clubs (Leadership, Art and Culture, Sports and Science) in the ASP by educational level.[6] The experiment was designed using each educational level as the stratification variable.

Since it is a voluntary program, only children interested in participating filled out a registration form with their personal and family information and an application form for the club. They were then assigned to a group based on their preferences, parent's authorization and the aggregated demand for the club category.

Clubs meet twice a week for approximately 1.5 hours each and take place just after school hours. Each session is divided into two parts: Social Skills Development and Club Curriculum. The first part

[6] Each educational level consists of three years of schooling: the first educational level is from 1st to 3rd grade, the second from 4th to 6th grade, and the third from 7th to 9th grade.

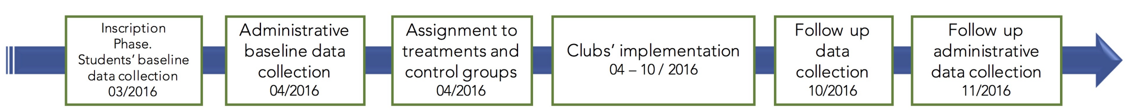
is common to all participants and all club categories, and includes the topics of conflict- and risk-management, school violence reduction, and soft skills. This part has a similar structure to Cognitive Behavioral Therapy (CBT) since instructors discuss concrete methods to regulate participants' violent behavior, such as experiential learning or role-playing. For instance, when the day's topic is conflict management, the tutor develops the following role-playing: She gives a ball to one of the students and encourages the rest of the class to provide alternatives for how to get the ball from the clubmate. Some of them suggest to forcibly retrieve it, either by hitting the ball or the clubmate. Then, the tutor discusses with the children other alternatives, such as negotiation or simply asking for the ball. It is worth mentioning that CBT has recently been studied in the U.S. context of vulnerable youth in Chicago with promising results on education and behaviors (Heller et al., 2017).

In the second part of the session, Club Curriculum, they develop activities related to each club category. For instance, in an *Art and Culture* club session students implement artistic performances or develop handicrafts. In a *Science* category club, the instructor implements and explains a "volcano eruption" experiment. The implementation of the Community Schools program was uniform across schools. For more details on the intervention, see Dinarte (2017).

Clubs are implemented by three types of volunteers: (i) *community volunteers,* who are local leaders living in the community; (ii) *corporate volunteers,* who are part of a firm that has joint projects with Glasswing; and (iii) *independent volunteers*, who are usually college students. The NGO assessed these volunteers, and their characteristics –gender, age and category– are balanced between treatment arms.

During the 2016 academic year, the NGO implemented this program in 5 additional schools in El Salvador, enrolling 1056 children.[7] They were willing to evaluate the impact of the intervention on neurophysiological outcomes through a randomized controlled trial. From the full sample, we randomly selected 598 children to collect their neurophysiological measures.

The timeline of the study is shown in Figure 1.



Timeline of the intervention and data collection.

Figure 2.1: Intervention and data collection timeline

[7]A comparison of school characteristics, programs and facilities or equipment between participant and non-participant public educational centers is presented in Table A1 in the Appendix. Comparing both groups of schools, we find that they are similar in most of the programs and characteristics except for two: participant schools have greater access to the breakfast program *Vaso de leche* and internet than non-participant public schools.

2.2 Experimental Design

2.2.1 Treatments

This paper provides experimental evidence for two gaps in economic literature. First, it measures the impact an ASP has on emotional regulation when implemented in the context of a developing and highly violent country. Second, this paper studies how group composition in terms of violence modifies the effectiveness of the intervention.

To address these questions, we rely on Dinarte's (2018) experimental design. According to the author, the first experimental design stage is to estimate a propensity for violence (IVV) in enrolled children. This IVV is a prediction generated by a model of violence and crime that uses a Two Sample Least Square strategy, as per Chandler et al. (2011). The author used an existing anonymized database of youths' violence and crime from El Salvador (FUSADES, 2015) and students' information collected during the registration phase.[8]

First, randomly assigning students to treatment and control groups allowed us to measure the overall effect of the program. Then, to find evidence on group composition effects, those who were granted participation in the ASP were randomly assigned to integrated or segregated groups according to their initial propensity for violence (IVV). Enrolled children were assigned to control (C, 25%), heterogeneous (HT, 25%) or homogeneous (HM, 50%) groups at the strata level. Then, students in homogeneous groups were ranked and assigned to subgroups according to their IVV level. All students with an IVV above the median at each 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 group (HM-Low, 25%).[9]

More specifically, treatments were structured as follows:

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

2. *Homogeneous-Low (HM-Low):* 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. *Homogeneous-High (HM-High):* Randomly selected students are assigned to participate in a club with high-violence peers, only if their IVV is greater than the median the HM group within their respective strata.

[8]A comparison of this study and FUSADES (2015) samples is presented in Table A2 in the Appendix section. Table A3 presents estimation results of the violence model. Summary statistics of the IVV are presented in table A4 of the Appendix. A deeper discussion about how the IVV is a good measure of violence can be found in Dinarte (2018). As a summary, that paper provides evidence that the IVV is a good proxy for students' misbehavior. For example, similar to previous studies (Klassen and O'connor, 1988; Chandler et al., 2011), the measure has a high predictive power for future misbehavior. Using data from students in the control group, the measure gives a positive and statistically significant correlation between IVV and their bad behavior at the end of the academic year. Additionally, this measure does not predict academic performance. We estimated the correlation between the predicted index and grades reported by teachers and found that it is not statistically significant.

[9]Additional details of the intervention, balance on observables, attrition, and descriptive statistics of the total sample can be found in Dinarte (2018). Some of those details are incorporated in this paper in the Appendix section.

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

Figure 2 shows the full sample of registered students (1056 children) assigned to each treatment arm and the control group. Numbers in parentheses indicate how many students were randomly selected from each group to collect neurophysiological data, generating a subsample of 598 children for this study.

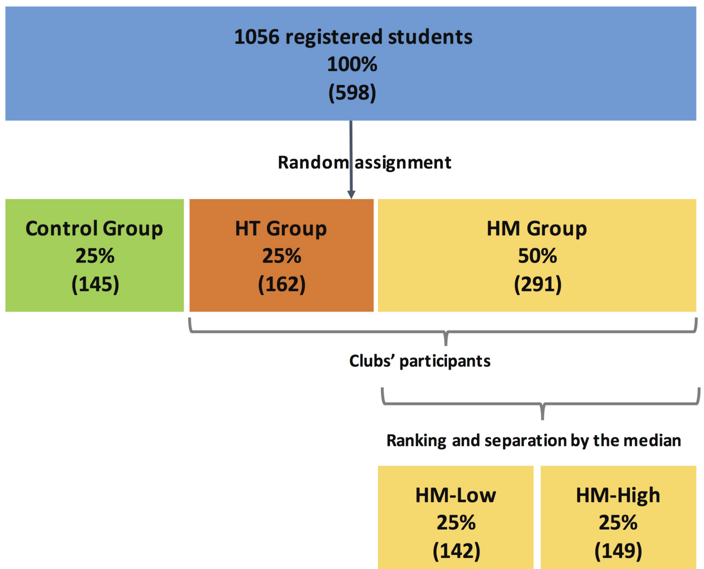


Figure 2.2: Experimental design and randomization process

As opposed to Duflo et al. (2011) and similar to Lafortune et al. (2016), neither instructors nor participants knew details regarding the assignment in order to capture the effects of the interactions between participants rather than other channels, such as curriculum adaptation.

2.3 Data Collection Process and Main Outcomes

After randomly selecting a subsample of 598 enrolled participants in the ASP (from both control and treatment groups), we followed Egana-delSol (2016) and established a lab-in-the-field setting to collect three streams of data for each student: pre-test resting emotional state; psychometric tests to measure non-cognitive, creative and cognitive skills; and emotional responsiveness to both positive and negative stimuli. These measures were taken at the end of October 2016, right after the ASP's completion.

As explained in Egana-delSol (2016), to proxy emotional regulation we use the emotion-detection theory from affective neuroscience literature, and use electroencephalogram (EEG) recordings to measure emotions. We acquired low-cost, portable EEG headsets to obtain a proxy measure of students' emotional states and responsiveness to stimuli in the arousal-valence locus (Ramirez and Vamvakousis, 2012). These devices are research graded and provide emotion-detection with an average accuracy (correct detection) of 79% (Egana-delSol, 2016; Martinez-Leon et al., 2016).

The experiment consisted of showing a series of images selected to generate positive and negative stimuli and elicit emotional responses from students' brains. There are three estimates of arousal and valence according to the type of stimuli: pre-test resting state, positive reaction and negative reaction.

The first part of the experiment was the pre-test on emotional state. This phase allowed us to estimate emotional arousal and valence indices at a resting state. The pre-test was constructed using EEG recordings while students watched a black cross in the center of a gray screen for a period of 30 seconds before taking the battery of psychometric tests. Emotional arousal and valence indices at resting state are estimated using these EEG recordings (our baseline measures). In this study, we use arousal as a proxy of children's stress, measured directly from her brain activity. Valence can be interpreted as a positive or negative mood, as well as an attitude of either approach or withdrawal towards/from a stimulus (Harmon-Jones et al., 2010; Kassam et al., 2013).

Then, in the second part of the experiment, students responded to the battery of psychometric tests, which included the Rotter Locus of Control Scale (Rotter, 1966), raven-like progressive matrices, Torrance's Test of Creative Thinking (TTCT), and the Cognitive Reflection Test (CRT). Our measure of locus of control indicates that children think that they are not able to control what happens in their lives. More specifically, a decrease in locus of control indicates that students feel they can manage their experiences, thus demonstrating an increase in self-efficacy. Raven is a measure of abstract reasoning and a non-verbal estimate of intelligence. It is implemented as a set of matrices in progressive order. TTCT measures creative skills by producing an index based on fluidity, flexibility and originality of answers when subjects are asked to relate seemingly unrelated objects or provide multiple uses of an object (i.e. disadvantages of a smartphone). Finally, CRT is a test designed to measure if an individual tends to automatically choose an initially incorrect response and then engage in a deeper reasoning to find a correct answer.

During the third part of the experiment, right after the students finished the battery of psychometric tests, we obtained emotional response intensity for negative and positive stimuli in terms of valence locus. Here, we exposed students to alternate series of images that elicit positive and negative emotional responses in order to estimate post stimuli valence indices.

We complement these neurophysiological measures with academic and misbehavior data provided by schools before the intervention (March 2016) and right after the end of the academic year (November 2016). We collected math, science, and reading grades, behavior reports, and school absenteeism.

2.4 Summary Statistics

Descriptive statistics of this study's sample (598 observations) and for each treatment and control groups are presented in Table 1. In Column 1 we exhibit statistics for the full sample. Columns 2-5 are for control (C), any treatment (T), and specific treatment groups (HT and HM) that were randomly selected from the 1056 enrolled students. Columns 6-7 show statistics for the two homogeneous subgroups.

In Panel A we present summary statistics of the IVV determinants. Participants are closer to 12 years old, 49% are male, and 75% live in an urban area. In terms of household composition, 91% of the students live with at least one parent, and 9% live with a relative or a non-related adult. Regarding mothers' education, 61% of students' mothers have an intermediate education level (between 7-12 years), and 31% have less than six years of schooling. In terms of risk exposure, only 5% of students reported being alone at home when they are not at school. However, they have to travel around 17.2 minutes to school on average. Moreover, 28% of students are enrolled in the afternoon shift, which increases the probability of being alone 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.

More IVV descriptive statistics for the full sample can be found in Dinarte (2018) and, for this paper subsample, in Table A5 of the Appendix. IVV standard deviation for the full subsample is 0.029 and its level ranges from 0.002 to 0.216.

We find a balance in all variables except for age, when we compare in the comparison between T and C groups. Then, when comparing HM and C groups, school year and reading grades differ. Finally, the average share of male students differs between HM and HT groups. We do not find any difference between HT and C groups.

Asterisks in column (7) indicate statistical differences in means between HM-H and HM-L groups. As is expected from this experimental design, there must be differences on most of the IVV determinants between these two subgroups. First, we find differences in individual student characteristics: HM-H groups are on average older and the share of boys is greater than in the HM-L groups. Second, most mothers of students assigned to HM-H groups have intermediate education and a lower share has basic education compared to mothers of students in the HM-L groups. Since women with basic education in El Salvador usually perform domestic activities at home, this may indicate that it's more likely for most students in the HM-L group to be under adult supervision most of the time. Finally, there are differences in all measures of exposure to risky environments except for being enrolled in the morning shift. On average, students in the HM-H group have to commute 39% longer times and have an 8% higher probability of being alone at home than children in HM-L groups.

Panel B shows academic scores and absenteeism for the first quarter of the 2016 school year. On a grade scale of 0-10, which requires a minimum passing grade of 5, enrolled students receive between 6.6 and 6.7 points. This score is similar to average grades at the national level. The mean absenteeism rate in the first quarter, before the intervention, was 6.3% (2.53 out of 40 days).

Finally, Panel C summarizes the clubs' characteristics: mean club size was 13 students, and community tutors ran approximately 31% of the clubs. The average take up was 57%. Moreover, the share of enrolled students in each club category is statistically similar between treatments, except for 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.

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	0.49	0.49	0.55	0.45	0.70	0.20***
Student's age	11.9	11.8	11.9*	12.1	11.9	12.5	11.2***
Student lives in urban area	0.75	0.74	0.75	0.73	0.76	0.79	0.72*
Student's household composition							
Student living with both parents	0.53	0.48	0.55	0.52	0.56	0.54	0.58
Student living with only one parent	0.32	0.37	0.31	0.35	0.29	0.31	0.27
Student living with one parent and step-parent	0.06	0.06	0.06	0.06	0.07	0.05	0.08
Student living with other relative /adults	0.08	0.10	0.08	0.07	0.08	0.09	0.07
Student's mother's level of education:							
Basic education (1-6 years)	0.31	0.34	0.29	0.26	0.31	0.23	0.39***
Intermediate education (7-12 years)	0.61	0.56	0.62	0.65	0.62	0.69	0.54***
University or higher (13 and +)	0.08	0.09	0.08	0.08	0.08	0.08	0.07
Student's travel time from house to school (min.)	17.2	15.8	17.6	16.5	18.2	21.1	15.2***
Student is alone at home after school	0.05	0.06	0.05	0.06	0.05	0.09	0.01***
Student's school year	5.72	5.62	5.75**	5.80	5.72	6.14	5.28***
Student enrolled in the morning shift	0.72	0.74	0.71	0.70	0.71	0.68	0.75
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	6.71	6.48	6.78	6.67	6.84	6.75	6.94
Math scores	6.55	6.52	6.56	6.59	6.55	6.38	6.73**
Science scores	6.65	6.50	6.71	6.67	6.74	6.55	6.95**
Behavior scores	7.40	7.27	7.44	7.42	7.46	7.47	7.45
Absenteeism Q1 2016	2.35	2.83	2.18	2.18	2.19	2.47	1.91
PANEL C: CLUB CHARACTERISTICS							
Average club size	-	-	13.3	13.3	13.3	13.1	13.6
Average club take up	-	-	0.57	0.57	0.57	0.56	0.59
Community tutors	-	-	0.31	0.29	0.32	0.33	0.31
Club category							
Leadership	-	-	0.29	0.14	0.16	0.18	0.13
Arts 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
Observations	598	145	453	162	291	149	142

Table 1 shows descriptive statistics of the available variables at baseline for the full sample. Panel A summarizes information obtained from the enrollment form 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 club characteristics. Take-up is estimated as number of hours attended by student i / max hours attended by any student in each club. There are no differences between HT and Control. In the comparison between HM = Control groups, there are statistically significant differences in school year and reading grades; in the comparison HM = HT, there are statistically significant differences in gender, and for T = C the only difference is in age. ***, **, * indicate differences statistically significant at 1%, 5% and 10% respectively between HM-Low and HM-High groups.

To generate an exogenous variation that allows for identification of the causal impact of group composition effects, this experimental design had to meet the following five requirements: (i) treatments and control groups must be balanced, (ii) most IVV determinants of the HM-High group should be greater than those of the HM-Low group,[10] (iii) HT groups must be more violence-diverse than any of the HM groups and the HT group's mean IVV must be between the HM-Low and HM-High levels, (iv) IVV distributions for the HT, HM and C groups must be similar at the baseline, but the distributions for the HT, HM-High, and HM-Low must differ, and (v) 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. Dinarte (2018) provides additional details of these five experimental design characteristics.

2.5 Attrition Analysis

The final sample, after filtering the EEG data and accounting for attrition is a total of 308 valid EEG recordings of students; i.e. an average attrition share of 48%. Table A6 in the Appendix shows summary statistics and balance tests for baseline characteristics of the 308 observations with valid EEG recordings.

In this subsection, we present some checks in order to verify if this attrition rate was quasi-random. First, in table A7 of the Appendix, we present summary statistics and balance tests of the baseline variables for the 290 students from which we have non-valid or missing EEG recordings. Comparing the average characteristics among the attrited and non-attrited samples (Tables A6 and A7), we find that they are similar in most of them except for the number of males and area of residence (both at 5%). Specifically, the number of boys and the share of students living in an urban area are 10% and 5% greater in the attrited subsample compared to the same shares in the non-attrited group.

Second, comparing the share of non-valid EEG recordings between treatments and control groups, we find that there are no statistical differences among the different groups.[11] This finding indicates that their missing status was quasi-random. Additionally, we have 13 attrited observations from the psychometric tests (locus of control, Raven and CRT) due to technical problems. As we show in Table A8 of the Appendix, the average characteristics of these 13 observations are similar to those of the other 295 observations, and are similarly distributed among treatments.

Finally, as we will describe later, we estimate our main specification using a Heckman's correction for selection bias h and then compare if the estimated coefficients are similar after accounting for the missing observations.

[10] See the above description of these differences.
[11] See the "Missing share" row at the end of Table A7.

3. Empirical Strategy

In this section, we present the empirical strategy used to generate evidence of the intervention's impact on emotional resilience and socio-emotional skills. We also describe the specifications used to measure heterogeneous effects by initial level of violence. Additionally, we study group composition's average effects on emotional resilience and how this heterogeneity interacts with children's initial propensity for violence.

3.1 ASP Average and Heterogeneous Effects on Emotional Regulation

To measure the overall impact of the intervention on emotional regulation, we exploit the random variation between treatment and control groups provided by the experimental design and estimate the following regression:

$$E_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 X_{ij} + S_j + \epsilon_{ij} \tag{2.1}$$

where E_{ij} are the emotional regulation and socio-emotional measures of the student i in school and education level j, expressed in standard deviations from the outcomes distributions of students in the control group. Then, T_{ij} is a dummy that indicates if student i has been assigned to a treatment group, S_j are school- and education-level fixed effects (stratification cells), and X_{ij} is a vector of controls, including a second order polynomial of students' predicted propensity for violence.

Due to measurement bias in the predicted propensity for violence, we estimate cluster bootstrapped at course-school level (Treiman, 2014). θ_1 is the ITT effect of being randomly assigned to participate in an ASP on each outcome E_{ij}, compared to the control group.

As mentioned before in the attrition analysis section, the valid EEG recordings were approximately 52% of the total subsample. This rate was not different across treatments groups and so, assuming that the process governing recordings validity follows a standard monotonicity property, and given that we are analyzing data from a randomized experiment, we can compare outcomes across treatments and control groups, which will be valid estimates of the impact of treatment on outcomes for the full randomly selected subsample (Lee, 2009). However, as a check, we computed Heckman selection-corrected regression results for the same subsample outcomes.

Since there is lack of evidence for the heterogeneous effects of ASP on socio-emotional skills, and no evidence of differences in the impact on emotional resilience by participants' initial propensity for violence, we included in specification (1) an interaction between the treatment dummy T_{ij} and a binary indicator IVV_high_{ij}. This dummy indicates that a student's IVV percentile at baseline is greater than the median at the group (C, HM, and HT) and stratum level, within the sample of this study. We also include this indicator as a control variable in the estimation.

More specifically, we estimated the following equation:

$$E_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times IVV\_high_{ij} + \theta_3 IVV\_high_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{2.2}$$

The other variables are defined as in specification (1). The effect of the intervention on low violence students is measured through θ_1, and $\theta_1 + \theta_2$ is the ITT effect on highly violent children.

We estimate further heterogeneous effects by initial level of violence analysis at different levels of the initial IVV distribution. These effects are estimated using an equation similar to specification (2), but including interactions with each quartile s of the predicted index. Each quartile is defined as 25% of students within each subgroup. The omitted category is the quartile of less violent children and its interaction with any treatment.[12]

3.2 Group Composition Average Effects on Emotional Regulation

In the line of studies that design classrooms experiments to analyze group composition effects (Duflo et al., 2011; Carrell et al., 2013; Lafortune et al., 2016), we can test the average effects of being treated in a particular composition of peers, thus exploiting the random variation generated directly from the experiment design. Additionally, we can exploit the random assignment of each participant to a group of peers with a certain mean and variability of violence. Finally, we use the discontinuity in the median of the IVV distribution for the HM group to evaluate the effect of tracking on the marginal participant. A comparison of the two sets of group composition measures will provide evidence as to whether the outcome is only affected by the average characteristics of peers, or if there is an interaction between a student's characteristics and that of her peers.

