Barcelona School of Economics

Master's Degree in Specialized Economic Analysis
Specialization in International Trade, Finance and Development

“Shaken Foundations: The Impact of the 2007 Perú Earthquake on Human Capital Accumulation”

Authors: Cathal Leslie, David Tillo, Harry Morris, Lakshmi Castillo, Luke Atazona and William Cheeseman

Supervisors: Fernando Broner and Geert Mesters

June 2023

ABSTRACT IN ENGLISH (100 words)**:**

Our paper investigates whether the 2007 earthquake in Ica, Perú had long lasting impacts through human capital accumulation in Perú. We conduct our analysis through a dynamic difference-in-difference and propensity score matching approach, constructing treatment rings based on actual earthquake intensity. We find that the earthquake had a significantly negative and persistent impact on several dimensions of education, including a 17 per cent increase in the student dropout rate. A common theme from our results is those who were youngest at the time of the earthquake are most impacted, consistent with the existing literature and 1000-day hypothesis. Our research highlights the importance of prioritizing limited resources on the young and most vulnerable in the aftermath of natural disasters.

ABSTRACT IN CATALAN/ SPANISH (100 words)

Nuestro artículo investiga si el terremoto de 2007 en Ica, Perú, tuvo impactos a largo plazo a través de la acumulación de capital humano en Perú. Realizamos nuestro análisis utilizando un enfoque de diferencia en diferencia dinámica y propensity score matching, construyendo anillos de tratamiento basados en la intensidad real del terremoto. Encontramos que el terremoto tuvo un impacto negativo y persistente significativo en varias dimensiones de la educación, incluyendo un aumento del 17 por ciento en la tasa de abandono escolar. Un tema recurrente en nuestros resultados es que quienes eran más jóvenes en el momento del terremoto son los más afectados, lo que concuerda con la literatura existente y la hipótesis de los 1000 días. Nuestra investigación resalta la importancia de priorizar los recursos limitados en los jóvenes y los más vulnerables en el período posterior a los desastres naturales.

KEYWORDS IN ENGLISH (3):

Natural Disaster, Human Capital Accumulation

KEYWORDS IN CATALAN/ SPANISH (3):

Desastres Naturales, Acumulación de capital humano

Shaken Foundations: The Impact of the 2007 Perú Earthquake on Human Capital Accumulation

Cathal Leslie, David Tillo, Harry Morris, Lakshmi Castillo,
Luke Atazona, and William Cheeseman

June 2023

International Trade, Finance and Development Master Project (15S101)

Abstract

Our paper investigates whether the 2007 earthquake in Ica, Perú had long lasting impacts through human capital accumulation in Perú. We conduct our analysis through a dynamic difference-in-difference and propensity score matching approach, constructing treatment rings based on actual earthquake intensity. We find that the earthquake had a significantly negative and persistent impact on several dimensions of education, including a 17 per cent increase in the student dropout rate. A common theme from our results is those who were youngest at the time of the earthquake are most impacted, consistent with the existing literature and 1000-day hypothesis. Our research highlights the importance of prioritizing limited resources on the young and most vulnerable in the aftermath of natural disasters.

JEL classification: O12, I25, J24
Keywords: Natural Disaster, Human Capital Accumulation

Acknowledgments

We would like to extend our heartfelt gratitude to the following individuals, without whom the completion of this master project would not have been possible: **Fernando Broner** and **Geert Mesters** - our esteemed Professors, whose guidance and expertise have been instrumental in shaping the direction of our research. Their valuable insights, and constructive feedback were been indispensable throughout this journey. **Ayah Bohsali** - our dedicated TA, whose unwavering commitment and tireless assistance have greatly contributed to the success of this thesis. Her patience, expertise, and willingness to go the extra mile in helping us navigate through challenges have been truly invaluable. **Gianmarco León-Ciliotta** - we would like to express our sincere appreciation to Prof Léon-Ciliotta for his generous assistance with guiding us through the various data sources, including the ECE Test Scores and the ENAHO Household Survey. His willingness to share his knowledge has significantly enriched our research, enabling us to delve deeper into our study. Finally, a special thanks to **Tellus Thorne** - we extend our gratitude to Lakshmi's mother for her invaluable contribution in downloading the data when access to the INEI website was not possible for those outside of Peru.

Contents

1 Introduction

Education's pivotal role in human capital accumulation and its transformative potential for social mobility is well-established in the field of Economics. Education can serve as a vehicle to lift millions out of poverty, yield higher lifetime incomes, and provide a range of non-monetary benefits from improved health to greater civic engagement and social cohesion.[1] The literature has primarily focused on the benefits of the duration, quality and accumulation of education. Less attention has been paid to the consequences resulting from these shocks to education.

