Afraid to go to School?
Estimating the Effect of Violence on Schooling Outcomes.*

Martin Foureaux Koppensteiner† Lívia Menezes‡

April 5, 2018

Abstract

Homicides are undoubtedly the crudest outcome of violence and crime. They may instil fear with observers and may lead to behavioural adjustments. We use a number of large administrative Brazilian datasets to estimate the causal effect of exposure to homicides in the public way on schooling outcomes. Within-school estimates show that violence in the surroundings of schools, at the residence of students, and on the walking path from residence to school has a negative effect on a number of measures of school achievement such as test scores, repetition, dropout and school progression. We also find that school attendance suffers following a homicide in the school surroundings. Exceptionally rich data allow us to investigate heterogeneous effects and explore the channels underlying these effects.

JEL Classification: I25, K42, O12

*We would like to thank Esteban Aucejo, Fernanda Brollo, Paolo Buonanno, Antonio Ciccone, David Figlio, Mark Hoekstra, Michael Lovenheim and Philip Oreopoulos and seminar participants in Reading, Leicester, the 2017 IZA Workshop on The Economics of Education, the 2018 Workshop on Human Capital at the Bank of Italy and the 2018 Royal Economic Society Annual Conference for their very useful comments and suggestions. We also thank Google for their support with geocoding the data. Our particular thanks goes to Kerstin Pittl for her work on creating school corridors. We also express our gratitude to the Brazilian Ministry of Health, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) and the State Secretariat for Education in São Paulo for the data used in this project. Foureaux Koppensteiner very gratefully acknowledges financial support received from British Academy grant SG150627. The usual disclaimer applies.

†University of Leicester, University Road, Leicester, LE1 7RH, UK. Email: mk332@le.ac.uk

‡University of Leicester, University Road, Leicester, LE1 7RH, UK. Email: flsm1@le.ac.uk

1 Introduction

After a decade of declining rates of crime and homicides, recently Brazil (and other countries in Latin America) has observed a steep increase in violent crime. Today, Brazil has one of the highest homicide rates in the world, according to statistics from the World Bank. In 2016, the intentional homicide rate in Brazil was 29.9 per 100,000 people, which is approximately 6 times the US rate and 30 times the UK rate. According to national security statistics, in 2016, 61,619 homicides were registered in the country. The Brazilian Institute of Applied Economic Research (Ipea) estimated that the cost of violence corresponds to more than 5 percent of the country's GDP, not including yet many intangible costs which are difficult to quantify (Cerqueira et al. (2007)). The pain, suffering and trauma caused by direct victimisation and exposure to violence in the local neighbourhood may negatively impact a variety of societal outcomes, among those educational production. Violence may affect school supply as well as the behaviour of students, parents, teachers and principals. In this paper, we want to estimate the effect exposure to violence on the performance of students in Brazil making use of unique novel dataset containing georeferenced information on all homicides occurring in the public way and combining this with very detailed information of student performance.

A number of qualitative studies by psychologists, psychiatrists and sociologists has found a range of adverse consequences in the behaviour of children after exposure to community violence: depression, anxiety, hyper vigilance, avoidance as well as aggressive behaviour, delinquency and deterioration of cognitive performance (Cooley-Quille et al. (1995), Smith and Tolan (1998), Fowler et al. (2009), Farrell et al. (2010), Sharkey et al. (2014)). Community violence can also affect students attendance at school. When a crime occurs in their neighbourhood of residence or in the proximity of their schools, parents may feel uneasy of sending their children to school. According to the 2012 edition of the Brazilian National Survey of School Health, almost 9 percent of the 9th grade students that answered the survey declared they had stopped going to school at least once in the 30 days preceding the survey, for not feeling safe on the way from residence to school. Low attendance caused by fear can potentially damage the learning process of the students. They fail to attend classes that form part of their curriculum and they are also deprived from the regular contact with their classmates. This will eventually lead to low scores in their exams and potentially impact a number of measures of school failure, including repetition and dropout. The exposure to homicides in the local neighbourhood may also reveal information to students and parents about likely victimisation and affect the expected returns to education and hence the optimal schooling

decision.

Because of the potential for such negative externalities, the cost of violence may go well beyond the cost of direct victimisation. Poor neighbourhoods with lower socio-economic status often register higher rates of violence and if this has also a negative effect on human capital accumulation, this could be a relevant channel leading to the perpetuation of poverty. The correlation between socio-economic conditions and crime rates nevertheless makes the estimation of the causal effect of exposure to violence on schooling outcomes difficult, as one needs to disentangle (unobserved) neighbourhood characteristics, which may be related to both high levels of violence and to worse schooling outcomes, from the underlying causal relationship.

This paper estimates the causal effect of violence on schooling performance, using a unique set of Brazilian microdata. We make use of access to information on the exact timing and the precise location of each homicide, and information on the location of the schools students attend, and their residence. We exploit variation of homicides across space and over time to estimate the effect of exposure to homicides on a number of educational outcomes, including test scores, repetition, dropout, school progression and attendance, while controlling for school and time fixed effects. Given the prevalence of high crime rates in many countries in Latin America and elsewhere, the findings from this analysis may be relevant for the understanding of the perpetuation of poverty in these countries.

There is a small number of studies estimating the relationship between exposure to violence and school performance (for example Grogger (1997), and Aizer (2008)) which generally, given the cross-sectional nature of the data used, cannot deal with the endogeneity problem arising from the fact that violence might be correlated with other sources of socio-economic disadvantages and school outcomes. A notable exception is Monteiro and Rocha (2017) who estimate the causal effect of gunfights between drug gangs in Rio de Janeiro's favelas (slums) on students' achievements using panel data for the city of Rio de Janeiro. They look at the effect of conflicts in favelas on students who study in schools located in favelas and in schools located within a 250 meter radius from a favela border and find that students' test scores in Math are lower in years in which they are exposed to drug battles. This paper sheds light on how these conflicts affect children's development in the context of conflicts associated with drug battles in poor neighbourhoods or their close proximity. The conflicts in and around favelas are often context specific, and for example related to battles between rival drug gangs, but violence in Brazil is a more widespread phenomenon. The measure of violence we use, homicides, will be able to capture the widespread nature of violence and allows

us to estimate the effect of violence on students' achievements in a much more general setup likely more representative for violence in Brazil.

In this paper, we introduce a unique set of microdata, which provides us with a measure of violence that is consistent across space and time: homicides in the public way. This is important as this allows to use variation over time and across space, including across vast areas and different administrative units for which consistent crime data, that includes information on violence, is rarely available. For these homicides we have available extremely granular address information which we geocode and combine with the georeferenced information on the addresses of schools and the address of residence of students attending these schools. This allows us to not only investigate the effect of violence in the surroundings of schools or the residence of students, but also to investigate in detail exposure to violence on the way to school for a period of 7 years. We focus most of our analysis on the city of São Paulo, which is the largest city in the Americas with a population of 12 million people. Within-school estimates show that violence at the surroundings of school and residence and at the walking path from residence to school has a negative effect on attendance and on a number of measures of educational achievements such as test scores, repetition, dropout and school progression.

We find that homicides in the surroundings of schools lead to a substantial deterioration of educational performance of schoolchildren, as measured by standardised test scores in Math and Portuguese Language. We find that one additional homicide in a 25 meter radius around schools reduces test scores in Math by about 5 percent of a standard deviation in test scores. Furthermore, we find that homicides also increases repetition and dropout rates and negatively impacts school attendance. Using rich information on the student background, we find that the effects are particularly pronounced among students from relatively poorer families, possibly suggesting that income works as a buffer against the negative effect of crime. We furthermore show that the effect cannot be explained by lower attendance rates alone and we provide suggestive evidence that exposure to homicides may deteriorate incentives to invest in human capital for boys, who are most likely to be victimised in homicides. To our best knowledge, this is the first paper to provide credible causal estimates on the effect of exposure to homicides on schooling outcomes that uses a generalisable measure of violence.

The remainder of the paper is organised as follows. Section 2 explains the institutional background. Section 3 details the datasets used in the analysis. Section 4 presents the identification strategy applied to estimate the causal effect of violence on educational outcomes. Sections 5, 6, 7

and 8 explain the results and Section 9 presents the final remarks.

2 Institutional Background

The Brazilian educational system is predominantly regulated by the federal government, which is also responsible for distributing resources to states and municipalities. These secondary layers of government not only manage the funds received, but are also allowed to implement state or municipality specific programs and policies. The education system is composed by two main levels: *Educação Fundamental* (basic education) - which comprises *Educação Infantil* (nursery), *Ensino Fundamental* (primary school), *Ensino Médio* (secondary education) - and *Educação Superior* (higher education).

Public primary education is offered at no cost for all, irrespective of the age, and it is mandatory for children between 6 and 14 years of age. It lasts 9 years[1] and it is divided in two stages: the first cycle which comprises 1st to 5th grade; and the second cycle which includes 6th to 9th grade. Public secondary school is also offered at no cost and lasts 3 years, it is not compulsory, but recent regulation pushes towards gradually making secondary education compulsory as well. To be able to enrol in secondary school, students must conclude primary school.

A school year contains at least 800 hours spread over at least 200 school days. The precise starting and ending day of the school year varies across schools and over the years. Figure A1 in the Annex exemplifies the school calendar in São Paulo for 2010. Every year São Paulo State's Secretariat of Education formally announces, by releasing a document called *Resolução*, the desirable starting day of the school year. In general, the first semester finishes on the last working day of June; second semester starts on the first working day of August and finishes on the last working day before Christmas. Each semester is composed by two bimesters, with roughly 50 days each, the precise ending dates of each bimester is school specific. This setup leads to semesters that are defined state-wide, and bimesters that are school-specific. Students may be retained in a grade at the end of the year in case they do not achieve adequate school performance and/or they do not meet the minimum level of attendance required by law, which is at least 75 percent of the school days in primary schools and 85 percent in secondary schools.

