Name: Lelys Dinarte-Diaz

email: ldinartediaz@worldbank.org

Affiliation: Development Research Group. The World Bank[1]

Date of Ph.D.: August 2018

Personal website: https://sites.google.com/view/lelys-dinarte/home

Summary of current research activity:

I am a research economist in the Human Development Team of the World Bank's Development Research Group. My primary research fields are development economics and economics of education, with a focus on violence and crime. I also study the impacts of psychology-based interventions on mental health and economic wellbeing.

Combining experimental and non-experimental approaches, one strand of my research agenda studies how educational interventions implemented in developing countries can modify at-risk youth performance, including socio-emotional skills, mental health, and violent behaviors. I also empirically analyze the interaction between crime and welfare, noting how criminal organizations usually harm countries' economic growth path. I have built original datasets from the ground up, by combining datasets from administrative and geographic records, with primary data collected by myself. I have additional ongoing joint research projects in Colombia, El Salvador, Guatemala, Honduras, Mexico, Jamaica, Peru, and Ecuador.

I obtained my Ph.D. and Master in Economics from Pontificia Universidad Católica de Chile in 2018 and 2016, respectively. I got my B.A. in Economics from ESEN in El Salvador, the country where I was born and raised.

Summer Institute sessions I'm most interested in attending

1. Education
2. Economics of Children
3. Crime
4. Development economics

[1] Note that although my affiliation is not a university, the research group at the World Bank works as if it were an Economics Department in a university. Specifically, we have a promotion process that is similar to the tenure track in the academia.

LELYS DINARTE DIAZ

Development Research Group. The World Bank.
1818 H St NW, Washington, DC 20433, USA.
e-mail: ldinartediaz@worldbank.org
https://sites.google.com/view/lelys-dinarte

FIELDS Development Economics, Economics of Education, Economics of Crime.

EDUCATION

Pontificia Universidad Catolica de Chile | Santiago, Chile
Ph.D. in Economics. *With distinction.* *2018*
M.Sc. in Economics. *With honors.* *2016*

Escuela Superior de Economia y Negocios | San Salvador, El Salvador
B.S. in Economics. *Summa Cum Laude* *2010*

EMPLOYMENT

The World Bank | Washington, DC. 2018 - date
Economist. Development Research Group.

UNICEF | El Salvador 2017-2018
Consultant of the project *"Returns of Early Childhood Interventions in El Salvador"*

UNDP | El Salvador 2013
Consultant of the project *"Post-2015 Human Development Agenda"*

WWU-Universitaet Muenster | Muenster Germany. 2012
Visiting Researcher at AFO-Research Office.

GIZ | El Salvador. 2010-2011
Economist

PUBLICATIONS

Montoya, Maria, Maria Susperreguey, Lelys Dinarte-Diaz, Frederick J. Morrison, Ernesto San Martin, Cristian A. Rojas-Barahona, and Carla E. Forster. "Executive function in Chilean Preschool Children: Do Short-term Memory, Working Memory, and Response Inhibition Contribute Differently to Early Academic Skills?". *Early Childhood Research Quarterly* 46 (2019): 187-200.

Bandiera, Antonella, Lelys Dinarte-Diaz, Sandra Rozo, Carlos Schmidt-Padilla, Micaela Sviatschi, Hernan Winkler. "The Unintended Consequences of Deportations: Evidence from Firm Behavior in El Salvador." *Economic Development and Cultural Change* Forthcoming.

Bendini, Magdalena, and Lelys Dinarte-Diaz. "Maternal Depression and Early Childhood Cognitive Development: Evidence from Peru." *Int. J. Environmental Research and Public Health* 2020 Oct; 17(19): 7248.

Ferreyra, Maria Marta, Lelys Dinarte-Diaz, Sergio Urzua, and Marina Bassi. "The Fast Track to New Skills: Short-Cycle Higher Education Programs in Latin America and the Caribbean." Washington, DC: World Bank. **Forthcoming**

Ferreyra, Maria Marta, Lelys Dinarte-Diaz, Sergio Urzua, and Marina Bassi. "Quality

Determinants of Short-Cycle Higher Education Programs in Latin America and the Caribbean." In *The Fast Track....* Washington, DC: World Bank. **Forthcoming**

Dinarte-Diaz, Lelys, and Betsabe Vasquez. "Youth Development in the Most Violent Contexts. The Case of Central America." *Journal of Youth Development.* **Forthcoming**

WORKING PAPERS

Dinarte-Diaz, Lelys, and Pablo Egana-delSol. "Preventing Violence in the Most Violent Contexts. Behavioral and Neurophysiological Evidence." *World Bank Policy Research Working Paper 8862.* ***Revision requested at REStat***

Dinarte-Diaz, Lelys. "Peer Effects on Violence. Experimental Evidence in El Salvador." *World Bank Policy Research Working Paper 9187.* ***Submitted***

Hola, Alaka, Lelys Dinarte-Diaz, Magdalena Bendini, and Iva Trako. "What makes preschool interventions effective? A Review of (quasi) Experimental Research." *World Bank Policy Research Working Paper 9723.*

Amaral, Sofia, Lelys Dinarte-Diaz, Patricio Dominguez, Santiago Perez. "Helping Families Help Themselves: Effects of an SMS Parental and Stress Management Intervention." *World Bank Policy Research Working Paper 9850.* ***Submitted***

Dinarte-Diaz, Lelys, Maria Marta Ferreyra, Sergio Urzua, Marina Bassi. "What Makes a Program Good? Evidence from Short-Cycle Higher Education Programs in Latin America and the Caribbean." *World Bank Policy Research Working Paper 9722.*

Dinarte-Diaz, Lelys, Maria Marta Ferreyra, Tatiana Melguizo, and Angelica Diaz. "Estimating the Contribution of Short-Cycle Programs to Student Outcomes: Evidence from Colombia" *World Bank Policy Research Working Paper 9424.*

Dinarte-Diaz, Lelys, Pablo Egana-DelSol, and Claudia Martinez. "Socioemotional Skills Development in Highly Violent Contexts. Measurements and Impacts." *World Bank Policy Research Working Paper 9957.* ***Submitted***

Amaral, Sofia, Lelys Dinarte-Diaz, Patricio Dominguez, Santiago Perez, and Steffanny Romero. "Talk or Text? Evaluating Response Rates by Remote Survey Method During COVID-19." *World Bank Policy Research Working Paper 9999.* ***Submitted***

Dinarte-Diaz, Lelys, David Jaume, Eduardo Medina, and Hernan Winkler. "Neither by Land nor by Sea. The Rise of Electronic Remittances in Mexico."

Baires, Wilber, Lelys Dinarte-Diaz, and Carlos Schmidt-Padilla. "Unintended Effects of Roads: Labor, Education and Crime in El Salvador."

WORK IN PROGRESS

Bandeira, Antonella, Lelys Dinarte-Diaz, Juan Miguel Jimenez, Micaela Sviatschi, and Sandra Rozo. "Rebel Governance and Development: The Long Term Effects of Guerrilla's in El Salvador"

Dinarte-Diaz, Lelys, Helen Henningham, Shawn Powers, Saravana Ravindaran, Manisha Shah. "Non-violent parenting and stress coping support to caregivers in Jamaica."

Dinarte-Diaz, Lelys, Francisco Haimovich, Mauricio Romero, and Marcela Silveyra.

"Socioemotional Skills in Preschool. Evidence from Mexico."

Christian, Paul, Lelys Dinarte-Diaz, Jonas Heirman, Saahil Karpey, Erin Kelley, Florence Kondylis, Gregory Lane, and John Loeser. "Gender Gaps in Earnings and Economic Empowerment: Experimental Evidence from 6 Countries."

Avitabile, Ciro, Lelys Dinarte-Diaz, and Renata Lemos. "Strengthening School Management in a School Reopening Context. Evidence from Peru."

Dinarte-Diaz, Lelys, Gaurav Kahnna, and Carlos Schmidt-Padilla. "Peace Dividend: Gangs Truce and Development"

Dinarte, Lelys, and Julio Galarraga. "Long-Term Effects of Roads on Economic Performance in Ecuador."

CONFERENCES

2022. IPA Webinar on Remittance Flows and COVID-19, IPA Webinar on Violence and Crime in LAC, *WALES 2022, *Advances in Field Experiments-Chicago (*Scheduled).

2021. SCPs-World Bank Authors'Workshop, RIDGE BRAIN, University Adolfo Ibanez Seminar, PUC Alumni Seminar, UDenver, AEFP 2021, (Scheduled) University of Munchen, AlCapone 2021, SEA 2021.

2020. ASSA Meeting, BREAD (Mental Health Workshop), Texas A&MU Crime Economics Online Seminar, GWU (Economics of Education Seminar).

2019. Brown University-Watson Institute, SEA 89th Annual Meeting, WADES, TX Economics of Crime Workshop (Texas A&MU).

2018. LACEA-LAMES, RIDGE Forum (Public Economics), Job Market Seminars at World Bank, IADB-RES, U. Javeriana, U. Los Andes, U. del Rosario

2017. NEUDC (Boston). FEN-Universidad de Chile, UDP, CESS-Universidad de Santiago and PUC Seminars (Santiago). SECHI Annual Meeting. 5th EH Clio Lab Annual Conference (Santiago). 1st Workshop in Urban and Regional Economics (BogotÃ¡). ESA World Meeting (San Diego, CA). 7th Annual Meeting of America Latina Crime and Policy Network AL CAPONE (Santiago). 10th Maastricht Behavioral and Experimental Economics Symposium and 3rd Maastricht Behavioral Economic Policy Symposium. PUC Seminar. Centre for Experimental Social Sciences Seminar. 5th Antigua Experimental Economics Conference (Guatemala).

2016: 1st Multidisciplinary Conference (Chile). International Summer School in Applied Microeconomics (Germany).

OTHER ACADEMIC ACTIVITIES

Referee Services: The Quarterly Journal of Economics, Economic Journal, EDCC, JDE, World Bank Economic Review, Latin American Journal of Economics (LAJE), IDB Working Paper Series.
Membership: LACEA, SECHI, AEA.

GRANTS, FELLOWSHIPS AND AWARDS

- World Bank Education Finance Thematic Group Trust Fund for project *"Understanding Principalâs Demands for School Reopening..."* US$31,000.
- World Bank Research Support Budget for project *"Guerrillas and Development"* US$25,000.

- World Bank Research Support Budget for project *"Virtual Parenting Support in Jamaica..."* US$88,000.
- World Bank Early Learning Partnership Fund for the project *"Non-violent parenting and stress coping support to caregivers in Jamaica..."* US$300,000.
- World Bank Research Support Budget for project *"Neither by Land..."* US$16,000.
- World Bank Research Support Budget for project *"Strengthening School Management..."* US$90,000.
- World Bank Research Support Budget for project *"Gang Truce and Development..."* US$18,900.
- World Bank Research Support Budget for project *"Addressing Parental Stress During the Pandemic..."* US$79,000.
- World Bank Early Learning Partnership Fund for the project *"Addressing Parental Stress During the Pandemic..."* US$9,100.
- IDB Research Support Budget for project *"Addressing Parental Stress During the Pandemic..."*. Co-PI. US$150,000.
- World Bank Research Support Budget for project *"Socioemotional Skills in Preschool. Evidence from Mexico"* US$94,800.
- EFO funding for project *"Socioemotional Skills in Preschool. Evidence from Mexico"* US$20,000.
- World Bank Research Support Budget–Data Collection Window–for project *"Distribution of Violent Crimes..."* US$30,000.
- World Bank LAC-Gender Innovation Lab Grant for project *"Crime Exposure and Women's Economic Decisions in the Most Violent Contexts."*
- IDB Research Support Funding for project *"Repatriations, Jobs and Firms..."*. US$18,000.
- WFP-DIME Grant for project *"Cash-Based Transfers and Women Economic Decisions. Experimental Evidence from El Salvador."*. Implementing Partner: World Food Program. US$290,000.
- Templeton World Charity Foundation Grant for project *"Building the Local Case for Character Development in High Conflict Settings in the Northern Triangle in Central America"*. Principal Investigator. Co-PI: Claudia Martinez and Pablo Egana. Implementing Partner: Glasswing International. US$233,783.57.
- Grants for Graduate Thesis 2016. CEDLAS, Argentina. Project: *Peer effects in After School programs. Evidence from El Salvador.* (2016-2017). US$6,000
- Fellowship for Research Dissemination. VRI Pontificia Universidad CatÃ³*licadeChile.*(2015, 201
- National Commission for Scientific and Technological Research (CONICYT) Scholarship. Chile (2014-2017).

LANGUAGES Spanish (Native), English (Fluent), German (Basic).

Peer Effects on Violence: Experimental Evidence from El Salvador*

Lelys Dinarte Diaz†

October, 2021.

Abstract

This paper provides experimental evidence of the effect of having peers with different propensities for violence in the context of an after-school program. By randomly assigning students to participate with a set of similar or diverse peers in terms of predicted violence, I measure the effects of segregation or integration on students' behavioral, neurophysiological, and academic outcomes. Results indicate that integration of students with different propensities for violence is better for both highly and less violent children than segregation. The intervention can have unintended effects on misbehavior and stress if highly violent students are segregated from their less violent peers.

