جامعة نيويورك ابوظبي
NYU ABU DHABI

MDPI

Article

On School Shootings, Racial Diversity & Academic Performance

Reem Hazim [1,‡], Alex Chae [1,‡], Kevishen Valeyatheepillay [1,‡], and Zachary Raye [1,‡]

[1] New York University Abu Dhabi
[‡] These authors contributed equally to this work.

Abstract: School shooting is a major social problem in the United States with wide-ranging implications. This study examines one such implication by investigating its relationship with academic performance. We first perform Coarsened-Exact Matching (CEM) to illustrate that compared to similar schools in the same school district, schools that experience shooting tend to have a greater decrease in academic performance after the shooting incident. Next, we delve deeper into the potential reasons behind the observed negative association of shooting and academic performance through a matching experiment. We demonstrate that high school students that enter schools that had a shooting before their enrollment tend to have lower academic performance in mathematics than students in the same schools before shooting, suggesting that shootings may be associated with a decrease in school educational quality. Lastly, we perform a regression analysis to show that among schools that had school shooting, an increase in racial diversity of students is associated with a larger decrease in academic performance after shooting. This study extends the literature of school shootings to conduct an analysis of schools across the entire United States over a decade and provides opportunities for further research into the impact of school shootings on education for different social groups.

Keywords: school shooting; academic performance; racial distribution; education; coarsened exact matching; regression

1. Introduction

In recent years, research has demonstrated that school shootings have a profoundly detrimental effect on students' well-being and can lead to diminishing academic performance [1]. In the USA, school shootings have only become more and more frequent over the past decade [2]. These worrisome findings have sparked an urgent need to delve deeper into the nuanced dynamics at work within this critical issue. While previous studies have made substantial contributions, a notable limitation is that many have focused on smaller data sets, primarily examining the impact of school shootings within a single state by analyzing only a select number of shooting incidents, such as the study performed on students related to Connecticut's Sandy Hook shooting [3]. Consequently, a comprehensive understanding of how these tragic events affect students nationwide, across varying school settings, remains elusive.

To address this literature gap, our research explores the multifaceted relationship between school shootings and students' academic performance on a broader scale. We recognize that the negative association of such events is a confirmed reality, but we seek to discern whether this association remains consistent across all schools or if some educational institutions are disproportionately affected. Furthermore, our investigation recognizes the existing disparities in access to quality education, with students of color facing barriers compared to their white peers. Our research rigorously examines the relationship between school shootings and the academic performance of students from different racial backgrounds, shedding light on how these tragic incidents interact with existing inequalities.

Citation: Hazim, R.; Chae, A.; Valeyatheepillay, K. ; Raye, Z. New York University Abu Dhabi. *Computers* **2023**, *1*, 0. https://doi.org/

Received: December 10, 2023
Accepted: December 10, 2023
Published: December 10, 2023

By investigating these critical questions, our research aspires to provide a more comprehensive and nuanced understanding of the relationship between school shootings and students' academic performance. Ultimately, this knowledge can inform policy, intervention strategies, and support systems to better address the needs of students in the aftermath of these traumatic events and work towards a safer and more equitable educational environment for all.

2. Literature Review

Researchers have studied extensively the effect of school shootings on many aspects of students' lives, including their mental health, their academic performance, their school's fiscal spending, their graduation outcomes, and their enrollment rates.

Some interdisciplinary studies have highlighted the neurological and physiological mechanisms by which exposure to violence in general can impact young people, affecting their biological stress systems and developing brains [3–7]. Suomalainen et al. have found that the likelihood of students developing mental disorders like PTSD post-shooting incidents increases with exposure to violence. Focusing on school shootings, Rossin-Slater et al. employs a difference-in-differences approach to examine the number of antidepressant prescriptions to youth in the area surrounding a school that recently experienced a shooting. They found a significant increase in antidepressant prescriptions and absenteeism among students post-shootings, along with a decrease in high school graduation and college enrollment rates.

Moreover, school shootings have been linked to adverse educational outcomes for students, which varies based on factors like grade and location. For instance, Beland and Kim noted a decrease in 9th-grade enrollment and test scores following shootings, with no significant effect on graduation outcomes, attendance, and suspensions. Cabral et al. found that students exposed to school shootings in Texas are more likely to be chronically absent, less likely to graduate high school and less likely to have jobs by the age of 24. These effects were more pronounced for non-Hispanic students, Black students, and students receiving free or reduced-price lunch [8].

Yang and Gopalan use difference-in-differences and event study approaches to quantify the effect of school shootings on school finance and student composition. They find that School shootings are associated with an increase school spending and a decline in student enrollment.

Case studies, such as Gershenson and Tekin's analysis of the Beltway Sniper Attacks, reveal that such traumatic events significantly reduce school proficiency rates, particularly in schools with higher proportions of minority and socioeconomically disadvantaged students. In addition, they examine how factors like race and socioeconomic status relate to academic performance after a school shooting [10]. However, their dataset is limited to one state and their scope is focused on the community of schools near the school where a shooting took place rather than the shooting itself.

Studies have established persistent gaps in educational outcomes among different racial and ethnic groups [11]. Although there are signs of the gaps narrowing in achievement and attainment, significant disparities remain, especially for less advantaged groups like African Americans, Hispanics, and Native Americans compared to more advantaged groups such as whites and Asian Americans. This research underscores the importance of considering pre-existing disparities when examining the effects of school shootings in a broader context.

While there is substantial evidence of the negative impact of school shootings on various aspects of student life, there remains a gap in comprehensive, nationwide data that considers the varied correlations based on racial composition. Our study aims to address this gap, contributing to a more inclusive understanding of the issue.

3. Methodology

3.1. Data

This project utilizes 2 main datasets to study the interaction between school shootings and academic performance: a dataset of school-level academic performance and a dataset of school shootings. We also use datasets on poverty rate and crime rate for specific parts of our study.

3.1.1. Academic Performance Data for U.S. Schools

We retrieved data on school-level academic performance from the EDFacts dataset published by United States Department of Education [1]. As part of EDFacts, an initiative to "collect, analyze, and promote the use of high-quality, pre-kindergarten through grade 12 data," the Department of Education publishes data on standardized assessments every academic year [12]. At the time of study (September 2023), there was data available for every academic year from 2009-2010 to 2020-2021. The data includes the Mathematics and Reading and Language Arts (RLA) assessment results for grades 3 to 8 and high school for every public school in the United States, where each school is identified via a unique school ID. The data set also provides information on the number of students who took the assessment each year overall and by racial groups.