Considering the nature of funding and administration of schools, they can be classified in four

[1] Previously, primary school began at age 7 and lasted eight years. In 2006, the government passed a law that expands primary school from 8 to 9 years and mandatory enrolment at 6 years old. States and municipalities had until 2010 to implement the new law.

types: federal, state, municipal and private schools. The first three are essentially public schools, maintained by the respective administrative units. In general, private schools are of better quality, however only a relatively small share of the population can afford the substantial school fees charged in these schools. At least 87 percent of the students go to public schools in Brazil, in São Paulo this number is slightly smaller, 80 percent. Schools may offer all or only specific levels of basic education, and there are schools which offer only primary, some only secondary and some offer both primary and secondary education.

Public school students are not bound to a specific school; they are able to enrol in any school with vacancies. In most cases, students attend schools located within walking distance of their residences. When this is not possible, they may qualify for school transport.

3 Data

We build a novel dataset by combining administrative data from three institutions: the Brazilian Ministry of Health, the Brazilian Ministry of Education and the São Paulo State's Secretariat of Education and link these datasets using school, class and individual identifiers and geographic information from the addresses.

3.1 Educational data

We have access to unique microdata of all students in primary and secondary school, collected by the Brazilian Ministry of Education that contains information on the addresses of students and their schools. From 2007, this dataset contains information from individual records on students, schools and teachers and their characteristics. In addition, it is possible to follow students over time and across schools through a unique student id, which allows us to construct some of the outcomes we use in the analysis: repetition, dropout, school progression and school transfers. Characteristics of students and teachers include date of birth, sex, race, grade for the students, and educational background for the teachers (among other).

Table 1 presents summary statistics of students and school characteristics for São Paulo over the period from 2007 to 2012. For consistency, we do not consider nursery schools[2] and any kind of special education, which is offered to students with special needs. The final dataset contains on average 1.8 million students a year spread over 3,164 schools. The majority of observations covers

[2]Pre-primary education has gone through a period of very rapid expansion over the last years and comprises a number of different levels across ages, which makes it difficult to come up with a consistent definition of pre-school type.

students in primary school (77 percent) and given universal primary school enrolment and the longer duration of primary school, demographic characteristics are roughly representative for the population at large. Measures of school efficiency, such as repetition and dropout reveal substantial problems in the Brazilian educational system. Close to 8 percent of schoolchildren repeat any given grade and almost 13 percent drop out of a given grade. There is a substantial number of students that change school after the school year. Only 73 percent of students carry on beyond compulsory education and enrol in secondary school. A small fraction (0.1 percent) of students changes school during the school year.

Of the schools in the sample, about a third is run by the state (mainly secondary schools) and about 17 percent are run by the municipality. The large fraction of private schools reveals that, given that only about 20 percent of students are enrolled in these, private schools are on average much smaller compared to state and municipal schools. Close to 60 percent of schools offer free school meals, an indication for students from poor households.

We use standardised test scores provided by São Paulo State's Secretariat of Education. The exam is applied every year and evaluates the performance of students in 4th, 5th, 7th and 9th grades of primary and at the 3rd grade of secondary school in Portuguese and Math. In order to be able to compare the results to national standardised exams, we focus on test scores for 5th and 9th grades of primary school and 3rd grade of secondary school. These coincide with the end of each of the educational cycles described above.

Attendance data is also provided by São Paulo State's Secretariat of Education. The dataset contains attendance record of all students at state schools in São Paulo at a bimonthly frequency. The data contains information on the number of schools days missed with some basic information about the reason of non-attendance. The attendance data can be merged to individual school records from the school census by the unique student identifier.

3.2 Violence data

We use microdata of official death records published by the Brazilian Ministry of Health. This dataset comes from the Mortality Information System, which compiles information from death certificates on all natural and non-natural deaths in Brazil. We use information from the ICD-10 coding of cause of non-natural deaths to identify victims of intentional homicides. In addition to cause of death, the death certificates contain characteristics of the deceased, such as date of birth, sex, race, occupation and the location of the occurrence of the homicide.

We have information on the precise location available only for homicides that occur in the public way. We believe these homicides are particularly salient for our analysis for two reasons: first, these homicides cause a lot of attention and are particularly visible to the population. Second, these homicides form a more homogeneous group of homicide (and for example largely exclude domestic homicides). We geocode homicide addresses using Google maps API's and restrict homicides geocoded at the street level, which correspond to 95 percent of all homicides in the public way.

Table 2 displays summary statistics of the victims of homicides for which the death occurs in the public way, as well as the description of characteristics of homicides. Approximately 70 percent of the homicides are a result of assault by gun discharge, and about 10 percent each by assault using sharp or blunt object. The majority of victims is in the age group between 19 to 50 years old, but there is substantial number (8.4 percent) of relatively young victims of homicide between the ages of 11 and 18. The vast majority of victims is male, and individuals from lower socio-economic background are over proportionally represented as victims of homicides, as indicated by very low levels of completed education. Figure A2 in the annex shows the distribution over time and space of the homicides in the public way in São Paulo. Darker shades of red represent areas more affected by homicides. In the paper, we make use of the variation of homicides over time and space depicted in the maps allowing us to disentangle the effect of violence from other correlates of socioeconomic variables and thus establishing causality between violence and education, as described in the next section.

4 Identification Strategy

Disentangling the effect of violence on education from confounding factors is not straightforward. In our case, poor neighbourhoods often register higher homicide rates and students from disadvantaged background are more likely to attain unsatisfactory results at school, hence it is necessary to deal with confounding factors that may lead to a positive association between levels of violence and poor educational performance. If, for example, areas with low socio-economic characteristics also exhibit high crime rates, and if pupils from relatively poorer households in these areas also perform worse at school, this would lead to a positive relationship in these variables even in the absence of any causal effect of violence on education.

In order to deal with these potential confounders, we use variation in homicides across space and time, where we are able to pinpoint the precise location of these homicides to the exact street

address, while applying school fixed effects, effectively dealing with unobservable characteristics of the school and neighbourhood. We also include time fixed effects to account for time trends in outcomes. The variation of homicides in each area associated with the measures of schooling performance of pupils whose neighbourhoods or schools were exposed to violence allow the estimation of the following model:

$$y_{ist} = \beta_0 + \beta_1 homicides_{st} + X_{ist}\beta_2 + Z_{st}\beta_3 + d_s + d_t + u_{ist} \tag{1}$$

y_{ist} is a range of different measures for schooling outcomes; $homicides_{st}$ is the number of homicides that lie in the close periphery of schools; X_{ist} are vectors of individual characteristics; Z_{st} are school and classroom time varying characteristics; d_s and d_t are school and time fixed effects, respectively; and u_{ist} is the error term.

We present an example of the variation we use in the maps in Figure A3 in the annex. Each individual map shows schools (white dots) and homicides in the public way (green circles) in a neighbourhood in São Paulo in a semester. The very precise information on school addresses and the address of occurrence of homicides allow us to construct very granular exposure points, and we focus on homicides occurring in a 25 meter radius around schools. The very granular geographic information helps us to minimise measurement error and to avoid that the measure of homicides of two schools in close proximity overlap.

As identifying assumption we assume that conditional on time and school fixed effects characteristics, the variation in the number of homicides in a very small geographic area is exogenous. In addition, we include a very rich set of individual, teacher, classroom and school characteristics to reduce sampling variability. We test for balancing of a large set of school and students characteristics by schools exposed and not exposed to violence, results in Table A1 show that the differences are very small and, from the very large set of characteristics, only three are statistically different.

To test how violence at different places of exposure impacts educational outcomes separately, we estimate the effect using measures for exposure at the school, the residence of students and on the way from the residence to school.

5 Results

In this section we present the results of the effect of exposure to violence on schooling outcomes. First, we investigate how violence affects test scores and attendance of children in Subsection 5.1 and

Subsection 5.2. We then look at some broader measures of school performance in Subsection 5.3.

5.1 Effect of violence in the school surroundings on standardised test scores

First, we are interested in whether exposure to violence affects educational achievement of students. We use standardised test scores in Math and Portuguese Language as measures of achievement. Table 3 presents the regression results of the effect of violence on Math and Language standardised test scores. All test scores are normalised at a (250,50) scale. The analysis includes students in 5th and 9th grades of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. The explanatory variable *Homicides* corresponds to the count of homicides within a 25m radius from school. We present robust standard errors clustered at the school level in parentheses. In order to account for possible spatial dependence between schools and for serial correlation, we compute Conley standard errors[3] Conley (1999), presented in brackets.

In the first column, we estimate the effect of homicides on standardised Math test scores without further individual controls (only school and time fixed effects); in the second column we include the rich set of student controls. In columns three, four and five we include, respectively, teacher characteristics, school characteristics and the classroom composition as controls.

Across specification we find a negative effect of homicides on Math test scores. Adding individual, teacher, school and classroom controls does not change the coefficients in any meaningful way, lending extra credibility to the identification strategy, but improve precision of the estimate. We find that an additional homicide in the surroundings of school during the year decreases Math test scores by 4.7 percent of a standard deviation.

In columns six to ten we repeat the exercise for Portuguese language scores; all test scores are normalised at a (250,50) scale. Across specifications, we find that exposure to homicides around schools has a negative effect on test scores. This is consistent with the findings of Monteiro and Rocha (2017), who find that the coefficients for Language, although negative, are generally smaller compared to the effects for Math test scores and not significant.

We also create indicator variables identifying high and low performers in these grades to investigate whether the effects are particularly driven by shifts in the lower or upper part of the test score distribution, Table A4 in the annex presents the results. The variables *Math high level* and *Language high level* indicate whether students reach the 'advanced' level in these subjects. Similarly, *Math low level* and *Language low level* show if the student's test scores are considered

[3]We compute Conley standard errors using a 25m cut-off distance. Results remain the same if we use 50m or 100m.

in the 'below the basic' level in these subjects. These variables provide an easy way to identify how students may be impacted differently at different parts of the skills distribution. We find that students are more likely to be classified as low level and less likely to be classified as high level in Math when they are exposed to violence around the school during the year. The coefficient for *Math low level* is higher compared *Math high level*, suggesting that low achieving students are more affected.