On the other hand, the literature on natural disasters tends to focus on the direct costs experienced in the aftermath of disasters. These typically consider costs to infrastructure or welfare loss associated with dampened economic activity Baez et al. (2010). This paper aims to contribute to the growing literature on the interaction of the natural disaster and education literature.

These shocks have received revived attention with the onset of the recent Covid-19 pandemic. However, they are far from uncommon, especially in developing economies. Despite the growing trend to increase access to education in the developing world, natural disasters, civil conflicts, and other forms of social unrest are just a few of the many catalysts for significant disruptions to schooling. These disruptions can profoundly impact children's formative educational years, potentially affecting their long-term outcomes. Understanding the leakage effects of such shocks could provide valuable information for policymakers who prepare for such, even if rare, events. Schooling interruptions and their products are both academically intriguing and crucial moments for enacting effective policies to support children and mitigate the long-term impacts on their educational outcomes.

This paper studies one significant disruption in August 2007 when a catastrophic earthquake struck Ica, Perú. This earthquake resulted in almost 600 deaths, over 2,000 injuries and extensive infrastructure damage.

We employ a dynamic difference-in-difference and propensity score matching approach, utilising a measure of earthquake intensity - peak ground acceleration, to define treatment status more accurately.

We find that the earthquake resulted in a 17 per cent increase in the student dropout rate, which was sustained over time. Concurrently, the earthquake induced a 20 per cent rise in academic failures in the year of its occurrence. These findings are supported in alternative datasets where we find an increase in the schooling gap and a fall in literacy. These impacts are particularly pronounced for individuals who were born in the years just before the earthquake, as is consistent with the 1000-day hypothesis. We carry out several robustness checks on these results, utilising alternative treatment definitions which provide mixed support.

The paper is structured as follows: Section 2 reviews the relevant literature. Section

[1] For example, Hanushek (2021) emphasizes the association between an economy's knowledge capital, particularly in math and science, and long-term economic growth. Angrist and Krueger (1991), document the effect of longer periods of education on wages, whilst Oreopoulos and Salvanes (2011) shows the positive relationship between education and not only financial benefits but also significant non-pecuniary returns, such as improved decision-making skills, better health and reduced risky behaviours.

3 presents the context of our analysis. Section 4 introduces the data sources utilized in our study, followed by a detailed explanation of our empirical strategy in Section 5. Our results are detailed in Section 6, and we discuss our findings in Section 7. We conclude in Section 8.

2 Literature Review

2.1 Dimensions of human capital accumulation

Perhaps the most straightforward measure of educational attainment is measuring the quantity of education through years of schooling. Shidiqi et al. (2023) examine the effect of exposure to the 2006 Yogyakarta earthquake on educational attainment. They show that exposure amongst school-aged students decreased years of schooling by 0.74 years by negatively affecting the probability of completing compulsory education (primary and junior high school) but with no impact on post-compulsory schooling rates. Similarly, Paudel and Ryu (2018) study Nepal's 1988 earthquake's impact on educational outcomes finding that the average school grade declined by 0.8. Finally, Caruso and Miller (2015) analyse the impact of the 1970 Ancash earthquake on human capital accumulation 37 years after the shock and observed that males exposed to the earthquake in utero completed 0.5 years less schooling while females completed 0.8 years less.

Another essential education outcome indicator in developing economies is the prevalence of child labour. However, this has received little attention in this literature. One exception is Baez et al. (2007) which examined the effects of a Hurricane that devastated Nicaragua in 1998, observing a significant increase in child labour (8.5 percentage points). While they observed no significant change in school enrollment, the number of children simultaneously enrolled and working more than doubled.

Also important is education quality, which has received relatively less attention in the literature. For example, one could assess schooling quality through test scores. A paper on the 2005 Pakistani Earthquake Andrabi et al. (2020) found that children aged three to eleven scored significantly worse on academic tests.

2.2 Heterogeneous impacts

It is clear from the literature that natural disasters have differential impacts on different groups. In the papers previously mentioned, Andrabi et al. (2020) found that children whose mothers were primary school educated did not experience negative impacts of the 2005 Pakistani earthquake, Paudel and Ryu (2018) found the greatest effects of Nepal's 1988 earthquake were on female and low caste groups. Caruso and Miller (2015) found larger impacts on women's schooling years. Furthermore, a study on the Tangshan 1976 earthquake Tian et al. (2022) find the highest impacts for those in utero, particularly those in the second trimester.