Our main dependent variable is the proficiency rate of students in Mathematics and RLA. Since each state has different school examinations, the raw test scores for schools in different states cannot be compared directly. Therefore, EDFacts provides proficiency rates as a standardized indicator of academic performance across states. Proficiency is the number of students who passed the state exam in that subject math or RLA), and it is determined by each state based on the students' test scores. Proficiency rate, which is provided for each school by the EDFacts dataset, is calculated as the number of students who attained proficiency divided by the number of students who completed the state assessment and for whom a proficiency level was assigned.

We manually downloaded the Assessment Proficiency data files for each academic year from from 2009-2010 to 2020-2021 and merged the data sets using a custom Python script. While the proficiency rate was mostly an integer value representing the percentage of the student body that passed the assessment, some schools chose to report a range of values to preserve anonymity (e.g. <50%, 50%-75%). In such cases, we used the median value of the range as the proficiency rate for that school. Proficiency rate and academic performance will be used interchangeably in this paper.

3.1.2. U.S. School Shootings Data

We retrieved data on school shootings in public schools from the dataset compiled by The Washington Post [2]. The dataset contains 392 acts of gunfire in primary and secondary schools across the United States. The data was collected from news articles, open-source databases, law enforcement reports, school websites and direct calls to schools and police departments. Among more than 1,000 incidents that were found, the dataset only includes acts of gunfire that happened on school campuses immediately before, during or just after classes and those that caused direct threat to the children in schools primary and secondary schools. The data set has been updated to include further incidents of gunfire since its original release on April 20, 2018 and now contains school shooting incidents from 1999 to 2023. The data set includes information on the school name, educational district, date and time of shooting, number of student enrollment at the time of shooting and aggregated demographic information about the students at the time of shooting. Each school can be identified via a unique school ID, assigned by each school district. The yearly distribution of school shooting incidents is shown in Figure 1a.

[1] The datasets are available here: https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html
[2] The dataset is publicly available here: https://github.com/washingtonpost/data-school-shootings

The creators of the Washington Post school shootings dataset manually annotated the academic year (e.g. 2009-2010) in which each shooting occurred. We used this academic year and the school ID to merge the school shootings dataset with the EDFacts academic performance dataset. Each school in the school shootings dataset was matched to its math and RLA results in the academic year when the shooting occurred. We assume that most states administer their public school assessments at the end of the academic year, between April and May. Therefore, it is highly likely that the students would have taken the state exam after the shooting occurred, and the effects of the shooting would be reflected in the assessment results of that academic year. However, there are very few states that administer their exams halfway through the academic year. We do not account for these states since they are very few, and this is one limitation of our study.

Since we only have academic performance data for the years between 2009 and 2021, we only kept the shootings that happened in that time period. Therefore, we were left with 208 school shootings out of the original 392. We further filtered the data to remove any schools that did not have test result data for the academic year when the shooting happened. We were finally left with 105 school shootings. Figures 1 and 2 display the geographic and temporal distribution of our filtered school shootings dataset.

Figure 1 suggests that most states had less than 6 shootings throughout this time period. Notably, Texas, Florida, and North Carolina had far more shootings than the rest of the states, with North Carolina recording the highest number of shootings (12).

Figure 2 reveals that the years 2014 and 2018 had the highest number of shootings within this time period. Most shootings occurred between September and May, which coincides with the academic year, but there are a few shootings that happened during the months of June and August as well.

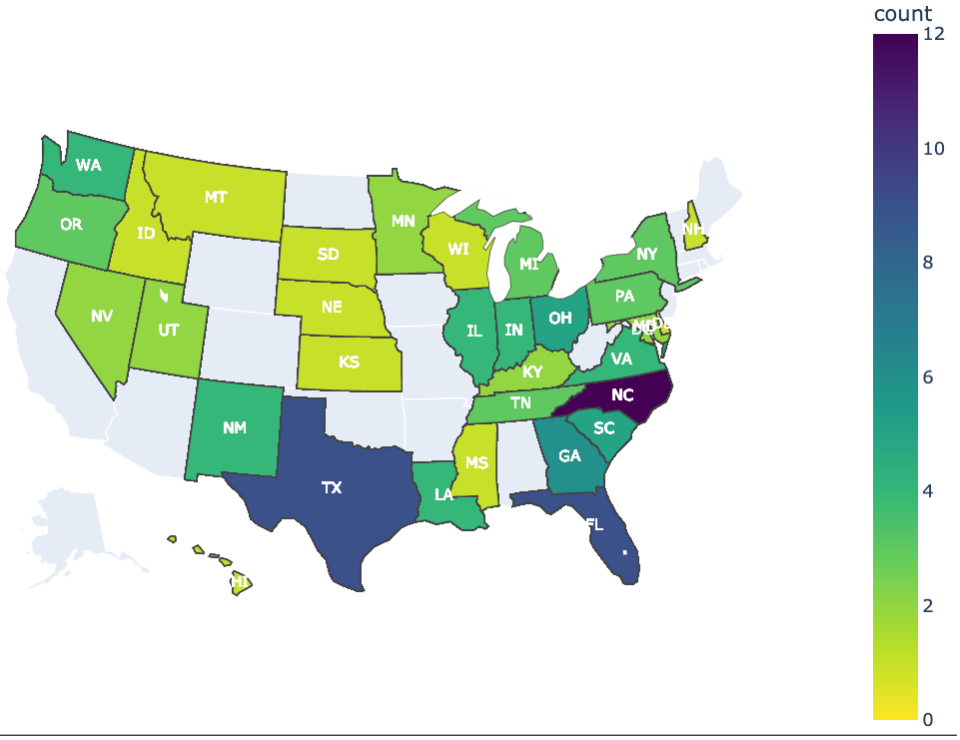
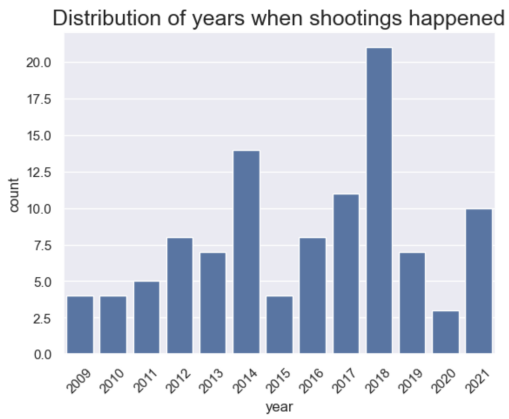
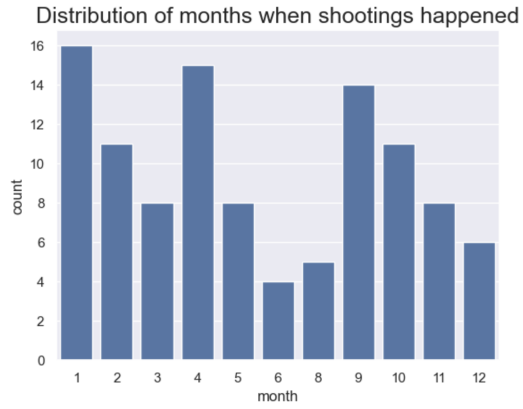


Figure 1. Geographic distribution of schools where shootings occurred across the United States from the filtered Washington Post data set.