3.2.1 Average and Heterogeneous Effects

Due to the random variation in composition of peers' propensity for violence that is generated by the experimental design, we can directly estimate the effect of group variability on emotional regulation by running the following equation:

$$E_{ij} = \alpha_0 + \alpha_1 HT_{ij} + \alpha_2 HM_{ij} + \alpha_4 X_{ij} + S_j + \epsilon_{ij} \tag{2.3}$$

where HT_{ij} and HM_{ij} are dummies indicating whether the student i in educational level j is assigned to a club constituted by more different (heterogeneous) or similar (homogeneous) peers to her predicted level of violence, respectively. The rest of variables are defined as in the first specification.

Considering this, α_1 and α_2 can be interpreted as the average effects of receiving an offer to participate in an ASP with a heterogeneous or homogeneous composition of peers' initial propensity for violence, compared to being assigned to the control group. Testing by the differences among α_1 and α_2, we also provide evidence for how group composition might change the ASP's effectiveness on emotional regulation.

[12]Finally, as previous studies indicate (Durlak et al., 2010), this type of ASP may impact boys and girls differently. 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, and following Dinarte (2018), we use an alternative specification to show that the differences found in this subsection are driven mostly by students' propensity for violence. A detailed description of the equation and estimations is presented in Appendix 1.

Similar to specification (1), we include as control a second order polynomial of the predicted propensity for violence per student. Since we included the estimated IVV in the main regression, we also estimate the standard errors using clustered bootstrap at the course level.

One may also be interested in comparing high- and low-violence children assigned to the HT treatment with children at the same violence level who were assigned to HM groups. This experimental design allows us to implement that comparison. We can exploit that the HM group is constituted by HM-H and HM-L subgroups. Thus, we can use a specification in which we separate the homogeneous treatment into two indicators, HMH_{ij} and HML_{ij}, and restrict the sample to treated participants. We can also include as additional controls the median of the IVV distribution at the HM-stratum level IVV_high_{ij}, and the IVV median $I\bar{V}V_j$ at the j level.

To sum up, the required equation should be the following:

$$Y_{ij} = \theta_0 + \theta_1 HMH_{ij} + \theta_2 HML_{ij} + \theta_3 IVV\_high_{ij} + I\bar{V}V_j + \theta_4 X_{ij} + \epsilon_{ij} \tag{2.4}$$

where HMH_{ij} and HML_{ij} are dummies indicating whether the student i in strata j was assigned to HM-High or HM-Low treatment arms, respectively. Thus θ_1 is an ITT estimator of assigning a child i, with a propensity for violence greater than the median, to a less diverse and highly vulnerable-to-violence group of peers, compared to assigning her to a more diverse group of peers. Also, θ_2 is an ITT estimator of assigning a child i,with low propensity for violence, to a less violence-diverse group of peers compared to a heterogeneously violent group.

Specification (4) allows us to directly analyze the effects of group composition. However, we were also interested in measuring the differences in the effects of the intervention by initial level of predicted violence. Dinarte (2017) finds a statistically significant ITT effect on academic outcomes of students treated in homogeneous groups, but violence-related outcomes are statistically significant for students treated in heterogeneous groups.

3.2.2 Effects of Tracking on the Marginal Student

Further analysis of group composition through this experimental design includes the study of the effects on marginal students, which in this setting are those with an IVV around the median. As in Dinarte (2018), these effects are relevant since having high-violence peers on average also means that the student is the least violent child in her group before the intervention. ON the contrary, 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, may face the a greater impact from tracking.

Restricting the sample to students randomly assigned to the HM treatment, we exploit the discontinuity around the median of each HM group's IVV distribution function to explore the effect of peers' violence within a tracking setting. Since students just above the median are similar regarding their propensity for violence to those at or below the median, we use a regression discontinuity design and mark the median in each strata as the discontinuity. 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.

To identify this effect, we estimate the following equation:

$$Y_{ij} = \lambda_0 + \lambda_1 HMH_{ij} + f(IVV_{ij}) + \lambda_2 S_j + \epsilon_{ij} \tag{2.5}$$

where $f(\text{IVV}_{ij})$ is a flexible second order polynomial of the percentile of the individual's IVV within each stratum, and HM-H_{ij} is an indicator of whether the participant i was in the HM-High group. We can interpret λ_1 as a LATE estimator that indicates how being assigned to similar peers with high levels of violence affects students' emotional regulation outcomes. To confirm that our results are robust to different specifications of the IVV polynomial, we also estimate specification (5), but control for third and fourth order polynomials.

4. Main Results

In this section we present the reduced form estimates of the ASP's impact on students' emotional regulation and socio-emotional skills. We also present estimations of heterogeneous ASP effects by students' initial propensity for violence. Moreover, we describe average and marginal group composition effects of the ASP on the dimensions of interest.

4.1 Average ITT Effects of ASP on Emotional Regulation

Table 2 shows the estimation results of specification (1). In Panel A we present the effects on our measures of emotional regulation and socio-emotional skills. In Panel B we describe the ASP indirect effects on academic grades, following Cunha and Heckman (2008) model, which states that cognitive outcomes are generated mainly by non-cognitive skills.

First, regarding emotional regulation, we find that ASP participants face a reduction in their valence outcome by 0.36 standard deviations compared to the control group, as we can see in column (2) Panel A. Additionally, the result in column (6) in Panel A indicates that there is also a reduction in participants' reaction to positive stimuli. In particular, the "positive valence difference" variable measures the variation in the valence index when the stimulus is positive net of the individual's baseline resting state valence level. This can be interpreted as a lower level of frustration of participants – they become more phlegmatic or cold headed– or that participants move towards a more withdrawal behavior or attitude, relative to the control group.

This improvement in emotional regulation is also complemented by students' self report on locus of control. As explained before, an increase in our measure of locus of control indicates that participants perceive they are unable to control what occurs in their lives. In that sense, we can see in column (3) of Panel A that treated children report a reduction in their locus of control by 0.25 standard deviations, compared to what children in the control group reported. Thus, treated students are perceiving that they can manage or control what happens in their lives by a greater magnitude than non-treated children. In the rest of the outcomes, we find no statistical difference between treated and control students at the conventional levels due to power limitations, as we show in the MDE estimations row.

These results shed light on the conclusion that the ASP directly affects students' reaction to some stimuli. The program is oriented to teaching them how to manage their automatic responses that give rise to violent behaviors across the different domains where children perform. Moreover, the results are consistent with recent evidence on how some interventions in CBT can positively impact behaviors (Blattman et al., 2015; Dinarte, 2017; Heller et al., 2017).

Although ASP activities are not directly related to academic outcomes, there is a positive correlation between academic performance and social skills. For example, as students learn to handle their automatic responses, they may behave better at school and are less disruptive during classes, thus facilitating the learning process. In this sense, their grades might improve. Using academic and behavior reports from schools, we find that the intervention does help reduce misbehavior at school by 0.34 standard deviations. Yet, we do not find any statistical differences on academic grades between treated and control children.

To verify whether the attrition on EEG recordings were quasi-random, we compute the Heckman selection-corrected regression as a check. These results are presented in Table A9 of the Appendix. We find these estimated ITT effects using the 598 observations. After correcting for the selection bias, we notice they are very similar to the results obtained from the 308 valid EEG recordings. Thus, the selection bias seems to be quasi-random or not related to the intervention. For that reason, we will only rely on the sample with valid EEG measures.[13]

4.2 ASP Heterogeneous Effects on Emotional Regulation

Due to the heterogeneous effects by propensity for violence, found in Dinarte (2017) for academic and violence-related outcomes, we exploit the IVV prediction to study if the ASP is affecting high and low violence children differently. Thus, we separate participants into subgroups of highly or less violent children, considering their predicted IVV level within each stratum and treatment group. From here, we create the indicator described in specification (2).

In Table 3 we present the estimated effects of specification (2) for emotional resilience measures. Coefficients in row [i] 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. Then, coefficients in row [iii] point to the difference in effects between high- and low-violence treated students. Finally, row [iv] indicates p-values for the test of difference in effects between high- and low-violent treated students.

[13]Results of estimations using the full sample are available upon request.

TABLE 2. OVERALL EFFECTS OF THE ASP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A. EMOTIONAL REGULATION							
	Arousal (stress)	Valence	Locus of control	CRT	Raven	Positive Valence Difference	Negative Valence Difference
Any treatment	0.067	-0.355**	-0.247**	0.003	-0.032	-0.446**	-0.243
	(0.119)	(0.159)	(0.115)	(0.128)	(0.116)	(0.195)	(0.186)
Observations	308	308	295	295	295	308	308
Mean control group	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD - control group	1.00	1.00	1.00	1.00	1.00	1.42	1.28
MDE T = C	0.21	0.21	0.21	0.21	0.21	0.21	0.21
PANEL B. ACADEMIC GRADES							
	Reading	Math	Science	Score	Absenteeism	Misbehavior reports	
Any treatment	-0.080	0.116	0.028	0.046	-0.715	-0.336**	
	(0.096)	(0.078)	(0.107)	(0.096)	(0.737)	(0.143)	
Observations	308	308	295	295	264	307	
Mean control group	6.42	6.05	6.26	6.24	6.95	6.63	
SD - control group	1.67	1.76	1.58	1.55	1.32	8.93	

***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. All outcomes have been standardized at the control-course level, with a mean of 0 and standard deviation 1.0. All regressions include as controls: a second order polynomial of student's propensity for violence, and *ciclo-school* fixed effects (stratification level). Differences in number of observations is due to variation in the response rate for each outcome.

When we compare low-violence students assigned to treatment or control groups (first row), we find that treated children are driving both the reductions in their withdrawal behavior and in their perception that they can control their circumstances by 0.46 and 0.49 standard deviations, respectively. Then, looking at the results in the second row, we also find that treating a highly violent child increases her stress level while also reducing their reaction to positive stimuli. There are no statistical differences in the remaining emotional regulation and socio-emotional skill outcomes.

Next, we tested if the effects on the highly violent treated group where different from the effects on the less violent treated groups. This evidence is presented in row [iii]. We find here that there are differences in the effects of the intervention, by initial propensity for violence, on stress and locus of control. Specifically, the difference on arousal –a proxy for stress– is approximately 0.43 standard deviations between both groups, indicating that high-violence treated children are facing an increase in their level of stress compared to their low-violence treated peers. Also, we find differences in the locus of control: the reduction for less violent students was greater than the reduction for highly violent students, as shown in column (3).

In summary, it seems that the ASP is benefiting both groups of students. Even so, the net effect on the highly violent one is not clear. First, even when the most violent treated children experience a greater reduction in positive valence difference, their stress level also increases. In other words, we might cause unintended neurophysiological effects on the highly violent children; those who were supposed to benefit most from the ASP. However, due to the novelty of this design, we also provide evidence that group composition is what's driving this unexpected result, which we will discuss later.

The reduction on valence for the less violent students is consistent with the heterogeneous effects found by Dinarte (2017) who studied the impacts of misbehavior on school outcomes using teachers reports. That paper shows that the reduction in the intensive margin of bad behavior at school was greater for low-violence students, indicating that students' violence trends might help explain this last result. It is possible that the automatic responses that generate bad behavior are harder to modify, particularly for those who are 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 towards violence may be harder.

Results of a more disaggregated analysis on heterogeneous effects, by initial propensity for violence, indicate that most of the differences appear in the comparison between most and least violent students (Q4 and Q1, respectively). We find that the effects on valence, locus of control and CRT are specifically greater for treated children with the lowest propensity for violence; when compared to the 25% most violent share of students in the IVV distribution. Moreover, there are no differences between treated children in the middle of the IVV distribution –those in the quartiles Q2 and Q3. These estimations are available upon request.

TABLE 3. HETEROGENEOUS EFFECTS OF THE ASP ON EMOTIONAL REGULATION BY PARTICIPANTS' INITIAL PROPENSITY FOR VIOLENCE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Arousal (stress)	Valence	Locus of control	CRT	Raven	Positive Valence Difference	Negative Valence Difference
i. Effect on low-violent treated students	-0.142	-0.459**	-0.490***	0.126	0.056	-0.395	-0.249
	(0.163)	(0.248)	(0.166)	(0.103)	(0.189)	(0.358)	(0.190)
ii. Effect on high-violent treated students	0.291**	-0.240	-0.004	-0.119	-0.120	-0.499*	-0.233
	(0.145)	(0.253)	(0.172)	(0.231)	(0.163)	(0.266)	(0.358)
iii. Difference of effects between high-low treated students [ii] - [i]	0.433**	0.219	0.486*	-0.245	-0.176	-0.104	0.016
	(0.210)	(0.368)	(0.267)	(0.254)	(0.262)	(0.489)	(0.428)
iv. p-value Ho: Difference = 0	0.039	0.552	0.069	0.336	0.502	0.831	0.971
Observations	308	308	295	295	295	308	308

***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. All outcomes have been standardized at the control-course level, with a mean of 0 and standard deviation 1.0. All regressions include as controls: a second order polynomial of student's propensity for violence, and *ciclo-school* fixed effects (stratification level). Differences in number of observations is due to variation in the response rate for each outcome.

Finally, as previous studies indicate (Durlak et al., 2010), this type of ASP may impact boys and girls differently. However, after implementing a typical heterogeneity analysis on the intervention effects by gender, we find that there is no statistically significant difference at conventional levels between treated girls and boys. Thus, this is evidence that the differences in the effects are driven by children's IVV. As an additional test, we follow Dinarte (2018) and use an alternative specification to show that the differences found in this subsection are mostly driven by students' propensity for violence. A detailed description of the equation and estimations is presented in Appendix 1.

4.3 Average Group Composition Effects on Emotional Regulation

Table 4 summarizes the main effect of average group composition obtained from specifications (3) and (4). In Panel A we present the ITT effects of being assigned to participate in the ASP with a particular composition of peers, compared to control group outcomes. In Panel B, we restrict the sample to treated students in order to identify pure differences of group composition regarding violence. Finally, Panel C presents the same results as Panel B, but separates the treatments into two groups by their initial propensity for violence: highly or less violent groups.

Using the full sample, we estimated specification (3) and obtained the results presented in Panel A. These estimations indicate that the results for students treated with similar violence-level peers are driving the effects on withdrawal behavior, locus of control and reaction to positive stimuli, compared to students assigned to the control group. Similarly, comparing students treated in violence-diverse peer groups to those in the control group, we find a reduction in their reaction to positive stimuli. We do not find statistical differences in the rest of the outcomes for both treatment arms.

Panel B summarizes the comparison in the effects among treatment arms. It's important to notice here that the only statistical difference between both HM and HT group compositions is in participants' stress. We are finding that when students are treated in homogeneous groups, their stress levels are greater than those of students treated in heterogeneous groups.

When combined, these two pieces of evidence indicate that both group compositions can have positive effects on different emotional regulation and socio-emotional skill measures; when compared to ASP non-participants. However, similar to Dinarte's (2018) findings, the second result particularly indicates that integration is a superior group composition, since tracking participants can increase their stress level.

Moreover, since there are two very different groups regarding violence in the HM treatment arm, specification (4) allows us to compare each HM subgroup with the corresponding HT subgroup by propensity for violence. These results are also reported in Panel C of Table 4. First, surprising but consistent with previous results (Dinarte, 2017), the increase in stress is greater for children treated in HM-High groups when compared to respectively similar children treated in heterogeneous groups. Second, looking at the rest of estimations on emotional regulation and the psychometric measures reported in column (3) Panel C, we can argue that the high-violence children in segregated groups *may* have worse performance than their similar peers treated in integrated groups.

To sum up the heterogeneous group composition effects, results indicate that tracking students can create unintended effects. These effects are caused since this group's composition increases stress for highly violent children treated in the segregated group; when compared to similar children treated in the integrated groups. This result explains one of Dinarte's (2018) findings in the group composition assessment section. Her results indicate that tracking increases the probability of receiving bad behavior reports for children with greater propensity for violence.

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

4.4 Group Composition Effects on Marginal Students

To directly measure the tracking effects (i.e to capture differences of being assigned to a homogeneous peer group of higher or lower propensity for violence) we can compare the two homogeneous subgroups and use specification (5).

In Table 5 we present estimations for the effects of tracking on marginal students. Panel A summarizes these estimations, controlling with a second order polynomial for a student's percentile in the IVV distribution (HM group) at each stratum. We find that assigning a marginal student to a group of peers with higher propensity for violence reduces her tendency for further reflect on response by 0.77 standard deviations. This finding compares to peers with similar propensity for violence, but who were enrolled with lower-violence peers. We do not find an effect in the rest of outcomes.

Panels B and C in Table 5 present estimations of the tracking effects on marginal students using third and fourth order polynomials. We can see in column (4) that the coefficients of CRT are lower than the estimations of a less flexible polynomial. Additionally, we find that there is a reduction in locus of control for marginal students assigned to the HM-High group compared to those assigned to less violent peers.

In summary, the marginal student is negatively affected by being assigned to a more violent group since they either increase their automatic responses or reduce their tendency to override incorrect responses while analyzing the correct ones. We also find that marginal students assigned to HM-High groups perceive that they can control their circumstances more than those assigned to HM-Low. However, this result may be explained by the fact that these participants are exposed to the most violent peers of their violence distribution function. Even after the intervention, such exposure increase their misbehavior at school (Dinarte, 2017).

TABLE 4. GROUP COMPOSITION EFFECTS ON EMOTIONAL REGULATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Arousal (stress)	Valence	Locus of control	CRT	Raven	Positive Valence Difference	Negative Valence Difference
FULL SAMPLE (Specific. 3)							
Homog. Group	0.163	-0.375*	-0.314***	0.017	-0.071	-0.458*	-0.278
	(0.124)	(0.205)	(0.103)	(0.139)	(0.108)	(0.240)	(0.204)
Heterog. Group	-0.099	-0.320	-0.131	-0.022	0.036	-0.424*	-0.185
	(0.128)	(0.202)	(0.154)	(0.148)	(0.174)	(0.229)	(0.300)
Observations	308	308	295	295	295	308	308
ONLY TREATED SAMPLE (Specific. 3)							
Homog. Group	0.257***	-0.048	-0.178	0.042	-0.114	-0.014	-0.084
	(0.071)	(0.257)	(0.128)	(0.114)	(0.129)	(0.283)	(0.343)
MDE Hom = Het	0.205	0.205	0.210	0.210	0.210	0.205	0.205
Observations	238	238	227	227	227	238	238
ONLY TREATED SAMPLE, BY HOMOGENEOUS SUBGROUPS (Specific. 4)							
Low Homog. Group	0.142	-0.263	-0.182	0.197	0.070	-0.184	-0.286
	(0.094)	(0.374)	(0.160)	(0.130)	(0.152)	(0.445)	(0.416)
High Homog. Group	0.336***	0.049	-0.181	-0.024	-0.245	0.025	-0.010
	(0.082)	(0.347)	(0.161)	(0.160)	(0.168)	(0.436)	(0.450)
Observations	238	238	227	227	227	238	238

***, **, * indicates that the effect of being treated in a HM (high or low) group compared to being treated in a HT group is significant at 1%, 5% and 10% respectively. 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 the 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 effects (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 5. EFFECTS OF TRACKING ON EMOTIONAL REGULATION
Only tracked subsample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Arousal (stress)	Valence	Locus of control	CRT	Raven	Positive Valence Difference	Negative Valence Difference
Second order polynomial specification							
Homog. Group	0.107	0.656	-0.279	-0.773**	-0.008	0.478	0.666
	(0.157)	(0.830)	(0.208)	(0.340)	(0.251)	(0.990)	(0.800)
Third order polynomial specification							
Homog. Group	0.147	0.768	-0.404*	-0.557*	0.026	0.631	0.757
	(0.139)	(1.078)	(0.227)	(0.316)	(0.328)	(1.288)	(1.153)
Fourth order polynomial specification							
Homog. Group	0.125	0.732	-0.418*	-0.599*	-0.003	0.580	0.726
	(0.147)	(1.107)	(0.238)	(0.340)	(0.361)	(1.329)	(1.207)
Observations	151	151	145	145	145	151	151

***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A presents 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 effects (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).

5. Conclusions

This study contributes to the understanding of channels relative to social programs that could be observed to foster socio-emotional skills among students of schools in highly violent communities. Consistent with previous studies, we show that social programs designed to foster non-cognitive skills are likely to affect emotional regulation. In other words, emotional regulation can be the mechanism explaining the educational and behavioral outcomes, which was observed in Dinarte (2017) for the same intervention.