Keywords: Peer effects, Segregation, Integration, School Behavior, After-School Program.

JEL Classification: I29, K42, Z13

*This paper has greatly benefited from helpful feedback from Sule Alan, Kathleen Beegle, Jennifer Doleac, Pablo Egana, Francisco Gallego, Sara Heller, Kirabo Jackson, Eeshani Kandpal, Jeanne Lafortune, David McKenzie, Claudia Martinez, Owen Ozier, Tomás Rau, Rodrigo Soares, Micaela Sviatschi, José Tessada, and Juan Vargas. I appreciate the comments and suggestions of participants at the ASSA 2020, WADES 2019, LACEA-LAMES 2018, 2017 NEUDC Conference, 2017 ESA World Meeting, 7th Annual Meeting of AL-CAPONE, 2017 SECHI Annual Meeting, and the 5th Antigua Experimental Economics Conference. Additionally, I appreciate comments from participants during seminars 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 truly grateful for the incredible support of Glasswing International as an implementer partner, especially to the M&E and education sector teams. I am also grateful to principals, teachers, students, and instructors of the five public schools in El Salvador. This study was funded by CEDLAS-IDRC (Canada), CONICYT Chile, and USAID's El Salvador Office. The contents of this study are the responsibility of the author and do not necessarily reflect the views of her employer, funding agencies, or governments. This study was registered in the AEA RCT Registry with unique identifying number AEARCTR-0001602. This research project's protocol was reviewed and approved by the Institutional Review Board (IRB) at the *Pontificia Universidad Catolica de Chile* on April 2016 with the ID protocol 160314001.

†Corresponding Author. The World Bank. Development Research Group. Email: ldinartediaz@worldbank.org. Postal Address: 1818 H St. NW, Washington DC, 20433. Phone: +1 202 873 5331

1. Introduction

Segregation of children and adolescents based on their socioeconomic background represents one of the most significant policy challenges in terms of inequality. Notably, segregation can deprive them of the benefits of diversity in all dimensions, including behavioral conduct. Existing evidence shows that peers can have important effects on individuals' economic decisions. During adolescence in particular, young people are exposed to different groups of peers. This exposure can influence subsequent human capital accumulation and behaviors,[1] and thus the effectiveness of social interventions, such as violence-reduction and educational programs. In the criminal justice policy space in particular, most violence-reduction programs are implemented in group settings.[2] However, little is known about how the composition of the groups affects the program effectiveness.[3]

From the existing literature on group composition and social programs, there are two important evidence gaps. First, most existing papers only identify the effect of exposing an individual to a group of peers with certain average characteristic. Yet, with some notable exceptions,[4] they do not directly compare average effects of different allocation of participants in the interventions. As a result, there is little consensus about what type of group composition—integration or segregation[5]—is more effective in improving relevant economic outcomes. Some studies find that integrated groups are preferable than segregated ones because the interactions with high-performing peers improves outcomes for more disadvantaged individuals, due to the enhancement of the learning experience.[6] However, another strand of the literature finds that grouping individuals with similar

[1] Bursztyn et al. (2019); Bursztyn and Jensen (2015)

[2] This can be determined by implementation costs or because it might be easier to convince at-risk individuals to engage in such programs in a group rather than one-on-one.

[3] For example, as noted by Davis and Heller (2020), existing evidence from the US show that only very expensive training programs for at-risk youth improve labor market outcomes and a very limited number reduce crime in the short term. A potential explanation could be that the negative peer effects from spending time with other hard-to-employ peers might cancel out any benefits from the training.

[4] Garlick (2018); Duflo et al. (2011); Lafortune et al. (2016)

[5] In this paper, I use the terms segregation and tracking interchangeably.

[6] Lafortune et al. (2016); Lavy et al. (2012); Rao (2019); Griffith and Rask (2014); Oreopoulos et al. (2017)

peers can generate better results, since individuals prefer to interact with similar peers or because it allows to a more specialized instruction.[7] Second, despite the exposure to violent peers can have important negative impacts on human capital accumulation and criminal behavior later in life,[8] with the exception of Davis and Heller (2020), there has been little rigorous analysis of the impacts of group composition in terms of propensity for violence on behaviors, economic outcomes, and programs' effectiveness.[9] This paper conducts a randomized controlled trial designed to address these two gaps.

We conducted an experimental evaluation of an After-School Program (ASP) and find that the intervention improves academic performance, behavior at school, and mental distress of participants (Dinarte and Egana, 2019). Then, I embedded a second experiment on top of this one to study group composition effects. Specifically, I experimentally manipulate whether students participated in the ASP in homogeneous or heterogeneous groups according to their initial predicted propensity for violence. The empirical design overcomes some issues related to the identification of peer effects, such as the reflection problem and the strong assumption of the separability of peer composition and other confounding effects within groups (Manski, 1993; Angrist, 2014). In addition to estimating average treatment effects, my study design allows me to exploit a discontinuity in the estimated propensity for violence to estimate effects on marginal group members. The analysis focuses on the differential impacts of exposure to a particular composition of peers on academic, neurophysiological, and behavioral outcomes, or if it changes students' effort at school. My results show that it is more beneficial to mix students with different levels of predicted violence rather than to segregate them into more and less violent groups for some behavioral and neurophysiological outcomes. To my knowledge, this paper provides the first experimental evaluation of the effects of

[7] Duflo et al. (2011); Carrell et al. (2013); Girard et al. (2015); Goethals (2001)

[8] Billings et al. (2019); Di Tella and Schargrodsky (2013); Bayer et al. (2009); Damm and Dustmann (2014).

[9] Most of the existing research examines group composition effects by income, academic or labor performance, or gender (Duflo et al., 2011; Rao, 2019; Lavy et al., 2012; Zimmerman, 2003; Bandiera et al., 2010; Lafortune et al., 2016; Gong et al., 2021; Burke and Sass, 2013; Carrell et al., 2018; Hoekstra et al., 2018; Barseghyan et al., 2019; Carrell et al., 2019, 2009, 2008)

peer composition by propensity for violence. It studies social dynamics that can operate in nonlinear ways and generates the empirical results that bring new evidence to the literature on how group composition determines the effectiveness of a social program.

The ASP I study in this paper consists of clubs implemented after school but within school facilities in El Salvador from April to mid-October during the 2016 academic year.[10] Students participated in two sessions per week that lasted 1.5 hours each. Every session combined: (i) a discussion oriented toward fostering children's conflict management, violence awareness, and social skills (similar to content used in cognitive behavioral therapy); and (ii) club curricula that 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 1,056 enrolled students ages 10–16 years from five vulnerable public schools, where children have a high risk of engaging in or becoming victims of criminal activities.

The Salvadoran context is relevant for this analysis because youth violence is of foremost importance. In 2015, El Salvador was one of the world's five deadliest places for young boys (UNICEF, 2017). In recent years, 18% of students reported that they dropped out of school due to delinquency within schools or in the surrounding neighborhoods (MINED, 2015). Thus important policy implications emerge from studying how to increase the effectiveness of an intervention designed to protect at-risk children and adolescents while teaching them how to address and control violent behaviors.

To measure group composition effects and to exploit the existence of excess demand for the program, within each participating school, I randomly assigned individual students according to their initial propensity for violence to a group with a heterogeneous or homogeneous combination of peers[11] or to a control group. Students in the homogeneous treatment were separated into two

[10] El Salvador's academic year begins in January and ends in November

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

subgroups by considering their percentile in the distribution of violence—that is, students whose predicted violence was higher (lower) than the median were assigned to a club with peers with high (low) predicted propensity for violence.[12] Participants in the control group went home after school time. This design was inspired by Duflo et al. (2011) and Lafortune et al. (2016). Randomization ensured 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 as well as enrolled children's exposure to risky spaces. I combined this self-reported information with neurophysiological evidence, particularly measures of stress and emotional regulation, from a random subsample of enrolled students. I used low-cost, portable electroencephalograms within an in-field lab setting. Finally, I also collected administrative records on grades, behavioral reports, and absenteeism data from the students. Schools provided these data before and after the intervention.

I report three sets of results. First, when I compare both types of group composition, integration is better than segregation. Results show that, on average, improvements in attitudes and misbehavior at school are larger when participants are in more diverse groups than when they are segregated. This is true for both high- and low-violence children. These results align with the evidence that interactions with diverse peers can generate differences in the learning experience (Lafortune et al., 2016).[13] In this sense, students in heterogeneous groups have the opportunity for exposure to both good behaviors they should follow and negative ones they should reject. These interactions are less available for students in the homogeneous group. In terms of neurophysiological outcomes, I find that students exposed to similar peers in terms of violence—that is, assigned to homogeneous groups—have stress levels greater than those of students placed in a heterogeneous setting. The

[12] Due to this design, the interpretation of results is as the effect of changing both variance and mean of peers propensity for violence.

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

increase in stress is particularly great for children treated in a homogeneous and highly violent group versus comparable children treated in heterogeneous groups.

Second, I study tracking effects on marginal students, who are just above or below the median of the propensity distribution function within each stratum. By virtue of my design, very similar students around this cutoff were assigned either to homogeneous highly or less violent groups. By exploiting the discontinuity around the median and using only the sample of children assigned to the homogeneous treatment, I find evidence that marginal students are negatively affected by being assigned to the most violent group in both academic outcomes and misbehavior at school. In this sense, segregation is not only worse than integration for the average, it is especially problematic for the marginal student. This result contributes to the literature on how segregation by initial violence may encourage the formation of networks of violence (Billings et al., 2019; Di Tella and Schargrodsky, 2013; Bayer et al., 2009) and affect individuals who were supposed to be the key beneficiaries of these types of interventions.

Third, both integration and segregation of students by their initial propensity for violence generate better average effects on behavioral, neurophysiological, and academic outcomes than no ASP at all. In summary, these three pieces of evidence on peer effects can indicate that having some highly violent peers can constitute a learning alternative for children less prone to violence because they can see the behaviors they should not emulate. However, the jump around the median in the tracking group also indicates that being exposed to a more significant share of "bad" peers can have the opposite effect. This implies there is an optimal "bad"-to-"good" peer combination that maximizes the program's overall impact.

This paper contributes with causal evidence to the discussion of segregation versus integration as an optimal strategy for assigning participants to a social intervention. The effects on academic and noncognitive outcomes as a result of integration that I present here accord with a body of microlevel evidence; it explains that these effects likely stem from the interaction between diverse

individuals within groups.[14]

My results are mainly similar to those of Rao (2019) and Davis and Heller (2020). The first paper finds an improvement in some social preferences outcomes such as generosity, prosocial behavior, and equity when there is an exogenous change in wealth heterogeneity in India. Davis and Heller (2020) emphasize the potential gains from more flexible approaches to study treatment heterogeneity in the context of a summer job program targeted to at-risk youth. My study is novel in its modification of the composition regarding violence and the inclusion of analysis of peer effects on additional noncognitive outcomes such as violence, misbehavior, and attitudes toward school and learning that are important in developing countries.

There is also a growing body of evidence that finds benefits from tracking. Theoretically, Lazear (2001) shows that, amid 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).[15] In contrast to those papers, my results indicate that tracking can have unintended effects on academic outcomes and misbehaviors when it targets 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 club curricula to their groups' needs. In fact, my results fit 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 levels to which teachers orient

[14] See Sacerdote et al. (2011) for a summary of literature on peer effects on student outcomes in educational settings. Further recent evidence studying peer effects on other cognitive and noncognitive outcomes are the works from Gong et al. (2019); Fletcher et al. (2019); Fletcher and Ross (2012); Burke and Sass (2013); Carrell et al. (2018); Hoekstra et al. (2018); Barseghyan et al. (2019); Carrell et al. (2019, 2009, 2008).

[15] 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 homophily along several dimensions.

instruction. Under this assumption, tracking by violence worsens outcomes for those above the median of the distribution of violence in the group to which they are assigned; but increases the performance for those below it.

The remainder of the paper is organized as follows: section 2 describes the intervention, data collection, and study design. Section 3 summarizes the specifications used to estimate the effects of the intervention on academic, behavior, and 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. Finally, the conclusions appear in Section 6. All appendix tables are at the end of this paper.

2. The Tracking-by-Violence Experiment

2.1 Intervention: After-School Clubs

I leverage an ASP program in El Salvador to study how the composition of groups (based on past violence attributes) impacts the outcomes. This ASP was implemented by an NGO, Glasswing International. I partnered with Glasswing to design and implement an experimental evaluation that (i) measures the impact of this ASP as is and (ii) identifies the best implementation approach to improve its effectiveness.