(a) Distribution of school shootings in The Washington Post dataset by year.



(b) Distribution of schools shootings in The Washington Post dataset by month.

Figure 2. Temporal distribution of school shootings from The Washington Post dataset.

3.1.3. Crime and Poverty Data sets

For the matching experiment in Section 4 section of the paper, we introduce two control variables. The first variable is poverty rate by district. We downloaded the poverty rate by school district in from 2012 to 2021 from the United States Census Bureau [3]. The data set provides information on the estimated population of children of age 5 to 17 in each school district, along with information on the estimated number of children in that age group who are in poverty. Poverty rate was calculated as the number of children of age 5 to 17 in poverty divided by the total number of children in that age group.

The second control variable used in the matching experiment is crime rate. Due to lack of availability of data of crime rate by school district, this study used crime rate by each state to control for the level of crime in each school each year. The study uses crime rate data provided by the Federal Bureau of Investigation (FBI) [4]. The FBI publicly reports the number of crime reported per 100,000 people in each state, where crime includes all violent crime, such as homicide, rape, robbery and aggravated assault.

3.2. Roadmap

This section provides an outline of the methodological approaches used in this paper to answer the research question. First, the researchers conducted a preliminary analysis of

[3] The raw data for poverty rates can be found here https://www.census.gov/programs-surveys/saipe/data/datasets.2009.List_1743592724.html#list-tab-List_1743592724

[4] The crime rate data can be found here https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads

the academic performance trends over time between schools with a shooting and schools without a shooting, before performing a more robust coarsened exact matching technique of schools that had a shooting in a given year to similar schools that did not have a shooting in the same year. The purpose of this matching was to determine whether the shooting is associated with a decrease in academic performance of the school. Findings demonstrate that schools that experienced a shooting had a larger decrease in academic performance the year after shooting, compared to the change in proficiency rate of similar schools in the same time period. This suggests that school shootings are associated with a decrease in students' academic performance.

Next, as an attempt to delve deeper into the potential reason behind the negative relationship between school shootings on academic performance, the study conducted a matching experiment to get insights into whether shooting is associated with a negative influence to the students themselves who firsthand experience shooting, or is associated with a decrease in the educational quality of the school. By comparing the proficiency rates of students in before the shooting with that of students who entered the same school after the shooting happened, the study provides suggestive evidence that shootings are linked to a decrease in the school's educational quality for high schools in mathematics, thus associated with a decrease in the academic performance of future students in the same schools.

Lastly, a regression analysis was conducted to examine the relationship between the racial distribution of the school's student body and the negative relationship between school shootings and academic performance. The analysis concludes with a finding that for RLA assessment scores, an increase in the student body racial diversity is associated with a decrease in test scores, while the results were inconclusive for Math assessment scores.

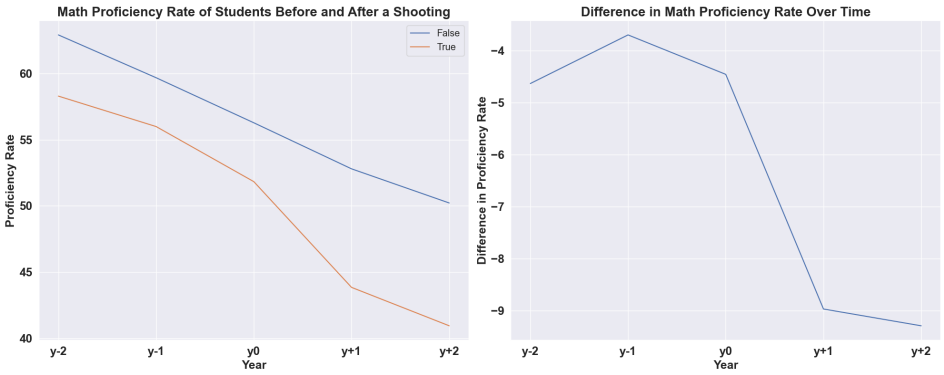
4. Results

4.1. Preliminary Insight into Proficiency Rate Trends

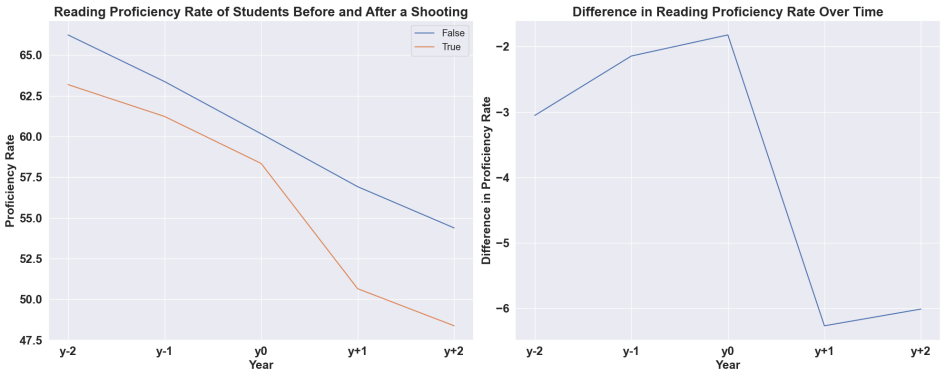
In this study, we sought to preliminarily examine the correlation between school shootings and academic achievement, specifically in mathematics and reading, prior to applying Coarsened Exact Matching. To accomplish this, we analyzed the average proficiency rates in these subjects over a designated time frame in relation to the occurrence of shootings. This analysis involved a comparative approach: for a school experiencing a shooting in a given year (denoted as year 0), we recorded its math and reading proficiency rates for two years preceding and following the incident (years -2, -1, 0, +1, +2). Similarly, we tracked the proficiency rates for schools within the same district that did not experience shootings during this period. As depicted in Figure 3, a trend emerges: although academic performance generally declines over time in both sets of schools, those that experienced a shooting exhibit a more pronounced decline in proficiency rates compared to their counterparts without such incidents. With this correlation, we sought to accomplish a more robust analysis with coarsened exact matching.