5.2 Effect of violence in the school surroundings on attendance

Next, we investigate whether homicides in the surroundings of the school affects students' attendance. In Table 4 we present the regression results of the effect of violence on attendance. We have access to the number of absences of each student for each bimester. As the ending dates of the bimesters are school specific and are not available from the data, we group the four bimesters in two semesters.[4] We then calculate the percentage of absences of each student in the entire year, in the first and in the second semester. We use the same routine to calculate the explanatory variables: *Homicides (year)* corresponds to the number of homicides within a 25m radius from school in the entire year; *Homicides (1st semester)* and *Homicides (2nd semester)* are the number of homicides within a 25m radius from school in the first and second semesters.

We find that one additional homicide in the year increases the number of absences by 1.4 percent. Each additional homicide around the school in the first semester increases the number of absences in the respective semester by 1.5 percent. In the second semester, one additional homicide in the surroundings of the school increases absences by 3.2 percent.

The coefficients for the second semester exceed the magnitude of the coefficients of the first semester. These results possibly could be explained by the dynamic incentives for students to attend over the year. As students can be retained if they fall below a 75 percent attendance threshold, earlier in the school year students may be more prudent regarding their attendance. Later in the year, when students have more control over their overall yearly attendance, they may be less prudent. In addition, the law regulating student attendance in São Paulo states that if a student has accumulated an excess of absences, the school must intervene and inform parents, so that they can take measures to remedy the problem. If parents are unsuccessful and the problem persists, the school must notify *Conselho Tutelar*, which is a local legal institution responsible for ensuring the well-being of children and adolescents. This is to make an effort and take measures

[4] We used the official starting and ending dates of each semester provided São Paulo State's Secretariat of Education.

during the year to avoid students' repetition due to absences. If students accumulate an excess of absences in the first semester, the school intervene and tries to remedy the situation. As a result of parents and school's effort, the effect in the first semester may decrease. In the second semester, closer to the end of the year, in the event of any negative shock that may impact student attendance, the school may not have time to intervene before the end of the year. Moreover, since it is the end of the year, students may find it harder to catch up missed classes and potentially miss even more school days.

Attendance may be one possible mechanism through which violence affects schooling performance. Aucejo and Romano (2016) found that a reduction in absences at school leads to an increase in both Math and reading test scores. Given the increase in absences and the large negative effect we find on test scores, attendance is possibly one of the reasons why violence affects test scores, therefore signalling the need for extra measures to deal with this problem.

5.3 Effect of violence in the school and residence surroundings on schooling performance

In addition to test score results, we are interested in additional educational outcomes as broad measures of educational achievement. We have these measures for a longer period, 2007 to 2013, and for all cohorts. Table 5 presents regression results of the effect of violence on these outcomes for all students in primary and secondary school, by place of exposure. *Panel A* and *Panel B* present the results for exposure in the school and residence surroundings, respectively, which are defined as the number of homicides that lie in a 25m perimeter from school (residence). We also look at the effect of exposure to violence in the walking path from residence to school for a fraction of students, for whom we have access to their postcode of residence. For that, we use Google maps API to calculate the shortest walking distance between school and residence geocoordinates. Along the walking path line, we build polygons of 50m width (25m to each side), which we call corridor, we show an example in Figure A4. Then, we calculate the closest orthogonal distance of each homicide, within the corridor, to the walking path line and estimate within corridor regressions, presented in *Panel C*.

Repetition is a dummy variable, which indicates whether the student has repeated the same grade as the current year in the coming year. We find positive coefficients for all types of exposure, but significant at the ten percent significance level in *Panel A*; this suggests an effect size, compared to the mean repetition rate, of 11.5 percent.

Dropout is a dummy variable, which captures whether a student drops out of school at the end of the school year (or indeed during the year). We find consistent positive coefficients, but significant and larger in magnitude for exposure in the school path. The variable *School progression* indicates whether students in the last grade of primary school progress to secondary school at the end of primary school. Although negative, as expected, the coefficients for this variable are not significant at conventional levels of significance.

We also investigate the effect of violence on school mobility. The variables *Between year transfer* and *Private school transfer* captures students who transfer to a different school in the following year and the latter specifies whether students transfer from a public to a private school. Results suggest that students are more likely to transfer schools between years in the event of a homicide near the school, and that parents are more likely to enrol their children in private schools; each additional homicide in the year rises transfers to private schools in the following year by 12.4 percent. The coefficients for the variable *In year transfer* indicate that transfers within the year are not affected by violence, which shows that violence does not change school composition within the year, reassuring the results on test scores found in the previous section.

6 Robustness checks

In this section we present a number of robustness checks.

6.1 Spatial correlation

Using spatial variation for identification has been identified as potentially problematic (Conley (1999)). In our context, dependent variables and our explanatory variable possibly are spatially correlated. In order to address this concern, we compute Conley standard errors using a weighted average of spatial covariances, using a cut point of 25 meters.[5] We report these standard errors in brackets for all specifications presented. In general, spatial standard errors are similar to regular clustered standard errors, confirming that spatial correlation likely plays no role in our context.

As schools are distributed very close to each other given the high-density urban setting of São Paulo, we use the very granular geographic measures provided by our data and consider homicides within a 25 meters radius from school as measure of exposure avoiding that exposure to the same homicides overlaps across different schools. As a robustness check, we estimated regressions in

[5] We also computed these standard errors at 100m cut-off, the results are unchanged.

Table 3 using homicides within a radius of 100 meters from school, estimates in Table A3 are consistent with our main results.

6.2 Selection

Given that we only consider students who attend the exams in our analysis, this may lead to a selection problem potentially introducing a bias to results we present in Table 3. If homicides in the school surroundings affect students' decisions to take the test and the propensity to attend differs systematically by student types, this could bias our results accordingly. Because this is low-stakes test and schools have generally little incentive to manipulate attendance of students at the test, the scope for selection is relatively small. Overall, approximately 87 percent of students attend the test.

In order to rule out the possibility that the students taking SARESP are selected, we test if violence in the school surroundings affect attendance at Math and Language tests. For this purpose, we estimate the effect of exposure to homicides in the school surroundings on an attendance indicator separately for Math and for Portuguese. The coefficients in Table A2 are small and not statistically different from zero, which reassures that the results in Table 3 are not biased by selection.

6.3 Teachers' attendance

As teachers are also exposed to the violence around the school, we test whether the effects are driven by teacher absenteeism rather than by a direct effect on students by including teacher attendance as a control in specifications in columns (5) and (10) in Table 3. We present the results in Table 13, the difference in the coefficients when including teacher attendance is minimal. This is contrary to Monteiro and Rocha (2017) who state that the effect of exposure to drug battles on educational outcomes they find is at partially caused by teacher absenteeism and turnover.

7 Heterogeneous effects

We are also interested in understanding how specific groups of students are affected differently by violence around the school. We investigate heterogeneity in the effect of violence on attendance and test scores by splitting the sample by cohort, by gender and by socio-economic status.

7.1 Analysis by cohort

In Table 6 we present the results of the effect of exposure to violence on Math and Language test scores for each of the three cohorts in our sample: 5th and 9th grades of primary school and 3rd grade of secondary school. All specifications include time and school fixed effects and the full set of controls. The coefficients for Math are significant and sizeable in magnitude, particularly for students in 5th grade of primary school, for whom an additional homicide in the surroundings of the school during the year implies in a reduction of 7.9 percent of a standard deviation of Math proficiency. The effect is slightly smaller for 3rd grade of secondary school, 6 percent of a standard deviation. Students in 9th grade are the least affected, compared to the other mentioned cohorts, but exhibit still a considerable effect, with a 3 percent of a standard deviation reduction in Math test scores. The coefficients for Language are also larger for 5th graders compared to the other cohorts, but not significant. This pattern is consistent with two different mechanisms for the different age groups being at work: a negative effect induced by fear for the younger students and a negative effect due to the disincentives to invest in human capital for the older students.

We also estimate how the attendance of each cohort is affected by violence, the results are in Table 7. Each additional homicide in the year increases absences of 9th grade students by 1.9 percent. The effect is slightly smaller for 5th graders 1.3 percent. The coefficients for secondary school are not significant at the conventional levels of significance.

7.2 Analysis by gender

In Table 8 we present results of the effect of violence in the school surroundings on Math and Language standardised test scores separately for boys and girls. All specifications include time and school fixed effects and the full set of controls. We find that boys are more affected than girls, for each additional homicide around the school in the year, boys Math proficiency decreases by 6.7 percent of a standard deviation. The effect on girls is about half this size, 3.1 percent of a standard deviation, and only significant at the 10 percent significance level when considering Conley standard errors.

Table 9 presents the effect of exposure to violence in the school surroundings on boys and girls' attendance in the year and in each semester. All specifications include time and school fixed effects and the full set of controls. Violence in the school surroundings affects both boys and girls, however, the coefficients for boys are larger, confirming a similar pattern documented for test scores.

7.3 Analysis by socio-economic status

We use information on parental income and educational background to look at heterogeneous effects by socio-economic status. First, we split the sample by income per capita and classify as *Low income* parents whose family income per capita is less than the median income in each year of the analysis and *High income* otherwise. Second, we analyse separately students whose both parents' level of education is at most primary school, defined as *Less educated* and students whose both parents have more than secondary school, defined as *More educated*.