In Dinarte and Egana (2019), we address the first objective. Our main estimations indicate that this ASP not only improved attitudes toward school but also enhanced learning and lessened misbehavior at school. These two effects translate into an improvement in participants' academic performance. We also estimated neurophysiologial measures of emotional regulation and stress using electroencephalogram recordings. We find evidence that control of emotions and automatic responses to stimulus is a potential mechanism for the effects on behaviors and academic attainment. By implementing the tracking-by-violence experiment I describe below, this paper addresses

the second objective. Therefore, I highlight in this section only the curricula structure and enrollment process that will help readers to better understand the experimental design and its results.

The ASP implementation is part of the NGO's program *Community Schools* in five public schools in highly vulnerable communities in El Salvador. According to the intervention approach, its main objective is to modify children's violence and attitudes through the acquisition of life skills, thereby improving their academic performance (Glasswing International, 2012). The NGO offers four categories of clubs—Leadership, Art and Culture, Sports, and Science—in the ASP by education level (*ciclos*). Considering this intervention structure, the experiment was designed by using the natural school-by-education-level organization as the stratification variable.

Clubs meet twice a week for approximately 1.5 hours each and take place immediately following school hours. This intervention occurred between April and October of 2016. Each session is divided into two sections according to the services an ASP can provide. The first section is social skills development, which relates to the learning service of an ASP (Taheri and Welsh, 2016; Durlak et al., 2010). This section is common to all participants and includes topics such as management of conflict and risk, school violence reduction, and soft skills. Some of this section's activities originate in exercises from Cognitive Behavioral Therapy and positive psychology. By implementing experiential learning and role playing, these activities try to make students aware of certain behaviors, to disrupt these patterns, and to promote better ones.

The session's second section relates to the protection—or incapacitation—aspect of an ASP (Gottfredson et al., 2007; Mahoney et al., 2001). During after-school hours, children are under adult supervision as a means of preventing them from being exposed to risks in their communities. In this section, instructors implement activities related to each club's category. For example, in the sports clubs, children play soccer and basketball, among other athletic pursuits. The sports element also attracts student participation and therefore increases ASP attendance.

At the beginning of the school year, the NGO visits schools to offer the program and enroll participants. Out of a total of 2,420 children from the five beneficiary schools, 1,056 students ages

10–16 years are recruited and enrolled to participate in the ASP. Children can self-enroll; they are only required to bring a parent's signed authorization to participate.

During the registration stage, children are asked to complete an enrollment form. It collects personal and family information I describe below. Then they are assigned to a group in light of their preferences, parent's authorization, and aggregated demand for each club category. Clubs are composed of 13 students from a single educational level.

2.2 Experimental Design

This paper aims to provide experimental evidence on the best targeting approach to increase the effectiveness of an ASP. After randomly assigning enrolled students to treatment or control groups, I create an additional exogenous variation on each student's peer composition in terms of violence. This is to test if integration or tracking is the best implementation strategy in the context of this program.

Therefore, this design has different steps: violence measure estimation per participant, random allocation of enrolled students to different treatment arms, collection of relevant data at different stages, and experimental design robustness checks. This last component includes other distributional criteria in addition to balance on observables across treatments before the program. I describe all of them in this subsection.

Propensity-for-Violence Index (IVV) Estimation

To assign enrolled students to each treatment group, I first needed to measure their propensity for violence. It was not possible to ask directly about this during the registration phase because we could not guarantee this personal information would be kept confidential. For example, either the local authorities or gang organizations might force the research team or the NGO to reveal information that identified each child, risking not only the intervention but also—most important—

the children's safety. Additionally, the inclusion of specific questions about gang membership and other associations with these organizations, which are highly correlated with crime and violence in El Salvador, might endanger both children and instructors.

Instead, I estimate a predictive model of violence and crime from existing data using a Two-Sample Least Squares strategy. First, using an existing anonymized database of youths' violence and crime from El Salvador (FUSADES, 2015),[16] I estimate the likelihood of having committed a violent act V_f as a function of a wide range of covariates:

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

where D_f is a vector of violence determinants of student f in the FUSADES dataset. Based on existing evidence, I include in this vector some variables that are associated with individuals' vulnerability to violence, such as student characteristics (e.g., age, gender, time spent alone at home, and education level); children's household variables (e.g., residence area, mother's education, and household composition); and school-level controls (e.g., school location and commuting time to school).[17] Descriptive statistics and comparison of means (p-values) between the FUSADES sample and the one from this study are in Appendix Table A1. 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 hours.

All estimated coefficients $\hat{\alpha}_1$ have the expected sign according to the literature of violence determinants, as shown in Appendix Table A2. For instance, boys are more likely to be violent than

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

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

girls, adolescents behave worse than children (Rodriguez-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. Overall, lack of parental supervision is the most important determinant of propensity for violence in this sample.

Then, exploiting the availability of these variables in the registration forms of enrolled students, I predict the measure of propensity for violence (IVV) for each child, using the vector of estimated coefficients $\hat{\alpha}_1$. Two features of this IVV are important to emphasize. First, since the variables included in the estimation relate to students' violence exposure at different domains (family, school, and community), this measure is a more accurate proxy of students' overall propensity for violence than are reports of students' misbehavior from school records. Second, this predicted index can be interpreted as a measure of students' *propensity* for violence rather than as an indicator of *effective* violence.

An important caveat of this IVV is that it is not a perfect measure of (effective) violence. However, I provide some evidence to document that it is the best proxy of propensity for violence considering the restrictions of data availability and collection 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),[18] these types of crime and violence models estimated from existing data have an acceptable predictive power.[19] As I mention below, in using misbehavior reports as the classification variable in the experimental design, estimations indicate that I would have had a similar classification in an important share of the total sample.

An additional concern is that this index might explain another factor like school performance. Thus, I estimate the correlation between the predicted index and grades reported by teachers, and I find that it is not statistically significant. Yet, I also find that the correlation between the predicted

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

[19] Klassen and O'Connor (1988) use a sample of adult males at risk for violent behavior who were admitted as inpatients at a community mental health center. They find that this model correctly classified 85% of the total sample.

IVV and misbehavior at school is positive and statistically significant at 1%. In Appendix Table A3, I present these estimations using different standardizations of academic grades and behavior reports.[20]

Finally, the IVV predicts both intensive and extensive margins of future misbehavior.[21] Using data from students in the control group, I find that the correlation between IVV and bad behavior at the end of the academic year is positive and statistically significant at 5%. The estimation strategy and main results are in Appendix Table A4.

Treatments

After estimating the IVV, enrolled children were randomly assigned to two groups—control (C, 25%) and treatment (T, 75%)—within each school-by-educational-level "block."[22] Then, in a second randomization stage, treated children were randomly assigned to two treatment arms—heterogeneous (HT, 25%) and homogeneous (HM, 50%)—as shown in Figure 1.

Next, students in the HM set 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 HM-Low and HM-High groups are defined using the median in each school-by-educational-level block because I wanted variation in the cutoff. Moreover, an uniform cutoff across all randomization blocks could generate differences in group sizes that would probably confound the effects of tracking with group size. This design permits me to judge if targeting helps to improve the ASP's effects and allows me to test the potential existence of peer effects and heterogeneity by initial propensity for violence in the program. In addition, this strategy utilizes a

[20] Another potential and valid concern raised by an anonymous referee is that the IVV is a measure of socioeconomic disadvantage. Using mother's education as a proxy for socioeconomic status, I estimate correlation between this variable and IVV. I find that this correlation is not statistically significant (p-value = 0.185).

[21] Intensive margin consists on changes in the misbehavior score. Extensive margin is measured as the probability of having a misbehavior report higher than 5, the official threshold that determines if a child will receive a disciplinary sanction.

[22] Each education level consists of three years of schooling: the first is from first to third grade, the second from fourth to sixth grade, and the third from seventh to ninth grade.

regression discontinuity (RD) design approach to measure the impact of tracking on the marginal student and contrast it to the average impact.

Treatments are described below:

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

2. *Homogeneous-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 stratum.

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

4. *Control:* This group of students was not selected to participate in the clubs during the 2016 academic year. They left school facilities after school hours. We were able to collect their information at follow-up because we gave them a "participation coupon" they could redeem to take part in the ASP the next year.

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 the effects of interactions among participants instead of other channels such as teaching or curriculum adaptation. To test for changes in teaching methodologies, I collected information from a trainers' survey and present the results in upcoming sections.

2.3 Data

The time line of the study is shown in Figure 2. As previously mentioned, after the NGO advertised the ASP in school facilities, a research team returned to schools to enroll and register participants.

In this stage, students were asked to bring a consent form signed by their parents or tutor and to complete a registration form. This instrument collected personal and family information such as age, gender, mother's education, and average commuting time, among others. These were used to estimate the IVV. I also collected school records of academic grades, absenteeism, and behavior reports for all children.

According to the ASP's theory of change, the program can directly affect behavioral and neurophysiological outcomes such as children's violence, misbehavior at school, and emotional regulation. It may also have some indirect effects on academic performance, since changes in noncognitive outcomes can affect cognitive skills (Cunha and Heckman, 2008). Since this paper investigates differences in those outcomes by group composition, I study the effects of tracking and integration in the same three categories of variables.

First, I collected short-term follow-up data on attitudes toward school and violent behaviors from enrolled participants in school facilities. Students completed a follow-up survey in classrooms set up especially for this purpose at the end of October 2016, after all clubs had finished implementing their curricula. Each survey took approximately 45–60 minutes. Most surveys were self-administered, with assistance from staff trained in the survey methodology. To increase statistical power to detect effects for outcomes within a family and to reduce the number of hypothesis tests, I construct indexes of variables that are expected to move in a similar direction (Haushofer and Fehr, 2014; Heller et al., 2017). Since self-reports might be affected by social desirability bias, I attempted to recheck and validate these behaviors and attitudes using proxies for these outcomes obtained from administrative data such as absenteeism figures and reports of misbehavior at school.

For the analysis of group composition on stress and emotional regulation, I use neurophysiological recordings collected for our work (Dinarte and Egana (2019)) from a random subsample of enrolled students.[23] Finally, to measure the effects of group composition on academic per-

[23]More details on the data-collection process for this third group of outcomes are discussed in Dinarte and Egana

formance, schools provided administrative records of math and science grades at the end of the academic year. Appendix 1 includes a description of all collected data and main outcomes.

As shown in Appendix Table A5, the average matching rate of administrative data of enrolled children was 94% at baseline and 97% at follow-up. All matching rates were balanced between treatments and C groups except for the fraction of math grades at baseline between the HM and C groups, significant at 10%; and in absenteeism between both tracking groups, also significant at 10%. To account for this difference, I include the imputed grade for missing observations at baseline and a missing value indicator in all specifications for the academic outcomes. Additionally, the average matching rate of administrative data of nonenrolled students was 85% at baseline and 98% at follow-up.

The share of initially enrolled students that filled out the follow-up survey after the intervention was 92%, on average; for the HM and HT groups, it was 91% and 94%, respectively. There were no statistical differences between treatment and C groups in overall attrition rates. Therefore, results are not driven by the absence of follow-up survey data for any group. For the neurophysiological measures, after filtering electroencephalogram (EEG) recording data, the average attrition share was 49%. In Dinarte and Egana (2019), we present several checks to verify that this attrition rate was not correlated with the intervention. We argue that attrition was caused mainly by the quality of the data recordings. For example, long, dense, or dirty hair and/or "frozen" computers were the most common troubles that the Matlab toolbox encountered in reading the EEG recordings.

2.4 Summary Statistics

Descriptive statistics of the full sample and each treatment and control group are shown in Table 1. Column 1 exhibits statistics for the control (C) group and columns 2 and 3 for the treatment (HT and HM) groups, respectively. Columns 4–5 show statistics for the two homogeneous subgroups.

(2019).

Panel A presents summary statistics of the violence determinants. Participants are on average 11.9 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 nonrelated adult. On average, 62% of students' mothers have an intermediate education level (7–12 years), and 31% have fewer than six years of schooling. Regarding risk exposure, only 5% of students report being alone at home when they are not at school. On average, they travel around 18 minutes to school.

Finally, the last row of panel A shows that the average propensity for violence for any treatment and C groups is 0.038, with a standard deviation of 0.029, ranging from 0.001 to 0.215. This average propensity for violence is 14 times the mean probability that a given student will be vulnerable to violence in Chicago (Chandler et al., 2011). Even when both estimations are not completely comparable (because I use fewer violence determinants than Chandler et al. (2011)), this difference sheds light on the tremendous propensity for violence of the children in this study.[24]

Panel B shows academic scores and absenteeism for the 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 had 6.5 points. The mean absenteeism rate in the first quarter, before the intervention, was 4.2% (1.69 out of 40 days).

2.5 Experimental Design Checks

This experimental design must meet five requirements to generate an exogenous variation that allows me to identify the causal effects of group composition in terms of violence. First, the treatment and C groups must be balanced. I find some differences in means between the treatment arms and the comparison group. p-values for all the tests of differences in means between each treatment—the HT, HM, HM-H, HM-L—and the C group are in Appendix Table A6.