4.2. The Correlation Between School Shooting and Academic Performance

Our first experiment tests the hypothesis that school shootings are correlated with worse academic performance. We test this hypothesis by matching schools where a shooting occurred (treatment schools) with other schools in the same district that did not experience a shooting (control schools) in the same year. We then determine whether the proficiency rate of a certain grade (e.g. grade 5) in the treatment school decreased more after the shooting occurred than the proficiency rate of that same grade in the control schools. Our dependent variable in this analysis is the difference between proficiency rate of a specific grade in the year before the shooting (t-1) and the year of the shooting (t0) ($\Delta prof_{t-1}$). We also repeat this analysis for 2 other dependent variables: the difference in proficiency between the year of the shooting (t0) and the year after the shooting (t+1) ($\Delta prof_{t+1}$), and the difference in proficiency between the year of the shooting (t0) and two years after the shooting (t+2) ($\Delta prof_{t+2}$). To perform this analysis, we filter our dataset to include only the



(a) Change in Math Proficiency Rates Over Time between Schools with Shooting and Schools Without Shooting.



(b) Change in Reading Proficiency Rates Over Time between Schools with Shooting and Schools Without Shooting.

Figure 3. Change in Proficiency Rates Over Time

schools where test score data was available for the year before the shooting (t-1), the year of the shooting (t0), the year after the shooting (t+1) and two years after the shooting (t+2). We were left with 43 shooting schools after this filter.

We use Coarsened Exact Matching (CEM) to match the treatment and control schools into bins. CEM is a technique invented by Iacus et al. to reduce the imbalance between the control and treatment groups. We rely on 5 control variables to match the treatment and control schools. First, we bin together schools in the same district to control for hidden confounders such as crime rate and poverty rate. Second, we match schools that had similar proficiency rates in the year before the shooting (t-1) as a proxy for school quality. We assume that schools that had similar test scores before the intervention would have a similar quality of students, and binning them together would account for this confounder. Since proficiency rate is a continuous variable, we divide it into 5 quantiles and match the control and treatment schools that fall in the same quantile. Third, we match the data points by grade to avoid comparing the proficiency rate of students in different grades (e.g. 4th graders versus high schoolers). Fourth, we match based on academic year to ensure that the academic performance of the students was measured in the same year for all the schools in that bin. Finally, we match schools based on number of test-takers as a proxy for school size.

We use the CEM Python package provided by Iacus et al. to bin the control and treatment schools and reweigh the $\Delta prof$ values. [5] We then compare the average weighted $\Delta prof$ of control and treatment schools using Welch's t-test. Figure 4a shows the result

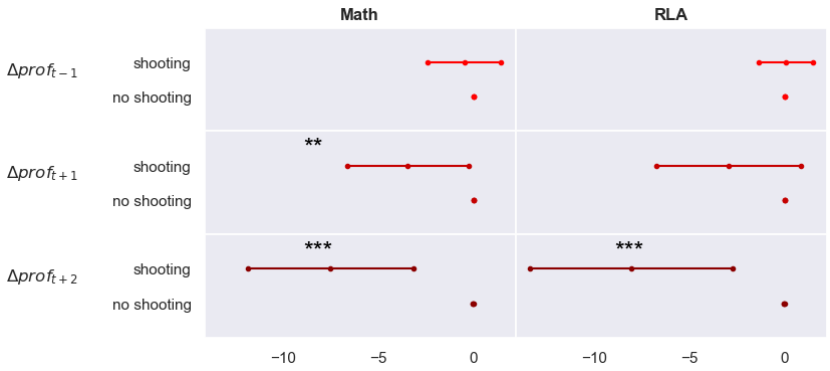
[5] The CEM Python package can be found here: https://github.com/lewisbails/cem.

of this analysis. There was no significant difference in $\Delta prof_{t-1}$ (difference between proficiency in the year of the shooting and the year before) between treatment and control schools for both math and RLA. One possible reason is that the shootings might have taken place after the state exam was administered in that year, so the effect of the shooting did not show in the proficiency rates for year 0. However, we found a significant difference in the math $\Delta prof_{t+1}$ between treatment and control schools. Schools that did not experience a shooting had almost no change in math proficiency rate between t0 and t+1 ($\Delta prof_{t+1} = 0$), whereas schools that did experience a shooting saw a significant drop in the year after the shooting ($\Delta prof_{t+1} < 0$). We also found that treatment schools witnessed a larger decrease in both math and RLA proficiency two years after the shooting ($\Delta prof_{t+2} < 0$). These results are consistent with other studies [1] that find that students' academic performance tends to decrease after a shooting.

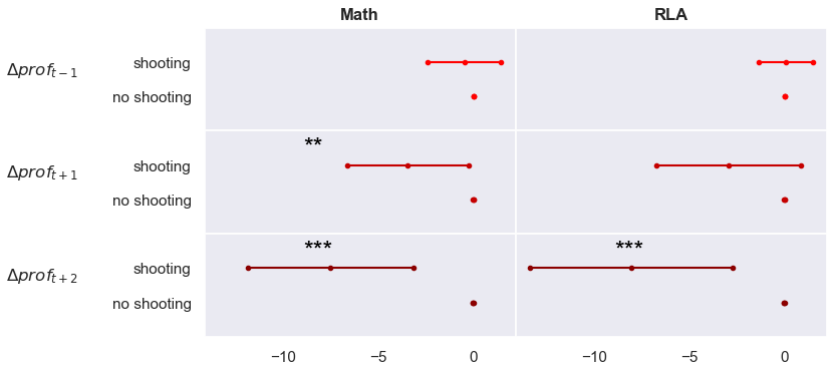
Since our filtered dataset only includes 43 treatment schools, every treatment school is in a unique school district, and each bin ended up having only one treatment school and several dozen control schools. We wanted to test if the effect we found would still hold even if we made a one-to-one matching of treatment and control schools. Therefore, we chose a random control school from each bin and matched it to the treatment school in that bin. We then ran the Student's t-test to compare the mean $\Delta prof$ for treatment and control schools. The results are shown in figure 4b. We found that our conclusion holds even after performing a 1-to-1 matching between treatment and control schools.