2.3 Methodology

In terms of the methodological approach of this literature, virtually all papers adopt a difference-in-difference framework. This includes Shidiqi et al. (2023), Paudel and Ryu (2018), Tian et al. (2022) and Baez et al. (2007). Shidiqi et al. (2023) also employ matching as a robustness check. There is a growing consensus within this literature of moving towards measures of earthquake intensity instead of using distance-based proxies as measures to capture treatment intensities. Shidiqi et al. (2023) use the modified Mercali Intensity Scale (MMI), Paudel and Ryu (2018) use the Modified Mercali scale, Tian et al. (2022) use a seismic intensity scale.

One notable exception is Andrabi et al. (2020), who explicitly argue against this trend. There measure of treatment intensity is distance to nearest Faultline. They also cite Andrabi and Das (2017) who argue that earthquake intensity are worse measures of treatment as they are in part a function of "characteristics that may be correlated with socioeconomic characteristics". Papers in this literature overwhelmingly use a binary approach to defining treatment and controls in their studies, with Andrabi et al. (2020) again a notable exception in using a continuous measure of distance.

2.4 Literature Contributions

Our paper makes valuable contributions to each of the three aspects of the literature discussed above.

Firstly, considering various educational outcomes from a variety of sources allows us to better understand the earthquake's impacts on human capital accumulation. This contrasts with the literature, which typically only utilises a small number of outcome variables. Furthermore, the fact that these papers find impacts on different dimensions of educational attainment suggests that contextual factors are important and thus analysing the setting as we do is important.

Secondly, a clear finding from the literature is the heterogeneous impacts of natural disasters on education across different population sub-groups. Therefore, our analysis across age, sex and other dimensions is an important contribution. In particular, many of our results find powerful impacts on those who would have been a few years old or in utero at the time of the earthquake. This is consistent with much work in the literature, including Tian et al. (2022) and the so-called first 1000 days hypothesis. This is the theory that the first 1000 days after conception are particularly important for a child's cognitive development.

Finally, our methodology makes several important contributions. As discussed later, our use of Peak Ground Acceleration is a particularly accurate measure of earthquake intensity, allowing us to define treated regions more accurately. Furthermore, the fact that we consider the intensity of treatment in these rings as a robustness check provides more granularity than most papers that only consider a binary treatment/control. We also use distance as a robustness check, another commonly used measure in the literature. Our use of matching is also relatively rare in this literature, with the aforementioned exception of Andrabi et al. (2020).

3 Setting: 2007 Earthquake in Perú

In this section, we outline the setting for our analysis. Specifically, we will discuss the earthquake and its immediate impacts, the Peruvian government's response to the earthquake and the education system in Perú.

3.1 Earthquake

On the 15th of August 2007 at 18:40 local time, an earthquake measuring 7.9 on the Richter scale occurred off the Peruvian coast, lasting for two minutes. The epicenter was located 60km west of Pisco city and 150km south-southeast of Lima, at a depth of 39km. Several aftershocks then took place the next day. It also generated a Tsunami both locally and on a deep pressure sensor off the coast of Chile (Fritz et al., 2008).

The earthquake caused severe damage to the department of Ica, especially to the cities of Pisco, Chincha and Ica. The earthquake also affected the city of Cañete in the department of Lima, and the Huancavelica department. Almost 600 people were killed, over 2,000 injured and over 70,000 families affected. There was extensive damage to government buildings, homes, schools, churches, roads and bridges, creating a consequent need for shelter, food, water, healthcare, sanitation and security (Elhawary & Castillo, 2008).

These destructive consequences of natural disasters - particularly earthquakes, are not a new phenomenon in Peru. The country's geological location, the Circum-Pacific Belt (also known as the Ring of Fire), accounts for about 81 percent of the world's largest earthquakes and seismic activities, making Perú one of the top ten earthquakes prone countries in the world (U.S.G.S, 2023). The US National Geophysical Data Center estimates that since 1950, more than 68,600 deaths have resulted directly from earthquakes in Perú.

3.2 Government Response

The Peruvian government led the response to the earthquake and was supported by the private sector as well as civil society and the international community. The initial earthquake response required searching for survivors, removing rubble and providing shelter, food, water, medical assistance and education. The government's response to the earthquake was poor despite sufficient resources and an elaborate formal system intended to heavily involve regional actors. Rather than support the regional system, the central government circumvented it by creating a parallel structure.

The initial response was characterized by poor coordination, especially in the early phases of the response. This in turn led to an absence of information, duplicating effort and poor needs assessment. Most of the relief was initially concentrated in Pisco, despite significant damage and need in Chincha and other remote rural areas. Despite these significant shortcomings, the response was nonetheless successful in some key respects. Most of the injured were evacuated promptly, there was no spread of disease and the majority of the affected population eventually received some form of support. However, this was to a large extent thanks to international actors. There was an initial well-coordinated,

rapid and generous relief effort by international NGOs and subsequent support by the UN system.