For example, in the comparison between the HM and C groups, there are differences in the

[24]More descriptive statistics of the predicted propensity for violence are presented below.

share of students who live with both parents or only with one of them. In addition, children in the HT group attend higher-level courses than students assigned to the C group. When I compare the HT and HM groups, there is a slight difference in the mean propensity for violence. Finally, average absenteeism is higher for students in the C group compared to any treatment arm. However, after adjusting p-values for multiple hypothesis testing of means and FWER, all differences are not statistically different from zero. Yet I account for all the differences by including these variables as controls in the estimations. Specifically, I control by the percentile of predicted IVV, household composition, absenteeism, and student's school year. Additionally, in academic outcomes specifications, I include the respective grades at baseline to account for the differences in academic performance before the intervention.

A second condition is that the HM-High group's IVV should be greater than that of the HM-Low group, also expressed in most of its determinants. As evident in columns 4 and 5 in Table 1, and which I can verify using tests for differences, 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 they live in an urban area, face greater travel time to school, and most of them spend time at home alone.[25] Finally, the average academic performance of students in the HM-High group is worse than that of children in the HM-Low treatment.

As the assignment to the HM and HT groups was defined over the predicted violence index, the third requirement is for the design to effectively generate changes in the propensity for violence of one's clubmates, depending on the random assignment. As I show in Table 2, consistent with the premise that nontracking groups are more violence diverse than any of the tracking groups, the standard deviation of the HT group was 0.010 and 0.018 points higher (35%–78%) than the same figure for the HM-H and HM-L groups, respectively. Additionally, a second premise held that the

[25]Most students in the HM-Low group have mothers with either basic or higher education. These results could be explained as follows: if the 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.

average violence level of the HT group must be between the HM-Low and HM-High levels. This design fulfills these conditions, as is apparent from the results in Table 2: the average HT group's IVV is between those of the HM-High and HM-Low.

The fourth requirement relates to three desired characteristics of the IVV distribution functions of the HT, HM, and C groups, before treatment. The first one is that these distributions must be similar at the baseline. Using the two-sample Kolmogorov-Smirnov test for equality of distribution functions, hypotheses are not rejected—p-values of 0.62, 0.89, and 0.68 for the HT-HM comparison, the HT-C comparison, and the HM-C comparisons, respectively. The similarity among distributions can be verified as well in Figure 3. The second characteristic is that the distributions of the HT, HM-High, and HM-Low groups must differ. As Figure 4 illustrates, there are differences among the three groups' distributions. Particularly, using the two-sample Kolmogorov-Smirnov test, I reject the hypothesis of equality of each comparison of distribution functions pairs at 1%. The last desired feature is that the distributions of the HM-High and HM-Low groups should not fully overlap in the full sample in order to have some variability between both HM subgroups. Had I not stratified, there would be no overlap between both groups. However, as the assignment was defined within each stratum, there is overlap in 67% of the sample, as shown in Figure 5.

The fifth condition is that there must be a sharp discontinuity at the fiftieth percentile for the HM subsample, consistent with the discontinuous assignment at the median IVV within each stratum. Figure 6 shows the predicted IVV median of a student's clubmates as a function of her own IVV, and the expected jump at the fiftieth percentile. Moreover, a RD-robust estimation using only this homogeneous subsample indicates that students assigned to the HM-High group are enrolled with peers with a mean IVV 0.8 points greater and statistically significant at 5%.[26]

Finally, I contend that this IVV is a good proxy for violence because even after using misbehavior reports as the classification variable for high and low propensity for violence, estimations

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

indicate a similar classification in 53% of the total sample. Crucially, there are no differences in the classification among treatments, as the last row of Appendix Table A7 shows.

3. Empirical Strategy

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

3.1 Group Composition (GC) Effects

This study design creates a direct experimental variation on GC regarding violence. Thus, 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:

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

where y_{ij} is the postintervention behavioral, neurophysiological, and academic outcome of student i in school and education level j. HM_{ij} and HT_{ij} are dummies that indicate whether student i in school level j is assigned to the HM or HT treatment, respectively. X_{ij} is a vector of control variables measured at or before baseline, including a second-order polynomial of student's IVV percentile.[27] For the academic outcomes regressions, I also include standardized grades at

[27]Main results are similar when we use a linear polynomial of student's IVV percentile. We prefer to use a second-order polynomial because it gives a better fit of our model to the data. Moreover, the association between crime and violence and other social variables is non-monotonic (Fonoberova et al., 2012). In fact, despite higher levels of crime can affect a particular social outcome, after certain threshold those effects grow at a decreasing rate due

baseline—including imputed values—and a missing baseline grades indicator as controls. Finally, I also control for "randomization blocks" with school-by-education-level fixed effects S_j. Due to the possible uncertainty in the estimation of the IVV, standard errors are adjusted using a cluster bootstrapped at the course-school level (Treiman, 2009). As a robustness check, I also estimate robust standard errors for the average effects.

In this setting, θ_1 (θ_2) can be interpreted as the effect on student i of receiving an offer to participate in the ASP with a homogeneous (heterogeneous) composition of violent peers, compared to effects of the C group. Testing for differences between the estimated coefficients θ_1 and θ_2 indicates the effects of group composition—tracking or integration—on the outcomes of interest.[28]

I can also exploit the variation in peer quality generated by the experiment. Since participants in the HT subsample were randomly assigned to a group in the ASP, they will have a random set of peers. I can restrict the sample to these groups and estimate the effect of a student's peers mean and variance baseline IVV. An alternative specification that can help to increase the efficiency in the estimation is to control for the mean and variance baseline IVV at the stratification-block level and to restrict the sample to treated students. Details of these estimation approaches are in Appendix 2. With these estimations, I can directly provide causal evidence of how student's i behaviors and/or academic outcomes are affected by the average or variance in the violence of her peers.

3.2 GC Heterogeneity by Initial IVV

By design, the HM group comprises two different subgroups: HM-High and HM-Low. This allows me to also analyze differential effects of group composition for children assigned to the lower and upper section. The assignment variable to those subgroups was the median of the IVV distribution at each HM-stratum level. Therefore, after controlling by the indicator IVV_high_{ij} and by the

to saturation.

[28] As an anonymous referee noted, θ_1 and θ_2 are in fact capturing the effect of changing not only the mean but both the variance and mean in the distribution of peer propensity for violence–or the elements that constitute the IVV.

IVV median at the j level, $I\bar{V}V_j$, I can directly compare the results of each HM subgroup with the respective HT treatment, estimating the following specification and using the T group only:

$$Y_{ij} = \theta_0 + \theta_1 HomH_{ij} + \theta_2 HomL_{ij} + \theta_3 IVV_high_{ij} + \theta_4 I\bar{V}V_j + \theta_5 X_{ij} + \epsilon_{ij} \quad (3)$$

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, with the rest of the variables defined as before.

Specification (3) allows me to compare both treatments within each half of the IVV distribution, which is equivalent to including in specification (2) an interaction between Hom_{ij} and IVV_high_{ij}. In the upper half, θ_1 is an ITT estimator of assigning a child i with higher propensity for violence to a low-violence diverse set of peers, compared to allocating her to a high-violence diverse group. Also, for the lower half of the IVV distribution, θ_2 is an ITT estimator of assigning a less violent child to a low-violence diverse group of peers compared to a heterogeneously violent group.

3.3 Effects of Tracking on the Marginal Participant

Results of previous equations enable 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 fiftieth percentile of the IVV distribution. Given that these just-above-the-median children have a similar propensity for violence to those at or below the median, I exploit their assignment to a group of high-IVV peers and compare them with others in a low-IVV set.

It is interesting to study effects on the marginal participant because having highly violent peers (on average) means she is the least violent child in her group before the intervention; 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. Therefore, they may experience greater tracking impact.

To identify this impact, HM groups provide a natural setup for an RD design, with the median of the IVV distribution in each stratum as the discontinuity. To maintain the validity of this strategy, the assumption is that nothing else changed discontinuously around the point of separation between the two groups, which holds true in this design. I estimate the following equation:

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

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

4. Results

In this section, I present reduced-form estimates of group composition and tracking effects on main outcomes of interest in the context of an ASP. I draw two main conclusions: first, better average effects emerge from mixing students by their initial propensity for violence than from segregating them on some behavioral and stress-related outcomes. Second, tracking can have detrimental effects compared to integration for students with a greater propensity for violence. Finally, being the marginal participant—that is, the least violent child within a highly violent group—negatively affects behaviors but does not necessarily harm academic performance.

4.1 GC average effect

Table 3 shows estimations of GC in terms of violence using specification (2). In panel A, I present the effects of group composition on attitudes toward school as well as on learning and violence-related outcomes. In panel B, I present the main neurophysiological outcomes estimated through electroencephalogram recordings. Finally, in panel C, I present the effects on children's academic performance. Columns 1 and 2 present the estimated impacts of being assigned to participate in the ASP with either a heterogeneous (HT) or homogeneous (HM) composition of peers, compared to the C group. Column 3 shows estimated differences on the impacts between the HT and HM treatment arms.

First, from the comparison between each treatment arm with the C group, estimations indicate that any group composition can be more effective than no treatment. For example, in panel A, I find that any group composition improves behavioral outcomes in relation to not participating in the ASP. Moreover, the estimated impacts are important in terms of magnitude. For example, the reduction in absenteeism was almost 25%. Moreover, teacher reports of student misbehaviors indicate a reduction of almost 0.14–0.17 standard deviations (intensive margin) and 6–11 percentage points in the probability of having a bad behavior report (extensive margin).[29]

In terms of emotional regulation (panel B), estimations indicate that most of the statistical differences come from the comparison between the HM and C groups. I estimate a reduction in the overreaction to stimuli (valence), particularly to positive ones (positive valence difference), and a move toward internal locus of control, which is a belief that their own actions determine the rewards they receive. The HT and C groups are also different in the overreaction to positive stimuli. The rest of the estimated coefficients are not statistically different from zero.

Finally, academic performance can be improved mostly by the homogeneous group composition, compared to no treatment. As panel C shows, the HT composition of peers can increase the

[29] All estimated coefficients are statistically significant at the conventional levels after adjusting by multiple hypothesis testing (MHT) following Jones et al. (2019), except for time spent doing homework, which is not different from zero for both treatment groups, and violent actions (Index) for the het group.

probability of a passing grade by almost 4 percentage points (column 1), while being assigned to similar peers can also reduce the probability of failing at least one grade by almost 3 percentage points, both compared to the C group (column 2).[30]

Does one particular group composition improve the ASP's effectiveness more than another? The estimated results indicate that the HT composition of peers has a better impact on some outcomes than tracking by violence. As presented in panel A, students assigned to the HM groups show a reduction of 0.16 standard deviations, on average, in positive attitudes toward school, compared to students assigned to heterogeneous groups. They also increase their probability of having a bad behavior report at school by 5.5 percentage points.

As shown in panel B, the only statistical difference between the HM and HT group compositions is in participants' stress. I find that when students are treated in the HM groups, their stress levels are 0.28 standard deviations greater than those of students treated in the HT group. Finally, I do not find statistical differences between both treatments in the rest of the academic outcomes.

The positive impacts of a more diverse composition of peers 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 students do best when they have a diverse group of classmates. Additionally, the results on misbehavior at school are suggestive evidence that treating students in violence-diverse groups reduces the probability of creating networks of violent children (Billings et al., 2019).

Appendix Table A8 presents estimated coefficients from specification (2), but estimating robust standard errors. Most of the results are similar in terms of statistical significance. Since bootstrapping to estimate standard errors is more conservative given the bias in the estimation of one of the control variables (IVV), I prefer to use this approach in the rest of the estimations.

[30] After adjusting p-values for MHT, only the probability of failing at least one course for the hom group relative to the C group is statistically significant at the conventional levels.

4.2 GC Heterogeneity by Initial IVV

A still-open question is which children within each treatment arm benefit from the composition of their peers. An argument in favor of segregation posits that when children are mixed, the least violent can be "contaminated" by the most violent ones. However, segregation limits the potential for highly violent children to learn good behaviors from nonviolent peers. Since the design includes two different subgroups regarding violence in the HM group, I can further explore differences in group composition by comparing each HM subgroup with the HT group, using specification (3). These results are in Table 4. Column 1 presents estimated differences between low-violence children assigned to the HM-L group and similar students assigned to the HT group. Column 2 shows the same differences but between highly violent students who were randomly assigned to either the HM or HT groups.

First, perhaps surprisingly, I find that the HM-Low set is driving the negative effect of group composition on attitudes toward learning. Compared with the HT group, students in the HM-Low set experience a reduction in their positive attitudes by 0.22 standard deviations (panel A, column 1). This unexpected result is related to Hoxby and Weingarth (2005)'s *invidious comparison* peer-effects model, which (in this context) implies that the exposure to only less violent—or well-behaved—students depresses the average performance of the group. An alternative explanation is that students in heterogeneous groups have the opportunity to be exposed both to good behaviors they should follow and negative ones they should not emulate. 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 by more than 0.09 percentage points for highly violent students when they are segregated, as shown in panel A, column 2.