We found evidence that the program is likely affecting mental state and self-reported test scores through changes in emotion. In fact, besides the positive qualitative evidence in favor of changing non-cognitive skills, it is not possible to correctly separate these effects. Particularly, we found that the program had a significant impact on emotional regulation and socio-emotional skills: treated participants face a reduction in the reaction to positive stimuli by 0.31 sd and in locus of control (a proxy of the increasing belief that one's life can be controlled) by 0.26 sd. This result is also interesting since the increase in withdrawal behavior – i.e. a reduction in positive valence – biases self-reported measures of non-cognitive skills downward. Such a phenomenon generates evidence that estimated results in behavior might fall into lower bounds of the intervention impact. In fact, the program might affect the self-perception of subjects while they are answering self-reported tests.

We also found heterogeneous effects by initial level of violence. Most of the effects might be driven by less violent children when compared to their respective control group. In terms of stress, the most violent students face an increase in stress compared to low-violence students. Similarly, for locus of control, both subsamples face a reduction though it is greater for low-violence students.

Additional interesting results emerge for group composition: there were greater effects in the homogeneous group compared to the control group on resting-state valence and locus of control; yet we do not find any effect on treated student in heterogeneous groups when compared to the control group. Moreover, the only difference for group composition is stress: treated students in the homogeneous group experience an increase in stress, compared to the heterogeneous group. These results are evidence of an increase in stress when children are with similar violence-level peers, which might be caused by an innate preference for peers diversity.

These behavioral findings are consistent with the asymmetrical impact on emotional responsiveness (DellaVigna, 2009a). We highlight the relevance of emotional disposition and modulation as a key mechanism to overcoming "Psychological Poverty Traps" (Haushofer and Fehr, 2014). According to Haushofer and Fehr (2014), "poverty causes stress and negative affective states, which may lead to short-sighted and risk-averse decision making; possibly by limiting attention and favoring habitual behaviors at the expense of goal-directed ones.

The methodology proposed in this paper has many benefits. First, it offers a way to incorporate emotion into the education and violence economics field. The importance of emotional regulation in life satisfaction has recently been highlighted for both developed and developing countries (Deming, 2017; OECD, 2015). This study shows that there are neurophysiological approaches to proxy emotional disposition and responsiveness with a high level of accuracy, and at a relatively low cost in violent communities. The results may also aid evaluation of similar programs oriented to improving non-cognitive skills.

Chapter 3

Unintended Effects of Public Infrastructure: Labor, Education and Crime Outcomes.

Abstract[1]

This paper studies the short-term impacts of infrastructure on economic activity, employment, education, and crime in the context of a developing country by analyzing the construction of a highway (NTH) in northern El Salvador. Using new data sources and two identification strategies – an instrumental variables approach and differences-in-differences – we find that this infrastructure improved specific economic activities in the region in the short-run. However, it also reduced participation in the formal labor sector of males and females between 15-19 years old, and increased drop out rates of boys between 12-16 years old. Our data allow us to provide evidence suggesting that these NTH's unintended effects are mainly driven by national gangs' arrival to newly connected and more prosperous municipalities. In fact, we find that districts near the NTH faced an increase in the short-term gang related crimes growth such as homicides and extortions.

Keywords: Public Infrastructure, Crime, Economic Activity, Labor, Education, Gangs.

JEL Codes: D72, H41, I1, I2, H7

[1]We are very grateful for the comments of Claudia Martinez, Francisco Gallego, Jeanne Lafortune, Francisco Pino, Micaela Sviatschi, as well as participants at the 1st Workshop in Urban and Regional Economics in Universidad Javeriana, 5th EH Clio Lab Annual Conference, SECHI Annual Meeting, and Universidad El Rosario. We also appreciate the support from information officers working in El Salvadoran government who provided us the data. All errors and omissions are our own.

1. Introduction

There is a wide strand of literature that analyzes public infrastructure as a booster for economic development. Empirical work, however, has just started testing this argument in a formal way. These papers have found that that access to road infrastructure reduces transportation costs, allowing economic agents to enter new and/or bigger markets. This opportunity then increase their trade opportunities and subsequently their income level or growth (Jacoby, 2000; Donaldson, 2018; Banerjee et al., 2012; Duranton et al., 2014; Faber, 2014; Jedwab and Moradi, 2016; Martincus et al., 2012). For example, Faber (2014) examines the impact of highway construction on industrial GDP in peripheral Chinese regions, finding positive effects of this road infrastructure. Additionally, Banerjee et al. (2012) explore the causal effect of proximity to transportation networks on per capita GDP levels across sectors in China, finding also positive long-term effects of infrastructure.[2]

However, there is no evidence that simultaneously studies the causal effect of road infrastructure on other developmental outcomes, such as education, labor and crime.[3] The simultaneous effect of proximity to road infrastructure on these three outcomes is not obvious. First, it could be argued that by reducing commuting costs, this infrastructure can facilitate the access to educational services. Likewise, it can make the formal or informal – criminal or non-criminal – labor markets more attractive for youth, increasing the opportunity cost of staying at school.

We also add "ugly" groups,[4] defined here as criminal organizations that finance their operations by exploiting the local labor force and generating economic inefficiencies. When we remove consideration of law enforcement in the agents' decision, the costs of committing crime become even lower and its benefits seems to be higher than those expected from the non-criminal labor sector. Then, the final equilibrium will be determined by the alternative that generates the maximum expected marginal benefit to individuals.

In this paper, we study the simultaneous short-term effects of public infrastructure on economic development outcomes such as economic activity, labor, education, and crime. We exploit the construction of a highway – the Northern Transnational Highway (NTH) – in a highly vulnerable region in El Salvador as source of exogenous variation in road infrastructure. Using both instrumental variables and differences-in-differences approaches to account for endogeneity in the relationship between development outcomes and distance to the road network, we identify and quantify the effects of this

[2]Further analysis of infrastructure's economic effects focuses on outcomes such as local trade (Duranton et al., 2014), individual sales and wages of high qualification workers (Michaels, 2008; Dinkelman, 2011), properties value (Agostini and Palmucci, 2008; Gonzalez-Navarro and Quintana-Domeque, 2016; Cellini et al., 2010) and firms' exports (Martincus, Carballo and Cusolito, 2012).

[3]Some papers study how infrastructure changes the demand of specific skills. For example, Michaels (2008) finds that a higher level of trade within cities near the new roads increased the demand for high-skill workers in manufacturing. Akee (2006) finds that road construction increased wages in the formal labor sector in rural areas and reduced self-employment in agriculture in Republic of Palau. Dinkelmann (2011) finds an increase in female employment caused by a massive electrification project in rural South Africa. Cattaneo et al. (2009) find evidence of the replacement of dirt floors with cement floors on children's cognitive development. Finally, Asahi and Dominguez (2016) find positive effects of subway proximity on robberies and larceny in Chile.

[4]This term has been inspired in Bruhn and Gallego's (2012) work, in which the authors follow Engerman and Sokoloff (1997, 2002) to classify colonial activities in three categories: "good," "bad," and "ugly". Specifically, they call "ugly" colonial activities as those rely primarily on the native population as an exploitable resource.

public infrastructure. Similarly to the existing literature, we find a positive effects of proximity to the highway on overall economic activity, driven by the relevant sectors in the connected municipalities. Unexpectedly, we also find that municipalities closer to the highway face: (i) a reduction in the labor supply growth of youth between 15 to 19 years old in the formal sector, (ii) an increase in drop out rates for boys between 12 to 16 years of age, and (iii) an increase in the growth of the number of homicides and extortions.

An additional novelty of this paper is that we provide causal evidence of the presence of "ugly" groups as a plausible mechanism that can be generating these effects. We argue that the greater economic activity due to road construction may have attracted local or external criminals, which in the Salvadoran context can be summed up as gangs. To test this argument, we compare gangs to non-gang crimes and find a difference between them. Then, considering that gangs recruit individuals between 12 to 25 years old (Aguilar and Carranza, 2008; Santacruz Giralt et al., 2001),[5] we argue that this criminal groups may have forced youths to either abandon the formal labor market or drop out school, driving results (i) and (ii) that were previously mentioned.

The NTH project was a partnership between the Millennium Challenge Corporation (MCC) of the U.S. and the government of El Salvador. The project's primary objective was to physically connect the most important municipalities in the northern region of El Salvador and to generate new economic opportunities for rural households (MCC, 2009). Specifically, this highway was supposed to reduce transportation costs, facilitating access to markets, both for agricultural and nonagricultural firms. Additionally, promoting households' productivity and increasing the diversification of their activities would generate an increase in the local income. Moreover, according to the project, households would have more access to public health services and education through the NTH, thus improving human capital and reducing poverty in the northern region of El Salvador (MCC, 2009).

Despite these good intentions, new infrastructure can also open areas to new crime, which can be driven by an increase in the local economic activity or resources availability (Asahi and Dominguez, 2016). These effects can be greater in contexts of weak institutions or lacking law enforcement like El Salvador. Specifically, crime and violence are highly relevant outcomes for it. Between 2002-2006, El Salvador's average homicide rate increased, ranging from 39.2 to 64.7 homicides per 100,000 inhabitants. During 2007-2008 there was a reduction, but it was still higher compared to other countries,[6] reaching 51.9 homicides per 100,000 inhabitants. There was another increase in the 2009-2012 period, where the country was the second most violent in Central America, with an average homicide rate of 69 deaths per 100,000 inhabitants (UNDP, 2013; IUDOP, 2015). Most of these homicides are officially attributed with local criminal organizations or gangs, which have been causing most of the violence in

[5]Crimes, specifically homicides, are very common for youths, particularly those living in countries with a similar contexts as that under analysis. For instance, the homicide rate for males aged 15-24 reaches 92 per 100,000 in Latin America, almost four times the regional average rate. Youth between 25 to 29 years of age, predominately males, are also the main perpetrators of crime and violence (World Bank, 2016)

[6]As a reference, the worldwide homicide rate is 6.2 per 100,000 inhabitants (UNODC, 2013). On average, Latin America and the Caribbean have an annual average of 24 homicides per 100,000 inhabitants. Note that the World Health Organization (WHO) considers a rate of 10 homicides per 100,000 inhabitants or higher to be characteristic of endemic violence.

most regions of the country during the last 20 years (Aguilar, 2007a).

To assess the causal effect of proximity to road infrastructure on development outcomes, the primary challenge we face is the existence of endogeneity issues. Estimating this relationship by OLS implies the assumption that the NTH was randomly assigned to municipalities in the northern region. However, this assumption does not hold given that the MCC project clearly established that the highway construction aimed to connect the most important municipalities in the region to other national markets (MCC, 2009). To address this issue, we implement the same Instrumental Variables approach (IV) as Faber (2014) and Morten and Oliveira (2016). We instrument the distance of the municipalities to the NTH using three different instruments: (i) Euclidian Spanning (ED), (ii) Least Cost Path (LCP) network, and (iii) Weighted Least Cost Path (WLCP). Similar to Faber (2014), the LCP network instrument is preferrable to the ED instrument because it yields more precise route predictions between any bilateral connection. Additionally, all our estimations include only municipalities that were in the path of the road but not intended to be connected.

Alternatively, exploiting the existence of the Pan-American Highway (PH) in the central-southern region of the country and the availability of data before and after the NTH construction, we use a Differences-in-Differences approach (D-in-D). We define as treated municipalities as those whose distance to the NHT is shorter than their distance to the Panamerican Highway and estimate the effect of this treatment on the same set of outcomes.

We construct an unique panel dataset of aggregated variables for the 262 municipalities in El Salvador. We collect and merge data for the 2006 to 2012 period – the years before and after the NTH construction – from different administrative sources. For instance, our data on highway construction was obtained from the MCC, the Technical Secretary of the government of El Salvador and the Ministry of Public Infrastructure. To address the fact that there is no GDP data to measure economic performance at the municipality level, we follow Michalopoulos and Papaioannou (2013) and use light intensity as a proxy for overall economic activity. We collect this data from the NASA's *Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS)* and geo-reference it at the municipality level. Additionally, we use tax revenues on commerce, services, and agrarian goods as a proxy of specific sectors' economic activity. This data was obtained from the Ministry of Finance. Educational data was collected at the school/municipality level from the educational censuses for the period of reference. The Salvadoran Social Security Office provided the formal sector employment data separated by gender and age groups. Finally, crime data was collected from the National Police of El Salvador (*Policía Nacional Civil*). They provided access to data on homicides, extortions, gangs detentions, robberies, thefts, and drug trafficking at the municipality level.

Our results indicate that municipalities closer to the NTH face an increase of both their overall economic activity and in the region's most important economic sectors, such as commerce and tourism. We find a short-term NTH-distance elasticity to economic activity and to taxes on commerce and tourism revenues of 0.31 and 0.56 respectively. These magnitudes are similar to existing evidence (Faber, 2009; Duranton, Morrow, and Turner; 2012).

We then turn to analyzing if a greater economic activity translates to an increase in labor demand, labor supply, or both. The argument here is that workers can experience a reduction in their transportation costs due to their closeness to the NTH and, therefore, can be more willing or able to participate in the labor market. However, we unexpectedly find that proximity to road infrastructure causes a reduction in youth labor supply, particularly for those between 15-19 years old. Specifically, we estimate a short-term NTH-distance elasticity to labor supply of 0.012 and 0.010 for boys and girls respectively.

This last result raises concerns of how the road infrastructure may have change this cohort's preferences for any of their outside options. The most plausible alternatives are participation in the education system, or enrolling in the informal – criminal or non-criminal – labor sector. First, we test whether they have enrolled in the educational system using growth in drop out rates. Using data at the school-municipality level, we find that municipalities closer to the NTH had increased drop out rate for boys between 12-16 years, with an estimated NTH-distance elasticity of 0.014. However we do not find any impacts on younger male students or girls at any age.

Thus, we move to test whether these youths may be increasing their participation in the informal criminal labor sector. This particular cohort has two alternatives: being an independent criminal or a gang member (Aguilar, 2007b). According to existing evidence, the first alternative is unlikely since in El Salvador – and the rest of countries in the northern triangle of Central America – local gangs have monopolized crime (Wolf, 2014; International Crisis Group, 2017). They are approximately 60,000 members, and they have indirect links with other 500,000 El Salvadorans. Additionally, we argue that being a gang member is more likely for this cohort due to gangs' recruiting process: they hire local teenagers and youth with around 15 years old who are living in neighborhoods where they operate (IUDOP, 2008; Aguilar, 2014).[7]

Although we do not have official data on gangs' territories or presence in different municipalities, we use characteristics of these criminal organizations to provide evidence of their presence and activities in the newly connected municipalities. For example, there is qualitative evidence that extortions and homicides are mostly committed by these groups (Aguilar, 2007) and this is how they finance their operations. This evidence also shows that after controlling a territory, gangs provide indirect protection to local communities from thieves (Aguilar and Carranza, 2008), and additional qualitative evidence indicate that they are more likely to be using drugs than trafficking them (Wolf, 2016; International Crisis Group, 2017).

Exploiting these factors, we compared gang and non-gang crimes, and find evidence that those municipalities closer to the NTH face a huge increase in gang related crime such as homicides and extortions rates. Our estimations indicate a NTH-distance elasticity to homicides and extortions of 0.14 and 0.31 respectively. Moreover, the NTH impact on other non-gang related crimes such as

[7]As we will explain later, gangs' emblematic crimes – homicides and extortions – are micro-territorial (Aguilar, 2007). Two main gangs characteristics that allow them to execute territorial control are size and recruiting process (Wolf, 2014; Crisis Group, 2017). They use violence to recruit or coerce those between 12 to 25 years of age as part of their workforce (Cruz, 2009; Santacruz Giralt et al., 2001). These specific gang targets allow us to evaluate their presence as a mechanism of the unintended effects of the NTH construction on development outcomes including crime.

robbery, thefts, and drug trafficking are not statistically significant. This result also supports the characterization of monopoly of criminality and violence exerted by these organizations.

Finally we provide evidence of some falsification tests to address potential concerns about: (i) the temporality of the NTH construction, and (ii) that there are not any differences between crime levels in the newly connected municipalities by the NTH compared to those previously connected by other roads. We do not find any effect of the NTH on the period previous to its construction or on communities that were already connected before the NTH.

Our results contribute to three strands of literature that analyze the impact of public infrastructure on different development outcomes. We further analyze these studies in the following section. The novelty of our paper is that we provide evidence of how road infrastructure can impact the topics of all three fields of literature. The first strand includes papers that measure the long- or short-term effects of the public infrastructure construction on economic activity and household income (Banerjee et al., 2012; Jacoby, 2000; Donaldson, 2018; Jedwab and Moradi, 2016; Duranton et al., 2014; Faber, 2014), house values (Agostini and Palmucci, 2008; Cellini et al., 2010), or trade of goods (Michaels, 2008).

We also contribute to a second category of evidence that studies the impact of public infrastructure on human capital such as children's cognitive development and health (Cattaneo et al, 2009), and labor markets (Akee, 2006; Michaels, 2008; Dinkelman, 2011; Das and Mocan, 2016; Gibbons et al, 2016).

Finally, this paper is also related to the literature on the link between employment and crime. The most recent research on this topic are the papers of Mishra (2012); Das and Mocan (2016), and Asahi and Dominguez (2016). Mishra (2012) focuses on infrastructure constraints in Delhi and how this influences criminal behavior. Das and Mocan (2016) evaluate rural public works program in India in 2006. They find a substitution effect between employment and crime; new jobs generated by the program had a negative impact on violent crime. Finally, Asahi and Dominguez (2016) find that closer proximity to subway stations in Chile increased both robbery and larceny in the short-term.

There are important policy implications of our results in the context of a developing and highly violent country. First, these results indicate that before implementing large investment in infrastructure in countries with high crime or low law enforcement levels, it is necessary to implement complementary policies oriented at strengthing institutions and lowering violence coming from criminal groups. Additionally, since we have learnt from the economics of institutions literature that the impact of institutions on long-run development is robust (Acemoglu, Gallego and Robinson, 2014; and Dell, 2010), reducing territorial expansion and persistence over time of criminal organizations on these countries' economic development is relevant.

The reminder of the paper is organized as follows. Section 2 presents a theoretical and empirical framework of road infrastructure effects on economic development, briefly describes the NTH's characteristics, and explains the gangs' structure and their features. Section 3 describes the data we use in this paper and our sample characteristics. Section 4 presents in more detail the two identification strategies we implement. In Section 5 we present our main results of the reduced form regressions using data of economic outcomes at the municipal level. A detailed evidence description of a plausible channel and falsification tests are presented in Sections 6 and 7 respectively. Finally, in Section 8 we present a brief discussion of our main results and conclude.

2. Background

2.1 Road infrastructure's theoretical effects on development.

Most of the existing evidence on infrastructure and economic development consistently indicates that access to roads is a key booster of economic growth and welfare. These papers have found that that access to infrastructure increase agents' – households, firms, regions or countries – access to new and/or bigger markets through the reduction of transportation costs. This then increases their trade opportunities and, subsequently, their income level or growth (Jacoby, 2000; Duflo, and Quian, 2012; Donaldson, 2010; Jedwab and Moradi, 2016; Martincus, Carballo and Cusolito, 2017).[8]

This evidence can be grouped by the time window used to study the effects. First, Banerjee et al. (2012) find positive long-term effects of proximity to better transportation networks on China's GDP.[9] In their setting, this economic improvement is explained by an increase on trade in a structure where capital is less mobile than traded goods. Similarly, Donaldson (2010) uses a general equilibrium trade model to study the long-term effects of a railroad network in India, finding that this infrastructure decreased trade costs, increased trade, and thereby regions' real income levels.[10] Second, regarding the intermediate- and short-term effects of road infrastructure, other papers also find that road infrastructure positively impact local or national exports both for developed (Duranton, Morrow, and Turner, 2012) and from developing countries (Faber, 2014).[11]

The proximity to roads may also affect further economic outcomes. Particularly, it can impact how individuals decide to distribute their time between acquiring human capital in the educational systems or increasing their participation in the employment sector. Also, within the labor market, infrastructure availability can change individuals preferences for being in the formal or informal[12] – criminal or non-criminal – sectors.[13]

[8]Particularly, Jacoby (2000) finds heterogeneous effects of infrastructure by impoverishment level in Nepal, with greater impact on the most deprived households.

[9]The authors evaluate a massive railroads projects built by the Chinese government and certain colonial powers to connect historical cities of China and particular treaty ports during the late 19th and early 20th century.

[10]In similar works that analyze long-term effects, Jedwab and Moradi (2016) find that railroads had large effects on the distribution of economic activity during the colonial period, and these effects have persisted to date, having large effects in poor countries.

[11]Additional studies analyze the causal impact of transport infrastructure on economic growth through increased property values (In Chile Agostini and Palmucci, 2006; and Gonzalez-Navarro and Quintana-Domeneque, 2016, for Mexico.)

[12]According to International Labor Organization statistics, the participation in the informal sector in developing countries is huge and shows a great heterogeneity. For example, the informal employment as a percent of employment for low and lower-middle income countries ranges between 17.6% up to almost 90%. These levels are mainly driven by the employment in the agricultural sector.