4.3. The Correlation Between School Shootings and School Quality

After confirming an association between school shooting occurrence and a larger decrease in academic performance, the matching experiment aims to delve deeper into the potential reasons behind the negative association between school shootings and students' academic performance. There are two potential reasons we want to test. First, shootings may have impacted the students who experienced the shooting firsthand. In other words, the negative association between shooting and academic performance may be derived from the mental trauma that students went through. Second, shootings may be associated with a decrease in the educational quality schools that experienced the shooting. To isolate the effect of the two explanations, we examined the academic performance of students who consequently enrolled in the schools a few years after the shooting incident occurred. If the first explanation that shooting is associated with a negatively influence on the student themselves is a stronger force, then we should not see a clear difference in the distribution of proficiency rate of students right before the shooting and that of students who entered the school a few years after the shooting and did not experience shooting firsthand. In contrast, if the second explanation that shooting is associated with a decrease in the school's educational quality is the stronger explanation, we would see a clear difference in the aforementioned two distribution of proficiency rates, as consequent students are receiving lower quality of education. As the proficiency rate data set differentiates elementary schools from high schools, we conducted separate matching experiments by type of school.



(a) The average change in math and RLA proficiency rate ($\Delta prof$) for treatment and control schools. We weighted the $\Delta prof$ values for control and treatment schools using CEM and compared them using Welch's t-test. Welch's t-test is a statistical test that can be used to compare the means of two groups with different variances. We use this test because the number of control schools was much larger than the number of treatment schools (only 43). The lines show the 95% confidence interval. The different colors are intended for clarity purposes only. (*) $p < 0.1$, (**) $p < 0.05$, (***) $p < 0.01$



(b) The average change in math and RLA proficiency rate ($\Delta prof$) for treatment and control schools after matching each treatment school with one random control school that falls in the same bin (has the same district, grade, academic year, baseline proficiency rate from the previous year, and number of test-takers). We do not use CEM for this analysis as the number of treatment and control schools are equal and there is no imbalance. We compare the means using Student's t-test for independent groups. The lines show the 95% confidence interval. The different colors are intended for clarity purposes only. (*) $p < 0.1$, (**) $p < 0.05$, (***) $p < 0.01$

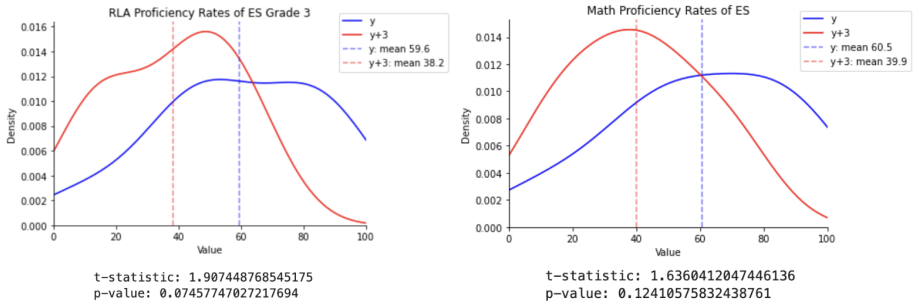
Figure 4. Results of the CEM Experiment.

4.3.1. Elementary Schools

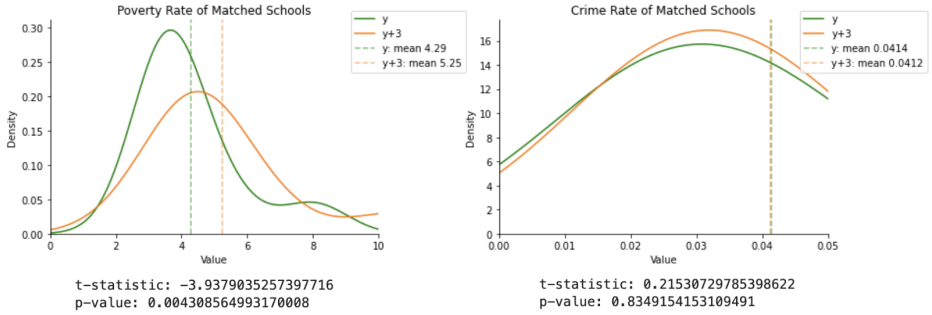
For elementary schools, our data set includes proficiency rate of grade 3 - 8 students for each academic year. We focus on the results for grade 3 students. If y represents the year in which shooting occurred at a particular school, the students that were in grade 3 in year y would be in grade 6 in year y+3. Moreover, in year y+3, the students in grade 3 would have joined the school in year y+1, and hence would not have experienced the shooting firsthand. Following this logic, we collect the student proficiency rate for grade 3 in y and y+3 and compare their distributions.

However, a cautionary note is required prior to presenting the results. After a filtering process of selecting 1) elementary schools, 2) whose shootings happened in or after 2009, 3) whose shootings happened in or before 2018, as the proficiency rate data set ranges until 2021, 3) whose data entries could be matched between the proficiency rate and school shootings data sets, we were left with only 9 schools to examine for RLA proficiency rate and 8 schools to examine for Mathematics proficiency rate. Hence, the analysis presented

here severely lack statistical power to draw out any conclusive evidence. Nonetheless, we present the result achieved in our study to aid future research on a similar topic.



(a) Distribution and Student's t-test analysis result for Grade 3 proficiency rates in the year of shooting and 3 years after the year of shooting



(b) Distribution and dependent t-test analysis result for control variables in the year of shooting and 3 years after the year of shooting

Figure 5. Matching experiment for elementary schools

Figure 5a shows the distribution of Grade 3 RLA and Math proficiency rates in the year of shooting and 3 years after the shooting. A Student's t-test was conducted to examine whether the two distributions are different with statistical significance (H_0: There is no significant difference in the proficiency rate distribution in year y and year y+3, H_a: There is a significant difference in the proficiency rate distribution in year y and year y+3). For RLA, we found significantly lower proficiency rates in year y+3, whereas the difference was not significant for the math proficiency scores. This result suggests that school shootings were associated with a decrease in the school's quality of education for RLA and hence is associated with a negative influence on students who joined the school three years after the shooting.

As a robustness check, we considered two variables other than school shooting that may have caused a change in student academic performance between the three years. Building on previous literature suggesting that poverty rate and crime rate may impact students' academic performance [4,7,14,15], we examine whether there were any significant changes in poverty rate and crime rate in the school districts in year y and year y+3 through a dependent t-test (H_0: There is no significant difference in the poverty/crime rate distribution in year y and year y+3, H_a: There is a significant difference in the poverty/crime rate distribution in year y and year y+3). The results in Figure 5b demonstrate that poverty rate significantly increased from year y to year y+3, while crime rate did not have a significant change. Therefore, the decrease in RLA proficiency rate from y to year y+3 that was observed in Figure 5a cannot be solely attributed to school shootings. Hence, the

matching experiment for elementary schools is inconclusive, due to potential confounding variables.