In Table 10 we present the results of the effect of violence around the school on test scores for each of the defined categories. All specifications include time and school fixed effects and the full set of controls. Columns (1) and (2) compare Math test scores of low and high income children. We find a much more pronounced and statistically significant negative effect for low income students. We find the same pattern for language proficiency which reveals a stronger effect for lower compared to higher income students, as shown in columns (5) and (6). In columns (3) and (4) we compare Math proficiency of students by educational background of their parents. Although not significant at usual significance levels, results suggest that students whose parents are more educated are more affected and the same is true for the language test scores.

We also look at absences of students considering the same categories. Results in Table 11 are consistent with the patterns we find for test scores. We first present absences in the year, then in each semester. All specifications include time and school fixed effects and the full set of controls. In columns (1) and (2), we compare absences of low and high income students in the year. Although not statistically significant, results suggest that low income students are slightly more likely to be absent when there is a homicide in the surroundings of school. This difference is driven by differences in attendance in the first semester; results for the second semester are relatively balanced. When comparing absences by levels of education, we find that students whose parents are more educated are more likely to be absent in the event of a homicide in the school surroundings.

These results suggest that socio-economic background plays an important role. Income seems to act as a buffer against the harmful effect of exposure to violence. Parents of higher SES may be better able to shield their children from the negative effect of the exposure to violence, for example through additional safety measures or by giving a sense of security by dropping and picking-up their children by car. This is also consistent with the large body of literature which has

documented that parents' socio-economic status may influence children's educational performance through their behaviour and beliefs. In particular, parents of a higher socio-economic status are in general more likely to actively engage in their children's educational process, they more engaged with teachers, spend more time with their children and provide more assistance and support for learning at home (Flouri and Buchanan (2004), Davis-Kean (2005), Dearing et al. (2006) Guryan et al. (2008), Houtenville and Conway (2008), De Fraja et al. (2010), Gelber and Isen (2013), Mora and Escardíbul (2018)).

The contrary effects by education seem at first surprising. As we simultaneously also control for parental income, these results possibly point to a different mechanism at work. For the results on attendance, more educated parents possibly may have a better perception of the risks involved, and in the event of a homicide, they might be more cautious in sending their children to school. They may also have better means in compensating for missed days at school by substituting educational inputs at school with their own input. Without further evidence, these results though call for a cautious interpretation.

8 Mechanisms

In this section we investigate potential underlying mechanisms through which violence affects student's performance at school.

8.1 Bereavement effect

In order to check whether the effect we find is driven by grief due to the death of a peer or a teacher at the same school, we use information on deceased students and teachers from São Paulo State's Secretariat of Education. We identified the cause of death by linking these data with information on death records from *Datasus*. From the student's data, we identified 510 deceased students in the period of 2010 to 2013. In order to be able to identify the cause of death, we had to drop 10 observations with the same year of death, sex and date of birth. From the 500 left, we could successfully identify the cause of death of 353 cases. From those, 41 cases were victims of homicides, but only 4 of them happened in the public way. None of these 4 cases nevertheless occurred in proximity of schools, and hence were not included in our explanatory variable. We repeated the same exercise for the teachers. From 2010 to 2013 we identified 220 deceased teachers and we could identify the cause of death of 128 of the cases. From those cases, none of them were

homicide victims. We are hence confident, that the effects are not due to grief of bereavement of peers or teachers of the students in our dataset.

To rule out the possibility that the variable *Homicides* is also capturing grief for the loss of a friend (who may live in the same neighbourhood, but may not attend the same school), we drop from the explanatory variable all the victims who are 18 years old or younger. We present the results for Math proficiency in Table A7; the specification for all entries follow the most satiated specification of columns (5) of Table 3. Column 1 shows the effect of homicides around the school including all the victims. In column 2 we exclude all 18 year old or younger victims. Column 3 considers only male victims in the explanatory variable and column 4 only gunshot victims. Results do not differ in any meaningful way. We hence can rule out a channel based on grief for the loss of an individual related to the students, either teacher, classmates or friends of the same age or younger.

8.2 Human capital accumulation

There is a literature linking life expectancy with individuals human capital investment decisions (Becker (1964), Ben-Porath (1967), Jayachandran and Lleras-Muney (2009), Oster et al. (2013)). The gender specific results presented here are consistent with differences in the disincentives to invest in education for boys and girls linked to the pronounced differences in the probability of direct victimisation in a homicide by sex. This relates to a literature that has analysed how shocks to life expectancy that differ by sex, such as health and violence shocks, affect investments in education. Gerardino (2015) shows that when male-biased violence is high, boys are less likely to enrol in secondary school relative to girls. The author proposes two channels which might be responsible for this result: an increase in the opportunity cost of attending school and a reduction in the returns to education.

In the previous section, we find that the effect on boys is profoundly larger on both test scores and attendance. Boys seem to react much more strongly to the homicide exposure in the school surroundings. Simultaneously, recall that the vast majority of homicide victims are male; indeed more than 90 percent are male, as shown in Table 2. In Brazil, homicide is a leading cause of death for boys up to their mid-twenties. The difference in victimisation rates in homicides by sex might affect the perception of safety of males and females differently. The underlying mechanism behind our results may include a component related to the perceived returns to education that may be affected by directly experiencing homicides in the neighbourhood. Essentially, a non-negligible risk

to die as homicide victim may impact the decision to invest in education.

When investigating the effects by grade, in Table 6 we found particularly pronounced effects for students towards the end of secondary school[6] and at the end of the first cycle of primary school[7]. The number of children victimised in homicides at the age of 11 is very small, but substantially larger at the age of 18, when possibly at least some of the children (mostly boys) may have had some exposure to criminal activities. One might therefore expect that differences by sex may be pronounced for the older cohort of students in secondary school if a human capital mechanism is at work.

To investigate this further and to separate a human capital effect from a mechanism that arises from the general fear of going to school, we break down the results by cohorts further by sex. We present the results in Table A5. Indeed, we find that results for Math in secondary school are much more pronounced for boys, whereas we find no effect for girls. In fact the estimate for girls is positive and close to zero. While there also exists a difference in the effects for the younger cohorts, the difference is much less pronounced and even inverted for 9th graders. For Portuguese language we find a pattern that is almost the inverse of the results for math, possibly indicating that the production of language and Math skills are fundamentally different. We also look at the effect on absenteeism by cohort and sex. While we confirm that the effect on absenteeism is not present for the secondary school students, we find that the effects do not differ strongly by sex, with slightly more pronounced effects for boys in both, 5th and 9th grade.

8.3 Students' attendance

In section 5.2 we show that attendance is – apart from an outcome in its own right - also one of the possible mechanisms through which violence in the surroundings of school affects student's performance. In order to tease out how much of the results on test scores can be explained by absenteeism alone, we estimate specifications in columns (5) and (10) in Table 3 including students attendance as a control. Results in Table 12, show a decrease of about 17 percent of the Math coefficient and 20 percent decrease of Language coefficient. Although one needs to be careful when including an endogenous variable on the right hand side, this exercise may shed some light on the potential mechanisms. Interestingly, the inclusion of student attendance in either Math or Portuguese reduces the coefficient on test scores only very mildly. We interpret this as evidence, that

[6]Students in the final grade of secondary school would largely be around 18 years of age

[7]Target age for students in 5th grade is 11 years

attendance is only partially responsible for the negative effect on test scores suggesting possibly an underlying human capital mechanism driving the results on test scores (and possibly simultaneously of attendance).

9 Final Remarks

This paper uses georeferenced data on homicides for Brazil, and links these data with measures of schooling performance to estimate the causal effect of exposure to violence on schooling outcomes.

We find that students exposed to violence have lower performance in Math with larger effects for students in 5th grade of primary school and the last grade in secondary school. The coefficients for Language proficiency are negative and similar in magnitude to the coefficients in Math, but not significant at conventional levels of significance. We create indicator variables which identify high and low performance students and find that in the event of a homicide in the year, low achievers in Math are more affected compared to high achievers.

Violence around school also affects attendance of the students at school, especially in the second semester. Estimations show that one additional homicide in the year increases absences by 1.4 percent. These findings point to the potential role of attendance as a mechanism through which violence affects students performance, calling attention to public policies aimed at improving safety conditions around the school.

We find heterogeneous effects of violence on test scores and attendance of boys and girls. In both cases, the effect is larger on boys, calling attention for another possible mechanism through which violence affects student's performance at school: a shift in the incentives to invest in human capital. We also provide evidence for parental income serving as a buffer against the negative effect of exposure to crime.

In addition, we look at the effect of violence in broader measures of school achievement for all the cohorts in primary and secondary schools. We find that one additional homicide in the school surroundings during the year increases the repetition rate by 11.5 percent. Dropout rates increase after exposure to violence in the school path from residence to school. We also find an effect on school transfers. Homicides around the school during the year increase transfers to alternative schools in the following year, and students are more likely to transfer to private schools, revealing a potential behavioural channel of parents as reaction to homicides in the school surroundings.

These results are important to quantify some of the costs of violence that go beyond the cost

of direct victimisation. Even though we only measure the short term impact of violence, the negative effect we find on measures of school performance suggest that violence also affects human capital accumulation. Since poor neighbourhoods are often more violent, violence is potentially one additional contributor for the socio-economic gradient we observe in many low and middle income countries plagued with high crime rates.

References

Aizer, Anna, "Neighborhood Violence and Urban Youth," *NBER Working Paper Series [National Bureau of Economic Research]*, 2008, *13773* (13773).

Aucejo, Esteban M and Teresa Foy Romano, "Assessing the Effect of School Days and Absences on Test Score Performance," *Economics of Education Review*, 2016, *55*, 70–87.

Becker, Gary., "Human Capital.," *Chicago: University of Chicago Press.*, 1964.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings.," *Journal of Political Economy*, 1967, *75* (4), 352–365.

Cerqueira, Daniel R C, Alexandre X. Y. Carvalho, Waldir J. A. Lobão, and Rute I. Rodrigues, "Análise dos Custos e Consequências da Violência no Brasil," *Texto Para Discussão, Ipea*, 2007, *1284*.