[13]There also exists evidence of the impact of construction or financing of infrastructure on other welfare outcomes such as health and political support. Cattaneo et al. (2009) find evidence of public financing for the replacement of dirt floors with cement floors on child health, the improvement of children's cognitive development, and a measure of self-reported happiness from adults. Regarding political support, Voigtländer and Voth (2014) analyze the construction of roads to assess their effectiveness in winning political support during the Nazi dictatorship. Using the same methodology we intend to exploit in this paper, they find increased political support for the Nazi ideology.

However, the differential effects of having access to road infrastructure on education and employment is not clear from the existing literature. One can argue that, by reducing commuting costs, this type of infrastructure can facilitate access to educational services. Notwithstanding, it can also make the labor market – criminal or non-criminal – more attractive for youths, increasing the opportunity cost of staying at school. The final equilibrium will be determined by the expected marginal benefit that the individual will obtain from both alternatives. Unfortunately, until now there is no rigorous evidence that simultaneously studies the causal effect of infrastructure on educational and labor outcomes.

Moreover, due to the weak institutions and the paucity of law enforcement in developing countries, the costs of committing crimes is low and the benefits can seem to be higher than those expected from the non-criminal labor sector. In that sense, people living in these countries face an important trade-off between participating in the criminal or non-criminal labor sectors, making crime a relevant issue in developing countries. For example, Das and Mocan (2016) evaluate a rural public works program in 2006 in India. They find that the employment generated by the program reduced violent crime and that this program operates as household insurance. Additionally, Asahi and Dominguez (2016) quantifies the effects of proximity to the subway network on crime outcomes in Chile. The authors find that closer proximity to subway stations increased both robbery and larceny in the public space within one year after the opening of the subway stations. Finally, from the existing literature we know that crime is associated with lower public schools enrollment for boys in several age groups (Dahbura et al., 2016).

If we add road infrastructure to the individuals' decision making process, there can be two results. First, roads can foster employment or education, and reduce the attractiveness of committing crimes. Second, this infrastructure availability can also enhance criminal activity, particularly because of the increase in the local economic activity.

All of the previous literature only addressed labor force swaps among sectors – formal or informal – or industries after an infrastructure shock. For example, Michaels (2008) analyzes the construction of U.S. Interstate Highway System on labor market outcomes and trade. He finds that a higher level of trade within cities near the new roads, increased the demand for high-skill workers in manufacturing, but reducing it in other places. Similarly, Akee (2006) finds that road construction increased wages in the formal rural sector and reduced self-employment in agriculture in Republic of Palau. Finally, Gibbons et al. (2016) find that new road infrastructure attracted transport intensive firms to an area but also negatively affected employment in incumbent firms.[14] However, there is no evidence at all of how infrastructure shocks modify people's preferences for participating in formal labor sector or committing criminal activities.

To isolate how important each of these channels is in explaining the impact of road infrastructure on development, it would be ideal to show what happens to: (i) economic activity, (ii) human capital

[14] Regarding electrification infrastructure, Dinkelmann (2011) investigates the impact of a massive electrification project in rural South Africa on employment growth. She finds heterogeneous effects regarding gender, specifically an increase in female employment explained by a rise in their entrepreneurship activities.

acquisition in the educational system, (iii) employment in formal sector, (iv) employment in the non-criminal informal sector, (v) employment in the formal criminal sector, and (vi) migration flows. However, data limitations only allow us to empirically analyze points (i), (ii), (iii), and (v) in this current paper.

We investigate in this paper whether road infrastructure improvements positively affect economic performance of a rural, poor and previously isolated region; whether it heterogeneously modifies the decision of participating in the educational system or in the formal employment for women and men; and how infrastructure indirectly affects individuals' decision to committing crimes in these areas. The results of these analyses substantially improve our understanding of infrastructure's impact in the context of a developing country with parallel criminal organizations.

2.2 Northern Transnational Highway (NTH)

In 2006, the US's Millennium Challenge Corporation (MCC) signed a five-year contract –between 2007-2012 – with the government of El Salvador. The main goal was to improve the lives of El Salvadorans through strategic investments in relevant development sectors, particularly with transportation infrastructure (MCC, 2009). One component of this contract was a connectivity project,[15] which sought to reduce travel costs and time within the northern zone.

Specifically, this project consisted of the rehabilitation and improvement of the Northern Transnational Highway (NTH), a two lane road with the aim of it functioning as a national transport artery connecting relevant municipalities in the northern region, and with its neighbors Honduras and Guatemala. The design was based on an existing project from 1976. Since then, here had been a goal to build a highway that connects all of the northern zone with these other two countries.[16] However, the final implementation did not include the sections that were going to connect with Honduras and Guatemala.

Northern El Salvador has been the poorest rural area in the country, particularly because it received most of the damages of the country's civil war during the 1980-1992 period.[17] Despite important economic growth after the Peace Agreement in 1992, there was no economic improvement in the northern zone, and its poverty rate was significantly higher than the national average. For example, the poverty rate in the northern region in 2011 was 48.4%, while in the rest of the country was 40.6%. The average schooling level in El Salvador in 2011 was 6.2 years, however in the northern zone it was only 4.7 years with a 21.9% of the population being illiterate, compared to a 12.8% at the national level. Additionally, 18.7% of northern households lived in extreme poverty in 2011, compared with

[15]Initially, the connectivity project included not only the construction of a full road network but also their connection with the NTH. However, due to budget redistribution and the existence of other interventions for connecting roads, this project was limited to the rehabilitation and improvement of the NTH (MCC, 2009).

[16]Between 1976 and 1977, the government of President Colonel Molina hired 10 companies to implement studies on the 10 sections that would form the NTH. Some of these sections were built, such as the highway part that connects Nueva Concepción with the deviation of Amayo in Chalatenango. However, due to the civil war, most of the plans were abandoned and were lost in the MOP archives. A decade ago, the National Commission for Development (CND) found the plans in archives of the Inter-American Development Bank (IDB) in Washington and coordinated with the same companies that had made the designs, managing to recreate the original design.

[17]Its mountainous territory was optimal to be the main staging ground for the conflict, thereby increasing violence and instability in the area and causing an exodus of large numbers of the region's inhabitants.

11.2% at the national level.

The NTH is 280.7 kilometers long and was built to increase the access to and connection with markets and the national economy for the most important municipalities in the region (MCC, 2009). In that sense, we argue that those non-targeted municipalities, located between the targeted ones, exogenously received an infrastructure shock, and we use this variation to measure the impact of infrastructure on development outcomes at the municipal level.[18]

According to the MCC's framework, the transportation costs would be reduced and the access to agricultural and nonagricultural markets would be increased. Thus, it would improve households' productivity and allow them access to a greater diversification of their activities. Additionally, households would have more access to public health services and education, which may generate an improvement in their human capital. These two results were expected to lead to an increase in households' income, alleviate poverty, and create regional economic growth (MCC, 2009).

We argue that the overall effect of the NTH's construction on economic development outcomes of the new connected municipalities is not clear. First, the NTH may have positively improved the economic activity of those municipalities closer to the highway, either by reducing transportation costs or by increasing demand for their products. However, this greater economic performance may have attracted gangs, which may have increased violence and criminality, committing homicides and extortions, and recruiting youths as a labor force for their criminal activities.

2.3 Gangs in El Salvador and their Illegal Operations.

El Salvador was the second most violent country in Central America during the 2009 to 2013 period. The average homicide rate was 69 deaths per 100,000 inhabitants (UNDP, 2013; IUDOP, 2015), almost three times the average homicide rate in Latin America. According to the WHO classification of violence as a health issue, El Salvador – and some other countries in the region – are categorized as being under endemic violence (WHO, 2012).

There is some heterogeneity in the distribution of homicides in the country. Figure 1 shows average homicides rate at the municipality level during the 2009-2013 period. This map highlights two characteristics in the regional distribution of homicides. First, municipalities in the central-southern region face greater rates than those in the northern region, which can be explained because they are closer to the main trading and commercial centers. Second homicide is uneven across the country; 92% of municipalities have homicide rates ranging from 10 to almost 194 homicides per 100,000 habitants.

As previously described, most of these homicides are officially attributed to local criminal organizations, which have been responsible for most crimes in most regions of the country during the last 20 years (Aguilar, 2007). Local gangs are rooted in the countries of Central America's Northern Tri-

[18]Qualitative evidence collected by a local newspaper indicates that effectively the original objective of the NTH was to connect specific municipalities. During an interview with Salvador Peña, major of Anamorós, he expressed that "The Fomilenio ends here, but in the beginning it was a connection, above all with Concepción de Oriente, passing through Nueva Esparta, connecting with Lislique. So, although they are blind spots, then, the influx of trade is occurring mostly with livestock and other items such as chemical fertilizers, concentrates and dairy products." (LPG, 2012)

angle.[19] A typical gang member in El Salvador is a young male around 25 years old who decided to participate into the gang at approximately the age of 15. Most of them are born into low-income and broken families, live in vulnerable neighborhoods (Cruz et al., 2016), do not have either secondary education or formal employment, and earn less than $250 USD per month (International Crisis Group, 2017). Their motivation for enrolling gangs are a need to belong to a particular group due to a sense of exclusion or lack of opportunities.

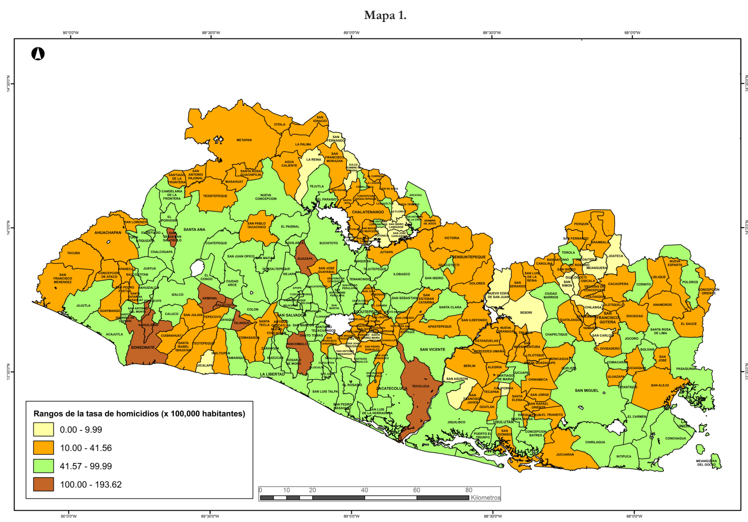


Figure 3.1: Average homicides rate in El Salvador at the municipality level.
Source: FUNDAUNGO (2015).

These criminal organizations emerged in El Salvador during the eighties. During that period, the country faced a civil war and economic crises that forced many El Salvadorans to emigrate to the U.S.[20] Once living in the U.S., immigrant youths banded together to protect themselves, creating the 18th Street gang (*Barrio 18*) and the *Mara Salvatrucha*, which later was know as MS-13.

In the mid-nineties, the U.S. increased deportations especially of gangs members. Unfortunately, during those years El Salvador was just emerging from its civil war, so the deportees faced limited school access and social services and no opportunities for reintegration into the formal sector, which made them look for illicit ways to obtain resources (Wolf, 2014).

Gangs have evolved into violent and complex criminal organizations. Currently the two largest and most violent groups are MS-13 and the two factions of 18th Street gang: *18-Sureños* and *18-revolucionarios*. They are better organized, have access to heavier weapons, and are more attractive

[19]The Northern Triangle is a geographic classification consisting of El Salvador, Guatemala and Honduras.

[20]The Central American immigrant population in the U.S. went from 354,000 in 1980 to 1.1 million in 1990. Most of them depended on low-wage job, and almost 21% lived below the poverty line. Many children and teenagers lived in disadvantaged urban neighborhoods, prominently in Los Angeles (International Crisis Group, 2017).

to recruits than the many smaller pre-existing street gangs, or *pandillas* (International Crisis Group, 2017).[21]

Gangs' emblematic crimes are micro-territorial (Aguilar, 2007). They extort businesses and individuals located in their territories as a source of revenue and impose deadly threats to reaffirm their control over those specific enclaves. There are two main characteristics of the gang structure that allow them to execute territorial control: size and recruiting process. First, estimations using official data show about 60,000 active gang members operating in El Salvador, and approximately 500,000 El Salvadorans – 8% of El Salvador's 6.2 million population – linked with these members as social support base (Wolf, 2014; Crisis Group, 2017).[22] They are also highly geographically dispersed. In 2008, each group had from 15 to 100 members, with an average of around 25 members.

Second, gangs use violence to recruit – voluntary or not – teenagers and youths as part of their workforce. They usually recruit men between 12 to 25 years old (Cruz, 2009; Santacruz Giralt et al., 2001), who constitute the labor force of the gang, and oriented to perform different functions from extortions to homicides. Women are often recruited when they are between 16 to 25 years old (IUDOP, 2010), but their main role in the gang is to be the "wife" of one of the gang members. Occasionally they could also work as labor force within the gang (IUDOP, 2010). These specific gangs targets allow us to evaluate their presence as mechanism of the unintended effects of the NTH construction on development outcomes including crime.

As the most important gangs' crime, extortion raises the most revenue since it is relatively easy to execute thanks to their enclave management, gangs' access to firearms, and lack of police control in those territories. The revenue level they are able to collect is massive: the MS-13 alone raises up to $31.2 million per year and extorts 70% of all the businesses in their territories (International Crisis Group, 2017). Most of their victims are small- and medium-sized business owners, informal tradespeople, and transport workers.[23] This extortion levels have been on the rise since 2015, and now affects 22% of firms in El Salvador, forcing to almost 67% of those businesses to close down each month as a result of inability to cover these additional "security" costs (FUSADES, 2016) .

Unlike gangs in Guatemala, El Salvadoran gangs are more associated with drug use than with drug trafficking according to existing qualitative evidence. Specifically, narco-traffickers employ them sporadically as muscle in some operations (Farah, 2011; Cruz et al., 2016).

[21] Although gangs may have been becoming more organized, there is no evidence that these groups are as specialized in their operations as are other transnational cartels.

[22] The social support base includes both active collaborators and ordinary citizens indirectly related to these groups, but who do not necessarily support them.

[23] Transport firms and their workers in particular have become targets of systematic intimidation and assassination, forced to pay for crossing gang-controlled territory. A total of 692 transportation workers were killed between 2011 and 2016 in El Salvador (International Crisis Group, 2017).

3. Data and Sample Characteristics

3.1 Data

For the main analysis of the road infrastructure effects on economic, labor, and crime outcomes, we construct an unique panel dataset of aggregated variables for the 262 municipalities in El Salvador. We collect and merge data for the 2006 to 2012 period – the years before and after the NTH construction – from different sources.

First, following the economic literature of infrastructure impacts on economic performance (Jacoby, 2000; Duflo, and Quian, 2012; Donaldson, 2010; Jedwab and Moradi, 2016; Martincus, Carballo and Cusolito, 2017), we needed a measure of economic performance such as GDP. However, the first challenge was that there is no disaggregated GDP data at the municipal level. To overcome this, we follow Michalopoulos and Papaionnou (2014) and Baires (2017) and collect high resolution satellite data of night light density as a proxy for municipal economic activity. This data comes from the images reported by the *Defense Meteorological Satellite Program's Operational Line-scan System (DMSP-OLS)*[24]. Then, using the ArcGIS software, we geo-reference the average pixels within the boundaries of the municipalities in El Salvador.

Additionally, previous evidence have found that the economic impact of the NTH may be greater on specific sectors. For instance, Jedwab and Moradi (2013) find that the effects of infrastructure was greater on the commercial and agriculture sectors in Ghana. In that sense, we may expect heterogeneous effects by economic relevance of particular sectors in the northern region. To measure these effects, the Ministry of Finance provided administrative data for taxes revenues on commerce, services, tourism, and agricultural production at the municipal level. Then, we use these data as proxies for sectorial economic activity. We only include those municipal taxes that the local authorities receive as a direct revenue source.[25]

To contribute to the literature studying the effect of road infrastructure on labor outcomes (Michaels, 2008; Akee, 2006; Gibbons et al, 2016), we add to our panel data on formal labor market participation at the municipality level. As explained before, the gangs' recruitment process is highly micro-territorial and targeted at specific groups of age. To exploit these features and to provide evidence that the expected displacement of youth labor force from the formal sector can be driven by the gangs' arrival in the northern municipalities after the NTH construction, we collect the data on the number of workers in formal labor sector separated by gender and age groups. This detailed and unique information was provided by the Salvadoran Social Security Office (ISSS, *Instituto Salvadoreño del Seguro Social)*.

Further exploring the effects of the NTH construction on educational outcomes, we add geo-coded data from all 3,344 public schools in El Salvador matched to each municipality. This data was obtained from the annual Educational Census for the 2006-2012 period, conducted by the Ministry of

[24] This data is captured at night at a height of 830 km. The measure is a six bits digital number (0-63) calculated for each 30 second output pixel that is averaged with respect to the overlapping input pixels and with all of the valid nights in during the year (Baires, 2017)

[25] In that sense, we excluded VAT and income taxes, which are collected at the national level by the Ministry of Finance.

Education.[26] Similar to our approach for labor outcomes, we want to further analyze any displacement of youth either from the formal labor sector or from the educational system. In that sense, we collect the educational data separated by gender and educational levels. Each annual census includes data on the number of students that left school and the number of enrolled students at the beginning of each academic year. Using these variables, we estimate drop out rates for each year as the share of students who left school compared to the total number of enrolled children in each school.

Our criminal data section come from different offices of the National Police of El Salvador (PNC, *Policía Nacional Civil*). First, we collected data on crimes separated by their likelihood of being perpetuated by gangs. Among those gang crimes, we include number of homicides and extortions at each municipality during the period 2006-2012. In the non-gang category of crimes, we include thefts, robbery and drug trafficking (Farah, 2011; Cruz et al., 2016). Specifically, as measures of drug trafficking we use number of drug confiscations, amount of drugs confiscated (in US$), and the number of captures related with drugs activities during the 2006-2012 period.[27]

To control for changes in population, we have estimated population growth rates at the municipal level using data from the National Household Survey (EHPM, *Encuesta de Hogares y Propósitos Múltiples*) for the 2006-2012 period and the 2005 Census. To control for geographic variables, we use data of elevation obtained from the National Registration Center (CNR, *Centro Nacional de Registros*). Finally, we control for crime trends and crime-deterrence policies using: (i) indicator variables of whether a municipality was part of the *"municipios santuarios"* policy implemented by the government of El Salvador during the gang truce in 2012, and (ii) homicides growth rate in the previous period.[28]

3.2 Sample Characteristics

Table 1 presents descriptive statistics of key variables used in our analysis. We present means and standard deviations for the 160 NTH connected districts in columns (1)-(2), respectively. These communities corresponds to those within 40 kilometers on both sides of the NTH, which we call the "newly connected municipalities". We grouped variable by outcome categories – economic activity, labor force, drop out rates, and crime – for the years before (2009) or at the end (2012) of the NTH construction.[29]

Municipalities in the sample are rural and most of them produce agrarian goods for self-consumption. Most of the municipal tax revenues come from commerce and tourism and services. On average, local governments collected more than $114.000 of tax on commerce and tourism and $25.000 of tax services before the NTH construction. In terms of participation in the formal labor sector, the average participation for both male and female in the 15-19 years cohort is small. The most important

[26]All datasets from the Educational Census for the 2006-2016 period are publicly available at the following link: http://www.mined.gob.sv/index.php/estadisticas-educativas/item/6116-bases-de-centros

[27]Since the coefficients estimated using the three measures are similar, in this paper we report only the effects of the NTH on the number of drug confiscations. The rest of estimations are available upon request.

[28]We also obtain data on early childhood health from the Ministry of Health of El Salvador for the same period of analysis. Specifically, they provide data on child mortality and morbidity in the child's mother's municipality of residence.

[29]A more complete description of the variables construction is provided in Appendix 1.

TABLE 1. DESCRIPTIVE STATISTICS OF MAIN OUTCOMES

	NTH PROJECT MUNICIPALITIES (N=160)	
	Mean	std. dev
	(1)	(2)
Economic Activity (in USD)		
Taxes on commerce and tourism, 2009	\$114,095	\$516,940
Taxes on services, 2009	\$25,269	\$116,959
Taxes on agriculture, 2009	\$809	\$9,033
Taxes on commerce and tourism, 2012	\$156,011	\$725,487
Taxes on services, 2012	\$30,600	\$118,827
Taxes on agriculture, 2012	\$301	\$1,933
Workers in the formal sector, by cohort.		
Male, 15-19 years old, 2009	8.2	22.8
Male, 15-19 years old, 2012	8.9	24.5
Female, 15-19 years old, 2009	5.5	16.9
Female, 15-19 years old, 2012	6.0	18.2
Male, 20-25 years old, 2009	428.5	1139.8
Male, 20-25 years old, 2012	461.4	1227.5
Female, 20-25 years old, 2009	321.4	964.1
Female, 20-25 years old, 2012	346.1	1038.3
Crime outcomes.		
Number of homicides, 2009	71.2	38,7
Number of homicides, 2012	41.2	16,1
"Sanctuary" municipalities	0,04	0,21
Number of robberies, 2009	24,2	84,7
Number of robberies, 2012	20,2	67,7
Number of thefts 2009	36,8	111,6
Number of thefts 2012	35,6	117,8
Drugs trafficking, 2009	0,9	2,8
Drugs trafficking, 2012	6,2	16,5

Table 1 shows descriptive statistics of the available variables for the 160 municipalities within 40 kilometers on both sides of the NTH, which we call the "newly connected municipalities".

participation in the formal labor sector before the NTH was from the 20-25 years old cohort, with an average of 428 and 321 male and female, respectively.