Moreover, as is evident in panel B, the increase in stress is greater for children treated in the HM groups when compared to respectively similar children treated in the HT group, independent of their propensity for violence at baseline. Finally, the effects of group composition on academic

performance are not clear. While tracking by violence seems to improve academic scores by 0.13 standard deviations (panel C, column 1)for the less violent children in the HM groups compared to the HT group, it also worsens the performance (by almost 0.9 standard deviations) of the highly violent children assigned to tracked groups, compared to those assigned to integrated ones.

In sum, implementing this ASP using a tracking-by-violence design instead of an integration approach can negatively impact behaviors and emotional regulation outcomes of all high- and low-violence participants; it can also worsen the academic performance of the most vulnerable ones. 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 stem mainly from the content of the club curricula. According to the ASP structure, it may be that more time was employed for the club curricula in less violent HM groups, and therefore the reinforcement of "academic" content was greater here.[31]

These results may have different interpretations. First, diversity regarding violence can be more beneficial because it allows highly violent students to be exposed to less violent children and to learn social skills and good behaviors from them. Similarly, low-violence children benefit from being exposed to "misbehaviors" they should avoid. However, students in a homogeneous group are losing the opportunity to learn from behaviors of the other tail of the violence distribution function.

An alternative explanation is that diversity is the social norm in the environments where these children usually perform. Thus, assigning them to similar peers may stress them more. In this sense, enrolling and treating only highly violent students together in these programs can produce an unintended effect. This result suggests that teaching solely socioemotional skills may not be enough to reduce misbehavior or violence by highly violent students, who would also benefit from

[31] A potential concerned is that one of the main drivers of the IVV measure is gender, considering all literature showing that peer gender composition affects student outcomes, and that these effects can be heterogeneous by gender. To address that concern, I included the share of peers that are female for the results in Table 4. The estimated coefficients do not differ from those presented in Table 4.

interacting with—and emulating—low-violence students.

Finally, since participants were randomly allocated to a group in the ASP, some variation in the group composition stems from the fact that assignment to the HM vs the HT group directly affects the mean and variance of one's peers. Following Lafortune et al. (2016) and Duflo et al. (2011), the identification assumption is that after controlling for strata fixed effects, the variance and mean IVV of peers arises entirely from the random assignment. Details of the estimation and summary of results are in Appendix 2. These results reinforce the previous findings using direct variation of the experiment. First, higher average clubmates' IVV increases absenteeism and reduces time to do homework. Moreover, exposure to a more violence-diverse group of clubmates reduces absenteeism and the probability of having a misbehavior report, and it improves positive attitudes toward school and time employed to do homework.

4.3 Effects of tracking on the marginal participant

An additional piece of evidence from this experiment is the effect of tracking on students in the middle of the distribution. To directly measure these effects, I can compare the two homogeneous subgroups using specification (4). This equation allows me to identify differences of being assigned to a group of homogeneous peers with higher propensity to violence.

The estimations of tracking effects on a marginal participant's behavioral and academic outcomes are summarized in Table 5. Column 1 presents estimated coefficients using all students assigned to the HM treatment. Following Duflo et al. (2011), I run specification (3) but restrict the sample to the eight students around the IVV median within each stratum. Results are reported in column 3. As I discuss below, this sample restriction 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.

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 participant to a group of peers with higher propensity for violence reduces her self-report of attitudes toward school and learning by 0.64 standard deviations, and her time to do homework by almost 1.6 hours. I do not find an effect on other behavioral outcomes due to the increase in standard errors.

On neurophysiological outcomes, estimations presented in panel B indicate that assigning a marginal participant to a group of peers with higher propensity for violence reduces her tendency for further reflection on responses—measured through CRT—by 0.89 standard deviations. This finding compares to peers who had similar propensity for violence but who were enrolled with lower-violence peers. I do not find an effect in the rest of the outcomes. Effects of tracking on academic outcomes for marginal participants are also negative. As panel C illustrates, assignment to a high-violence group increases the probability of failing at least one course by 0.048 points. As before, there is an increase in standard errors, and some coefficients are not statistically significant.

In summary, being the least violent member of a highly violent group negatively affects behavioral and emotional regulation-related outcomes and academic performance. 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 group members and follow the group social norms of violence and negative attitudes, which indirectly impacts their academic performance.

5. Discussion

In this section, I present a summary of the group composition main impacts plus further discussion on the results of this paper to evaluate why the integration of students can have greater impact on most behavioral and neurophysiological outcomes. I also discuss why tracking can positively impact the academic performance of low-violence children in this particular context.

Better together: GC effects

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

As I briefly noted before, my results relate mostly to those of Rao (2019), who provides the first evidence of how changes in peer composition at school can shape a student's social preferences through an improvement in her generosity, prosocial behavior, and equity. My paper contributes to these results by providing additional experimental evidence that is especially relevant for the developing world. I test how exposure to diversity regarding violence positively impacts additional noncognitive outcomes such as violence, approval of peers' antisocial behavior, misbehavior, and attitudes toward school and learning. An additional outstanding characteristic in Rao (2019) is that the author uses well-constructed measures of social preferences. In my study, I collected measures of noncognitive outcomes from students' self-reports and administrative data provided by schools. These two sources of information enable me to contrast and validate the results.

Moreover, these results are consistent with a body of microlevel 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 (2019); Griffith and Rask (2014); Lafortune et al. (2016); Chetty et al. (2016); Oreopoulos et al. (2017). Overall, these papers find positive impacts of exposure 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).

Additional evidence from my experimental design regards the tracking effects for marginal participants. For example, an individual at the median in the violence distribution who is assigned to a high-violence group can be "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). By 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 were 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 tracking can have unintended effects on academic and noncognitive outcomes when it targets only the most violent students. This result reinforces the main conclusion of the benefits of diversity regarding violence, since it allows highly violent students to be exposed to less violent children and to learn social skills and good behaviors from them.

Why does integration generate (in general) better results?

In this subsection, I provide suggestive evidence to understand how these group composition impacts on average and marginal participants may operate. I start by exploring peer effects in *social skills learning*. Students in heterogeneous groups benefit from being exposed to both "good behaviors" they should follow and "misbehaviors" they should 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 good 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 more comfortable due to familiarity. 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 at-

tendance at the ASP between each HM group compared to the HT group and present the results in Appendix Table A9. Due to an increase in the standard errors, I find a small but not significant reduction in club attendance by both HM groups. Despite this lack of statistical significance, this result can shed light on preferences for diversity.

Further evidence to support the preference-for-diversity mechanism is presented in a previous work that studies the direct and spillover effects of this ASP on behavioral and neurophysiological outcomes (Dinarte and Egana, 2019). For the same sample, we use data from nonenrolled students that were exposed to a particular share of treated classmates. We find different effects regarding proximity to misbehavior between these two groups of students. Specifically, the results are higher for students whose bad behavior at school is between 1 and 2 standard deviations from the average misconduct of treated students from their classroom. Notably, the effects of this intermediate proximity are more significant on bad behavior reports.

The last mechanism that may underlie 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., 2019; Bayer et al., 2009). The implementation of interventions for groups comprising only highly or less violent students 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 from ending school tracking on the performance of students from lower-ability backgrounds in Finland. Another possible interpretation of the results is that when students are tracked, it feels more like a punishment to the disadvantaged kids: being put together with other disruptive youth could make the program feel remedial or for misbehaved students.

Such a limited effect of tracking is an unexpected finding, especially because the scope is large for targeting the learning experience of school children. In this particular context, qualitative evidence collected from focus groups with instructors indicate that there was no difference in adjusting the content of the club to the specific composition of the groups. Only 42% of instructors indicated that they employed different methods or activities in the clubs. Most of those activities

were not related to particular needs of the participants, but with lack of materials, adequacy of spaces to implement the activities, reorganization of the games, and others. This may indicate that it is relevant to separate the responses of instructors and the direct interaction of effects between peers when evaluating tracking in other contexts. In addition, the differences in the results from this paper with those from Duflo et al. (2011) also suggest the lack of incentives to club instructors, compared to school teachers, to target material to the particular needs of the students.

Explaining the puzzling outcomes for less violent children

It is puzzling that the effects on academic outcomes for low-violence students are greater under tracking even when mixing improves their attitudes toward school and learning. One explanation is that the time dedicated to 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.

6. Conclusions

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

By exploiting the direct variation from the experimental design, I find that in terms of behavioral and stress-related outcomes, tracking generates adverse attitudes toward school and learning, particularly for low-violence students, and increases the probability of bad behavior reports, especially for ex ante highly violent students. It also increases the stress levels of all participants, independent of their initial propensity for violence. These results are confirmed using the ex-

ogenous variation in the peer composition. Additionally, being the marginal participant—that is, the least violent within a highly violent group—negatively affects behaviors and harms academic performance.

These results have implications for public policy discussions regarding interventions oriented to improve academic outcomes and reduce violence within schools. First, this paper takes a first step in understanding the relevance of group composition in an ASP. It shows that within this context, peer effects are an important mechanism that can improve relevant outcomes, directing attention to the implementation of these interventions in heterogeneous groups. However, this is likely that these results might be relevant in other social and educational programs that aim to modify participants' behaviors and attitudes by allowing them to interact and learn from each other.

Second, in the existing literature of interventions aimed at reducing crime and violence, one important aspect of these programs concerns the development of new, healthier social ties that foster a sense of belonging for participants and positively influence their identity (Heller et al., 2017). However, there is still a lack of evidence for policy discussions about how these interventions can improve if students participate in the program within their closer network, exploiting their preferences for similar peers.

Will these results persist over time? Glasswing International's donors, in exchange for financially supporting this impact evaluation, required the research team to allow students in the control group to participate in the intervention the following year. This will make it difficult to measure the ASP's long-term effect. Nonetheless, this experimental design can potentially serve in other contexts where the implementing practitioner or policy maker would like to evaluate the usefulness of targeting while maintaining coverage of all initial beneficiaries.

References

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

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

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

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

Bandiera, O., Barankay, I., and Rasul, I. (2010). Social Incentives in the Workplace. *The Review of Economic Studies*, 77(2):417–458.

Barseghyan, L., Clark, D., and Coate, S. (2019). Peer preferences, school competition, and the effects of public school choice. *American Economic Journal: Economic Policy*, 11(4):124–58.

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

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

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

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

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.

Burke, M. A. and Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1):51–82.

Bursztyn, L., Egorov, G., and Jensen, R. (2019). Cool to be smart or smart to be cool? understanding peer pressure in education. *The Review of Economic Studies*, 86(4):1487–1526.

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

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

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

Carrell, S. E., Hoekstra, M., and West, J. E. (2019). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy*, 11(4):159–82.

Carrell, S. E., Malmstrom, F. V., and West, J. E. (2008). Peer effects in academic cheating. *Journal of human resources*, 43(1):173–207.

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

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

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

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. *American Economic Review*, 106(4):855–902.

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

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

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

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

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

Dinarte, L. and Egana, P. (2019). Reducing violence in the most violent contexts: Behavioral and neurophysiological evidence. *Policy Research Working Paper*, (WPS8862).

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

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

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

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

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

Fonoberova, M., Fonoberov, V. A., Mezic, I., Mezic, J., and Brantingham, P. J. (2012). Nonlinear dynamics of crime and violence in urban settings. *Journal of Artificial Societies and Social Simulation*, 15(1):2.

FUSADES (2015). El Salvador Youth Survey Database.

Garlick, R. (2018). Academic peer effects with different group assignment policies: Residential tracking versus random assignment. *American Economic Journal: Applied Economics*, 10(3):345–69.

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

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

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

Goethals, G. R. (2001). Peer Effects, Gender, and Intellectual Performance among Students at a

Highly Selective College: A Social Comparison of Abilities Analysis. WPEHE discussion paper series, Williams College, Williams Project on the Economics of Higher Education.

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

Gong, J., Lu, Y., and Song, H. (2021). Gender peer effects on studentsâ academic and noncognitive outcomes: Evidence and mechanisms. *Journal of Human Resources*, 56(3):686–710.

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.

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

Hoekstra, M., Mouganie, P., and Wang, Y. (2018). Peer quality and the academic benefits to attending better schools. *Journal of Labor Economics*, 36(4):841–884.

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

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

Jones, D., Molitor, D., and Reif, J. (2019). What do workplace wellness programs do? evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics*, 134(4):1747–1791.

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.

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

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

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

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.

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

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.

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

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

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

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.

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

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

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.

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.

UNICEF (2017). A Familiar Face: Violence in the lives of children and adolescents.