Conley, T G, "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92*, 1–45.

Cooley-Quille, M R, S M Turner, and D C Beidel, "Emotional impact of children's exposure to community violence: a preliminary study.," *Journal of the American Academy of Child and Adolescent Psychiatry*, oct 1995, *34* (10), 1362–8.

Davis-Kean, Pamela E., "The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment," *Journal of Family Psychology*, 2005, *19* (2), 294–304.

De Fraja, Gianni, Tania Oliveira, and Luisa Zanchi, "Must Try Harder: Evaluating the Role of Effort in Educational Attainment," *Review of Economics and Statistics*, 2010, *92* (3), 577–597.

Dearing, Eric, Holly Kreider, Sandra Simpkins, and Heather B. Weiss, "Family involvement in school and low-income children's literacy: Longitudinal associations between and within families," *Journal of Educational Psychology*, 2006, *98* (4), 653–664.

Farrell, Albert D, Steven E Bruce, D Albert, and Stwen E Brracc, "Journal of Clinical Child Psychology Impact of exposure to community violence on violent behavior and emotional distress among urban adolescents," 2010, (August 2014), 37–41.

Flouri, Eirini and Ann Buchanan, "Early father's and mother's involvement and child's later educational outcomes," *British Journal of Educational Psychology*, 2004, *74* (2), 141–153.

Fowler, Patrick J, Carolyn J Tompsett, Jordan M Braciszewski, Angela J Jacques-Tiura, and Boris B Baltes, "Community violence: a meta-analysis on the effect of exposure and mental health outcomes of children and adolescents.," *Development and psychopathology*, jan 2009, *21* (1), 227–59.

Gelber, Alexander and Adam Isen, "Children's schooling and parents' behavior: Evidence from the Head Start Impact Study," *Journal of Public Economics*, 2013, *101* (1), 25–38.

Gerardino, Maria Paula, "The effect of violence on educational gender gap," *ProQuest Dissertations and Theses*, 2015, p. 103.

Grogger, Jeffrey, "Local Violence and Educational Attainment," *The Journal of Human Resources*, 1997, *32* (4), 659–682.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney, "Parental Education and Parental Time with Children," *Journal of Economic Perspectives*, 2008, *22* (3), 23–46.

Houtenville, Andrew J. and Karen Smith Conway, "Parental Effort, School Resources, and Student Achievement," *The Journal of Human Resources*, 2008, *94* (May 2007), 71–89.

Jayachandran, Seema and Adriana Lleras-Muney, "Life expectancy and human capital investments: Evidence from maternal mortality declines.," *Quarterly Journal of Economics*, 2009, *124(1)*, 349–397.

Monteiro, Joana and Rudi Rocha, "Drug Battles and School Achievement: Evidence from Rio de Janeiro's Favelas," *Review of Economics and Statistics*, 2017, *99* (2), 213–228.

Mora, Toni and Josep Oriol Escardíbul, "Home Environment and Parental Involvement in Homework During Adolescence in Catalonia (Spain)," *Youth and Society*, 2018, *50* (2), 183–203.

Oster, By Emily, Ira Shoulson, and E Ray Dorsey, "Limited Life Expectancy , Human Capital and," *American Economic Review*, 2013, *103* (5), 1977–2002.

Sharkey, Patrick, Amy Ellen Schwartz, Ingrid Gould Ellen, Johanna Lacoe, Jesper Sørensen, and Stephen L Morgan, "High Stakes in the Classroom, High Stakes on the Street:

The Effects of Community Violence on Students' Standardized Test Performance," *Sociological Science*, 2014, *1*, 199–220.

Smith, Deborah Gorman and Patrick Tolan, "The role of exposure to community violence and developmental problems among inner-city youth," 1998, *10*, 101–116.

Table 1: Education summary statistics

	Students characteristics		
	Mean	Std.Dev.	Obs
Age			
06-10	0.315	0.464	10,400,046
11-15	0.462	0.499	10,400,046
16-18	0.201	0.400	10,400,046
18+	0.022	0.148	10,400,046
Demographics			
White	0.366	0.482	10,400,046
Black	0.031	0.173	10,400,046
Mixed	0.189	0.392	10,400,046
Male	0.503	0.500	10,400,046
School performance			
Repetition	0.075	0.263	8,873,385
Dropout	0.126	0.332	10,400,046
School progression	0.730	0.444	1,052,547
Private school transfer	0.005	0.072	10,400,046
Between year transfer	0.161	0.368	8,873,385
In year transfer	0.001	0.030	10,400,046
	School characteristics		
	Mean	Std.Dev.	Obs
General characteristics			
Federal	0.001	0.025	3,164
State	0.340	0.474	3,164
Municipal	0.165	0.371	3,164
Private	0.495	0.500	3,164
Has principal's office	0.889	0.262	3,164
Has teachers' office	0.943	0.190	3,164
Has computer lab	0.773	0.362	3,164
Has science lab	0.345	0.425	3,164
Has library	0.421	0.353	3,164
Number of school rooms in use	14.914	8.670	3,164
Has internet access	0.967	0.113	3,161
Number of staff members	60.777	45.636	3,164
School meals	0.583	0.456	3,164

Note: The table includes students from 1st grade of primary school to 3rd grade of secondary school over the period of 2007 to 2012. *Dropout* is a dummy variable which captures if the student drops out school in the successive year. *Private school transfer* captures students who transfer from a public to a private school in the following year. *Between year transfer* and *In year transfer* indicate whether the student transfers to a different school between or within the school year, respectively. *School progression* indicates if the students in the last grade of primary school progress to secondary school in the subsequent school year, for that reason, this variable is calculated only for students at the final grade of primary school.

Table 2: Homicides characteristics

	Homicide victims characteristics	
	Mean	Std.Dev.
Age		
02-10	0.002	0.048
11-15	0.018	0.131
16-18	0.066	0.248
19-25	0.225	0.418
26-50	0.495	0.500
50+	0.194	0.395
Demographics		
Male	0.925	0.263
White	0.420	0.494
Black	0.100	0.300
Mixed	0.456	0.498
Single	0.642	0.480
Married	0.122	0.328
Separated	0.027	0.161
Education		
None	0.013	0.115
01-03 years	0.094	0.292
04-07 years	0.391	0.488
08-11 years	0.266	0.442
12+ years	0.032	0.176
	Homicide characteristics	
	Number	Percent
Assault by gun discharge	1,813	69.093
Assault by sharp object	282	10.747
Assault by blunt object	275	10.480
Assault by bodily force	152	5.793
Assault by other means	102	3.887
Total	2,624	100.000

Note: The table includes all homicides for which the death occurs in the public way in São Paulo over the period of 2007 to 2013, which were geocoded at the street level.

Table 3: Effect of exposure to violence in the school surroundings on standardised test scores

	Math proficiency					*Language proficiency*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Homicides	−2.556	−2.281	−2.399	−2.428	−2.349	−2.111	−1.271	−1.266	−1.299	−1.188
	(2.553)	(1.036)**	(0.933)**	(0.945)**	(0.967)**	(2.969)	(1.015)	(1.006)	(1.027)	(0.977)
	[1.823]	[0.846]***	[0.783]***	[0.788]***	[0.808]***	[2.003]	[0.991]	[0.955]	[0.947]	[0.930]
Observations	666,718	666,718	666,718	666,718	666,718	666,453	666,453	666,453	666,453	666,453
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Teacher controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
School controls	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Classroom controls	No	No	No	No	Yes	No	No	No	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 4: Effect of exposure to violence in the school surroundings on attendance

	Absences year		*Absences 1st semester*		*Absences 2nd semester*	
	(1)	(2)	(3)	(4)	(5)	(6)
Homicides (year)	0.016					
	(0.005)***					
	[0.003]***					
Homicides (year)		0.014				
		(0.004)***				
		[0.004]***				
Homicides (1st sem.)			0.017			
			(0.006)***			
			[0.004]***			
Homicides (1st sem.)				0.015		
				(0.004)***		
				[0.004]***		
Homicides (2nd sem.)					0.036	
					(0.004)***	
					[0.005]***	
Homicides (2nd sem.)						0.032
						(0.005)***
						[0.006]***
Observations	726,215	726,215	726,215	726,215	726,215	726,215
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Dependent variables are the percentage of absences in the year and in each semester. Explanatory variables *Homicides (year)* corresponds to the number of homicides within a 25m radius from school in the entire year; *Homicides (1st semester)* and *Homicides (2nd semester)* are the number of homicides within a 25m radius from school in the first and second semesters. All regressions include time and school fixed effects. Controls include individual characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 5: Effect of exposure to violence on schooling performance

Panel A: Exposure in the school surroundings

	Repetition		*Dropout*		*Private school transfer*		*Between year transfer*		*In year transfer*		*School progression*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Homicides	0.011**	0.009*	0.011	0.009	0.001**	0.001**	0.042*	0.043*	0.000	0.000	−0.031*	−0.026
	(0.005)	(0.005)	(0.010)	(0.010)	(0.001)	(0.001)	(0.022)	(0.022)	(0.000)	(0.000)	(0.018)	(0.017)
Schools	2,680	2,680	2,682	2,682	1,594	1,594	2,677	2,677	2,688	2,688	1,833	1,833
Observations	7,698,069	7,698,069	8,580,404	8,580,404	6,237,778	6,237,778	6,897,926	6,897,926	8,944,932	8,944,932	1,047,110	1,047,110
R^2	0.050	0.124	0.022	0.082	0.010	0.012	0.065	0.248	0.002	0.002	0.052	0.122
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Panel B: Exposure in the residence surroundings