4.3.2. High Schools

For high schools, our data set includes an aggregated proficiency rate of all high school students in a school, rather than the proficiency rate of each individual grade. If y represents the year in which shooting occurred at a particular school, the students that were in grade 9 in year y would have graduated in year $y+4$. Moreover, in year $y+4$, the students in grade 9 would have joined the school in year $y+1$, and hence would not have experienced the shooting firsthand. Therefore, the proficiency rate result in $y+4$ reflects the academic performance of students that have not experienced the shooting firsthand. Following this logic, we collect the student proficiency rates in y and $y+4$ and compare their distributions.

After a similar filtering process as the one for elementary schools, selecting 1) high schools, 2) whose shootings happened in or after 2009, 3) whose shootings happened in or before 2017, 3) whose data entries could be matched between the proficiency rate and school shootings data sets, we were left with 41 schools to examine for RLA proficiency rate and 39 schools to examine for Mathematics proficiency rate.

Figure 6a and 6b show the distribution of high school RLA and Math proficiency rates in the year of shooting and 4 years after the shooting. A similar Student's t-test as before was conducted to examine whether the two distributions are different with statistical significance (H_0: There is no significant difference in the proficiency rate distribution in year y and year $y+4$, H_a: There is a significant difference in the proficiency rate distribution in year y and year $y+4$). For RLA, we did not find any significant difference in proficiency rate in the two years. However, for Mathematics, we did find significantly lower proficiency rates in year $y+4$. This result suggests that school shootings were associated with a decrease in the school's quality of education for Mathematics and hence is associated with a negative influence on students who joined the school four years after the shooting.

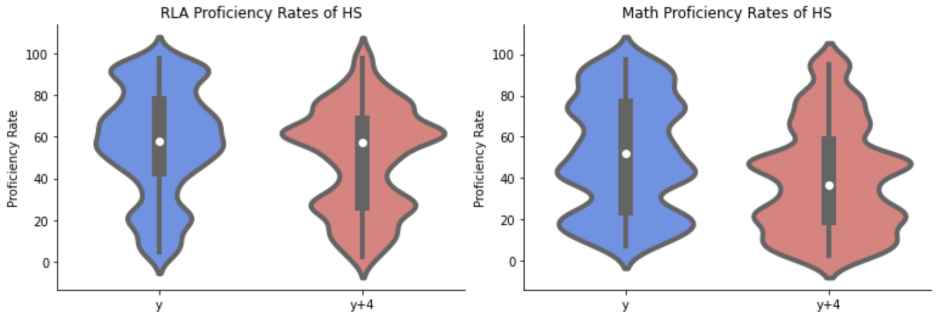
We performed a similar robustness check for the observed result for high schools by examining the change in poverty rate and crime rate in the same time period through a paired t-test (H_0: There is no significant difference in the poverty rate/crime rate distribution in year y and year $y+4$, H_a: There is a significant difference in the poverty rate/crime rate distribution in year y and year $y+4$). The results in Figure 6c demonstrate that poverty rate nor crime rate had a significant change from y to year $y+4$. The p-values from the paired t-tests were both insignificant, suggesting that the two control variables did not change. Therefore, we have suggestive evidence that school shootings is associated with a decrease in the school's educational quality for Mathematics, which is associated with a decrease in student academic performance for Mathematics.

To conclude, the results of the matching experiment suggests that high school shootings may have been associated to a decrease in school's educational quality for Mathematics, which led to a decrease in student academic performance even after the shooting incident, in the absence of a significant change in other factors such as poverty rate or crime rate. This provides an insight into the potential mechanisms through which school shootings have a negative relationship with academic performance.

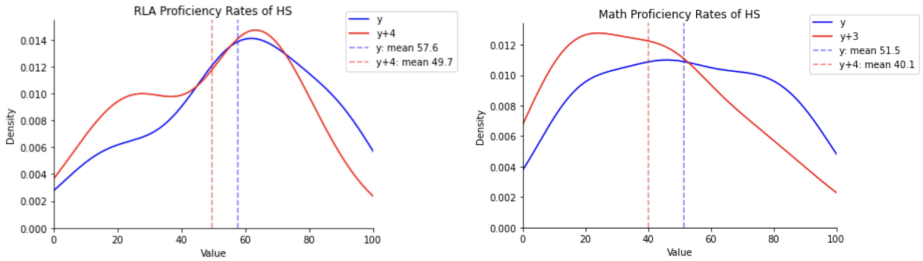
4.4. Relationship between Student Racial Distribution and Degree of Negative Influence of Shooting

To investigate the potential relationship about whether the racial distribution of students determine how much schools are affected by school shootings, we adopt a regression analysis.

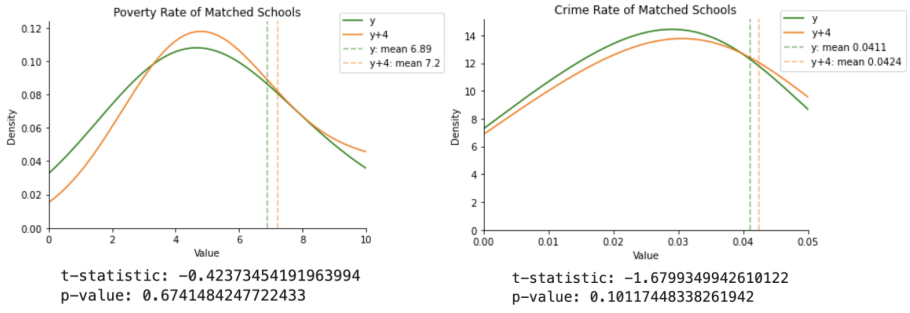
We want to lay emphasis to how we narrowed down our population of interest to only schools that had shooting to provide a simpler and more insightful regression. Here we run two separate regression - one having the mean RLA scores as dependent variable and the other one looking at mean Math score. Since we have a panel data, we also have a dummy variable - *After* - which takes the value 1 for the school's period after which a



(a) Distribution of RLA and Mathematics proficiency rates in year y and year $y+4$ as violin plots



(b) Distribution and Student's t-test result of RLA and Mathematics proficiency rates in year y and year $y+4$



(c) Distribution and dependent t-test analysis result for control variables in the year of shooting and 4 years after the year of shooting

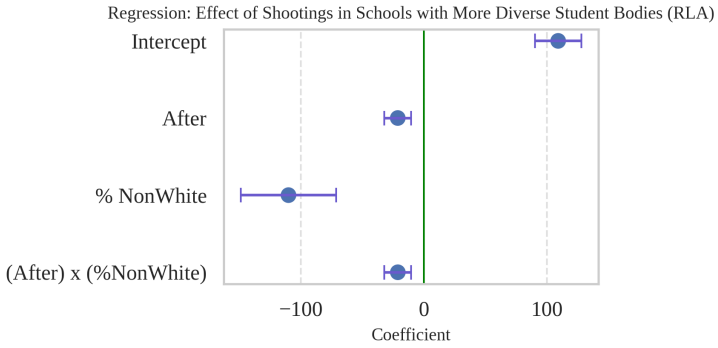
Figure 6. Matching experiment for high schools

shooting happened. We also include a variable called *%NonWhite* which represents the percentage of non-white students at a school at any time t. Briefly, this variable is calculated as number of non-white students that took the assessment in a given academic year, divided by the total number of students who took the assessment that year. To control for school district inherent differences and characteristics - that can be numerous, at the same time non-observable and hard to control for - we also decided to add a school district fixed affects (in the form of the categorical variable called *SchoolDistrict*).