Finally, in the last part of Table 1 we present means and standard deviations of crime measures. On average, previous to the NTH construction, the average number of homicides and robberies were 71 and 24, respectively. Only 4% of these municipalities were part of the anti-crime *municipios santuarios* Program.

4. Empirical Strategy

As pointed out by Dinkelman (2008), an important challenge when evaluating the economic impacts of infrastructure is to control for confounding effects of the new infrastructure, such as changes in the regions' economic performance. More specifically, we may think that simple correlations between the distance to the NTH and development outcomes could produce biased estimates of the causal effect of infrastructure on development outcomes since it is likely that the roads were constructed in places with either higher population densities and higher economic activity or in less connected and poorer municipalities. In this section we outline two different empirical strategies that deal with endogenous infrastructure placement and these confounding factors in alternate ways: instrumental variables (IV) and differences-in-differences (D-in-D) approaches.

First, we use the data described in the previous section to estimate the short-term effects of NTH on changes of outcomes of interest between 2009 to 2012 period. The baseline estimation strategy is a specification of the form:

$$\ln(Y_{ip}^{2012}) - \ln(Y_{ip}^{2009}) = \beta \ln(\chi_{ip}) + r_p + X_{ip}\eta + \epsilon_{ip} \tag{3.1}$$

where Y_{ip}^{T} is an economic, labor, education or crime outcome in municipality i in region p in the period $T \in \{2009, 2012\}$. $\ln(\chi_{ip})$ stands for the logarithm of distance to the NTH segment in 2009 measured from the centroid of each municipality unit. We also include a set of region fixed effects, r_p, so that all comparisons across connected and non-connected districts occur for communities in the same local areas. X_i is a vector of municipality control variables that includes demographic, economic, geographic, and crime controls, and the growth of the dependent variable in the previous period i.e., between 2006-2009–. Finally, ϵ_{ip} is the municipality specific error term. To control for heteroskedasticity, we implement the robust Huber-White estimator for the standard errors.

4.1 IV Estimation Strategy

Estimating specification (1) by OLS implies the assumption that the NTH was randomly assigned to municipalities in the northern region. However, this assumption would be wrong given that the MCC project clearly established that the highway was built to increase the access to and connection with markets and the national economy for the most important municipalities in the region (MCC, 2009).

In that sense, if the planners indeed targeted the NTH to communities with a greater (lower) economic growth potential or that were politically (less) important, then the outcome growth NTH distance elasticity, β_{OLS}, would be biased upwards (downwards).

We try to address these concerns including in specification (1) the controls we previously described. We also estimate it excluding the municipalities that were selected to be connected by the planners. However, even after including these two modifications, confounding municipality level economic potential and unmeasured political factors that could affect NTH placement are still of concern.

To address these challenges, we construct two hypothetical minimum spanning highway networks as instruments for actual route placements: Euclidian Spanning (ED) and Least Cost Path (LCP) networks following Faber (2009) and Morten and Oliveira (2016). These instruments correspond to the questions of which routes central planners would have constructed if the policy objective had been to connect relevant municipalities in a single network while keeping construction costs to a minimum (Faber, 2009).

Following Faber (2009), to compute the LCP we first estimate the least cost highway construction path between the targeted pairs of municipalities using layers on land cover and elevation, to predict a cost function (Jha et al., 2001; Jong and Schonfeld, 2003). Then, we implement a Dijkstra's optimal route algorithm to construct least cost construction paths between all possible bilateral municipalities. Finally, we incorporate the estimated costs in a Kruskal's minimum spanning tree algorithm to obtain the LCP connecting the six municipalities. Figure 2 shows the results after applying this algorithm. As we can see, the estimated LCP spanning network optimally connects targeted municipalities surrounding elevated spaces and water bodies.

We also construct the ED computing euclidian distances between the targeted municipalities. Then, we estimated the Kruskal's algorithm subject to the minimization of total network distance. In our setting, the final solution were simply the number of segments. Both segments are shown in Figure 3.

Then, as in Faber (2009), Attack et al. (2009), and Morten and Oliveira (2016), we instrument the distance from each municipality i's centroid to the NTH with their respective distance to either the LCP (η_{ip}^{LCP}) or ED (η_{ip}^{ED}). This will address the concern of non-random local highway placements on the way between targeted municipalities centroids.

An additional concern is that the community's centroid usually concentrates most of its economic activity. Therefore, instead of measuring the average effect of the road infrastructure at the municipality level, we can be capturing its upper bound (Galarraga and Dinarte, 2015). That can be an important concern in huge or highly heterogeneous districts. However, communities in El Salvador are pretty small – 262 geographic divisions in a 21,000 Km^2 territory – and internally highly homogeneous. However we also calculated a weighted distance to their LCP as an additional instrument η_{ip}^{WLCP} to address this concern. This WLCP was estimated as the average of all minimal distances between each municipalities' surface pixels units and the NTH.

To sum up, our IV strategy will include a system of equations to be estimated, in which specification (1) will be our second stage and our first stage will be:

$$\ln(\chi_{ip}) = \gamma_1 \ln(\eta_{ip}^H) + r_p + X_{ip}\gamma_2 + \mu_{ip} \tag{3.2}$$

where $H \in \{ED, LCP, WLCP\}$. We include the same set of control variables X_{ip} and regional fixed effects r_p as in the second stage.

It is important to notice that the exclusion restriction could be violated if locations along the LCP or ED paths are correlated with economic, demographic, or criminal characteristics due to history and sorting. To address this concern, we include a set of controls, X_{ip}, for pre-existing municipal level crime status or economic performance.

Finally, as we can see from Figure 3, and similar to Faber (2014), the LCP network instrument is preferred to the ED instrument because it yields more precise route predictions between any bilateral connection due to its use of land cover and elevation data.[30]

Summing up, the baseline identifying assumption is that the distance of each non-targeted municipality to the LCP affects economic development and criminal activity growth only through municipalities' distance to the NTH, conditional on baseline municipalities characteristics and regional fixed effects. As in Faber (2009), this implies that those municipalities that were in the path of the road, but were not specifically chosen to be so, are those that receiving randomly this infrastructure and capture the impact we want to measure. Therefore, conditional on instrument validity, the estimated coefficient β from specification (1) captures the local average treatment effect (LATE) of road infrastructure on municipal level development and crime outcomes growth.

4.2 D-in-D Estimation Strategy

We complement the IV strategy with an alternative D-in-D estimation approach taking advantage of the availability of a panel dataset. In this estimation, we consider the pre-NTH construction period to be 2006-2009 and the post-NTH is the 2009-2012 period. Then, to define which municipalities are in the treatment or comparison groups, we exploit the existence of the Pan-American Highway (PH) in the central-southern region of the country.[31] Before the NTH, the PH was the main road artery in the country, connecting cities whose main economic activity was commerce and trade of goods such as Santa Ana, San Miguel, and Santa Rosa de Lima.

In Figure 4 we present the two main highways crossing all the country. Municipalities closer to the PH were already connected before the NTH construction and it is highly unlikely that the NTH would increase their ability to transport their goods or to commute to main financial locations. Under this assumption, we define treatment as the relative closeness of each municipality to the NTH. More specifically, a municipality i will be classified as treated if its Euclidian Distance to the NTH is shorter that the same measure to the PH.

[30]Moreover, in the Appendix we provide further evidence that our main estimations are more precise when we use LCP compared to ED.

[31]The Pan-American highway connects 13 countries from north to south. Its extension is approximately 48,000 kilometers long. The section that passes through all the Central American countries is called CA-1. In El Salvador, it passes through the cities of Santa Ana, Santa Tecla, Antiguo Cuscatlán, San Salvador, San Martín, San Miguel, and Santa Rosa de Lima. This is the main artery moving South American goods northwards.

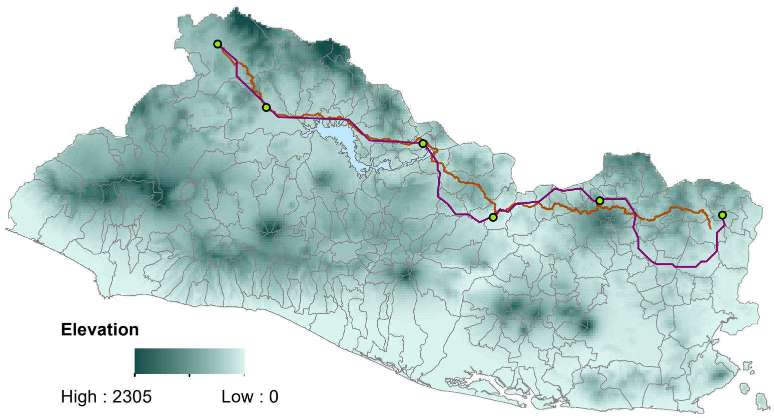


Figure 3.2: **Elevation and water bodies in El Salvador.**
Red line indicates the NTH network and purple line shows the LCP spanning tree.
Source: Authors' estimation using data provided by the Ministry of Public Infrastructure (MOP, 2009) and the ArcGis software.

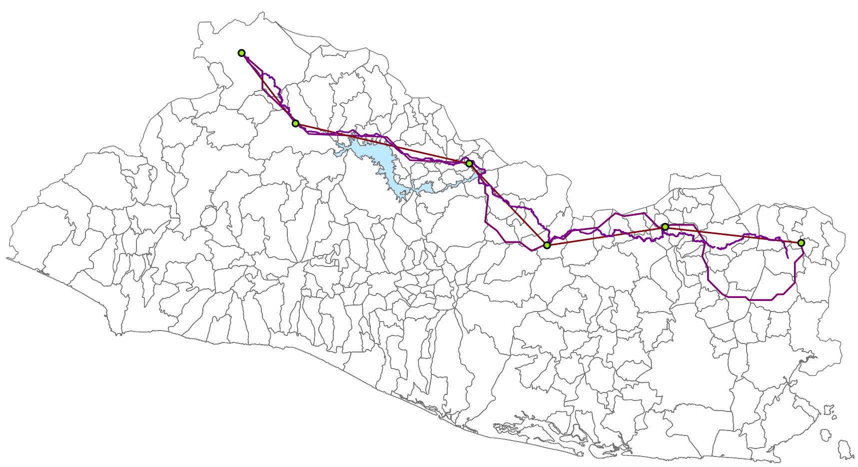


Figure 3.3: **NTH network, ED and LCP spanning trees.**
NTH road in the northern region of El Salvador (light purple line), Euclidian distance path (red line), and Least Cost Path spanning tree (dark purple line).
Source: Authors' estimation using data provided by the Ministry of Public Infrastructure (MOP, 2009) and the ArcGis software.

Considering these definitions, we regress each development and crime outcome (G_{ip}) growth during the pre- and post-NTH construction periods using the following specification:

$$\ln(Y_{ip}^{t+1}) - \ln(Y_{ip}^{t}) = \alpha_0 + \alpha_1\tau + \alpha_2 T_{ip} + \alpha_3(\tau \cdot T_{ip}) + X_{ip}\eta + \epsilon_{ip} \tag{3.3}$$

where τ is a dummy indicator of the post-NTH construction period; T_{ip} is relative closeness treatment indicator, according to our definition previously described; and X_{ip} is a vector of municipality level control variables in which we include the same demographic, economic, geographic and crime measures as in the IV estimation strategy.

The identification strategy is that before the NTH construction, connected and non-connected municipalities had similar trends in their economic, labor, educational, and crime outcomes. Under that assumption, α_3 indicates the average treatment effect on development and crime outcomes of being relatively closer to a new road infrastructure compared to being previously connected to a already existing one.

In Table A1 in the Appendix we present baseline characteristics of municipalities in the treatment or control groups. Both groups are similar on taxes revenues and in being part of the anti-crime national program *municipios santuarios*. However, as may have been expected, previously connected districts have a greater participation of individuals in the formal sector and, on average, their crimes levels were higher than those in the newly connected municipalities, before the NTH construction.

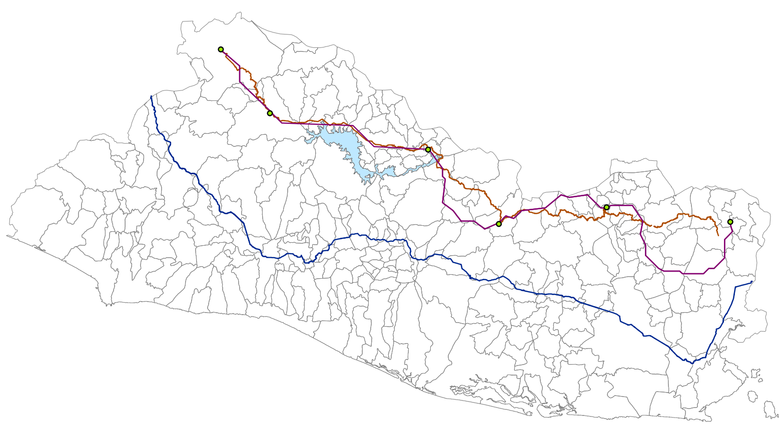


Figure 3.4: **NTH network, Pan-American highway and LCP spanning tree.**
NTH road in the northern region of El Salvador (red line), the Pan-American highway (blue line) and the spanning tree of our preferred instrument LCP (light purple line).
Source: Authors' estimation using data provided by the Ministry of Public Infrastructure (MOP, 2009) and the ArcGis software.

5. Main Results

This section reports results of specification (1) using OLS and IV approaches for different municipality level outcomes. We first present first stage IV results of our preferred instrument (LCP). First stage coefficients using alternative estimated instruments are presented in the Appendix. We then present estimated coefficients using OLS and second stage IV, separated in two outcome categories: (i) economic activity measures; and (ii) human development outcomes such as participation in the formal labor sector and school drop out rates. Estimated coefficients using D-in-D approach are presented in the Appendix.

5.1 IV Estimation: First Stage

For the IV approach, we show the first stage results for the LCP network instrument – specification (2) – in Table 2. We report first stage results for the log distance of each municipality centroid to the nearest NTH segment.

We try several models in order to test the robustness of our first stage. The coefficients shown in columns (1) - (3) were estimated including targeted municipalities. Then we drop them from the estimation and present the coefficients in columns (4) - (6). These last three specifications are our preferred ones because they include the sample of municipalities that quasi-randomly received the road infrastructure as in Galarraga (2015) and Attack et al. (2009).

Coefficients in columns (1) and (4) are estimated without controls, and those presented in columns (2) and (5) were calculated including some of the following control variables: economic activity in the previous period using light density growth, population growth, log of elevation, and indicator of a national anti-crime policy. Finally, we estimate the same specifications as before, but we also add regional fixed effects. These estimations are shown in columns (3) and (6).

We present the coefficients of all the previously described models using the LCP network. As we can see from Table 2, and similar to our graphic assessment of ED and LCP spanning trees, the LCP network is strongly significant as a within municipality predictor of the actual NTH road conditional on the full set of pre-existing demographic, geographic, and economic municipal characteristics. Moreover, coefficients remain very similar when we add controls and fixed effects.

After dropping targeted municipalities, the size of the coefficient does not change substantially with the addition of more controls or region fixed effects, while the precision of the estimate improves. We also can see that the LCP instrument is a more precise predictor of networks on any given bilateral connection, similar to Faber (2009). In Table A2 in the Appendix we present first stage estimated coefficients using ED and WLCP as instruments.

Coefficients of the preferred model provide evidence that regions selected to receive infrastructure meet the features defined by the MCC project (MCC, 2009). First, even when we do not find evidence that poorer regions are benefiting from this infrastructure, the sign of the economic activity coefficient indicate that more wealthy municipalities may be less likely to benefit by the connectivity project. Moreover, regions with greater elevation – a characteristic of the northern region – are more likely to benefit from the NTH. Finally, municipalities that were on the path of the NTH were less likely to be part of the *municipios santuarios* policy, which are targeted to the most violent communities in the country. These results are relevant for our analysis for two reasons: (i) using OLS estimation strategy may provide downwards biased estimates because it seems that economically disadvantaged communities are benefiting from the infrastructure; and (ii) this estimation provides evidence that before the NTH, these municipalities had lower crime levels and therefore it is more likely that the wealth generated after the highway construction attracted criminal activity.

TABLE 2. FIRST STAGE REGRESSIONS

Dependent variable: Distance of municipalities to the NTH

	(1)	(2)	(3)	(4)	(5)	(6)
Least cost path IV	0.596***	0.594***	0.571***	0.725***	0.732***	0.719***
	(0.108)	(0.111)	(0.113)	(0.058)	(0.053)	(0.056)
Economic activity $(t-1)$		-0.023	0.042		-0.001	0.043
		(0.114)	(0.100)		(0.105)	(0.094)
Population growth		3.370***	3.313***		3.217***	3.060***
		(0.905)	(0.905)		(0.835)	(0.817)
Log(Elevation)		-0.196***	-0.232**		-0.251***	-0.255***
		(0.075)	(0.094)		(0.063)	(0.083)
Sanctuary municipalities		0.175	0.064		0.184**	0.075
		(0.146)	(0.129)		(0.084)	(0.123)
Obs	160	160	160	154	154	154
R^2	0.517	0.574	0.592	0.595	0.658	0.678
First stage F-Stat	29.88	12.08	26.17	152.16	47.70	42.57
Region FE	NO	NO	YES	NO	NO	YES
Connecting municipalities	YES	YES	YES	NO	NO	NO

*, **, ***, significant at 10%, 5% and 1%. Robust Standard Errors in parenthesis. All the estimations include only the municipalities located within 40km to the NTH. In columns (1) and (4) we estimate an OLS without controls. In the rest of estimations, we include as controls: municipality's population growth rate during the 2009 - 2012 period, political ideology of the major in office, geography control (log elevation) and a dummy indicating whether the district was part of an anti-crime program called *municipios santuarios* implemented by the El Salvadoran government. In some specifications we include regional fixed effects. In columns (4) - (6) we exclude 5 municipalities which were intended to be connected by the highway. Economic activity is measured using light density (Michalopoulos and Papaioannou, 2011).

5.2 Economic Effects of Road Infrastructure: OLS and IV Results

Table 3 presents OLS and IV coefficients of regressing short-term log changes of municipal level economic outcomes on log distance of the municipality?s centroid to the NTH. Columns (1) to (4) in the table present OLS results, and columns (5) to (8) present the IV estimations. We measure economic performance using two set of outcomes. To calculate the overall economic activity, we use changes in light density as a proxy. In the IV estimation, we find that, conditional on baseline characteristics at the municipality level and region fixed effects, communities closer to the NTH face an increase in their economic activity. As we can see in column (5), the estimated short-term NTH-distance elasticity is 0.31, similar to existing evidence (Faber, 2009; Duranton, Morrow, and Turner, 2012).

We then turn to study the short-term effects of road infrastructure on specific sectors - -similar to Jedwab and Moradi (2013) – using log changes of taxes revenues on commerce, services, tourism and agricultural production at the municipal level. We find that due to the construction of the NTH, there is an increase in tax revenues from the tourism and commerce sector, with a NTH-distance elasticity of tax revenues of 0.56 and no impact in other economic activities such as services and agriculture.

An important pattern that bears mentioning emerges from Table 3: IV point estimates of NTH-distance elasticity are greater than the OLS estimates. This was somehow expected after looking at the first stage estimates. We have found that being on the path of the highway was negatively correlated with economic performance and positively with crime, therefore using OLS may estimate downwards biased coefficients.

Summing up, these results indicate that the NTH increased the overall economic activity in the northern region. Moreover, it brought an additional demand for goods and tourism in this region, which, before its construction was less exposed to external demand, from other municipalities. This may have been due to a reduction in transportation costs. These results remain when using the D-in-D approach, except for a change in tax revenues from services. Using D-in-D, we find a positive and statistically significant causal effect of road infrastructure on taxes from services. These estimations are presented in Table A2 in the Appendix.

No effects on agricultural tax revenues were expected because households in this region produce agricultural goods only for self-consumption. It is also plausible that there was a substitution between agricultural and tourism and commerce activities. Even when this region was agrarian, the reduction in transportation costs may modify preferences for land use, changing the use of land to other more profitable activities than agricultural goods production, such as constructing hotels and stores.

5.3 Road Infrastructure Effects on Human Capital: OLS and IV Results

Due to the increased economic activity, it may be expected an increase on labor demand, labor supply, or both. Workers can experience a reduction in their transportation costs cause by their closeness to the NTH and, therefore, can be willing to participate in the labor market. Simultaneously, as we showed before, closeness to NTH increases economic activity and hence can also increase labor demand. In any case, we may expect an increase in the number of workers participating in the labor market.