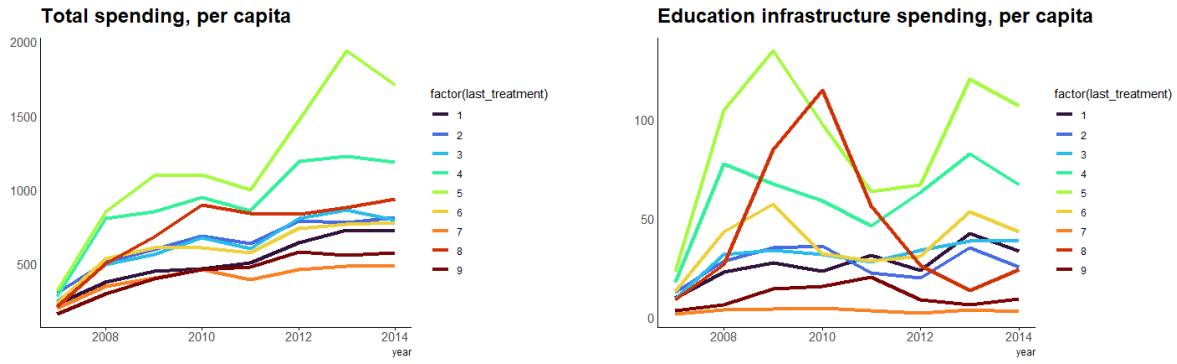
Other factors also enabled these successes, such as the proximity of an airport and port to the affected area and the relative lack of damage to the main road linking the region to the capital, Lima (Elhawary & Castillo, 2008).

3.3 Local government expenditure

3.3.1 Expenditure

Using local government expenditure data at the district level we investigate the relationship between how impacted[2] districts by the earthquake and their local government expenditure as we can see in Figure 1. We analyze data on total expenditure and education infrastructure expenditure.

The lack of a strong relationship between the most impacted districts and government expenditure is clear from simple plots of the relationship between treatment status and local government expenditure. This is consistent with the qualitative evidence of local governments being overridden by the central government in response to the earthquake, as well as a lack of capacity at local government level (Elhawary & Castillo, 2008).



(a) (b)

Figure 1

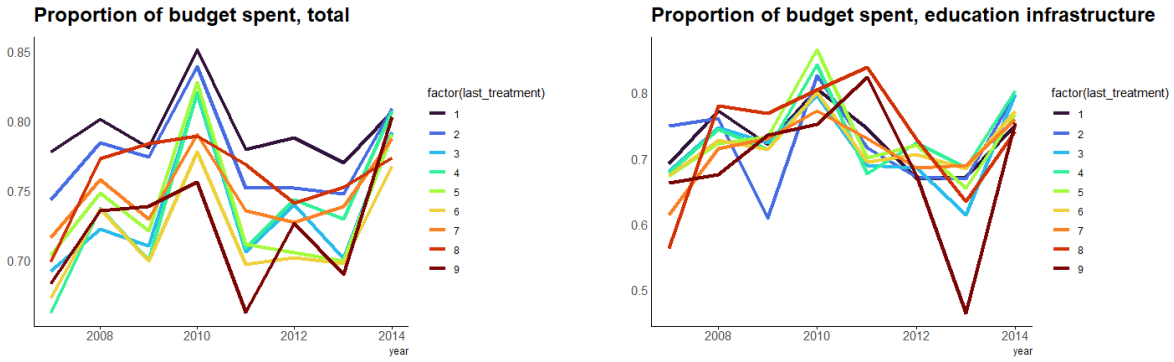
3.3.2 Proportion of allocated budget spent

We also consider what proportion of local governments allocated budgets ended up actually spending being spent, both overall and for education infrastructure spending as shown in Figure 2. What proportion of their allocated budget that local governments spent could be considered a measure of the competence of local government in terms of utilising the budget allocated to it. [3] The plot again shows that there is no strong relation-

[2] measured by the intensity of this impact where 9 is more impacted and 1 is less impacted

[3] Although the initial budgets are likely a function of how much money the local government initially asks for, complicating interpretation of this measure.

ship between these variables.



(a) (b)

Figure 2

3.4 Education in Perú

3.4.1 Test scores and literacy rates

Based on the Programme for International Student Assessment (PISA) which tests 15 year olds on reading, mathematics and science, Peruvian students score poorly when compared to the international average. However, they have also increased significantly over time. In 2018, for reading literacy Peruvian students scored 401 (OECD average was 487), up from 327 in 2000. They scored 400 in mathematics, (OECD average was 489) up from 365 in 2009. They scored 404 in science (OECD average was 489), up from 369 in 2009.[4]

In terms of literacy rates, using data from the UNESCO Institute for Statistics it is clear that they improved across this time period despite a lack of data available in many years. The increases are particularly noticeable for youth, increasing from just over 87.5 per cent in 2004 to just below 95 per cent in 2020.