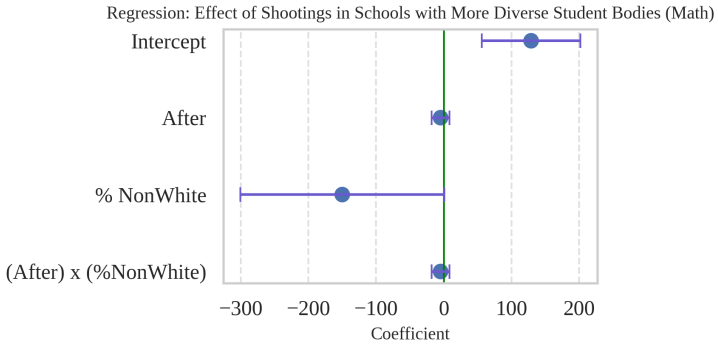
The regression model we use is as follow:

$$TestScores = \alpha + \beta_1(After) + \beta_2(\%NonWhite) + \beta_3(After).(\%NonWhite) + \beta_4(\%NonWhite) + \gamma(SchoolDistrict) + \epsilon$$

We again want to lay emphasis here that the variable *TestScores* is used to both mean RLA test scores and mean Math test scores. To interpret this regression, we need to emphasize mainly on the coefficient β_3. This coefficient shows the incremental effect of higher student body diversity (higher percentage non-white students) for schools that have had shooting after the shooting happened (*After* = 1). This is our coefficient of interest. Moving on to our estimation of the regression



(a) Using mean RLA scores as dependent variable



(b) Using mean Math scores as dependent variable

Figure 7. Coefficients of the two regressions. Confidence intervals including 0 imply insignificance

From Figure 7a, our first regression using mean RLA score as dependent variable, we find that our coefficient β_3 is equal to -21.13 and it highly significant to the 1% level with a p-value of 0.0 and a t-score of 5.5. This can be interpreted as: for schools that had shootings, for the period after the shooting, a 1% increase in the school's student body racial diversity tend to cause a decrease of 21.13 in the mean RLA test scores, all other things equal. It is also not surprising to note that both the coefficients of *After* and *%NonWhite* are significant to the 1% level, the first one of which is as was expected.

Moving on to the interpretation of Figure 7b - our regression with Math scores - we find that most of our coefficients are insignificant, with only *%NonWhite* being significant to the 10% level. Nevertheless, the coefficients for *After*, *%NonWhite*, and *(After).(%NonWhite)* are of the magnitude of -4.74, -150.13, and -4.73 respectively. This is a stark contrast with the effect we just found with the mean RLA scores.

The regression tables can be found in Appendix A.1.

5. Discussion

5.1. Conclusion

We first used Coarsened Exact Matching (CEM) and 1-to-1 matching techniques to match schools that experienced a shooting in a given year with similar schools that did not have a shooting. We found that both math and reading proficiency in shooting schools dropped significantly more than non-shooting schools two years after the shooting, and our results confirm previous studies [1]. Secondly, to identify whether shootings are correlated with a decrease in the quality of the educational institution, we conducted another matching experiment. Here we found that for high schools, math proficiency rate for students joining four years after the shooting was lower than the math proficiency rate right before shooting happened. Concerning RLA proficiency, we got insignificant results from matching the same group. When it comes to elementary schools, our results are inconclusive due to our very small sample size and potential confounders that we identified (such as poverty rate). Finally, we performed a regression analysis to examine whether a correlation exists between the racial diversity of the student body and the effect of school shootings on academic performance. We found that when it comes to RLA proficiency rates, treatment schools that had a more diverse student body after the shooting happened tend to have lower proficiency rate and this result was statistically significant at the 1% level. On the other hand, when it comes to math proficiency scores, our regression analysis did not find any statistically significant relation between the proficiency score and the racial diversity of the student body of schools affected with shootings.

5.2. Limitations

Our study is not without limitations. First, the sample size from which the study was conducted was small and hence cannot provide high statistical power. After merging the two main data sets on school shooting and academic performance by school, we were left with 105 schools. Only 43 of these schools were used in the first experiment (Section 4.2) due to the availability of academic performance data in the year of shooting, one year before shooting, one year after shooting and two years after shooting. For the school quality experiment (Section 4.3), to get insights on the potential reasons of the negative relationship between shootings and academic performance, the sample size for elementary schools (n = 8) was too small to draw a meaningful insight. Even when it comes to our conclusion concerning high schools, our sample size was only 39. Hence, a study with a bigger sample size may find more statistically significant result that may bring another layer of conclusions.

Furthermore, in the matching analyses, we assume that students took the standardized assessment used to compute proficiency rate in our data set after the shooting happened. Hence, we assume that in academic years during which a shooting happened, students took the assessment test after the shooting happened. While this is likely the case for most of the assessments that were administered, there may be exceptions for schools that administer

assessment tests towards the beginning of the academic year. This information was not readily available for every school in the EDFacts dataset that was used for this study. Hence, with more precise data on the assessment dates for each school, future studies could yield more accurate results.

Moreover concerning our regression model, problem of endogeneity can also be present. There may exists other potential controls that form part of the true model that we did not account for here. In addition, future studies can try to collect data for each grades and for each year after the shooting. This would allow us to have separate result for different grades for both Math and RLA separately.

Nonetheless, this study provides meaningful insights into the detrimental influence of school shootings on students across the United States. Our results suggest that school shootings are linked to a decrease in student academic performance and that the effect is more pronounced for schools with more non-white students. These results indicate an urgent need for policies that improve and expedite the recovery process of schools from a shooting incident to mitigate any impact on education quality. Furthermore, schools with a higher percentage of non-white students may require more support than other schools because they experience a larger decline in academic performance.