We explore the short-term effect of highway construction on the growth of jobs in the formal sector separated by gender and age cohorts. Results are shown in Table 4. As explained before, columns (1) - (4) are the coefficients estimated using OLS and columns (5) - (8) present the coefficients estimated through IV approach. All estimations include the same baseline controls as previously described and regional fixed effects.

TABLE 3. EFFECTS OF ROAD INFRASTRUCTURE ON ECONOMIC ACTIVITY

Dependent variable: Δ_t economic activity: light density and municipal taxes revenues.

	OLS REGRESSION COEFFICIENTS				IV REGRESSION COEFFICIENTS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Economic activity (light density)	Taxes on services	Taxes on Commerce and Tourism	Taxes on agriculture	Economic activity (light density)	Taxes on services	Taxes on Commerce and Tourism	Taxes on agriculture
ln($Dist$-NTH)	0.040***	0.018	0.492**	-0.030	0.307***	0.704	0.558**	0.016
(−)	(0.007)	(0.293)	(0.197)	(0.066)	(0.115)	(0.723)	(0.233)	(0.068)
Baseline controls?	Y	Y	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	4166	154	154	154	4166	154	154	154

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. Estimations include only the municipalities located within 40Km to the NTH. Columns (1) - (4) are estimations using OLS and (5) -(8) are estimated using IV. All models exclude connecting municipalities and include regional fixed effects and the following controls: municipality's population growth rate during the 2009 - 2012 period, political ideology of the major in office, geography control (log elevation) and a dummy indicating whether the district was part of an anti-crime program called *municipios santuarios* implemented by the El Salvadoran government.

Unexpectedly, IV results indicate that municipalities closer to the NTH face a reduction in the male and female labor force participation in the formal sector for the cohort between 15 to 19 years of age. The estimated short-term NTH-elasticities are similar between men and women (0.012 and 0.010, respectively), and lower than existing evidence (Michaels, 2008; Akee, 2006; Gibbons et al., 2016; Dinkelmann, 2011).[32] There is not any statistically significant result for older cohorts.

Using a D-in-D strategy, the results of the interaction between benefiting from the road are similar in sign for the estimated elasticities for both boys and girls between 15 to 19 years of age. The only difference using this approach is increased growth in the number of workers from the cohort aged 20 to 29 years old. However, this result does not invalidate the previous conclusion, but may indicate a swap in labor force between cohorts; the older cohort is off setting the reduction in the labor supply of the younger group. These results are described in Table A4 in the Appendix.

We then turn our analysis to ask about the other available alternatives for this cohort in the northern region after the NTH construction. Theoretically, the reduction in transportation costs could be attractive to them then education. Moreover, the increased economic activity may have attracted them to working in the informa, but non-criminal, labor sector such as working on entrepreneurial activities, opening small stores, etc. However, according to existing evidence (Aguilar, 2007) El Salvadoran youth between 10 to 19 years old are highly susceptible to gang recruiting.[33]

To evaluate if the NTH modified adolescents' and youths' preferences for staying in the educational system, we measure the NTH effect on drop out rates at the school level and estimated the standard errors using clusters at the municipal level. Estimated coefficients using both OLS and IV approaches are presented in Table 5, in columns (1)-(4) and (5)-(8) respectively. We include regional fixed effects and baseline controls at the municipality level as before. Differences in the number of observations are because some schools do not include all educational levels.

We find in our IV estimation that municipalities closer to the NTH faced an increase in the drop out rate of boys enrolled in 6th to 9th grade, i.e. between 12-16 years old. We found no impact on drop out rates for younger male students and for girls at any age. The estimated short-term NTH elasticity of the drop out rate is 0.014 for the 12 to 16 years old male cohort.[34]

[32]Despite the small magnitudes of the coefficients, these results have important implications in the El Salvadoran context. Using a household survey (EHPM, 2012), we estimate a total of 34,151 working youth in the formal sector between 15 to 19 years of age living in the 160 municipalities under analysis. An elasticity of 0.012 indicate that municipalities 10% closer to the NTH face a reduction of almost 410 individuals. Assuming the extreme case that all these individuals enter to their criminal option, this may correspond to an increase by almost 3% in the number of gangs members at that ages in these communities. If they exert their "initiation process" consisting in murdering an individual according to gang's interest (Aguilar, 2008), this implies a total of 410 homicides, an 8% increase in the 5,500 homicides per year, approximately.

[33]A fourth alternative for youth at those ages is emigration. We do not have data to test this last alternative because most of the emigration from El Salvador to the U.S. – the main destination for El Salvadorans – is undocumented. Additionally, when we separate the data by gender and age, estimations can be even more noisy. However, we are currently looking for data to test this last channel.

[34]Using data for older youth enrolled in high school, we find similar results: an increase in the drop out rate for males between 17-19 years old. In the female cohort, we find also a marginal increase in the drop out rate, but it is statistically non significant at conventional levels.

TABLE 4. EFFECTS OF ROAD INFRASTRUCTURE ON LABOR OUTCOMES

Dependent variable: Δ_t employment in the formal sector

	OLS REGRESSION COEFFICIENTS				IV REGRESSION COEFFICIENTS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	15-19 years old male	20-29 years old male	15-19 years old female	20-29 years old female	15-19 years old male	20-29 years old male	15-19 years old female	20-29 years old female
$\ln(Dist\text{-}NTH)$	-0.010***	0.002	-0.008***	-0.005	-0.012***	0.003	-0.010***	0.000
$(-)$	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)
Baseline controls?	Y	Y	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	154	154	154	154	154	154	154	154

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. Estimations include only the municipalities located within 40Km to the NTH. Columns (1) - (4) are estimations using OLS and (5) -(8) are estimated using IV. All models exclude connecting municipalities and include regional fixed effects and the following controls: municipality's population growth rate during the 2009 - 2012 period, political ideology of the major in office, geography control (log elevation) and a dummy indicating whether the district was part of an anti-crime program called *municipios santuarios* implemented by the El Salvadoran government.

Similarly, using a D-in-D approach, we find similar results of being "treated" by the new road infrastructure on changes in drop out rates for both boys and girls between 12 to 16 years of age. These estimated coefficients are presented in Table A5 in the Appendix.

These results lead to some important conclusions. First, despite the positive effects of the NTH construction on economic activity, it also generated unintended effects on labor and educational outcomes of vulnerable youths living in these poor communities. Second, youth who are abandoning the formal labor sector are not necessarily enrolling school. On the contrary, our results indicate that they are dropping out school. An argument can be that they are abandoning school to work in NTH construction. However it is unlikely due to project's restrictions in terms of labor force. Construction firms were not allowed to hire individuals younger than 16 years old, MCC (2009). Additionally, if that were the case, that additional labor force would be reflected in the estimations of labor force in the formal sector.

5.4 Road Infrastructure Effects on other Economic Outcomes

Theoretical and empirical economic evidence indicate that infrastructure affects the supply of health services and the demand for them (Agénor and Moreno-Dodson, 2006; Agénor, 2008). Specifically, roads can facilitate the access to health facilities through the reduction of transportation costs. To contribute to that literature, we also measure the impact of the NTH on health outcomes. As show in Table A6 in the Appendix section, we find no impact of the road infrastructure on health related issues, such as child mortality and morbidity.

This lack of impact can be explained by the expansion of health services throughout the country since 2008 through the program ECOS *(Equipos Comunitarios de Salud Familiar)*. They are groups of 4-5 health specialists visiting mainly households located in rural areas.[35] The program's objective is to guarantee the right of effective access to health of the population closer to their home, preventing need to travel long distances and incurring additional expenses. It also tries to reduce inappropriate demand for hospital bed. Since the NTH was constructed after the ECOS program was implemented, it might not generate any effect in these outcomes.

6. Channels at Work

The preceding section has presented empirical evidence suggesting that NTH has led to increased overall economic activity and in specific economic sectors that are relevant to the municipalities in the northern region of El Salvador. However, some surprising additional results have been found, including that proximity to the NTH has lead to: (i) a reduction in the growth of the male labor supply for individuals between 15 to 19 years of age, and to (ii) an increase in the growth of the drop out rate for boys between 12 to 16 years old enrolled in public schools.

[35]The ECOS program is a model of health care, with a focus on family health. They are a section of the health reform that the Ministry of Public Health promotes with the support of the Pan-American Health Organization.

TABLE 5. EFFECTS OF ROAD INFRASTRUCTURE ON EDUCATIONAL OUTCOMES

Dependent variable: Δ_t drop outs rate (9-16 years old)

	OLS REGRESSION COEFFICIENTS				IV REGRESSION COEFFICIENTS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	4-5th grade (9-11 yo) male	4-5th grade (9-11 yo) female	6-9th grade (12-16 yo) male	6-9th grade (12-16 yo) female	4-5th grade (9-11 yo) male	4-5th grade (9-11 yo) female	6-9th grade (12-16 yo) male	6-9th grade (12-16 yo) female
ln($Dist$-NTH)	-0.004	-0.007*	0.006	0.003	-0.001	-0.008	0.014**	-0.000
$(-)$	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)
Baseline controls?	Y	Y	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	3330	3344	1994	2002	3330	3344	1994	2002

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. All the estimations include only schools located in municipalities within 40km to the NTH. For the estimations in columns (1) and (5), we use a sample of male and students in 4-5 grade; and in columns (2) and (6) we use female students in the same grades. In columns (3) and (7) we use a sample of male students in 6-9 grade and in columns (4) - (8) a female sample in the same courses. All models exclude connecting municipalities and include regional fixed effects and the following controls: municipality's population growth rate during the 2009 - 2012 period, political ideology of the major in office, geography control (log elevation) and a dummy indicating whether the district was part of an anti-crime program called *municipios santuarios* implemented by the El Salvadoran government.

This section provides additional estimation results to further investigate the channels responsible in this context. The main question that we address is whether the observed effects – on economic activity, employment, and education – are driven by gangs.

As we previously discussed, gangs' emblematic crimes are micro-territorial (Aguilar, 2007). They execute territorial control recruiting individuals living in communities that they want to extort or where commit their crimes. Cruz and Portillo (1998) provide qualitative evidence explaining that these organizations employ individuals living in the communities where they want to operate because those people know strategic information about their neighbors, which allows for more efficient predation of the community. Additionally, and more important in terms of territorial control, a local labor force is able to identify foreigners who do not belong to the community.

These specific gangs targets and our previous results on labor and educational outcomes allow us to evaluate their presence as mechanism of the unintended effects from the NTH.

Summing up, our aim is to answer if prosperity in non-targeted connected municipalities has made them attractive to criminal organizations specifically gangs. If gangs are recruiting, this can explain our results of young men abandoning both the formal labor sector and school.

To test whether these results have been driven by gangs' presence in recently connected municipalities, we estimate the NTH impact on two categories of crimes: (i) attributed to gangs operations specifically homicides and extortions; and (ii) other crimes specifically robberies, thefts, and drug trafficking. We present the estimated coefficients in Table 6, separated by category of crime – gangs crimes are in panel A and non-gangs crimes are presented in Panel B – and using both OLS and IV approaches as estimation strategies.

We argue that the last three crimes are not related to gangs because the extortion works as indirect insurance for households. More specifically, if a family or business is paying extortion, the criminal organization is assuring them that other thieves or robbers will not be allowed to operate in their communities. Additionally, gangs in El Salvador are more related to drug use than with drug trafficking according to existing evidence (Farah, 2011; Cruz et al., 2016).

Regarding the results, we find short-term growth in homicides in municipalities closer to the NTH, and an increased number of extortions. We find a NTH distance elasticity of the homicide rate and extortions of 0.138 and 0.311 respectively. These results are presented in Panel A in Table 6. As explained before, extortions are the mechanism used by gangs to finance their activities, and homicides occur in order to protect their territory or as threats to control behavior in their territory.

These coefficients have important implications in the context of a highly violent country. They suggest that in all of the previously non-connected municipalities with a average homicides rate of 71 per 100,000 habitants in 2009, municipalities 10% closer to the NTH faced an increase in homicides rate by 13.8%, i.e. 10 additional homicides per district, approximately.

Additionally, we have to provide evidence that the growth on crimes not related to gangs in non-targeted connected municipalities is negative or zero. The estimated coefficients after controlling by baseline variables at the municipality level and regional fixed effects are presented in Panel B in Table

6. Supporting existing qualitative evidence of the indirect security services provided by gangs and their lack of direct involvement in drugs operations (Farah, 2011; Cruz et al., 2016), these additional estimations show no impact of the closeness to the NTH on thefts and drug trafficking growth at the municipal level.

Moreover, we estimate a NTH-distance elasticity on robberies of 0.218, which indicates that municipalities closer to the highway face an reduction in the growth of the number of robberies. This result was to some extent expected. As we briefly explained before, gangs finance their activities through two specific crimes: extortions and homicides. This high level of violence guarantees their control over criminal activities in their territories. For this reason, it is not surprising to find a reduction in crimes that are usually committed by individual offenders. In this sense, gangs block other criminal activity.

TABLE 6. EFFECTS OF ROAD INFRASTRUCTURE ON CRIME

Dependent variable: Δ_t Crimes

	OLS REGRESSION COEFFICIENTS			IV REGRESSION COEFFICIENTS		
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A. GANGS CRIMES						
	Homicides	Extortions	Detentions	Homicides	Extortions	Detentions
$\ln(Dist\text{-}NTH)$	0.160**	0.254*	0.196***	0.138*	0.311**	0.238***
$(-)$	(0.073)	(0.131)	(0.054)	(0.083)	(0.157)	(0.071)
PANEL B. NON-GANGS CRIMES						
	Robberies	Thefts	Drug trafficking	Robberies	Thefts	Drug trafficking
Least cost path IV	-0.168**	-0.070	-0.079*	-0.218**	-0.105	-0.087
$(-)$	(0.073)	(0.070)	(0.048)	(0.089)	(0.091)	(0.062)
Baseline controls?	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y
Obs.	154	154	154	154	154	154

*, **, ***, significant at 10%, 5% and 1% respectively, Robust standard errors in parenthesis. All the estimations include only the municipalities located within 40Km to the NTH. All models exclude connecting municipalities and include regional fixed effects and the following controls: municipality's population growth rate during the 2009 - 2012 period, political ideology of the major in office, geography control (log elevation) and a dummy indicating whether the district was part of an anti-crime program called *municipios santuarios* implemented by the El Salvadoran government. Dependent variables description are summarized in Appendix A1.

We also estimated the NTH's closeness effects using a D-in-D strategy and present the results in the Table A6 in the Appendix section. The results of the interaction between benefiting from the road after its construction are similar in sign to the estimated elasticities for gangs, columns (1) to (3), and non-gang crimes, columns (4) to (6).

Considering all of these results due to weak institutions, gangs, and the lack of law enforcement in El Salvador, the costs of committing crimes are low and its benefits may seem to be higher than those expected from the non-criminal labor sector. In that sense, people living in these counties face an important trade-off between participating in the criminal or non-criminal labor sectors.

7. Falsification tests

In addition to the results reported in the previous section, we estimate falsification tests concerning the estimated average NTH effects on homicide growth. The additional results address potential concerns about: (i) the temporality of the NTH construction, and (ii) a lack of differences between crime levels in the newly connected municipalities by the NTH compared to those previously connected by other roads.

To address the temporality concern, we use data on homicide rates during the years before the NTH construction, 2006 to 2009, and estimate specification (1) and (2). The aim is to identify if there were differences in the growth of homicide rates during the previous period of road construction. Our estimations included the same control variables described before and regional fixed effects. Estimated coefficients are presented in columns (1) and (2) in Table 7.

Using both OLS and IV, the point estimates are not statistically significant at conventional levels. This result indicates that before the construction of the road infrastructure in the northern region, closeness of municipalities' centroid to the area where the NTH would be built had no effect on growth in homicide rates. More specifically, the effects on our main crime variable exist only after the NTH construction.

We then turn to study if changes in homicides were similar between previously and newly connected municipalities. As we discussed before, we restricted our sample to 155 municipalities located within 40 kilometers of the NTH in all our OLS and IV estimations. Our aim was to capture only the effects on the newly non-targeted connected communities. In that sense, we should find that the effects of closeness to the NTH on previously connected municipalities are closer to zero. Hence, the second placebo test we implement was to estimate only using those municipalities far away from the NTH as a sample. Those municipalities are already connected by older El Salvadoran highways, such as the Pan American or *El Litoral* highways.

As we show in columns (3) and (4) in Table 7, the estimated point NTH distance elasticities are not statistically different from zero after controlling by demographic, economics and security variables, and after incorporating regional fixed effects using both OLS and IV approaches.

Summing up these results, we can argue that the effects we find in our main crime variable took place after the NTH construction and not at any period before. Additionally, the impact of the road infrastructure is occurring only for municipalities that did not previously benefit from road infrastructure until the NTH was built.

TABLE 7. FALSIFICATION TESTS
Dependent variable: Δ_t Gangs homicides

	Homicides before the NTH (2006-2009)		Homicides in previously connected municipalities	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
ln($Dist$-NTH)	0.009	-0.026	-0.047	0.191
$(-)$	(0.075)	(0.079)	(0.484)	(0.518)
Baseline controls?	Y	Y	Y	Y
Region FE	Y	Y	Y	Y
Obs.	154	154	101	101

*, **, ***, significant at 10%, 5% and 1%, Robust standard errors in parenthesis. All the estimations include some controls. Column (1) and (2) are estimations in the period before the NTH construction. Columns (3) and (4) are estimations using municipalities from 40 Km distance away. Dependent variables definitions are summarized in Appendix A1. In all models we included regional fixed effects and excluded connecting municipalities.

8. Discussion and Concluding Remarks

This research aims to contribute to the literature that studies the effects of road infrastructure on economic development. Additionally, this paper present the first evidence that simultaneously analyze road proximity's effects on other novel welfare outcomes of special interest in developing countries. Specifically, we study the effects of new road infrastructure on labor, educational, and crime outcomes by exploiting the construction of a highway in the northern region of El Salvador (NTH).

The highway construction project's main objective was to reduce transportation costs for the most important communities located in the impoverished northern region, and connect them to other prosperous municipalities. Thus these communities could face a greater demand for their good and services and improve their economic performance (MCC, 2009). In that sense, the highway location was not completely exogenous, but oriented to connect specifically targeted municipalities.

To address the possible endogeneity of the highway construction, we estimate the specifications using both OLS and an IV approach. We exploit as source of exogenous variation the fact that highways are generally constructed following a Least Cost Path (LCP) between some initial and terminal points – targeted municipalities in this context. This path is generally exogenous determined by geographical features such as mountains and water bodies. In that sense, the distance between the LCP and the centroids of municipalities can be treated as randomly assigned. Therefore, we instrument the actual distance between each municipality's centroid and the NTH with the distance between each centroid and the LCP.

Similar to the existing evidence of effects of road infrastructure on economic performance, we find that once we instrument for the LCP distance, the municipalities near the new highway increased their economic activity measured both using municipal tax revenues and light density at night.

However, unexpectedly, we also find unintended effects of road infrastructure on labor and educational outcomes in these poor regions. Specifically, proximity to the NTH reduces 15 to 19 years old youth participation in the formal labor sector and increase school drop out rates of 12-16 years old boys. All these results are sustained using a D-in-D approach. These novel results are relevant in the way that the context in which the road infrastructure investment was made may be of the first order in interpreting the mechanisms behind them.

Regarding the mechanism, the El Salvadoran context and the effects we find on these particular cohorts allow us to provide evidence that the greater economic activity attracted criminal organizations to these communities. Exploiting the specific characteristics in the gangs' recruitment process, we argue that gangs target these youth between 12-20 years of age, recruiting them as their local labor force. This would explain why these cohorts were displaced from the formal labor sector and educational system despite increased economic opportunities. They were diverted into the criminal labor sector.

To test this mechanism, we compare typical gang crimes with other criminal activities. Results indicate that municipalities 10% closer to the highway face an increase in the homicide rate by 14% and increased growth in extortions by 31%. The magnitudes of these effects are economically important and statistically significant, especially in the context of developing and highly violent countries such as those in the northern triangle of Central America.

How important is the monopolizing of violence by these criminal groups? Besides the extortions and homicides they undertake to finance their operations, they are also working as an alternative source of protection or security for people living in their territories. According to our estimations and arguments, municipalities 10% closer to the NTH, compared to the mean distance, face a 21% reduction on robberies. This result is likely to being explained according to the mechanism we have identified: the NTH attracted gangs that exert control over territories, restricting the crimes that committed by non-organized offenders.

The policy implications of these results are of first order. First, with these results we are not arguing that policy makers should not promote road infrastructure in impoverished countries with high levels of crime. On the contrary, similar to the literature of the effects of institutions on economic growth (Acemoglu et al., 2014; Bruhn and Gallego, 2012; Dell, 2010), the main conclusion that can be drawn from this paper is that before implementing large infrastructure investment in countries with weak institutions and lack of law enforcement, it is necessary to strength institutions and reduce gangs' power and control.