3.4.2 Enrollment rates, pupil to teacher ratio and expenditure

Since 2000 to 2018 secondary school enrolment rates have substantially increased. The primary school enrolment rate declined until 2015, before partially recovering over the next three years.[5] Government education expenditure has increased substantially over this time in absolute terms. As a percentage of GDP it also increased, albeit by less. Finally, the primary school ratio has fallen from 28.74 in 2000 to 17.39 in 2018, while the secondary school ratio fell from 20.36 to 14.20 over the same time period.

[4]Test scores are only available from 2009 for mathematics and science

[5]The gross enrolment rate is the total enrolment regardless of age, relative to the population of the age group that corresponds to the level of education. The net enrolment rate is the ratio of children of school age who are enrolled in school to the population of the corresponding school age.

4 Data

4.1 Encuesta Nacional de Hogares (ENAHO)

This paper utilizes data collected from the *Encuesta Nacional de Hogares* (ENAHO) for the years 2004 through 2019. The ENAHO is a comprehensive annual household survey implemented by Perú's *Instituto Nacional de Estadística e Informática* (INEI). It is the principal source of socio-economic information at the household level in Perú. It collects detailed data on household characteristics, education, health, employment, and income, among others. Appendix A.2.1 provides more information on the different modules used for our analysis.

The ENAHO uses a complex survey design, employing stratification, clustering, and multistage sampling. This design enables it to produce reliable estimates at national, urban-rural, departmental, and provincial levels. The survey is representative of the Peruvian population and is designed to ensure comparability over time. Whilst there is a panel version of the ENAHO, it does not capture longer time frames adequately and has a small sample. Due to these limitations, we have constrained our analysis to the use of repeated cross-sections of the entire sample.

The ENAHO questionnaire has been subjected to several revisions to improve the quality of the data it collects. However, a significant change in the survey methodology was implemented in 2004. To ensure the consistency and comparability of the data used in this study, only the ENAHO data collected after this methodological change, from 2004 onwards, is utilized. Data for the year 2020 was intentionally excluded from this study. The onset of the COVID-19 pandemic in 2020 had a significant and wide-ranging impact on many aspects of society, including education. Using 2020 data could thus introduce confounding factors that may potentially distort the estimated impact of the 2007 Ica earthquake on educational outcomes and human capital accumulation. The period from 2004 to 2019 provides a solid temporal frame for analyzing the impacts of the Ica earthquake. This timeframe allows for three years of pre-earthquake data for establishing baseline trends and for twelve years of post-earthquake data to analyze the long-term impacts of the disaster.

4.2 Censo Escolar (Educational Census)

The Educational Census is an annual statistical process that gathers detailed information from both public and private educational institutions, as well as Non-School Programs across the entire country. By conducting the Educational Census, the Peruvian Ministry of Education (MINEDU) can track progress in various aspects such as student enrollment, levels of academic delay, promotion rates, repetition rates, dropout rates, number of teaching and administrative staff, educational infrastructure, and more. This census started on the year 2000 and has been conducted annually since except for the year 2003.

The data used in our study has been restricted to primary and secondary schools between the years 2004 to 2015 from the modality of Regular Basic Education (EBR)[6], this

[6]This modality encompasses the educational levels of early childhood, primary, and secondary education.

data has been provided by the MINEDU.

4.3 Test score data from the Evaluación Censal de Estudiantes (ECE)

The *Evaluación Censal de Estudiantes* (ECE) is a standardized annual evaluation conducted across the country, which provides a valuable assessment of the proficiency of students in crucial academic areas, such as reading comprehension and mathematics. The ECE has been conducted annually since 2007, thus offering a rich longitudinal dataset that enables us to track changes in student performance over time. By exploiting this temporal variation, we can identify potential deviations in academic performance associated with the earthquake.

We incorporate student-level test scores data provided by the MINEDU to draw comparisons between the educational outcomes of students before and after the earthquake shock. Whilst the first iteration of the ECE was in 2007, this was before the earthquake shock, providing us with one observation before treatment. The dataset is especially suited for this purpose, as it not only records students' test scores, but also includes a range of demographic and school-level variables which we can link using the School Census. This allows us to control for potential confounding factors that might otherwise bias our estimates of the earthquake's impact. Figure 4 shows the geographic distribution of PGA values in Table 5.

The ECE data utilized in our study is restricted to primary school students between 2007 and 2014, specifically for the second grade, on average per year 50,000 of second grade students undertake this test. While this limits the scope of our analysis, it is worth noting that these grades represent critical stages in a child's educational trajectory. Therefore, changes in performance at these levels can have far-reaching consequences for a student's long-term academic and career prospects. Since this is a confidential data that we requested the MINEDU every student is assigned a random identification code .