	Repetition		*Dropout*		*Private school transfer*		*Between year transfer*		*In year transfer*		*School progression*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Homicides	0.010*	0.002	0.034	0.041	0.000	0.000	0.069	0.003	−0.000	−0.000	0.018	0.007
	(0.005)	(0.005)	(0.023)	(0.030)	(0.001)	(0.001)	(0.044)	(0.020)	(0.000)	(0.000)	(0.018)	(0.017)
Neighbourhoods	9,213	9,213	10,638	10,638	7,963	7,963	8,391	8,391	11,186	11,186	2,356	2,356
Observations	1,646,763	1,646,763	1,848,197	1,848,197	1,643,719	1,643,719	1,357,756	1,357,756	1,850,684	1,850,684	259,129	259,129
R^2	0.020	0.059	0.035	0.108	0.021	0.022	0.067	0.644	0.014	0.014	0.051	0.106
Neighb./time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Panel C: Exposure in the school path

	Repetition		*Dropout*		*Private school transfer*		*Between year transfer*		*In year transfer*		*School progression*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Homicides	0.006	0.004	0.048*	0.050**	0.000	0.000	0.014	0.002	−0.001	−0.001	0.003	−0.004
	(0.004)	(0.004)	(0.024)	(0.024)	(0.001)	(0.001)	(0.024)	(0.011)	(0.000)	(0.000)	(0.025)	(0.020)
Corridors	11,196	11,196	12,134	12,134	10,050	10,050	10,465	10,465	12,453	12,453	2,331	2,331
Observations	1,558,943	1,558,943	1,747,582	1,747,582	1,556,877	1,556,877	1,286,191	1,286,191	1,749,409	1,749,409	245,698	245,698
R^2	0.024	0.063	0.036	0.108	0.017	0.018	0.091	0.672	0.103	0.103	0.053	0.107
Corridor/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered (at the school level in Panel A, at the neighbourhood level in Panel B and at the corridor level in Panel C) in parentheses.

Note: The analysis includes students from 1st grade of primary school to 3rd grade of secondary school over the period of 2007 to 2012. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variable *Repetition* is a dummy variable which indicates whether the student has to repeat the same grade as the current year in the coming year. *Dropout* is a dummy variable which captures if the student drops out school in the successive year. *Private school transfer* captures students who transfer from a public to a private school in the following year. *Between year transfer* and *In year transfer* indicate whether the student transfers to a different school between or within the school year, respectively. *School progression* indicates if the students in the last grade of primary school progress to secondary school in the subsequent school year, for that reason, regressions for this outcome include only students at the final grade of primary school. All regressions include time and school fixed effects. Controls include individual characteristics, school characteristics and classroom composition. **Individual controls** are: age, sex and race fixed effects; **Classroom controls** are: share of black students, share of girls and share of students above the appropriate age. **School controls** are: number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals.

Table 6: Effect of exposure to violence in the school surroundings on standardised test scores - heterogeneous effects by cohort

	Math *5th grade* *(primary school)*	*Math* *9th grade* *(primary school)*	*Math* *3rd grade* *(secondary school)*	*Language* *5th grade* *(primary school)*	*Language* *9th grade* *(primary school)*	*Language* *3rd grade* *(secondary school)*
	(1)	(2)	(3)	(4)	(5)	(6)
Homicides	−3.970	−1.492	−3.039	−2.518	−0.185	−1.907
	(1.904)**	(0.757)**	(1.837)*	(2.866)	(1.660)	(2.576)
	[1.529]***	[0.917]	[1.468]**	[2.265]	[1.370]	[2.253]
Observations	237,000	308,311	121,407	236,735	308,311	121,407
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 7: Effect of exposure to violence in the school surroundings on attendance - heterogeneous effects by cohort

	Absences 5th grade (primary school)	*Absences 9th grade (primary school)*	*Absences 3rd grade (secondary school)*
	(1)	(2)	(3)
Homicides	0.013	0.019	−0.004
	(0.006)**	(0.004)***	(0.006)
	[0.005]**	[0.006]***	[0.006]
Observations	247,122	337,888	141,205
School/time fe	Yes	Yes	Yes
Controls	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Dependent variables are the percentage of absences in the year. Explanatory variable *Homicides* correspond to the number of homicides within a 25m radius from school in the year. All regressions include time and school fixed effects. Controls include individual characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 8: Effect of exposure to violence in the school surroundings on standardised test scores - heterogeneous effects by gender

	Math proficiency (boys)	*Language proficiency (boys)*	*Math proficiency (girls)*	*Language proficiency (girls)*
	(1)	(2)	(3)	(4)
Homicides	−3.352	−1.522	−1.570	−0.978
	(1.292)***	(1.312)	(1.030)	(1.445)
	[1.088]***	[1.133]	[0.864]*	[1.209]
Observations	329,159	328,903	337,559	337,550
School/time fe	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 9: Effect of exposure to violence in the school surroundings on attendance - heterogeneous effects by gender

	Absences (year - boys)	*Absences (year - girls)*	*Absences (1st sem. - boys)*	*Absences (1st sem. - girls)*	*Absences (2nd sem. - boys)*	*Absences (2nd sem. - girls)*
	(1)	(2)	(3)	(4)	(5)	(6)
Homicides (year)	0.019					
	(0.005)***					
	[0.004]***					
Homicides (year)		0.010				
		(0.004)***				
		[0.004]***				
Homicides (1st sem.)			0.019			
			(0.006)***			
			[0.005]***			
Homicides (1st sem.)				0.013		
				(0.004)***		
				[0.004]***		
Homicides (2nd sem.)					0.034	
					(0.005)***	
					[0.008]***	
Homicides (2nd sem.)						0.028
						(0.005)***
						[0.005]***
Observations	360,828	365,387	360,828	365,387	360,828	365,387
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Dependent variables are the percentage of absences in the year. Explanatory variables *Homicides* corresponds to the number of homicides within a 25m radius from school in the year. All regressions include time and school fixed effects. Controls include individual characteristics, school characteristics and classroom composition. **Individual controls** are age and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 10: Effect of exposure to violence in the school surroundings on standardised test scores - heterogeneous effects by socioeconomic status

	Math proficiency				*Language proficiency*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low income	*High income*	*Less educated*	*More educated*	*Low income*	*High income*	*Less educated*	*More educated*
Homicides	−2.556	−0.058	−0.434	−2.780	−2.376	0.203	0.608	−1.626
	(1.852)	(1.470)	(1.629)	(1.977)	(1.666)	(1.779)	(1.390)	(2.105)
	[1.316]*	[1.534]	[1.308]	[1.704]	[1.255]*	[1.368]	[1.338]	[1.729]
Observations	181,704	200,614	200,043	102,982	181,667	200,491	199,944	102,992
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. We coded as *Low income* parents whose income per capita is below the median income in each year and *High income* otherwise. *Less educated* include only cases in which both parents have only primary school and *More educated* cases in which both parents have more than primary school. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 11: Effect of exposure to violence in the school surroundings on attendance - heterogeneous effects by socioeconomic status

	Absences year				*Absences 1st semester*				*Absences 2nd semester*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Low income	*High income*	*Less educated*	*More educated*	*Low income*	*High income*	*Less educated*	*More educated*	*Low income*	*High income*	*Less educated*	*More educated*
Homicides	0.004	0.001	0.001	0.005								
(year)	[0.003]	[0.003]	[0.003]	[0.003]								
	(0.003)	(0.002)	(0.003)	(0.002)**								
Homicides					0.006	0.003	0.004	0.006				
(1st semester)					(0.002)***	(0.002)	(0.003)	(0.004)				
					[0.002]***	[0.002]*	[0.003]*	[0.004]*				
Homicides									0.021	0.024	0.022	0.025
(2nd semester)									(0.005)***	(0.011)**	(0.010)**	(0.010)**
									[0.005]***	[0.009]**	[0.008]***	[0.008]***
Observations	188,656	207,610	208,297	106,194	188,656	207,610	208,297	106,194	188,656	207,610	208,297	106,194
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Dependent variables are the percentage of absences in the year and in each semester. Explanatory variables *Homicides (year)* corresponds to the number of homicides within a 25m radius from school in the entire year; *Homicides (1st semester)* and *Homicides (2nd semester)* are the number of homicides within a 25m radius from school in the first and second semesters. We coded as *Low income* parents whose income per capita is below the median income in each year and *High income* otherwise. *Less educated* include only cases in which both parents have only primary school and *More educated* cases in which both parents have more than primary school. All regressions include time and school fixed effects. Controls include individual characteristics, school characteristics and classroom composition. **Individual controls** are age and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 12: Effect of exposure to violence in the school surroundings on standardised test scores: the role of students attendance

	Math proficiency		*Language proficiency*	
	(1)	(2)	(3)	(4)
Homicides	−2.251	−1.864	−1.285	−1.029
	(0.935)**	(0.843)**	(0.938)	(0.882)
	[0.789]***	[0.753]**	[0.931]	[0.918]
Observations	651,471	651,471	651,216	651,216
School/time fe	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes
Classroom controls	Yes	Yes	Yes	Yes
Student attendance	No	Yes	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table 13: Effect of exposure to violence in the school surroundings on standardised test scores: the role of teachers attendance

	Math proficiency		*Language proficiency*	
	(1)	(2)	(3)	(4)
Homicides	−2.349	−2.558	−1.188	−1.227
	(0.967)**	(1.061)**	(0.977)	(0.991)
	[0.808]***	[0.869]***	[0.930]	[0.945]
Observations	666,718	666,718	666,453	666,453
School/time fe	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes
Classroom controls	Yes	Yes	Yes	Yes
Teacher attendance	No	Yes	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Annex