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

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

	(1)	(2)	(3)	(4)	(5)
	Control	Treatments		Tracking groups	
	Group (C)	Heterogen. group (HT)	Homogen. group (HM)	Homog. High (HM-H)	Homog. Low (HM-L)
PANEL A: IVV DETERMINANTS					
Student is male	0.51	0.48	0.49	0.76	0.22
Student's age	11.9	12.0	11.93	12.4	11.4
Student lives in urban area	0.72	0.73	0.74	0.78	0.70
Student's household composition					
Student living with both parents	0.47	0.53	0.56	0.53	0.59
Student living with only one parent	0.37	0.34	0.30	0.33	0.26
Student living with one parent and stepparent	0.07	0.06	0.07	0.06	0.07
Student living with other relative/adults	0.09	0.07	0.08	0.08	0.07
Student's mother's level of education:					
Basic education (1–6 years)	0.34	0.27	0.31	0.22	0.40
Intermediate education (7–12 years)	0.59	0.65	0.62	0.72	0.52
University or higher (13 and +)	0.07	0.08	0.07	0.06	0.08
Student's travel time from house to school (min.)	17.0	17.8	17.9	19.6	16.1
Student is alone at home after school	0.05	0.07	0.04	0.08	0.01
Student's school year	5.67	5.81	5.76	6.03	5.49
Student's violence index	0.04	0.04	0.04	0.05	0.02
PANEL B: ACADEMIC OUTCOMES					
Academic score Q1 2016 (Baseline)	6.51	6.66	6.67	6.78	6.54
Behavior score	7.28	7.42	7.43	7.36	7.50
Absenteeism Q1 2016	2.22	1.59	1.48	1.71	1.26

Table 1 shows descriptive statistics of the available variables at baseline. Panel A summarizes information obtained from the enrollment form that was used as determinants in the IVV estimation. Panel B presents administrative data provided by schools from students who had consented. These academic data are from the first quarter of school year 2016, before the clubs were implemented. The scale of grades in El Salvador is 0–10 points. In columns (1), (2), and (3), I present control (C), heterogeneous (HT), and homogeneous (HM) groups' average characteristics, respectively. The last two columns show the average for the two homogeneous subtreatments (HM-H and HM-L). Unadjusted *p*-values are presented in **Table A6** in the Appendix. There are statistical differences in two categories of household composition and in average absenteeism between HM and C, in the average course and absenteeism between HT and C, and in the IVV score between HT and HM. Most of the differences between HM-H and HM-L groups are statistically relevant, which is desired in this particular experimental design.

TABLE II: DESCRIPTIVE STATISTICS OF THE IVV BY TREATMENT GROUP

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

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. For that prediction, I estimated the specification (1) denoted by $V_f = \alpha_0 + \alpha_1 D_f + \epsilon_f$. In FUSADES (2015) survey, 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) brought 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, mother's education, etc. According to this experimental design, I expect that the average IVV of the HT group should be between the HM-H and HM-L groups' IVV. Moreover, the IVV variation in the HT group should be greater than the respective figures of both groups.

TABLE III. EFFECTS OF ASP GROUP COMPOSITION

Full sample. Results from specification (2)

	Het. group (1)	Hom. Group (2)	Diff Het-Hom (3)	Mean C group (4)	Observations (5)
			PANEL A: BEHAVIORS		
Attitudes toward school and learning					
Positive attitudes toward school	0.305***	0.137*	0.168**	-0.12	948
	(0.067)	(0.083)	(0.085)		
Time to do homework (hours)	0.388***	0.290**	0.098	2.12	935
	(0.143)	(0.121)	(0.145)		
Absenteeism (days)	-1.790***	-1.434***	-0.357	7.16	836
	(0.355)	(0.301)	(0.390)		
Violence and Misbehaviors					
Violent actions (Index)	-0.097*	-0.148***	0.050	0.00	956
	(0.059)	(0.047)	(0.054)		
Approval of peers' antisocial behavior	-0.101***	-0.102***	0.001	0.17	956
	(0.028)	(0.025)	(0.016)		
Behavior reports (-)	-0.139**	-0.172***	0.033	0.00	1010
	(0.057)	(0.052)	(0.054)		
Probability of Behavior reports (-)	-0.111***	-0.060*	-0.055***	0.73	1010
	(0.029)	(0.023)	(0.017)		
			PANEL B: EMOTIONAL REGULATION		
Arousal (stress)	-0.120	0.164	-0.284***	0.05	308
	(0.136)	(0.116)	(0.098)		
Valence	-0.334	-0.396**	0.062	0.00	308
	(0.234)	(0.198)	(0.235)		
Locus of control	-0.148	-0.363**	0.215	0.00	295
	(0.153)	(0.144)	(0.133)		
			PANEL C: ACADEMIC OUTCOMES		
Grades					
Academic score	0.016	0.053	-0.037	0.00	1023
	(0.051)	(0.040)	(0.037)		
Prob. passing grade	0.036**	0.016	0.020	0.87	1023
	(0.016)	(0.016)	(0.013)		
Failing at least one course	-0.017	-0.028***	0.011	0.07	1023
	(0.011)	(0.010)	(0.009)		

This table presents the effects of tracking and integration on the outcomes of interest, which were estimated using specification (2). Panel A exhibits effects on behaviors and attitudes toward school and learning. Panel B presents results on emotional regulation outcomes and psychometric tests. Panel C shows estimated effects of each group composition on academic performance. Description of outcome variables is available in Appendix 1. Outcomes come from data collected during the follow-up, at the end of school year. 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. All regressions are estimated using models of specification (2). Sample size in each specification varies according to the amount of data available for each output. ***, **, * indicates that the effect of being treated in a specific group composition (HT or HM) compared to the C group is significant at 1%, 5%, and 10%, respectively. Bootstrapped standard errors are in parentheses. Table A8 in the Appendix section presents the same models but with robust standard errors. As I discuss, almost all results remain statistically significant at the conventional levels.

TABLE IV. HETEROGENEOUS EFFECTS OF GROUP COMPOSITION

Only Treated Subsample. Results from specification (3)

	Low-HM	High-HM	N
	(1)	(2)	(3)
	PANEL A: BEHAVIORS		
Attitudes toward school and learning			
Positive attitudes toward school	-0.224**	-0.105	716
	(0.106)	(0.110)	
Time to do homework (hours)	0.099	-0.258	707
	(0.257)	(0.183)	
Absenteeism (days)	0.235	0.456	631
	(0.586)	(0.747)	
Violence and Misbehaviors			
Violent actions (Index)	-0.061	-0.044	722
	(0.079)	(0.082)	
Approval of peers' antisocial behavior	0.003	-0.006	720
	(0.026)	(0.019)	
Behavior reports (-)	-0.023	-0.046	762
	(0.068)	(0.086)	
Probability of Behavior reports (-)	0.023	0.091***	762
	(0.023)	(0.031)	
	PANEL B: EMOTIONAL REGULATION		
Arousal(stress)	0.228*	0.256**	238
	(0.128)	(0.108)	
Valence	-0.171	0.016	238
	(0.386)	(0.344)	
Locus of control	-0.206	-0.265	227
	(0.182)	(0.184)	
	PANEL C: ACADEMIC OUTCOMES		
Grades			
Academic score	0.131*	-0.087**	771
	(0.067)	(0.040)	
Probability of passing	-0.016	-0.025	771
	(0.019)	(0.018)	
Failing at least one course	-0.014	-0.005	771
	(0.011)	(0.013)	

***, **, * 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 are in parentheses. Panel A exhibits effects on behaviors and attitudes toward school and learning. Panel B presents results on emotional regulation outcomes and psychometric tests. Panel C shows estimated effects on academic performance. Outcomes come from data collected during the follow-up, at the end of school year. All regressions are estimated using the models of specification (3). 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. Sample size in each specification varies according to the amount of data available for each output.

TABLE V. ASP GROUP COMPOSITION EFFECTS ON MARGINAL PARTICIPANTS

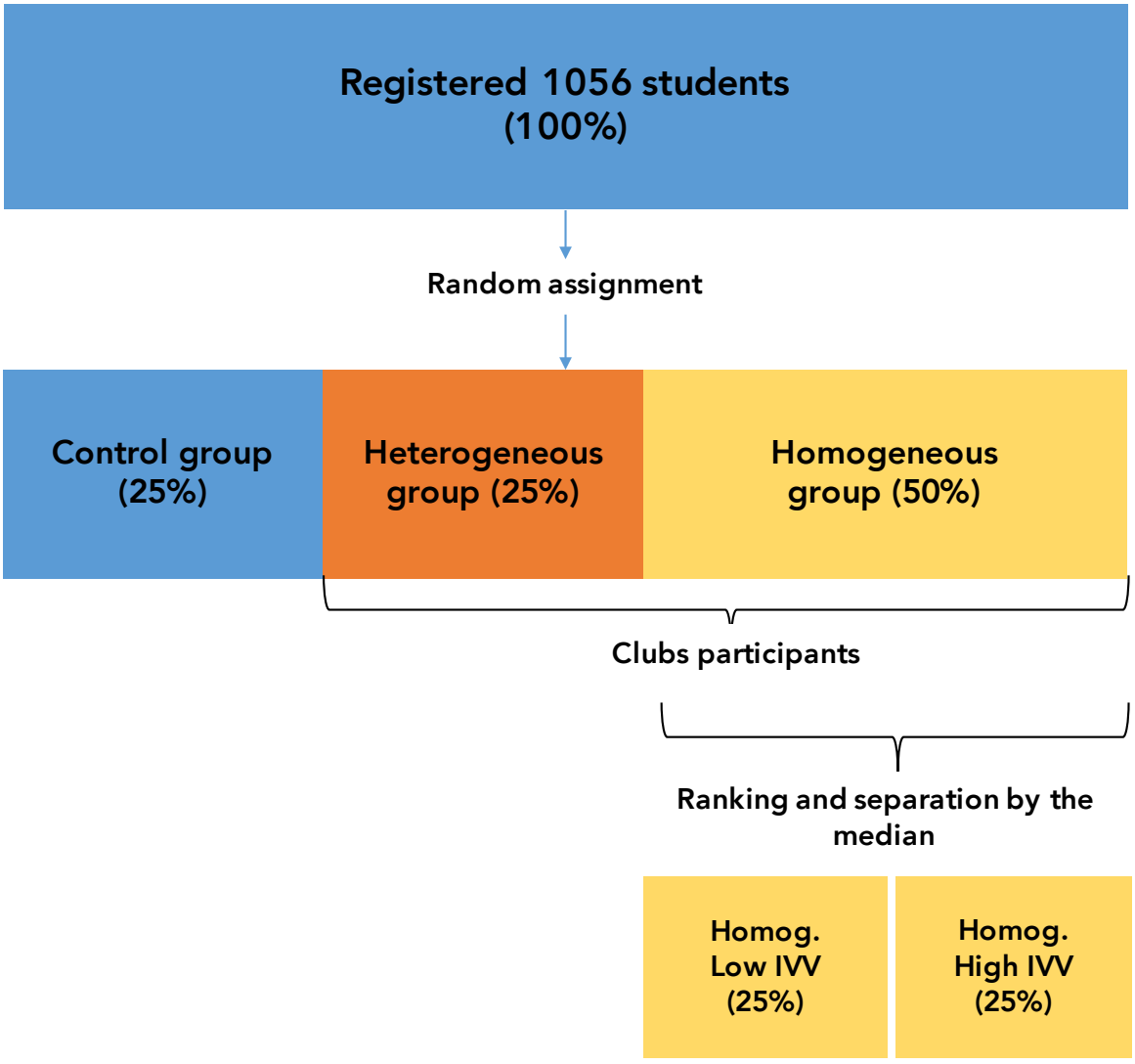
Tracking groups only. Results from specification (4)

	All HM Groups		Only around cutoff	
	High-HM	N (students)	High-HM	N (students)
	(1)	(2)	(3)	(4)
		PANEL A: BEHAVIORS		
Attitudes toward school and learning				
Positive attitudes toward school	-0.079	472	-0.645*	106
	(0.232)		(0.364)	
Time to do homework (hours)	-0.420	468	-1.596**	106
	(0.410)		(0.620)	
Absenteeism (days)	-0.027	423	0.294	92
	(1.078)		(1.764)	
Violence and Misbehaviors				
Violent actions (Index)	0.180	476	0.250	102
	(0.158)		(0.212)	
Approval of peers' antisocial behavior	0.021	474	-0.018	108
	(0.042)		(0.053)	
Behavior reports (-)	-0.233	511	-0.369	114
	(0.154)		(0.254)	
Probability of Behavior reports (-)	-0.090	511	-0.132	114
	(0.068)		(0.118)	
		PANEL B: EMOTIONAL REGULATION		
Arousal(stress)	0.083	151	0.623	30
	(0.203)		(0.433)	
Valence	0.368	151	1.001	30
	(0.672)		(1.950)	
Locus of control	-0.240	145	0.235	30
	(0.304)		(0.796)	
		PANEL B: ACADEMIC OUTCOMES		
Grades				
Academic score	-0.054	516	0.026	115
	(0.105)		(0.209)	
Probability of passing	-0.051	516	0.007	115
	(0.040)		(0.052)	
Failing at least one course	0.048**	516	0.031*	115
	(0.024)		(0.018)	

***, **, * indicates that the effect of being treated in a HM-H group compared to being treated in a HM-L is significant at 1%, 5%, and 10%, respectively. Bootstrapped standard errors are in parentheses. Panel A exhibits effects on behaviors and attitudes toward school and learning. Panel B shows effects on emotional regulation outcomes and psychometric tests. Panel C shows the effects on academic outcomes. All regressions include—in addition to the main controls mentioned above—a second-order polynomial IVV. Estimations first use the homogeneous groups subsample—results in columns (1) and (2)—and then the 8 students around the cutoff—results in columns (3) and (4).