4.4 Treatment Areas: US National Geological Survey

As noted in Section 2, a common measure of treatment is the distance from the epicentre. However, this is not always an appropriate proxy for earthquake intensity or damage. An earthquake's wave energy does not travel evenly in all directions from its epicentre; instead, a multitude of factors including the type and condition of the rock or sediment it travels through, the depth of the earthquake, and the type of fault that caused it (USGS).

While some studies have successfully leveraged distance metrics, the use of precise, on-site measurements of earthquake intensity can serve as a superior variable where available.

In this context, we harness data from the US Geological Survey (USGS). The USGS operates a comprehensive network of seismic readers scattered throughout Peru, documenting the local seismic activity caused by earthquakes. We selected Peak Ground

Its focus is on providing an educational experience that is appropriate to the physical, emotional, and cognitive development of children and adolescents, starting from birth. The organization of this modality is structured through levels, cycles, categories, and various modalities, ensuring a comprehensive educational journey for students.

Acceleration (PGA) as our metric of choice, which delineates the maximum horizontal ground movement during an earthquake tremor. This measurement is a standard tool for engineers seeking to gauge the potential risk and damage an earthquake might inflict on buildings and infrastructures.

The selection of this variable proved integral to our study, as the correlation between distance and earthquake damage was found to be weak. Under a distance-focused analysis, Lima—a densely populated area only mildly affected by the earthquake—would have been disproportionately weighted due to its geographical proximity to the earthquake's epicentre. Hence, the PGA metric enabled a more accurate and meaningful analysis.

4.5 Variables

As mentioned previously, we have several variables to measure educational outcomes. Using the school census data we have a comprehensive account of Peruvian primary and secondary schools from 2004 to 2015, including variables such as open/closed status, location and student characteristics. Table 1 shows 560,344 observations across our census datasets. It reveals that our school data are almost equally divided between rural and urban areas, and that most schools are public (76 per cent). This aligns with MINEDU annual reports Marcos, Vásquez, et al. (2018). Notably, we see substantial variation in the failure and dropout rates of students across schools, which we will exploit for our analysis.

On the variable students enrolled we see that the minimum is zero, this reflects that for schools that were not open the number of students enrolled is set to zero.

	Mean	Median	SD	Min	Max	Count
Treatment ring	3.99	4.00	2.28	1.00	9.00	560344
Students Enrolled	133.35	61.00	217.34	0.00	5437.00	560344
Students failed	9.27	3.00	24.00	0.00	1531.00	560344
Students dropout	5.63	2.00	11.12	0.00	634.00	560344
Students transfer	1.41	0.00	5.33	0.00	370.00	560344
Students death	0.04	0.00	0.60	0.00	200.00	560344
% Students drop	5.34	2.23	7.86	0.00	100.00	560331
% Students failed	7.25	4.55	8.51	0.00	100.00	560331
Teachers	10.97	6.00	15.50	1.00	436.00	560344
Administration	0.24	0.00	0.43	0.00	1.00	560344
Rural	0.52	1.00	0.50	0.00	1.00	560179

Table 1: Summary Statistics: Educational Census

In Figure 3 we separate the data into samples where affected are the schools in municipalities that were badly affected by the earthquake and unaffected for the remainder. In Figure 3a as presumed the earthquake shows an important increase of 5% in the school closing on the 2007. In Figure 3b we see no immediate building new of schools, despite

expecting the creation of temporary schools. Disproportionately higher building of new schools in affected areas does eventually commence, albeit at a very modest rate. The building of new schools of the affected areas starts in 2009. In Figure 3c we can see how the drop out rate increases continuously until 2008 and in Figure 3d the failure rate spikes in 2007. From this we can infer that the earthquake had a significant immediate impact on the children's educations due to the school year being affected by the closure of schools. Also, we observe that this impact could have some long lasting impacts due to the slow rebuilding of school's.

From the ENAHO data set our variables of interest are the schooling gap, literacy rate, and whether the child works. We define the schooling gap as the difference between their age and years of schooling. This could increase if they were to be retained in a school year or if they dropout, in which case we impute their last school year completed. An increase in the schooling gap is therefore considered a worse educational outcome.