Appendix A

Appendix A.1

Regression tables for analysis on section 3.3.

OLS Regression Results

Dep. Variable:	pct_mean	R-squared:	0.914
Model:	OLS	Adj. R-squared:	0.833
Method:	Least Squares	F-statistic:	11.36
Date:	Thu, 07 Dec 2023	Prob (F-statistic):	3.27e-21
Time:	10:16:38	Log-Likelihood:	-480.89
No. Observations:	144	AIC:	1102.
Df Residuals:	74	BIC:	1310.
Df Model:	69		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	109.2681	9.478	11.529	0.000	90.383	128.153
after	-21.1896	5.511	-3.845	0.000	-32.170	-10.209
per_nonwh	-110.0575	19.496	-5.645	0.000	-148.903	-71.212
after_per_nonwh	-21.1349	5.497	-3.845	0.000	-32.087	-10.183

Omnibus:	62.950	Durbin-Watson:	1.999
Prob(Omnibus):	0.000	Jarque-Bera (JB):	651.571
Skew:	-1.187	Prob(JB):	3.26e-142
Kurtosis:	13.147	Cond. No.	2.12e+16

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.22e-31. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

(a) Using mean RLA scores as dependent variable

OLS Regression Results

Dep. Variable:	pct_mean	R-squared:	0.844
Model:	OLS	Adj. R-squared:	0.733
Method:	Least Squares	F-statistic:	7.591
Date:	Thu, 07 Dec 2023	Prob (F-statistic):	3.33e-19
Time:	10:17:09	Log-Likelihood:	-636.95
No. Observations:	167	AIC:	1414.
Df Residuals:	97	BIC:	1632.
Df Model:	69		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	129.0891	36.681	3.519	0.001	56.288	201.890
after	-4.7406	6.587	-0.720	0.473	-17.814	8.332
per_nonwh	-150.1292	75.930	-1.977	0.051	-300.829	0.571
after_per_nonwh	-4.7284	6.570	-0.720	0.473	-17.768	8.311

Omnibus:	31.136	Durbin-Watson:	2.012
Prob(Omnibus):	0.000	Jarque-Bera (JB):	76.519
Skew:	0.773	Prob(JB):	2.42e-17
Kurtosis:	5.934	Cond. No.	2.55e+16

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.33e-31. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

(b) Using mean Math scores as dependent variable

Figure A1. OLS Regression. Tables generated on Google Colab

References

1. Beland, L.P.; Kim, D. The Effect of High School Shootings on Schools and Student Performance. *Educational Evaluation and Policy Analysis* **2016**, *38*, 113–126. Publisher: American Educational Research Association, https://doi.org/10.3102/0162373715590683.
2. Katsiyannis, A.; Rapa, L.J.; Whitford, D.K.; Scott, S.N. An Examination of US School Mass Shootings, 2017-2022: Findings and Implications. *Advances in Neurodevelopmental Disorders* **2023**, *7*, 66–76. https://doi.org/10.1007/s41252-022-00277-3.
3. Levine, P.B.; McKnight, R. Exposure to a School Shooting and Subsequent Well-Being, 2020. https://doi.org/10.3386/w28307.
4. Gimenez, G.; Barrado, B. Exposure to crime and academic achievement: A case study for Costa Rica using PISA data. *Studies in Educational Evaluation* **2020**, *65*, 100867. https://doi.org/10.1016/j.stueduc.2020.100867.
5. Suomalainen, L.; Haravuori, H.; Berg, N.; Kiviruusu, O.; Marttunen, M. A controlled follow-up study of adolescents exposed to a school shooting–psychological consequences after four months. *European Psychiatry: The Journal of the Association of European Psychiatrists* **2011**, *26*, 490–497. https://doi.org/10.1016/j.eurpsy.2010.07.007.
6. Rossin-Slater, M.; Schnell, M.; Schwandt, H.; Trejo, S.; Uniat, L. Local exposure to school shootings and youth antidepressant use. *Proceedings of the National Academy of Sciences* **2020**, *117*, 23484–23489. Publisher: Proceedings of the National Academy of Sciences, https://doi.org/10.1073/pnas.2000804117.
7. Boxer, P.; Drawve, G.; Caplan, J.M. Neighborhood Violent Crime and Academic Performance: A Geospatial Analysis. *American Journal of Community Psychology* **2020**, *65*, 343–352. https://doi.org/10.1002/ajcp.12417.
8. Cabral, M.; Kim, B.; Rossin-Slater, M.; Schnell, M.; Schwandt, H. Trauma at School: The Impacts of Shootings on Students' Human Capital and Economic Outcomes, 2020. https://doi.org/10.3386/w28311.
9. Yang, L.K.; Gopalan, M. The Effects of Campus Shootings on School Finance and Student Composition. *Education Finance and Policy* **2023**, *18*, 277–301. _eprint: https://direct.mit.edu/edfp/article-pdf/18/2/277/2075864/edfp_a_00350.pdf, https://doi.org/10.1162/edfp_a_00350.
10. Gershenson, S.; Tekin, E. The Effect of Community Traumatic Events on Student Achievement: Evidence from the Beltway Sniper Attacks. *Education Finance and Policy* **2018**, *13*, 513–544. Publisher: MIT Press ERIC Number: EJ1192644, https://doi.org/10.1162/edfp_a_00234.
11. Kao, G.; Thompson, J.S. Racial and Ethnic Stratification in Educational Achievement and Attainment. *Annual Review of Sociology* **2003**, *29*, 417–442. _eprint: https://doi.org/10.1146/annurev.soc.29.010202.100019, https://doi.org/10.1146/annurev.soc.29.010202.100019.
12. The EDFacts Initiative | U.S. Department of Education, 2023. Publisher: US Department of Education (ED).
13. Iacus, S.M.; King, G.; Porro, G. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* **2012**, *20*, 1–24. https://doi.org/10.1093/pan/mpr013.
14. Lacour, M.; Tissington, L.D. The effects of poverty on academic achievement.
15. Chioke, S. Influence of Poverty on Academic Performance: A Systematic Study of Secondary School Students in Enugu East LGA. *Archives of Educational Studies (ARES)* **2021**, *1*, 19–36. Number: 1.

Author Contributions: The authors contributed equally to the paper.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We would like to express sincere gratitude towards Professor Talal Rahwan and Instructor Sara Mumtaz for supporting us throughout the semester and providing us with detailed and constructive guidance. We greatly enjoyed the class and the group project. Thank you so much!

Conflicts of Interest: The authors declare no conflict of interest.