Figures

Figure I. Experimental Design.



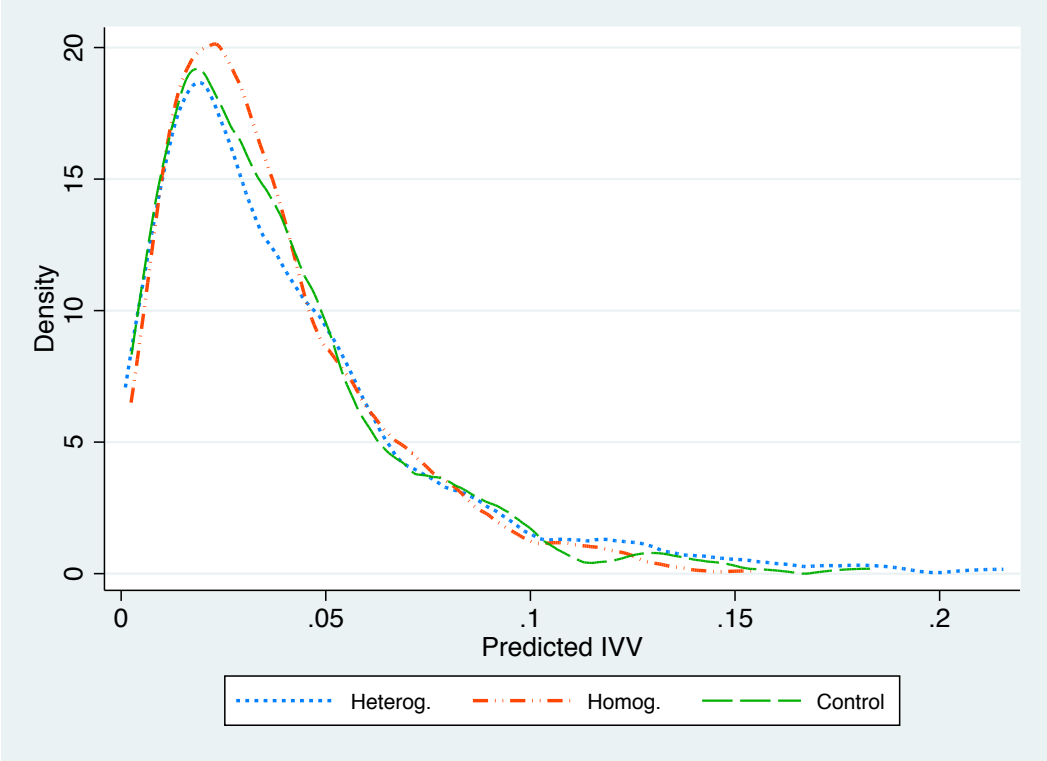
This figure shows the sample composition and randomization procedure applied in this design. From the total of enrolled children in each educational level $\{1,2,3\} \in$ school A, I randomly assigned 25% to C and 25% to HT treatment arm and 50% to HM groups. The same procedure was implemented in the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

Figure II. Timeline of the Intervention.



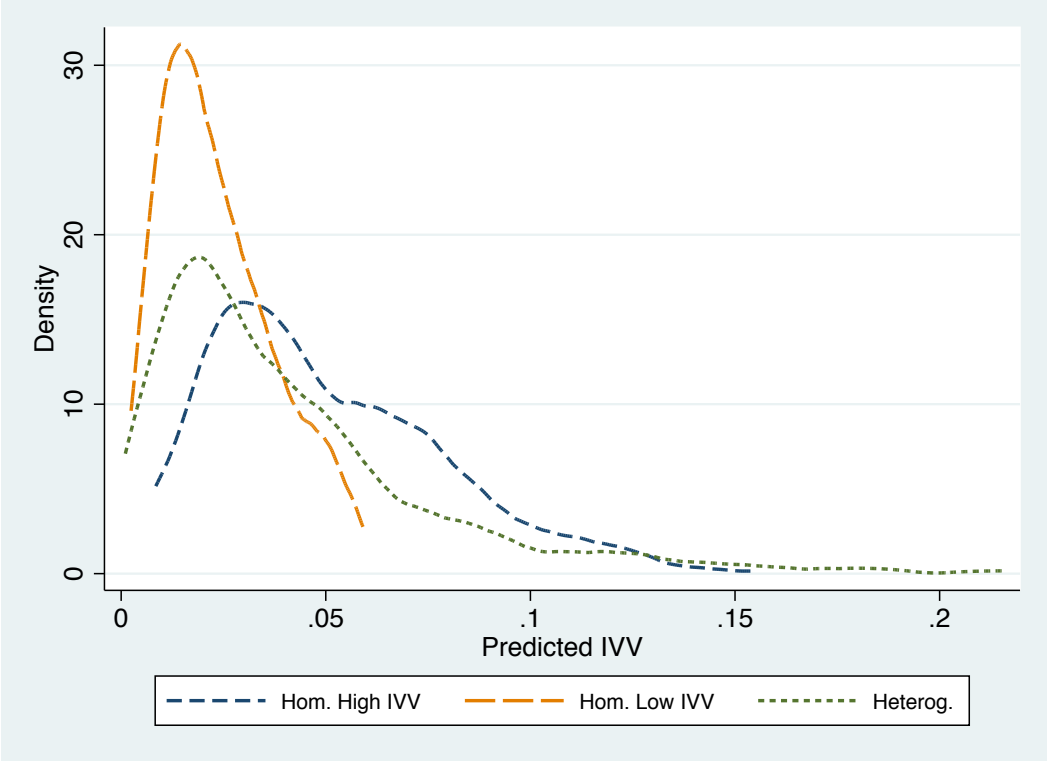
Timeline of the intervention and data collection.

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



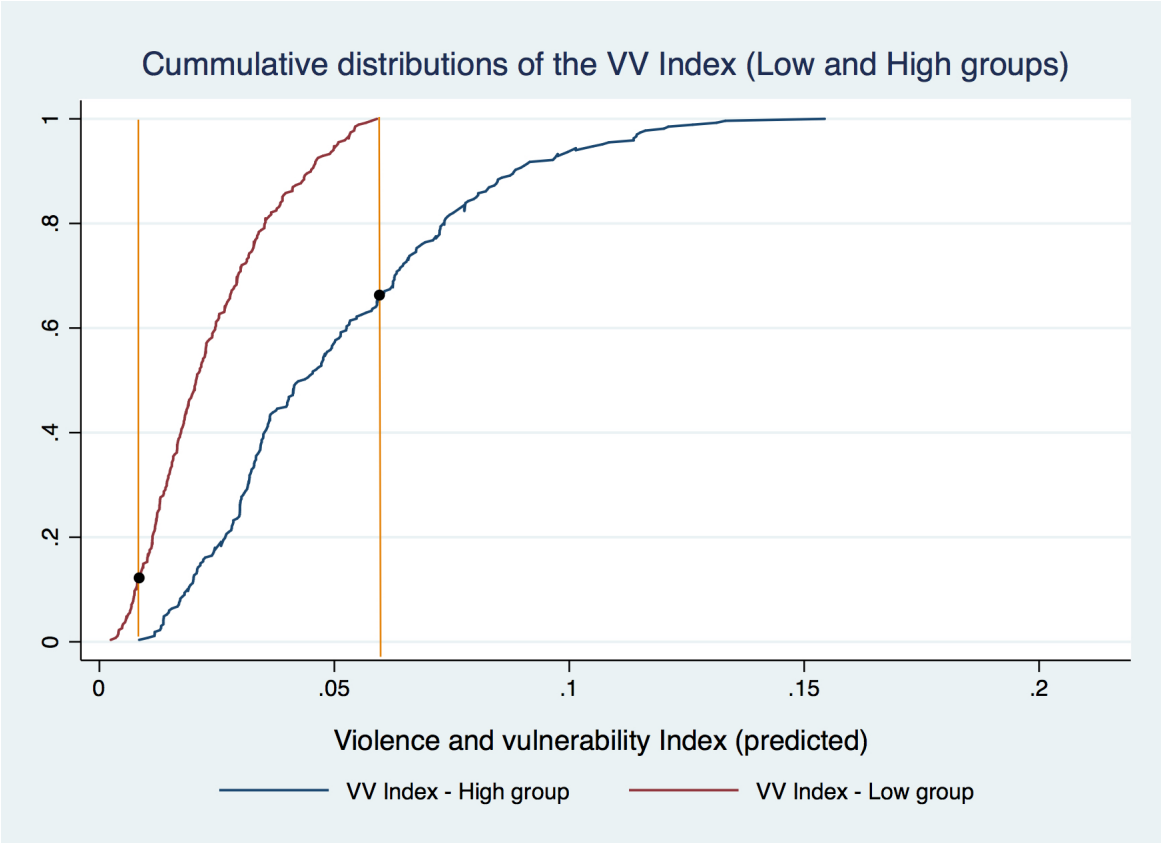
Predicted IVV distribution functions for the control and any treatment (homogeneous and heterogeneous) groups prior to treatment, for the whole study sample.

Figure IV. IVV Distribution Functions of Treated Groups.



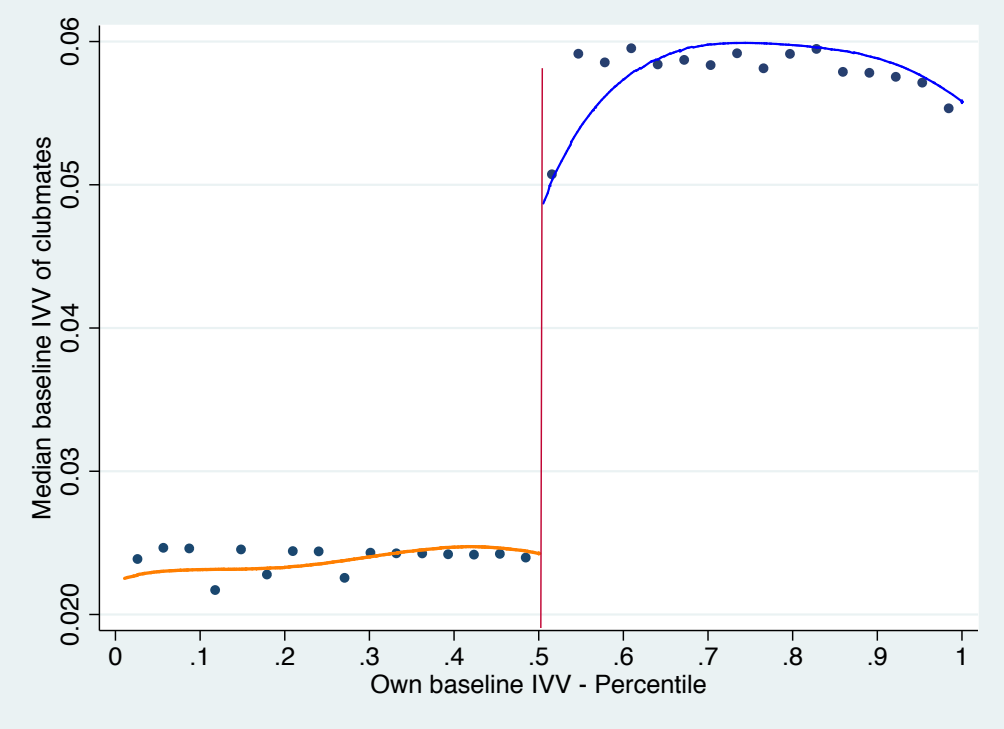
Predicted IVV distribution functions generated by the experimental design for the heterogeneous treatment group and each of the homogeneous subgroups (High and Low IVV) in the whole study sample.

Figure V. IVV Cumulative Distribution Functions of Homogeneous Subgroups.



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

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



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

Appendix - For Online Publication

Appendix 1. Description of Outcome Variables.

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

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

A. Behavior and Academic Outcomes

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

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

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

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

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

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

B. Neurophysiological Outcomes

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

2. *Valence*: pre-test resting state measure estimated directly from participants' brain activity using EEG recordings. As the arousal measure, this recordings were estimated while children were watching a black cross in the center of a gray screen for a period of 30 seconds. This variable can be interpreted as a positive or negative mood, as well as an attitude of either approach or withdrawal towards/from a stimulus (Harmon-Jones et al., 2010; Kassam et al., 2013).

3. *Locus of control*: Psychometric test developed by (Rotter, 1966). People with an internal locus of control believe that their own actions determine the rewards that they obtain, while those with an external locus of control believe that their own behavior doesn't matter much and that rewards in life are generally outside of their control. Scores range from 0 to 13. A low score indicates an internal control while a high score indicates external control. In other words, a reduction in the score indicates moving towards an internal control.

C. Academic Outcomes

1. *Reading, math and science scores*: Variables that indicate performance on each course. It is a 0-10 scale, where 0 is the worst performance and 10 is the best. We have standardized these

values from control groups at the school-grade level. Academic score is an average of the three courses. Intensive margin measures the increase on average academic score for each course.