Additionally, since we have learnt from the economic of institutions literature that the impact of institutions on long-run development is robust (Acemoglu, Gallego and Robinson, 2014; and Dell, 2010), reducing the territorial expansion and persistence over time of these criminal organizations on these countries' economic development is relevant.

Finally, criminal organizations are a plausible mechanism driving the unusual effects we find. However, there are also at least two outside options for these youths. First, using light density at night, we measure the NTH effect on both formal and informal economic activity. In that sense, it is plausible that these youth dropped out school and abandoned the formal labor sector to move to the informal economy. Second, migration is another likely alternative for this cohort, especially in the northern region (OIM, 2012) and its effect on the situation is not clear. On the one hand, greater economic activity may have increased households' income, and hence generated the resources to cover migration costs. However, it is also likely that their migration response could be lower since internally they are generating resources. Unfortunately, due to data limitations, we are not yet able to rule out any of these mechanisms, but they are part of our upcoming estimations in this paper.

Bibliography

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.

Acemoglu, D., Gallego, F., and Robinson, J. (2014). Institutions, human capital, and development. *Annu. Rev. Econ.*, 6(1):875–912.

Agostini, C. A. and Palmucci, G. A. (2008). The anticipated capitalisation effect of a new metro line on housing prices. *Fiscal studies*, 29(2):233–256.

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

Aguilar, J. (2007a). las maras o pandillas juveniles en el triángulo norte de centromérica. mitos y realidades sobre las pandillas y sus vínculos con el crimen. *Universidad Centroamericana José Simeón Cañas*.

Aguilar, J. (2007b). las maras o pandillas juveniles en el triángulo norte de centromérica. mitos y realidades sobre las pandillas y sus vínculos con el crimen. *Universidad Centroamericana José Simeón Cañas*.

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.

Akee, R. (2006). The babeldaob road: the impact of road construction on rural labor force outcomes in the republic of palau.

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.

Asahi, K. and Dominguez, P. (2016). Closer proximity to the subway network increases robbery and larceny. Technical report, Escuela de Gobierno.

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., Duflo, E., and Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in china. Technical report, National Bureau of Economic Research.

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.

Bruhn, M. and Gallego, F. A. (2012). Good, bad, and ugly colonial activities: do they matter for economic development? *Review of Economics and Statistics*, 94(2):433–461.

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.

Cattaneo, M. D., Galiani, S., Gertler, P. J., Martinez, S., and Titiunik, R. (2009). Housing, health, and happiness. *American Economic Journal: Economic Policy*, 1(1):75–105.

Cellini, S. R., Ferreira, F., and Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1):215–261.

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.

Cruz, J. M. (2009). Global gangs in el salvador: Maras and the politics of violence. In *Ponencia presentada en el Global Gangs Workshop, Centre on Conflict, Development, and Peacebuilding, Génova, mayo*, pages 14–15.

Cruz, J. M., Fonseca, B., and Director, J. D. (2016). The new face of street gangs: The gang phenomenon in el salvador. *IRB*, 16:0322.

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.

Dahbura, J. N. M. et al. (2016). The short-term impact of crime on school enrollment and school choice: Evidence from el salvador. Technical report, Institute for Economics Studies, Keio University.

Damasio, A. R. (1994). Descartes' error: Emotion, reason, and the human brain.

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.

Das, S. and Mocan, N. (2016). Analyzing the impact of the world?s largest public works project on crime. Technical report, National Bureau of Economic Research.

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

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica*, 78(6):1863–1903.

DellaVigna, S. (2009a). Psychology and economics: Evidence from the field. *Journal of Economic literature*, 47(2):315–72.

DellaVigna, S. (2009b). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2):315–372.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

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

Dinarte, L. (2017). Peer effects in after-school programs. experimental evidence in el salvador.

Dinarte, L. and Egana, P. (2017). Emotional regulation and after-school programs in highly violent communities: Neuro-physiological evidence from el salvador.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from south africa. *American Economic Review*, 101(7):3078–3108.

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.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

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

Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and trade: Evidence from the us. *Review of Economic Studies*, 81(2):681–724.

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.

Egana-delSol, P. (2016). Can we trust self-reporting as a measurement of non- cognitive skills?

Engerman, S. L. and Sokoloff, K. L. (1997). Factor endowments, institutions, and differential paths of growth among new world economies. *How Latin America Fell Behind*, pages 260–304.

Engerman, S. L. and Sokoloff, K. L. (2002). Factor endowments, inequality, and paths of development among new world economics. Technical report, National Bureau of Economic Research.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Farah, D. (2011). *Organized crime in El Salvador: The homegrown and transnational dimensions*. Woodrow Wilson International Center for Scholars, Latin American Program.

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.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.

Fudenberg, D. and Levine, D. K. (2006). A dual-self model of impulse control. *American economic review*, 96(5):1449–1476.

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

FUSADES (2015). El Salvador Youth Survey Database.

FUSADES (2016). Extorsiones a la micro y pequeña empresa. Technical report, Fundación Salvadoreña para el Desarrollo Económico y Social.

Fuster, J. M. (2013). *The neuroscience of freedom and creativity: Our predictive brain.* Cambridge University Press.

Galarraga, J. and Dinarte, L. (2015). Los caminos en la mitad del mundo: Efectos de largo plazo de construcción de carreteras en el desarrollo económico cantonal en ecuador.

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

Gibbons, S., Lyytikäinen, T., Overman, H. G., and Sanchis-Guarner, R. (2016). New road infrastructure: the effects on firms.

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.

Goleman, D. (2010). Emotional intelligence: Why it can matter more than iq. *Bantam Books.*

Gonzalez-Navarro, M. and Quintana-Domeque, C. (2016). Paving streets for the poor: Experimental analysis of infrastructure effects. *Review of Economics and Statistics*, 98(2):254–267.

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.

Harmon-Jones, E., Gable, P. A., and Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology*, 84(3):451–462.

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

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour economics*, 19(4):451–464.

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.

International Crisis Group (2017). El salvador's politics of perpetual violence. Technical report, International Crisis Group.

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.

Jacoby, H. G. (2000). Access to markets and the benefits of rural roads. *The Economic Journal*, 110(465):713–737.

Jaitman, L., Caprirolo, D., Granguillhome Ochoa, R., Keefer, P., Leggett, T., Lewis, J. A., Mejía-Guerra, J. A., Mello, M., Sutton, H., and Torre, I. (2017). The costs of crime and violence: New evidence and insights in latin america and the caribbean. *Inter-American Development Bank*.

Jedwab, R. and Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: Evidence from africa. *Review of economics and statistics*, 98(2):268–284.

Jha, M. K., McCall, C., and Schonfeld, P. (2001). Using gis, genetic algorithms, and visualization in highway development. *Computer-Aided Civil and Infrastructure Engineering*, 16(6):399–414.

Jong, J.-C. and Schonfeld, P. (2003). An evolutionary model for simultaneously optimizing three-dimensional highway alignments. *Transportation Research Part B: Methodological*, 37(2):107–128.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., and Just, M. A. (2013). Identifying emotions on the basis of neural activation. *PloS one*, 8(6):e66032.

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.

Lakoff, G. (2008). *The political mind: why you can't understand 21st-century politics with an 18th-century brain*. Penguin.

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.

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

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.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review*, 90(2).

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.

Martincus, C. V., Carballo, J., and Cusolito, A. (2012). Routes, exports, and employment in developing countries: Following the trace of the inca roads. *Inter-American Development Bank, mimeograph.*

Martinez-Leon, J.-A., Cano-Izquierdo, J.-M., and Ibarrola, J. (2016). Are low cost brain computer interface headsets ready for motor imagery applications? *Expert Systems with Applications*, 49(1):136–144.

MCC (2009). Evaluation design. mcc connectivity project in el salvador. the northern transnational highway. Technical report, Millenium Challenge Corporation.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701.

Michalopoulos, S. and Papaioannou, E. (2013). National institutions and subnational development in africa. *The Quarterly Journal of Economics*, 129(1):151–213.

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.

Morten, M. and Oliveira, J. (2016). Paving the way to development: costly migration and labor market integration. Technical report, National Bureau of Economic Research.

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.

OECD (2015). Skills for social progress: The power of social and emotional skills. Technical report, OECD Center for Research and Innovation (CERI).

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.

Ramirez, R. and Vamvakousis, Z. (2012). Detecting emotion from eeg signals using the emotive epoc device. *Brain Informatics*, pages 175–184.

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.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1):1.

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.

Salzman, C. D. and Fusi, S. (2010a). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annual review of neuroscience*, 33:173–202.

Salzman, C. D. and Fusi, S. (2010b). Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex. *The Annual Review of Neuroscience*, 33:173–202.

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.

Santacruz Giralt, M. L., Concha-Eastman, A., et al. (2001). *Barrio adentro: La solidaridad violenta de las pandillas*. IUDOP.

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.

Treiman, D. J. (2014). *Quantitative data analysis: Doing social research to test ideas*. John Wiley & Sons.

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

Voigtländer, N. and Voth, H.-J. (2014). Highway to hitler. Technical report, National Bureau of Economic Research.

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.

Wolf, S. (2014). Central american street gangs: Their role in communities and prisons.

Wolf, S. (2016). Drugs, violence, and corruption: Perspectives from mexico and central america. *Latin American Politics and Society*, 58(1):146–155.

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

Appendix A

Appendix to Chapter 1

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:

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

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:

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

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:

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

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:

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

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:

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

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:

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

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:

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

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:

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

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)
				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)
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)	(2)
	Sessions attended	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)

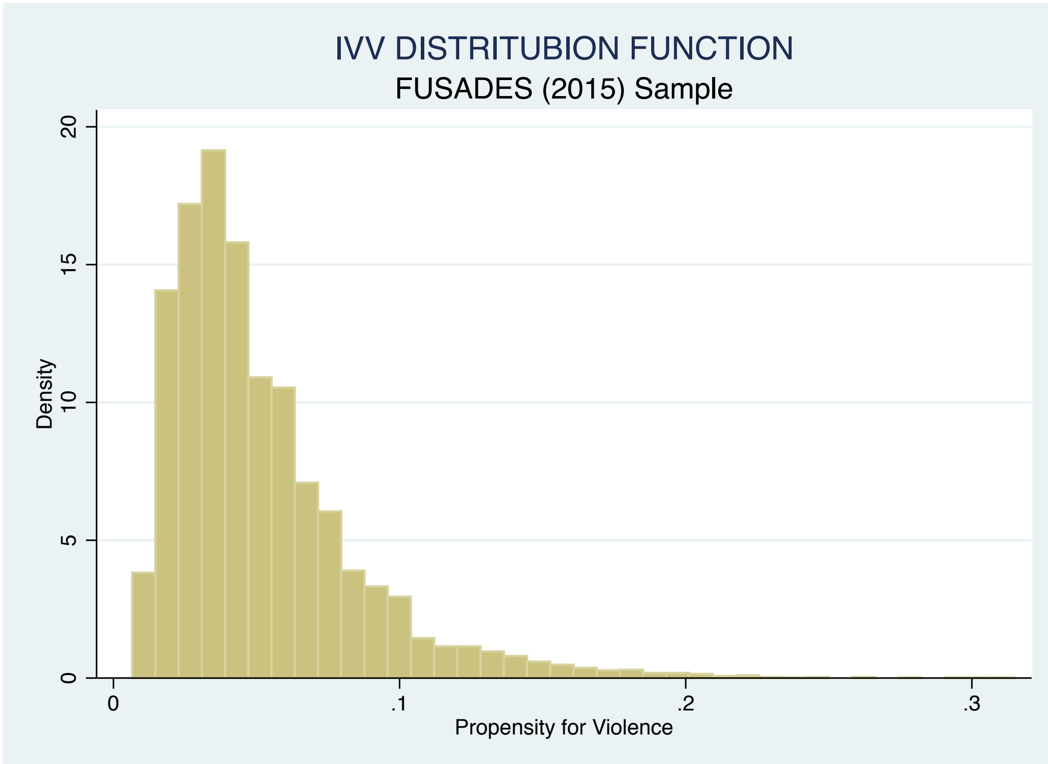
***, **, * indicate that the comparison between HM-Low vs HT at their respective category of course is significant at 1%, 5%, and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. Sample is restricted to N = 771 treated students. All regressions include as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

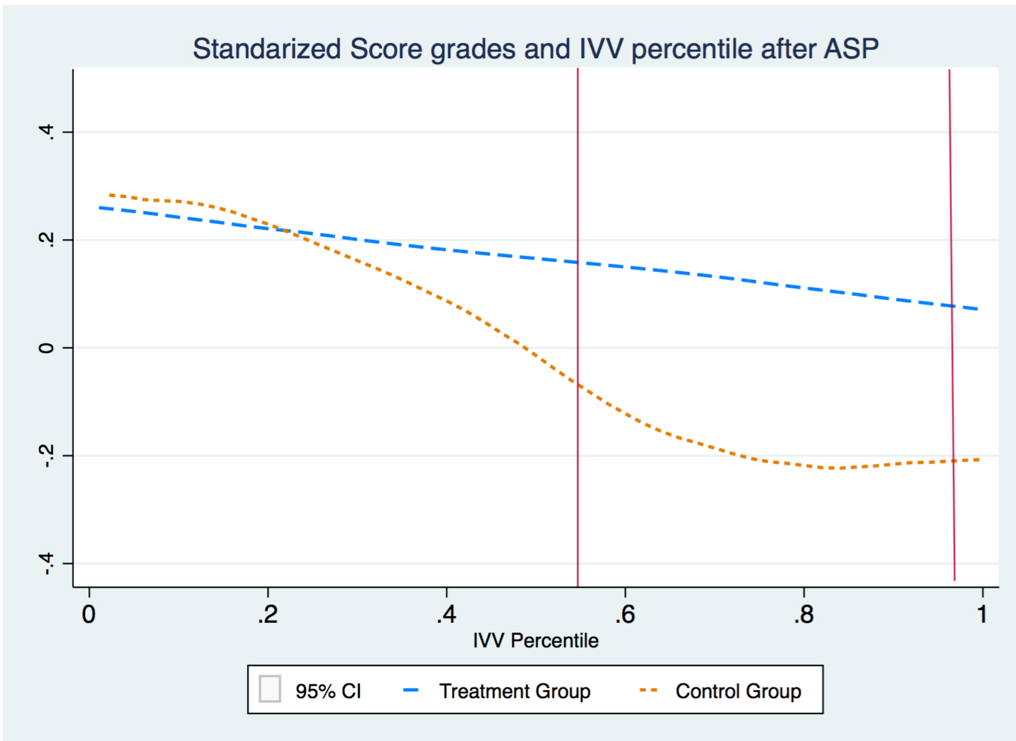
***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A is effects on non-cognitive outcomes. Panel B present results on academic outcomes. All regressions include an interaction between treatment and the share of clubmates 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



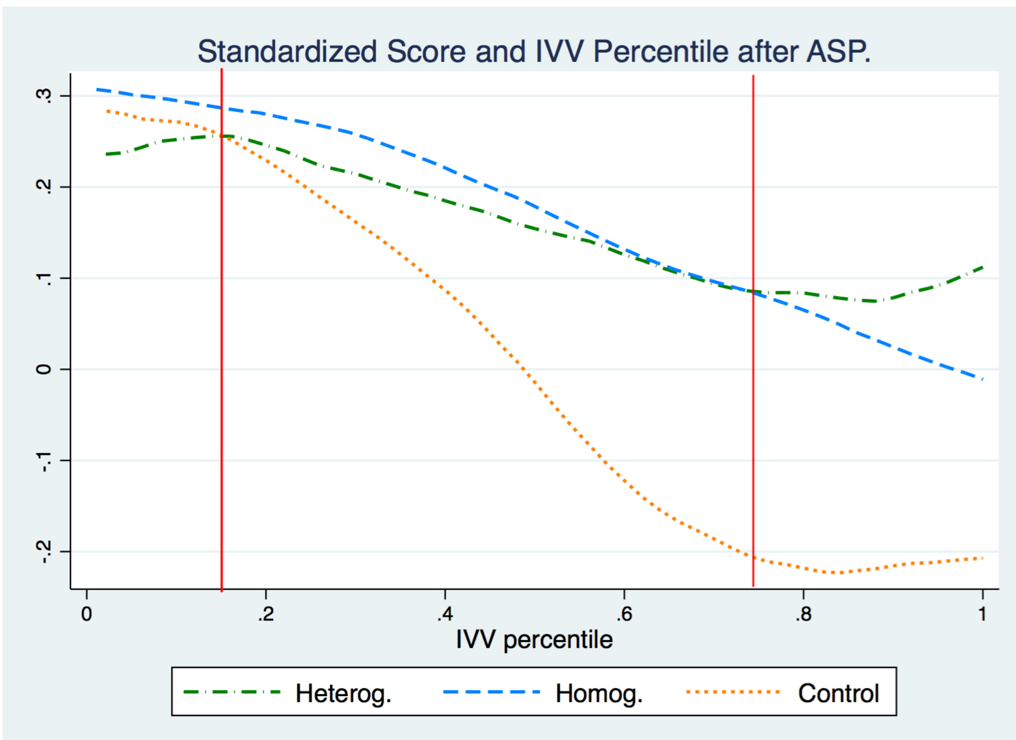
Distribution function of the estimated propensity for violence using the available determinants in the FUSADES (2015) dataset.

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.

Appendix B

Appendix to Chapter 2

Appendix 1. Gender vs. IVV 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, since the IVV estimation includes gender as a determinant, the difference in the effects between 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. To verify which of the measures are generating the differences, we use the following alternative specification:

$$E_{ij} = \theta_0 + \theta_1 T_{ij} + \theta_2 T_{ij} \times G_{ij} + \theta_3 T_{ij} \times IVV\_high_{ij} + \theta_4 X_{ij} + S_j + \epsilon_{ij} \tag{B.1}$$

where θ_2 indicates the difference of the ASP effects by gender (boys vs. girls) and θ_3 shows the difference of the impact by the propensity for violence (highly- vs. less- 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 A11 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 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 A11 here**]

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

	(1)	(2)	(3)	(4)	(5)
Panel A. School characteristics					
	Urban Area	Most violent municipality	Number of students	Indigenous students	Additional revenues
Participant school	0.125	0.000	130.41	-0.003	575.973
	(0.182)	(0.000)	(107.25)	(0.003)	(2,109.54)
Constant	0.245***	0.093	256.14***	0.041***	1,798.58***
	(0.000)	(0.000)	(0.104)	(0.000)	(2.054)
Panel B. School programs					
	EITP Program	School kits	Vaso de leche Program	Food Program	Psychological care
Participant school	-0.109	-0.184	0.191***	0.034	0.148
	(0.070)	(0.134)	(0.043)	(0.024)	(0.115)
Constant	0.149***	0.979***	0.572***	0.983***	0.035***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel C. School facilities or equipment					
	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)
Constant	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

Source: Data from 2015 Educational Census of El Salvador, Ministry of Education. Estimations from Dinarte (2017).
***, **, * significant at 1%, 5% and 10% respectively. Robust standard errors at course-school level are in parentheses.

TABLE A2. SUMMARY STATISTICS. 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	0.05
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	0.08
Student's course	5.75	2.71	5.5	2.52	0.29
N	1056		6641		

Source: Data from FUSADES (2015) and Dinarte (2017). Estimations from Dinarte (2017).
The table provides means and standard deviations of the main variables from the Study Sample and FUSADES (2015) Sample. These variables were used to estimate the IVV for each student in the study sample. Column 5 shows the p-value for the comparison of means between both samples. ***, ** and * denotes difference significant at the 1%, 5% and 10% level respectively when comparing the means.

TABLE A3. 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 with only 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's 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 in morning shift	-0.002
	(0.087)

Source: Data from FUSADES (2015) and Dinarte (2017). Estimations from Dinarte (2017). ***, **, and * denotes statistically significant difference at the 1%, 5% and 10% levels, respectively. Standard error in parentheses. Mother's education omitted category: mother has basic education (1-6th grade). Household composition omitted category: children living with both parents.

TABLE A4: 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

Source: Data from FUSADES (2015) and Dinarte (2017). Estimations from Dinarte (2017).
The table provides summary statistics for the Vulnerability and Violence Index (IVV) predicted using FUSADES (2015) dataset and variables available during the clubs enrollment phase.

TABLE A5: DESCRIPTIVE STATISTICS OF THE IVV BY TREATMENT GROUP. RANDOMLY SELECTED SUBSAMPLE

	(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)
Mean	0.038	0.037	0.039	0.042	0.037	0.050	0.022
Std. Dev	0.029	0.027	0.030	0.036	0.026	0.027	0.014
Median	0.030	0.029	0.030	0.029	0.031	0.044	0.019
Min	0.002	0.003	0.002	0.006	0.002	0.008	0.002
Max	0.216	0.151	0.216	0.216	0.133	0.133	0.057
N	598	145	453	162	291	149	142

Source: Data from Dinarte (2017).
The table provides summary statistics for the Vulnerability and Violence Index (IVV) predicted using FUSADES (2015) dataset and variables available during the clubs' enrollment phase. We restricted estimations to the randomly selected subsample for this study.