	Control Group	Treatment Group
Number of individuals	390,293	77,730
Age completed (mean (SD))	11.15 (3.68)	11.21 (3.71)
Sex (male) (%)	198,768 (50.9)	39,069 (50.3)
Attending school (%)	283,458 (92.6)	58,147 (92.8)
Household w/o access to running water (%)	71,897 (18.8)	6,146 (8.0)
Knows how to write (%)	155,763 (69.5)	30,459 (73.4)
Schooling Gap (mean (SD))	5.75 (1.44)	5.45 (1.14)

Table 2: ENAHO Dataset - Individuals aged 5-17

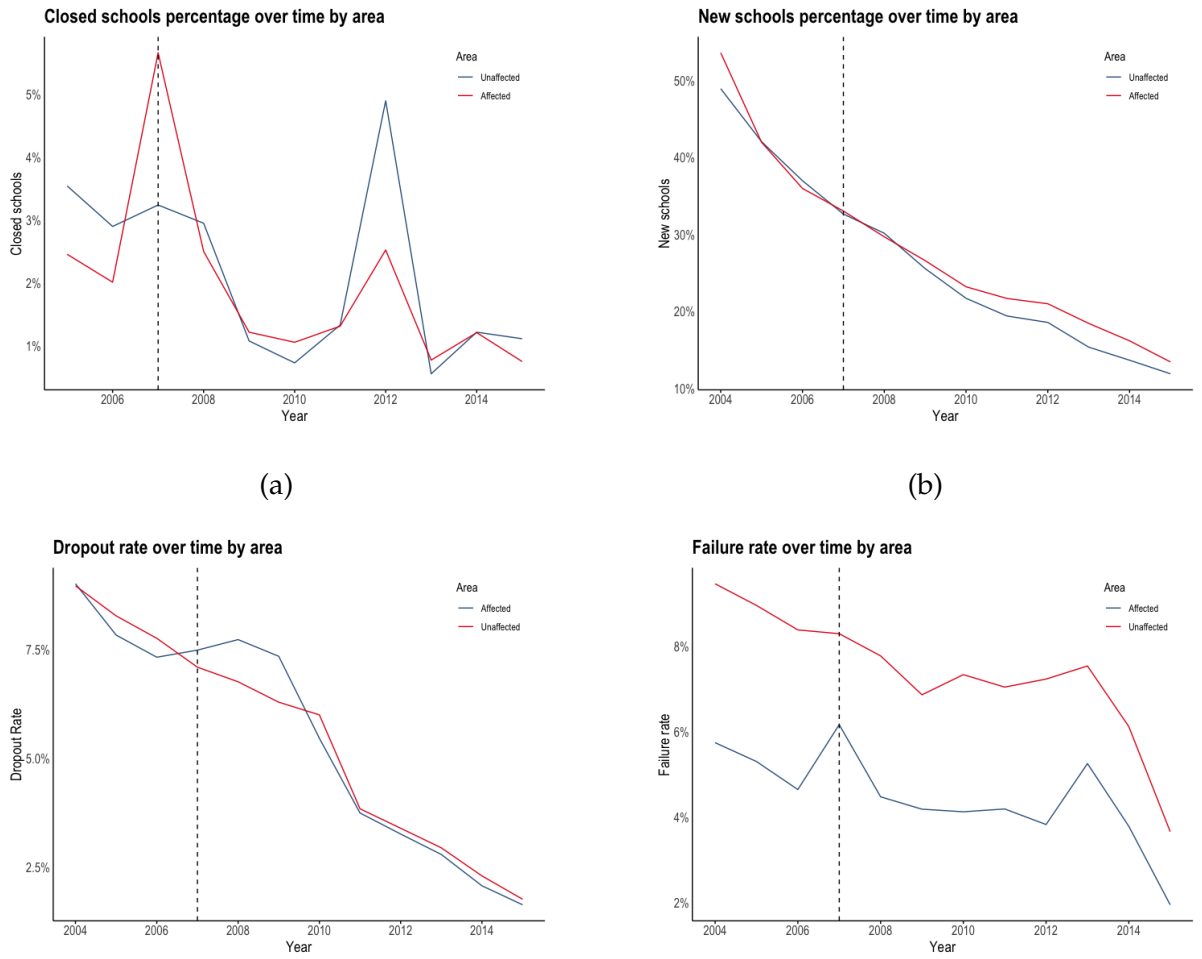


Figure 3

5 Empirical Strategy

We explore four alternative model specifications, choosing the best suited specification based on the respective dataset's characteristics.

Firstly, we note the exogenous nature of natural experiments and utilise a simple regression model on both treated and control regions in the periods following the earthquake. However, this approach assumes no differences between the treated and control regions pre-earthquake. Considering that the earthquake epicentre struck 30km off the coast of Perú, and the endogenous nature of location sorting, it is likely that pre-treatment differences existed between the treatment and control regions. Subsequent analysis of the means in both groups shows that such differences do exist post-treatment; however, they are also present in pre-treatment periods. Therefore, any significant differences found using a simple regression model would be biased.