Figure A1: School Calendar in São Paulo

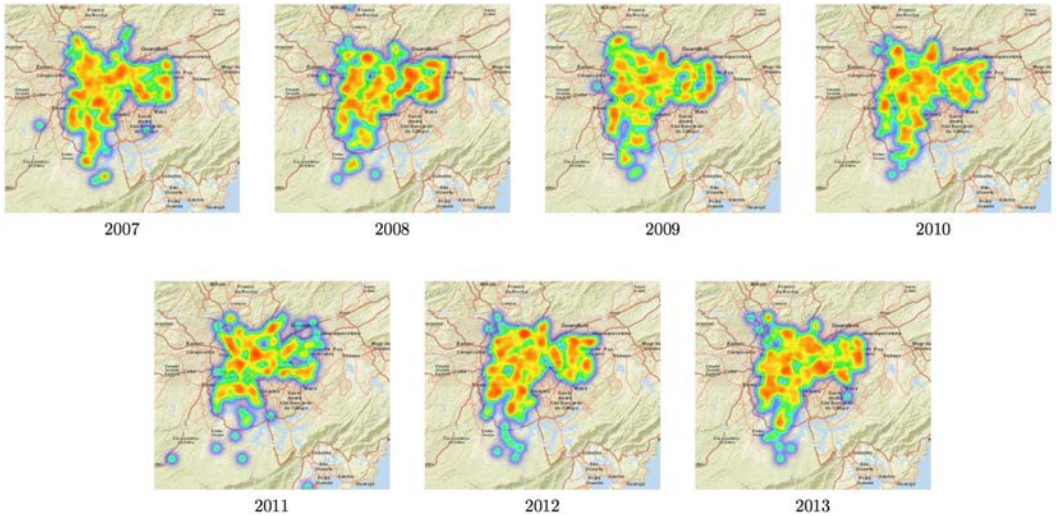


Figure A2: Homicides in the public way in São Paulo

(a) 2007s1

(b) 2007s2

(c) 2008s1

(d) 2008s2

(e) 2009s1

(f) 2009s2

(g) 2010s1

(h) 2010s2

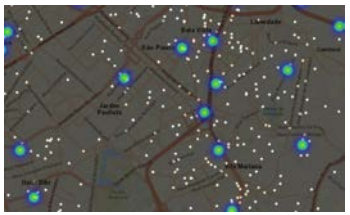
(i) 2011s1

(j) 2011s2

(k) 2012s1

(l) 2012s2

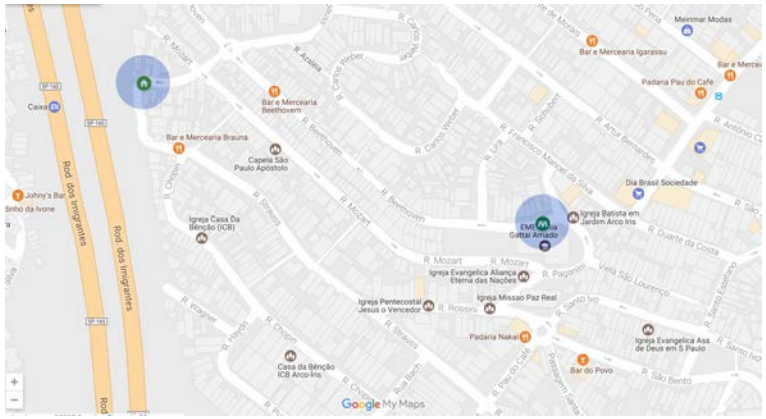
(m) 2013s1



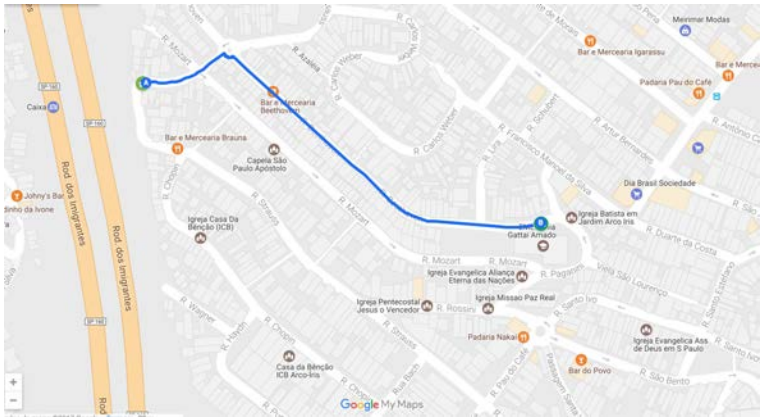
(n) 2013s2

Figure A3: Homicides and schools in a São Paulo neighbourhood

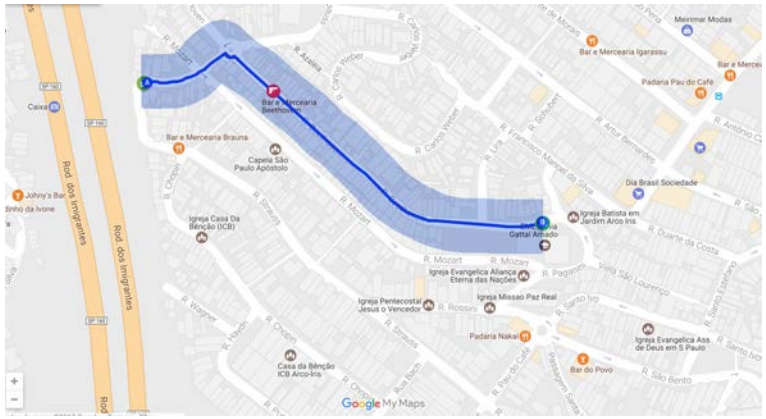
Note: Each individual map shows schools (white dots) and homicides (green circles) in a São Paulo neighbourhood in a semester.



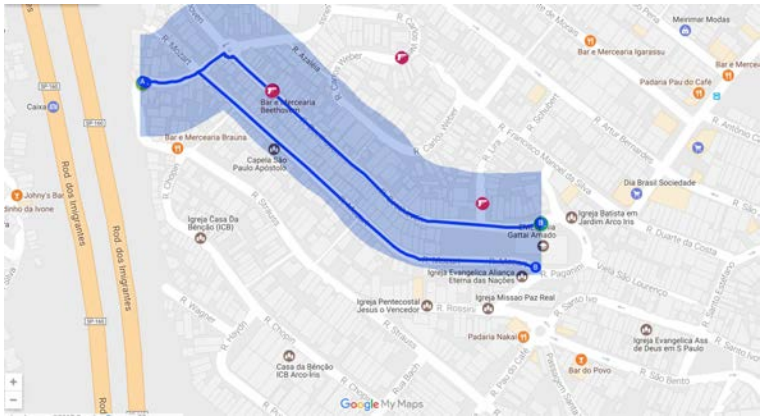
(a) School and residence radius



(b) Shortest walking distance from residence to school



(c) Corridor 1



(d) Corridor 2

Figure A4: Walking path from residence to school - Corridors

Table A1: Balancing tests

	Mean(Homicides=1)	Mean(Homicides=0)	Diff.	Std. Error	Obs.
Students characteristics					
Age	14.2171	13.1497	-1.0675*	0.5734	4,045
Female	0.4830	0.4978	0.0149	0.0120	4,045
White	0.5470	0.5750	0.0280	0.0285	4,045
Black	0.0634	0.0572	-0.0063	0.0079	4,045
Mixed	0.3734	0.3564	-0.0170	0.0273	4,045
Income	1620.1821	1595.2596	-24.9225	66.8630	4,016
Own home	0.5024	0.4622	-0.0402	0.0307	4,020
Rent home	0.4976	0.5378	0.0402	0.0307	4,020
Father's education: low	0.5930	0.6067	0.0137	0.0258	4,019
Father's education: mid	0.2791	0.2785	-0.0007	0.0218	4,019
Father's education: high	0.0614	0.0504	-0.0110	0.0095	4,019
Mother's education: low	0.5678	0.5700	0.0022	0.0287	4,019
Mother's education: mid	0.3381	0.3464	0.0083	0.0238	4,019
Mother's education: high	0.0658	0.0558	-0.0100	0.0105	4,019
Father's employment: has a job	0.4290	0.4388	0.0098	0.0283	4,045
Father's employment: has a temp. job	0.1273	0.1530	0.0257*	0.0133	4,045
Father's employment: has no job	0.0355	0.0358	0.0002	0.0050	4,045
Mother's employment: has a job	0.3358	0.3715	0.0357	0.0255	4,045
Mother's employment: has a temp. job	0.1094	0.1251	0.0157	0.0113	4,045
Mother's employment: has no job	0.1282	0.1185	-0.0097	0.0122	4,045
Travel time from home to school (in min.)	34.9722	34.5635	-0.4087	0.9999	4,019
Number of people in the house	4.4453	4.4750	0.0297	0.0711	4,019
Has at home: newspapers	0.2182	0.2224	0.0042	0.0139	4,019
Has at home: magazines	0.3262	0.3337	0.0075	0.0166	4,018
Has at home: dictionary	0.8668	0.8636	-0.0032	0.0190	4,019
Has at home: books	0.8225	0.8040	-0.0184	0.0184	4,018
Has at home: scientific books	0.7640	0.7605	-0.0035	0.0159	4,019
Has at home: water supply	0.9712	0.9734	0.0023	0.0094	4,019
Has at home: sewage supply	0.8795	0.8893	0.0098	0.0198	4,019
Has at home: electricity supply	0.9734	0.9730	-0.0004	0.0067	4,019
Has at home: gas supply	0.2238	0.2346	0.0109	0.0224	4,018
Has at home: waste collection	0.9226	0.9247	0.0021	0.0111	4,019
Has at home: television	0.9620	0.9637	0.0016	0.0064	4,018
Has at home: radio	0.8135	0.8146	0.0011	0.0171	4,018
Has at home: bathroom	0.9241	0.9152	-0.0089	0.0124	4,018
Has at home: car	0.4838	0.4646	-0.0192	0.0263	4,017
Has at home: maid	0.0686	0.0795	0.0109	0.0085	4,018
Has at home: vacuum cleaner	0.3526	0.3424	-0.0102	0.0266	4,017
Has at home: washing machine	0.8547	0.8580	0.0033	0.0170	4,018
Has at home: DVD player	0.8655	0.8818	0.0163	0.0123	4,018
Has at home: refrigerator	0.9277	0.9313	0.0036	0.0099	4,018
Has at home: freezer	0.4785	0.4813	0.0028	0.0291	4,018
Has at home: telephone	0.7080	0.6659	-0.0421	0.0277	4,017
Has at home: computer	0.7611	0.7303	-0.0308	0.0322	4,017
Has at home: cable TV	0.5164	0.5225	0.0061	0.0352	4,018
Has at home: microwave	0.7707	0.7634	-0.0073	0.0219	4,018
Schools characteristics					
Computer lab	0.8947	0.9384	0.0437	0.0554	4,045
Science lab	0.3684	0.3023	-0.0661	0.1057	4,045
Library	0.1053	0.1006	-0.0047	0.0692	4,045
Internet	1.0000	0.9766	-0.0234	0.0347	4,030
School meals	1.0000	1.0000	0.0000	0.0000	4,045
Staff members	92.5263	75.2588	-17.2675**	7.1247	4,045
Number of school rooms in use	17.3684	15.4367	-1.9318	1.3975	4,045

Note: Levels of education are coded as low for parents with up to 8 years of education; mid for parents with secondary school or incomplete high education; and high for parents with complete high education. Employment situation is coded as 'has a job' if parents either have a job, or own a business, or are retired; 'temp. job' if they work independently doing some services, or only do temporary jobs; and 'no job' if they are unemployed.