TABLE A6. SUMMARY STATISTICS: MEANS OF BASELINE VARIABLES BY TREATMENT GROUP SUBSAMPLE WITH VALID EEG RECORDINGS

	(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	0.44	0.49	0.43	0.46	0.42	0.63	0.16***
Student's age	13.08	13.00	13.10	13.05	13.14	13.62	12.58**
Student lives in urban area	0.72	0.71	0.73	0.70	0.74"	0.77	0.71*
Student's household composition							
Student living with both parents	0.50	0.47	0.51	0.48	0.53	0.55	0.51
Student living with only one parent	0.33	0.34	0.33	0.39	0.29	0.29	0.29
Student living with one parent and step-parent	0.06	0.07	0.05	0.03	0.07"	0.05	0.09
Student living with other relative /adults	0.10	0.12	0.11	0.10	0.11	0.10	0.11
Student's mother's level of education:							
Basic education (1-6 years)	0.32	0.33	0.32	0.30	0.32	0.26	0.41
Intermediate education (7-12 years)	0.59	0.56	0.60	0.63	0.59	0.67	0.49*
University or higher (13 and +)	0.09	0.11	0.08	0.07	0.09	0.07	0.10
Student's travel time from house to school (min.)	16.97	16.57	17.09	16.16	17.63	19.88	14.96
Student is alone at home after school	0.05	0.03	0.06	0.07	0.05	0.08	0.00**
Student's school year	6.84	6.74	6.87	6.79 ˆ	6.89	7.16	6.61
Student enrolled in the morning shift	0.64	0.67	0.63	0.64	0.63	0.59	0.68
Student's violence index	0.043	0.042	0.044	0.046	0.043	0.057	0.027***
PANEL B: ACADEMIC OUTCOMES							
Academic scores Q1 2016 (Baseline)							
Reading scores	6.67	6.46	6.73	6.76ˆ	6.71""	6.54	6.88
Math scores	6.48	6.41	6.51	6.46	6.49	6.52	6.44
Science scores	6.62	6.46	6.67	6.62	6.54	6.63	6.55
Behavior 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: CLUB CHARACTERISTICS							
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
Observations	308	70	238	87	151	82	69

Table 1 shows descriptive statistics of the available variables at baseline for the full sample. Panel A summarizes information obtained from the enrollment form 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 club characteristics. Take-up is estimated as number of hours attended by student i / max hours attended by any student in each club. ˆ 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 A7. SUMMARY STATISTICS: MEANS OF BASELINE VARIABLES BY TREATMENT GROUP SUBSAMPLE WITH NON-VALID EEG RECORDINGS

	(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	0.54	0.51	0.54	0.65*	0.49”	0.77	0.23***
Student's age	10.6	10.5	10.6	11.0**	10.5	11.2	9.89***
Student lives in urban area	0.77	0.71	0.77	0.77	0.77	0.82	0.73
Student's household composition							
Student living with both parents	0.56	0.48	0.59	0.57	0.59	0.54	0.64
Student living with only one parent	0.31	0.37	0.29	0.29	0.29	0.36	0.24
Student living with one parent and step-parent	0.07	0.07	0.05	0.08	0.07	0.06	0.08
Student living with other relative /adults	0.06	0.08	0.05	0.05	0.05	0.07	0.03
Student's mother's level of education:							
Basic education (1-6 years)	0.29	0.36	0.27*	0.23*	0.29	0.19	0.38***
Intermediate education (7-12 years)	0.62	0.56	0.65	0.67	0.64	0.72	0.58**
University or higher (13 and +)	0.08	0.08	0.08	0.10	0.06	0.09	0.04
Student's travel time from house to school (min.)	17.4	15.2	18.2*	16.9	18.8*	22.6	15.5**
Student is alone at home after school	0.06	0.08	0.05	0.05	0.05	0.09	0.01*
Student's school year	4.53	4.57	4.52	6.65*	4.45	4.91	4.02**
Student enrolled in the morning shift	0.80	0.80	0.79	0.77	0.81	0.81	0.81
Student's violence index	0.03	0.03	0.03	0.04	0.03’	0.04	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
Behavior 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
Missing share	48%	51%	47%	46%	48%	45%	51%
Observations	290	75	215	75	140	67	73

Table 1 shows descriptive statistics of the available variables at baseline for the full sample. Panel A summarizes information obtained from the enrollment form 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 club characteristics. Take-up is estimated as number of hours attended by student i / max hours attended by any student in each club. ^ 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 A8. p-VALUES OF MISSING VARIABLES FOR 13 OBSERVATIONS

	(1)	(2)	(3)
	Locus of control	CRT	Raven
Student is male	0.457	0.457	0.457
Student's age	0.573	0.573	0.573
Student lives in urban area	0.524	0.524	0.524
Student's travel time from house to school (min.)	0.544	0.544	0.544
Student's violence index	0.041	0.041	0.041
Student's school year	0.189	0.189	0.189
Student's household composition			
Student living with both parents	0.851	0.851	0.851
Student living with only one parent	0.320	0.320	0.320
Student living with a parent and a step-parent	0.632	0.632	0.632
Student living with other relative /adult	0.381	0.381	0.381
Student's mother's level of education:			
Basic education (1-6 years)	0.784	0.784	0.784
Intermediate education (7-12 years)	0.765	0.765	0.765
University or higher (13 and +)	0.963	0.963	0.963
Homog. Vs Control	0.181	0.181	0.181
Heterog. Vs Control	0.129	0.129	0.129
Homog. Vs Heterog.	0.522	0.522	0.522
Homog. High vs Homog. Low	0.627	0.627	0.627
Arousal	0.775	0.775	0.775
Valence	0.162	0.162	0.162
Positive Valence	0.213	0.213	0.213
Negative Valence	0.397	0.397	0.397
Positive Valence Difference	0.248	0.248	0.248
Negative Valence Difference	0.166	0.166	0.166

***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. All outcomes have been standardized at the control-course level, with a mean of 0 and standard deviation 1.0. All regressions include as controls: a second order polynomial of student's propensity for violence, and ciclo-school fixed effects (stratification level). Differences in number of observations is due to variation in the response rate for each outcome.

TABLE A9. OVERALL EFFECTS OF THE ASP ON EMOTIONAL REGULATION
Heckman correction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Arousal (stress)	Valence	Locus of control	CRT	Raven	Positive Valence Difference	Negative Valence Difference
Any treatment	0.048	-0.330**	-0.240**	-0.004	-0.006	-0.415**	-0.237
	(0.129)	(0.172)	(0.133)	(0.131)	(0.135)	(0.196)	(0.251)
Observations	598	598	585	585	585	598	598
Lambda	-0.425	0.429	0.213	0.177	0.648	0.445	0.126
	(0.423)	(0.978)	(0.506)	(0.510)	(0.654)	(0.971)	(1.192)
rho	-0.528	0.258	0.213	0.182	0.636	0.243	0.062
sigma	0.804	1.658	1.000	0.973	1.018	1.827	2.022

***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. All outcomes have been standardized at the control-course level, with a mean of 0 and standard deviation 1.0. All regressions include as controls: a second order polynomial of student's propensity for violence, and *ciclo-school* fixed effect (stratification level). Differences in number of observations is due to variation in the response rate for each outcome.

TABLE A10. HETEROGENEOUS EFFECTS OF THE ASP ON EMOTIONAL REGULATION OUTCOMES BY GENDER

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Arousal (stress)	Valence	Locus of control	CRT	Raven	Positive Valence Difference	Negative Valence Difference
i. Effect on treated girls	-0.174	-0.563***	-0.252	-0.059	0.115	-0.548	-0.512**
	(0.155)	(0.208)	(0.166)	(0.157)	(0.172)	(0.343)	(0.209)
ii. Effect on treated boys	0.331	-0.143	-0.234	0.068	-0.177	-0.340	0.027
	(0.210)	(0.237)	(0.154)	(0.206)	(0.162)	(0.278)	(0.308)
iii. Difference of effects between treated boys and girls [ii] - [i]	0.506*	0.420	0.017	0.127	-0.293	0.207	0.540
	(0.291)	(0.316)	(0.220)	(0.248)	(0.236)	(0.488)	(0.383)

***, **, * significant at 1%, 5% and 10% respectively. Bootstrapped standard errors at the course-school level are in parentheses. All outcomes have been standardized at the control-course level, with a mean of 0 and standard deviation 1.0. All regressions include as controls: a second order polynomial of student's propensity for violence, and *ciclo-school* fixed effect (stratification level). Differences in number of observations is because of variation in the response rate for each outcome.

TABLE A11. 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*] Any treatment	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*] Boy x Any T	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.

Appendix C

Appendix to Chapter 3

APPENDIX 1: Description of variables and sources.

1. Municipal homicides growth rate: Percentage change of the homicides rate (homicides per 100,000 habitants) between the periods 2006-2009 (pre NTH construction) and 2009-2012 (post NTH construction) at the municipal level. Source: Policia Nacional Civil.
2. Number of residents:Number of Salvadoreans during each year. Source: Population Census.
3. Percentaje of sanctuary municipalities: Percentage of municipalities members of the program "Municipios Santuarios" implemented by the government to reduce the high levels of homicides rates during the period 2009-2010. Source: www.elfaro.net (online newspaper).
4. Elevation: Number of meters above sea level where the municipality is located. Source: Centro Nacional de Registros (CNR).
5. Municipal taxes revenues: Municipal taxes collection in US$ from services, tourism, commerce, agriculture and transportation. Soure: Ministry of finance.
6. Robberies: Percentage change of the number of robberies between the periods 2006-2009 (pre NTH construction) and 2009-2012 (post NTH construction) at the municipal level where the robbery occurred. Source: Policia Nacional Civil.
7. Municipal thefts growth rate: Percentage change of the number of thefts between the periods 2006-2009 (pre NTH construction) and 2009-2012 (post NTH construction) at the municipal level where the theft occurred. Source: Policia Nacional Civil.
8. School drop out rate: Constitutes the difference in the quotients of drop out and initial enrollment in 2009 and 2012 at the municipal level (where the school is located). Source: Ministry of Education
9. Male participation in the formal sector: Percentage change of the number of men working in the formal sector between the periods 2006-2009 (pre NTH construction) and 2009-2012 (post NTH construction) in the municipality they live. Source: Instituto Salvadoreno del Seguro Social.
10. Child mortality rate: Percentage change of the number of infants and children death under the age of five at the municipal level (where the child's mother live) Source: Ministry of Health.

TABLE A1. DESCRIPTIVE STATISTICS OF BASELINE VARIABLES (2009) TREATED AND CONTROL MUNICIPALITIES

	TREATED		CONTROL		p-value
	Mean	std. dev	Mean	std. dev	
	(1)	(2)	(3)	(4)	(5)
Economic Activity (in USD) and demography					
Taxes on commerce and tourism	\$29,340	\$67,472	\$197,582.9	\$638,773.9	0.027
Taxes on services	\$4,551	\$13,253	\$35,708	\$184,550	0.152
Taxes on agriculture	\$153	\$1,069	\$2,394	\$16,846	0.257
Population	9,037	9,722	29,030	47,386	0.001
Workers in the formal sector					
Male, 15-19 years old	0.6	1.3	10.7	22.3	0.000
Female, 15-19 years old	0.5	1.1	7.1	15.2	0.001
Male, 20-25 years old	42.6	81.1	512.7	1,006.5	0.000
Female, 20-25 years old	33.7	58.7	378.1	777.6	0.000
Male, 26-35 years old	54.2	101.2	576.8	1,193.2	0.000
Female, 26-35 years old	41.0	73.4	418.3	947.3	0.001
Crime outcomes.					
Homicides rate	3.0	4.7	23.9	50.1	0.001
Extortions	4.5	12.4	23.9	79.9	0.042
“Sanctuary” municipalities	0.01	0.11	0.05	0.23	0.145
Number of robberies	5.8	8.1	27.6	71.3	0.011
Number of thefts	12.1	18.9	39.6	92.0	0.014
Municipalities	86		74		

Table A1 shows descriptive statistics of the available variables at baseline for municipalities within 40 Km of the NTH.

TABLE A2. FIRST STAGE REGRESSIONS

Dependent variable: Distance of municipalities to the NTH

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: EUCLIDIAN DISTANCE						
ln(Euclidian Distance)	0.739***	0.713***	0.698***	0.738***	0.710***	0.694***
	(0.067)	(0.065)	(0.066)	(0.075)	(0.073)	(0.073)
Economic activity $(t-1)$		-0.025	0.031		0.010	0.057
		(0.085)	(0.076)		(0.088)	(0.078)
Population growth		2.091***	2.170***		2.047**	2.059***
		(0.785)	(0.752)		(0.801)	(0.772)
ln(Elevation)		0.018	-0.025		0.017	-0.013
		(0.050)	(0.060)		(0.050)	(0.061)
Sanctuary municipalities		0.248***	0.228**		0.167***	0.116**
		(0.079)	(0.098)		(0.052)	(0.056)
R^2	0.792	0.813	0.822	0.788	0.808	0.818
First stage F-Stat	29.88	12.08	26.17	152.16	47.70	42.57
PANEL B: WEIGHTED LEAST COST PATH						
ln(weighted LCP)	0.749***	0.753***	0.743***	0.744***	0.748***	0.742***
	(0.041)	(0.038)	(0.039)	(0.043)	(0.040)	(0.041)
Economic activity $(t-1)$		0.188**	0.223**		0.198**	0.218***
		(0.096)	(0.086)		(0.096)	(0.083)
Population growth		1.908***	1.931***		1.869***	1.793***
		(0.494)	(0.528)		(0.485)	(0.511)
ln(Elevation)		-0.236***	-0.247***		-0.223***	-0.204***
		(0.056)	(0.068)		(0.052)	(0.060)
Sanctuary municipalities		0.091	0.118		0.025	0.041
		(0.064)	(0.087)		(0.051)	(0.075)
R^2	0.747	0.789	0.797	0.766	0.808	0.819
First stage F-Stat	29.88	12.08	26.17	152.16	47.70	42.57
Obs	160	160	160	154	154	154
Region FE	NO	NO	YES	NO	NO	YES
Connecting municipalities	YES	YES	YES	NO	NO	NO

*, **, ***, significant at 10%, 5% and 1%. Robust Standard Errors in parenthesis. All the estimations include only the municipalities located within 40km to the NTH. Panel A and B show the estimated coefficients of specification (2) using ED instrument and WLCP instruments. WLCP was estimated as the average of all distances between the LCP and all pixel-units at each municipality. In columns (1) and (4) we estimate an OLS without controls. In the rest of estimations, we include as controls: municipality's population growth rate 2009 - 2012, political ideology of the major in office, geography control (log elevation) and a dummy indicating whether the municipality was part of a program from the Government to reduce violence, called *municipios santuarios*. In some specifications we include regional fixed effects. In columns (4) - (6) we exclude the 6 municipalities that were intended to be connected by the NTH. Economic activity is measured using light density (Michalopoulos and Papaioannou, 2011).

TABLE A3. EFFECTS OF ROAD INFRASTRUCTURE ON ECONOMIC ACTIVITY DIF-IN-DIF REGRESSIONS COEFFICIENTS

Dependent variable: Δ_t economic activity

	Economic activity (light density) (1)	Taxes on services (2)	Taxes on Commerce and Tourism (3)	Taxes on agriculture (4)
Treatment	-0.226***	-0.601	-1.874***	0.006
	(0.021)	(0.618)	(0.602)	(0.150)
Post	-0.082***	-1.673***	-3.280***	-0.009
	(0.016)	(0.571)	(0.487)	(0.241)
Post×Treatment	0.052*	1.272*	2.063***	-0.156
	(0.028)	(0.708)	(0.618)	(0.290)
Baseline controls?	Y	Y	Y	Y
Region FE	Y	Y	Y	Y
Obs.	27,570	310	310	310

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. All models include regional fixed effects and controls: municipality's population growth rate 2009 - 2012, geography control (log elevation) and a dummy indicating whether the municipality was part of a program from the Government to reduce violence, called *municipios santuarios.*

TABLE A4. EFFECTS OF ROAD INFRASTRUCTURE ON LABOR OUTCOMES DIF-IN-DIF REGRESSIONS COEFFICIENTS

Dependent variable: Δ_t employment in the formal sector

	15-19 years old male (1)	20-29 years old male (2)	15-19 years old female (3)	20-29 years old female (4)
Treatment	-0.004*	-0.015***	-0.003	-0.014***
	(0.002)	(0.003)	(0.002)	(0.004)
Post	0.018***	0.028***	0.014***	0.036***
	(0.005)	(0.003)	(0.005)	(0.003)
Post×Treatment	-0.017***	0.019***	-0.012***	0.010
	(0.005)	(0.006)	(0.005)	(0.006)
Baseline controls?	Y	Y	Y	Y
Region FE	Y	Y	Y	Y
Obs.	310	310	310	310

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. All models include regional fixed effects and controls: municipality's population growth rate 2009 - 2012, geography control (log elevation) and a dummy indicating whether the municipality was part of a program from the Government to reduce violence, called *municipios santuarios*.

TABLE A5. EFFECTS OF ROAD INFRASTRUCTURE ON EDUCATIONAL OUTCOMES DIF-IN-DIF REGRESSIONS COEFFICIENTS

Dependent variable: Δ_t drop outs rate (9-16 years old)

	4-5th grade (9-11 yo) male (1)	4-5th grade (9-11 yo) female (2)	6-9th grade (12-16 yo) male (3)	6-9th grade (12-16 yo) female (4)
Treatment	-0.007	0.019	-0.395***	-0.403**
	(0.019)	(0.017)	(0.151)	(0.158)
Post	-0.054***	0.075***	-0.327**	-0.325**
	(0.014)	(0.012)	(0.154)	(0.153)
Post×Treatment	-0.014	-0.028	0.545***	0.514**
	(0.020)	(0.018)	(0.202)	(0.202)
Baseline controls?	Y	Y	Y	Y
Region FE	Y	Y	Y	Y
Obs.	3330	3344	1994	2002

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. All models include regional fixed effects and controls: municipality's population growth rate 2009 - 2012, geography control (log elevation) and a dummy indicating whether the municipality was part of a program from the Government to reduce violence, called *municipios santuarios*.

TABLE A6. EFFECTS OF ROAD INFRASTRUCTURE ON OTHER OUTCOMES

Dependent variable: Δ_t children's mortality and morbidity

	OLS REGRESSION COEFFICIENTS		IV REGRESSION COEFFICIENTS	
	(1)	(2)	(3)	(4)
	Child morbidity rate	Child mortality rate	Child morbidity rate	Child mortality rate
$\ln(Dist\text{-}NTH)$	-0.020	-0.044	-0.005	-0.087
$(-)$	(0.022)	(0.067)	(0.023)	(0.075)
Baseline controls?	Y	Y	Y	Y
Region FE	Y	Y	Y	Y
Obs.	155	154	155	155

*, **, ***, significant at 10%, 5% and 1%, Robust Standard Errors in parenthesis. Estimations include only the municipalities located within 40Km to the NTH. Columns (1) - (4) are estimations using OLS and (5) -(8) are estimated using IV. All models exclude connecting municipalities and include regional fixed effects and the following controls: municipality's population growth 2009 - 2012, geography control (log elevation) and a dummy indicating whether the municipality was part of a program from the Government to reduce violence, *municipios santuarios*.

TABLE A7. EFFECTS OF ROAD INFRASTRUCTURE ON CRIME
DIF-IN-DIF REGRESSIONS COEFFICIENTS
Dependent variable: Δ_t Crimes

	GANGS CRIMES			NON-GANGS CRIMES		
	Homicides	Extortions	Gangs Detentions	Robberies	Thefts	Drugs trafficking
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.076	-0.128	-0.814***	0.150	-0.031	-0.051
	(0.118)	(0.135)	(0.207)	(0.130)	(0.109)	(0.043)
Post	-0.495***	-0.625***	-0.004	-0.278**	-0.202**	0.313
	(0.112)	(0.126)	(0.194)	(0.111)	(0.098)	(0.079)
Post×Treatment	0.302*	0.371**	0.259	-0.356**	0.021	-0.087
	(0.156)	(0.174)	(0.269)	(0.164)	(0.146)	(0.092)
Baseline controls?	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y
Obs.	310	310	297	306	310	306

*, **, ***, significant at 10%, 5% and 1% respectively, Robust Standard Errors in parenthesis. All the estimations include only the municipalities located within 40Km to the NTH. All models exclude connecting municipalities and include regional fixed effects and the following controls: municipality's population growth 2009 - 2012, geography control (log elevation) and a dummy indicating whether the municipality was part of a program from the Government to reduce violence, called *municipios santuarios*. Dependent variables description are summarized in Appendix A1.