2. *Probability of passing course*: Is a dummy variable that takes the value of 1 if student has been promoted to the following course and 0 otherwise. Extensive margin indicates if student pass or not the course.

3. *Failing at least one course*: Is a dummy variable that takes the value of 1 if student failed one or more courses and 0 otherwise.

Appendix 2. Exploiting the random allocation of peers

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

$$y_{ij} = \gamma_0 + \gamma_1 \bar{x}_{-ij} + \gamma_2 \bar{X}_{-ij} + \gamma_3 var(x_{-ij}) + \gamma_3 Var(x_{-ij}) + \gamma_4 X_{ij} + \epsilon_{ij} \tag{5}$$

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. In addition, $\bar{X}_{-ij}$ and $Var(x_{-ij})$ are the IVV mean and variance of all treated students, also excluding child i's IVV. The vector of control variables X_{ij} includes participant's own baseline IVV. The rest of variables are defined as before. The estimated coefficients of interest are γ_1 and γ_3. They reflect the causal effect of student i's peers' mean and variance violence on her own results.

In addition, 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 et al. (2011); Boozer and Cacciola (2001); Lyle (2007). The estimating equation for the sample of students selected to participate in the ASP is:

$$Y_{ij} = \gamma_0 + \gamma_1 \bar{x}_{-ij} + \gamma_2 var(x_{-ij}) + \gamma_3 S_j + \gamma_4 X_{ij} + \epsilon_{ij} \tag{6}$$

All variables are defined as in the previous specification. With this equation 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.

Results from both specifications are presented in table A10 in the appendix tables section. Panel A exhibits effects on behaviors and attitudes towards school and learning. Panel B presents estimated effects of each group composition on emotional regulation, stress and psychometric out-

comes. Finally, panel C shows the estimated effects of both specifications on academic performance. Columns 1-3 present the results using model 1 and columns 4-6 shows similar estimations obtained from model 2. Using these alternative estimation approaches, I obtain similar results as using the direct variation on group composition generated by the experiment. 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 self reported time spent doing homework and reduces absenteeism. In terms of violence, I do not find an effect from either the mean or standard deviation of clubmates' IVV. Despite the lack of power from neurophysiological outcomes and psychometric tests, I find that greater diversity reduces stress levels of participants. Finally, the effects on academic performance indicate that a greater level of peers' violence can increase academic performance, which is in line with the inviduous comparison model of Hoxby (2000). However, greater diversity can also improve the extensive margin of academic grades.

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

	(1)	(2)	(3)	(4)	(5)
	Study Sample		FUSADES(2015) Sample		
	Mean	Std. Dev.	Mean	Std. Dev.	*p*-value
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 with only one of his/her parents	0.32	0.47	0.30	0.46	0.19
Student living with one parent	0.06	0.25	0.08	0.27	0.02
Student living with other relative	0.08	0.27	0.07	0.26	0.25
Student's travel time from house to school (minutes)	17.64	14.37	17.25	12.98	0.37
Student's mother's level of education	0.31	0.46	0.4	0.49	0.40
Student is alone at home after school	0.05	0.22	0.11	0.31	0.00
Student's age	11.95	2.95	13.87	1.67	0.09
Student's course	5.75	2.71	5.5	2.52	0.29
N	1056		6641		

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

TABLE A2. IVV ESTIMATION RESULTS AND DETERMINANTS USING 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 nonrelative 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 is enrolled on morning shift	-0.002
	(0.087)

I estimated the specification (1) denoted by $V_f = \alpha_0 + \alpha_1 D_f + \epsilon_f$. 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) brought 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, mother's education, etc. ***, **, * indicates if estimated coefficients α_1 are statistically different from zero. Standard errors are in parentheses. Mother's education omitted category: mother has basic education (1st–6th grades). Household composition omitted category: children living with both parents.

TABLE A3. CORRELATION BETWEEN IVV, ACADEMIC GRADES, AND MISBEHAVIOR REPORTS AT BASELINE

	(1)	(2)	(3)	(4)	(5)
	Grades				Misbehavior
	Reading	Math	Science	Score	reports
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. Nonstandardized 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 are in parentheses.

TABLE A4. 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 were 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 A5. MATCHING RATE WITH ADMINISTRATIVE DATA AND ATTRITION RATE

	(1)	(2)	(3)	(4)	(5)
		Treatments		Tracking groups	
	Control Group (C)	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	0.97	0.98	0.98	0.98	0.98
Math scores	0.89	0.90	0.92+	0.92	0.93
Science scores	0.94	0.96	0.96	0.96	0.96
Behavior scores	0.91	0.94	0.94	0.94	0.94
Absenteeism	0.68	0.68	0.67	0.65*	0.69
Fraction of students with matched administrative data, Q4 2016					
Academic scores	0.98	0.97	0.96	0.96	0.97
Behavior scores	0.96	0.95	0.95	0.95	0.96
Absenteeism	0.79	0.80	0.80	0.76*	0.83
Number of students at baseline and follow-up					
Number of students present at baseline	258	263	535	267	268
Number of students present at follow-up	237	248	483	239	244
Retention rate (1-attrition)	0.92	0.94	0.91	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 who could be matched with administrative data from schools. In comparing HM and C, + denotes difference significant at the 10% level. A similar notation is used to indicate statistically significant differences between HM-High and HM-Low (*).

TABLE A6. p-values OF DIFFERENCES BETWEEN TREATMENT AND CONTROL GROUPS

	(1)	(2)	(3)	(4)
	p-values			
	Control = Heterog.	Control = Homog.	Heterog. = Homog.	Homog. High = Homog. Low
PANEL A: IVV Determinants				
Student is male	0.543	0.740	0.616	0.000
Student's age	0.163	0.405	0.216	0.000
Student lives in urban area	0.949	0.504	0.563	0.102
Student's household composition				
Student living with both parents	0.228	0.050	0.300	0.137
Student living with only one parent	0.590	0.056	0.248	0.340
Student living with a parent and a stepparent	0.548	0.793	0.672	0.907
Student living with other relative /adult	0.516	0.632	0.780	0.844
Student's mother's level of education:				
Basic education (1–6 years)	0.105	0.481	0.147	0.002
Intermediate education (7–12 years)	0.131	0.566	0.369	0.000
University or higher (13 and +)	0.432	0.997	0.625	0.630
Student's travel time from house to school (min.)	0.539	0.501	0.991	0.036
Student is alone at home after school	0.189	0.835	0.143	0.000
Student's school year	0.089	0.120	0.368	0.012
Student's violence index	0.232	0.716	0.051	0.000
PANEL B: Administrative data				
Academic scores Q1 2016	0.267	0.197	0.844	0.042
Behavior scores	0.139	0.097	0.840	0.055
Absenteeism Q1 2016	0.055	0.006	0.478	0.046

This table shows unadjusted *p*-values of balance tests for all available variables at baseline. Panel A presents *p*-values of tests for variables obtained from the enrollment form and panel B presents similar values for variables from administrative data provided by schools from students who had consented. In columns (1) and (2), I present *p*-values of balance tests between (C) and each treatment arm. Column (3) presents similar values for the comparison between treatment arms, and the last column shows the results of the test between the two homogeneous subtreatments (HM-H and HM-L).

TABLE A7. 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 a 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 A8. EFFECTS OF ASP GROUP COMPOSITION

Full sample. Results from specification (2)

	Het. group (1)	Hom. Group (2)	Diff Het-Hom (3)	Mean C group (4)	Observations (5)
			PANEL A: BEHAVIORS		
Attitudes toward school and learning					
Positive attitudes toward school	0.305***	0.137	0.168+	-0.12	948
	(0.096)	(0.097)			
Time to do homework (hours)	0.388**	0.290*	0.098	2.12	935
	(0.198)	(0.158)			
Absenteeism (days)	-1.790***	-1.434***	-0.357	7.16	836
	(0.609)	(0.496)			
Violence and Misbehaviors					
Violent actions (Index)	-0.097	-0.148**	0.050	0.00	956
	(0.080)	(0.063)			
Approval of peers' antisocial behavior	-0.101***	-0.102***	0.001	0.17	956
	(0.032)	(0.029)			
Behavior reports (-)	-0.139*	-0.172***	0.033	0.00	1010
	(0.077)	(0.063)			
Probability of Behavior reports (-)	-0.111***	-0.060**	-0.055*	0.73	1010
	(0.037)	(0.027)			
			PANEL B: EMOTIONAL REGULATION		
Arousal (stress)	-0.120	0.164	-0.284***	0.05	308
	(0.115)	(0.118)			
Valence	-0.334*	-0.396**	0.062	0.00	308
	(0.183)	(0.185)			
Locus of control	-0.148	-0.363***	0.215*	0.00	295
	(0.166)	(0.124)			
			PANEL C: ACADEMIC OUTCOMES		
Grades					
Academic score	0.016	0.053	-0.037	0.00	1023
	(0.063)	(0.051)			
Prob. passing grade	0.036*	0.016	0.020	0.87	1023
	(0.021)	(0.019)			
Failing at least one course	-0.017	-0.028**	0.011	0.07	1023
	(0.014)	(0.011)			

***, **, *, + indicates that the effect of being treated in a specific group composition (HT or HM) compared to the C group is significant at 1%, 5%, 10%, and 10%, respectively. Robust standard errors are in parentheses. Panel A exhibits effects on behaviors and attitudes toward school and learning. Panel B presents results on emotional regulation outcomes and psychometric tests. Panel C shows estimated effects of each group composition on academic performance. Description of outcome variables is available in Appendix 1. Outcomes come from data collected during the follow-up, at the end of school year. All regressions are estimated using the full sample and models of specification (2). Sample size in each specification varies according to the amount of data available for each output.

TABLE A9. ASP ATTENDANCE OF TREATED STUDENTS

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

***, **, * indicates that the club attendance from the HM (high or low) group compared to being treated in a HT group is significant at 1%, 5%, and 10%, respectively. Bootstrapped standard errors at course-school level are in parentheses. Two measures of attendance are number of sessions and days. Regressions are estimated using only treated group and models of specifications (2).

TABLE A10. FURTHER MEASURES OF GROUP COMPOSITION EFFECTS

	Nontracking Group Only. Results from specification (5)			*All treated participants. Results from specification (6)*		
	mean(IVV) of clubmates	sd(IVV) of clubmates	N	mean(IVV) of clubmates	sd(IVV) of clubmates	N
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: BEHAVIORS						
Attitudes toward school and learning						
Positive attitudes toward school	0.166	0.022	244	-0.016	0.146**	716
	(0.110)	(0.122)		(0.066)	(0.068)	
Time to do homework (hours)	-0.456**	0.591**	239	-0.130	0.178	707
	(0.201)	(0.293)		(0.088)	(0.119)	
Absenteeism (days)	1.480	-2.370**	208	0.088	-0.392*	631
	(0.939)	(0.942)		(0.230)	(0.218)	
Violence and Misbehaviors						
Violent actions (Index)	-0.194*	0.196	246	-0.016	0.031	722
	(0.079)	(0.143)		(0.045)	(0.051)	
Approval of peers' antisocial beh.	-0.025	-0.004	246	0.000	0.001	720
	(0.023)	(0.028)		(0.010)	(0.011)	
Behavior reports (-)	-0.137	-0.034	251	0.024	0.014	762
	(0.088)	(0.098)		(0.033)	(0.039)	
Probability of Behavior reports (-)	-0.046	-0.017	251	-0.012	-0.038**	762
	(0.035)	(0.042)		(0.014)	(0.015)	
PANEL B: EMOTIONAL REGULATION						
Arousal (stress)	0.052	0.091	87	0.082	-0.097*	238
	(0.084)	(0.114)		(0.052)	(0.055)	
Valence	0.043	0.081	87	0.039	0.063	238
	(0.286)	(0.229)		(0.099)	(0.118)	
Locus of control	0.279*	-0.027	82	0.059	0.080	227
	(0.150)	(0.177)		(0.067)	(0.075)	
PANEL C: ACADEMIC OUTCOMES						
Academic score	0.156**	-0.057	255	-0.020	0.018	771
	(0.042)	(0.055)		(0.024)	(0.023)	
Prob. passing grade	0.015	-0.001	255	-0.014	0.024**	771
	(0.017)	(0.021)		(0.010)	(0.012)	
Failing at least one course	-0.012	0.007	255	-0.001	0.003	771
	(0.016)	(0.018)		(0.007)	(0.007)	

***, **, * indicates that the estimated coefficient is statistically significant at 1%, 5%, and 10%, respectively. Bootstrapped standard errors at the course-school level are in parentheses. Panel A exhibits effects on behaviors and attitudes toward school and learning. Panel B presents results on emotional regulation outcomes and psychometric tests. Panel C shows estimated coefficients on academic outcomes that come from schools' academic data. All outcomes come from data collected during the follow-up, at the end of school year. All regressions are estimated using the sample of students according to specifications (5) and (6). Details are presented in Appendix 2. Sample size in each specification varies according to the amount of data available for each output.