Table A2: Attendance at Math and Language tests

	Attendance at Math test	*Attendance at Language test*
	(1)	(2)
Homicides	−0.005	−0.001
	(0.008)	(0.007)
	[0.008]	[0.007]
Observations	767,069	767,069
School/time fe	Yes	Yes
Controls	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Attendance at Math test* and *Attendance at Language test* indicate whether the student attended the respective exam or not. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table A3: Effect of exposure to violence in the school surroundings on standardised test scores - 100 meters radius

	Math proficiency					*Language proficiency*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Homicides	−1.478	−1.365	−1.364	−1.382	−1.459	−0.986	−0.737	−0.694	−0.719	−0.802
	(1.133)	(0.770)*	(0.756)*	(0.745)*	(0.749)*	(1.188)	(0.670)	(0.660)	(0.654)	(0.644)
	[0.895]*	[0.622]**	[0.612]**	[0.599]**	[0.601]**	[0.855]	[0.566]	[0.573]	[0.567]	[0.561]
Observations	666,718	666,718	666,718	666,718	666,718	666,453	666,453	666,453	666,453	666,453
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Teacher controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
School controls	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Classroom controls	No	No	No	No	Yes	No	No	No	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 100m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 100m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table A4: Effect of exposure to violence in the school surroundings on standardised test scores - levels of proficiency

	Math high level	*Math low level*	*Language high level*	*Language low level*
	(1)	(2)	(3)	(4)
Homicides	−0.005	0.022	−0.002	0.001
	(0.002)**	(0.012)*	(0.003)	(0.011)
	[0.002]***	[0.009]**	[0.003]	[0.010]
Observations	666,718	666,718	666,453	666,453
School/time fe	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Variables *Math high level* and *Language high level* are dummy variables indicating whether the students reach the 'advanced' level in these subjects. Variables *Math low level* and *Language low level* show if the student's test scores are considered in the 'below the basic' level in these subjects. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table A5: Effect of exposure to violence in the school surroundings on standardised test scores - heterogeneous effects by cohort and gender

	Math *5th grade* *(primary school)*		*Math* *9th grade* *(primary school)*		*Math* *3rd grade* *(secondary school)*		*Language* *5th grade* *(primary school)*		*Language* *9th grade* *(primary school)*		*Language* *3rd grade* *(secondary school)*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Homicides	−6.011	−1.872	−1.317	−1.799	−7.979	0.483	−5.707	0.982	0.547	−1.110	−2.245	−1.842
	(2.476)**	(2.024)	(1.468)	(0.692)***	(1.988)***	(3.295)	(3.018)*	(3.537)	(2.199)	(1.383)	(2.054)	(4.347)
	[2.036]***	[1.697]	[1.311]	[1.005]*	[1.856]***	[2.453]	[2.511]**	[2.793]	[1.674]	[1.446]	[2.617]	[3.120]
Observations	120,724	116,276	156,260	152,051	52,175	69,232	120,468	116,267	156,260	152,051	52,175	69,232
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* corresponds to the number of homicides within a 25m radius from school. Dependent variables *Math proficiency* and *Language proficiency* are Math and Portuguese standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table A6: Effect of exposure to violence in the school surroundings on attendance - heterogeneous effects by cohort and gender

	Absences 5th grade (primary school)		*Absences 9th grade (primary school)*		*Absences 3rd grade (secondary school)*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Homicides	0.014	0.011	0.022	0.017	0.004	−0.010
	(0.008)*	(0.006)*	(0.006)***	(0.003)***	(0.007)	(0.008)
	[0.006]**	[0.005]**	[0.006]***	[0.006]***	[0.006]	[0.007]
Observations	127,356	119,766	173,302	164,586	60,170	81,035
School/time fe	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides* correspond to the number of homicides within a 25m radius from school in the year. All regressions include time and school fixed effects. Controls include individual characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Table A7: Effect of exposure to violence in the school surroundings on standardised test scores - specific groups of victims

	Math proficiency			
	(1)	(2)	(3)	(4)
Homicides (all victims)	−2.349			
	(0.967)**			
	[0.808]***			
Homicides (18+ victims)		−2.419		
		(1.019)**		
		[0.848]***		
Homicides (male victims)			−2.761	
			(1.009)***	
			[0.782]***	
Homicides (gunshot victims)				−2.691
				(1.251)**
				[0.938]***
Observations	666,718	666,718	666,718	666,718
School/time fe	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the school level in parentheses. Conley standard errors computed at the 25m cutoff distance in brackets.

Note: The analysis includes students in 5th, 9th of primary school and 3rd grade of secondary school, over the period of 2010 to 2013. Explanatory variable *Homicides (all victims)* corresponds to the number of homicides within a 25m radius from school; *Homicides (18+ victims)* corresponds to the number of homicides within a 25m radius from school for which the victims are 18 or older; *Homicides (male victims)* corresponds to the number of homicides within a 25m radius from school for which the victims are males; *Homicides (gunshot victims)* corresponds to the number of homicides within a 25m radius from school for which the victims were gunshot; Dependent variable *Math proficiency* are standardised test scores normalised at a (250,50) scale. All regressions include time and school fixed effects. Controls include individual characteristics, teachers' characteristics, school characteristics and classroom composition. **Individual controls** are age, sex and race fixed effects, grade, dummies indicating whether at home the student has access to daily newspaper, magazines, dictionary, novels, poetry and short stories books and encyclopaedias; commuting time from residence to school; age, race, education and employment status of the father and the mother; income and number of people in the house; if parents own the house or rent it; if the house has supply of energy, water, gas, sewage and garbage collection; number of tv's, radios, bathrooms, cars, maids, vacuum cleaners, washing machines, dvd's, fridges, freezers, telephone, computers, cable tv, microwave and internet. **Teacher controls** are sex, age and race of the Portuguese and Math teachers. **School controls** are number of staff members, number of school rooms in use and dummies indicating whether the school has computer room, science lab, library, internet connection and if the school offers school meals. **Classroom controls** are share of black students, share of girls and share of students above the appropriate age.

Geographic coordinates and school-residence corridors

In order to define the measures of exposure to violence we use, it is necessary to geocode the addresses of the schools, residences and homicides. For the schools, we have available the precise address, including street and house number. For the residences, the street and house number are confidential information and cannot be accessed. However, we were granted access to the postcodes and neighbourhoods. In São Paulo, postcodes are quite small units and, in some cases, even more precise than the street names, as streets are typically broken up in several postcodes. For the homicides we also have the precise location for each case.

We used Google maps API to geocode the addresses. There are five possible geocoding outcomes, which vary depending on the amount of information used in the process: street, neighbourhood, municipality, state and not found. If the address is geocoded at the street level, it means that the returned result is a precise geocode, for which Google has information down to street address precision. When street level information is not available, the returned geocoded addresses are approximations, either interpolated between two precise points, or the geometric centre of a result such as a polyline (for example, a street) or polygon (region).

In our analysis, we use only returned addresses geocoded at the street level. Hence, even though we have different levels of information on the addresses of schools, residences and homicides, the geocoding accuracy level for all these three units is the street level. From the addresses we geocoded, 96 percent of the schools and 97 percent of the residences were geocoded at the street level, 95 percent of the homicides in the public way were also geocoded at the street level.

We also used Google maps API's to calculate the corridors from residence to school. We used Google Directions API and calculated path polylines of walking transport mode for each school/residence pair, which we call *Homicide Exposure Point (HEP)*. For each pair, we went through all the homicide points and calculated the nearest distance between a homicide and that particular polyline. We also calculated walking and straight distance from residence to school and from residence to the *HEP*.

In order to make those calculations feasible and limit the time necessary to run them, we defined some filter rules:

- Define the threshold distance between the homicide points and the path polylines to 500m.
- Ignore walking mode if straightline distance is greater than 15km;

- Define $double - distance = max(straightline - distance * 2, 500 * 2)$: if $double - distance$ is greater than 100km, ignore homicide point outside the circle with radius $double - distance/2$ and centre as the middle of straight-line between school and residence; if $double - distance$ is less than or equal to 100km, ignore homicide point if the straight-line distance between homicide point and either of school location and residence location is greater than the double distance.

To avoid billions of unnecessary API requests, straight-line distance calculations, distances along the path of walking distance transport mode polylines and nearest distance between homicide points to polylines were all calculated with Google's code without invoking Google API's. Overall, we used approximately two billion API requests to geocode our data and to generate the corridors for our analysis.