Secondly, we consider a simple difference approach which would compare the outcomes of cohorts within the same treatment area. This approach also yielded results, however, it does not account for national trends where you would expect to see improved educational outcomes over time, and thus downward bias our results. Therefore, we primarily employ a dynamic difference-in-difference approach, which allows us to study the impact profile of the earthquake over time between treated and control regions.

The crucial identifying assumption in the difference-in-difference approach is that in the absence of treatment the difference between treatment and control regions remains constant. This allows the implicit calculation of a counterfactual for the treatment regions in post-earthquake periods. While the assumption is fundamentally unprovable, it can be supported by the presence of parallel trends between treatment and control regions in pre-treatment periods. This is shown in the following section, with the interaction term between treatment and pre-treatment periods being insignificant.

5.1 ENAHO

Since the ENAHO dataset is a series of cross-sections rather than panel data we cannot observe the same individuals in pre-earthquake periods. We therefore follow a strategy similar to Duflo (2001) where earlier cohorts of children are used as the pre-treatment baseline in treated and control regions.

A further assumption required to ensure a valid counterfactual is established is that no other external factor solely affects the treatment group. While there have been a number of earthquakes off the coast of Perú in subsequent years which could be correlated with our treatment measures, none of these caused a level of destruction which could be expected to significantly bias results. Furthermore, while there have been educational policies implemented post-2007[7], these have been targeted at rural areas which are presented in both treated and control regions, and have been additionally controlled for.

Our main specification for the dynamic difference-in-difference approach employed with the ENAHO dataset is given below:

$$Y_{itj} = \alpha + \sum_{t=2004}^{2019} \beta_t \cdot Treatment_j + \tau_t + \delta_j + \gamma \mathbf{X}_{it} + \varepsilon_{it} \tag{1}$$

where Y_{itj} denotes the outcomes of interest: schooling gap, literacy rate, whether the child works. Treatment is either a dummy variable equal to one if an individual is in a treated region, or a continuous measure of treatment intensity.

The impact of the treatment is estimated at each time period in an event study style approach, with 2006 used as the reference period. Time and district fixed effects are captured by τ_t and δ_j respectively. Finally, $\mathbf{X_{it}}$ represents the controls which include the following dummies: rural, sex, access to running water, is of high economic dependence, state school, and the interaction of rural and a post-2014 dummy. This final control captures the national compensation policy for teachers in rural areas which has been found to increase the teacher quality and subsequently test scores in these areas (Bobba et al.,

[7]Such as a rural compensation bonus for teachers from 2014

2021). We check for the possibility that the dummies have access to running water and are of high economic dependence are bad controls but find no increase in treated regions following the earthquake.

Our main results are estimated using a binary treatment measure based on the PGA intensity rings adopted from the US National Geological Survey, where the outer unimpacted rings 1-6 are defined as controls and inner rings are defined as treatment. Our robustness section explores the impact of employing four alternative continuous measures of treatment.

5.2 School Census

Similarly, for the School Census analysis, we utilise the same methodology but adjusted to reflect that it is panel data. This dataset's outcomes of interest Y_{itj} are: dropout rate and failure rate, both on the entire sample and subdivided by gender and primary/secondary status. We also weight our observations to reflect that there a large variety of school sizes in the dataset.

In addition to data cleaning, we have removed the year 2013 from our regressions on the failure rate. In 2013 there is a substantial spike in failure in treated and untreated areas. However, there is no obvious cause for why this would occur, which in addition to a number of other strange results for variables not considered in this paper (classroom sizes), we suspect there are data quality issues. We have not excluded this year for dropout rate analysis as the numbers seem steady for the year 2013 (and our results do not change substantially with its exclusion).

5.3 Test Scores

The ECE dataset, which captures individual test score data, has only one pre-treatment period available. Therefore, we are unable to provide support for the parallel trends assumption for test scores and cannot utilise the dynamic difference-in-difference approach above with confidence. So as to still take advantage of this rich dataset, we employ a propensity score matching approach to compare the mean test score outcomes of similar schools in treated and control regions. Treated schools are defined as those within the treatment rings 7-9, and they are matched based on their 2007 observables taken from both the ECE dataset, ENAHO and School Census data. the mean, variance and skewness of the standardized test score distributions of its students, average household income of the district, school size, if the school is rural, whether it is a private or state school, and the male-to-female student ratio. After estimating the propensity school of treatment, we employ a nearest-neighbour matching algorithm in order to obtain a suitable treatment and control groups. We use the same specification in equation 1 where we include the matching covariates to ensure it is doubly robust. Balance checks from the unmatched and matched samples can be found in the appendix section A.3.3 where the matched sample appears well balanced.

6 Results

6.1 School Census

We find that the 2007 earthquake in Peru had lingering effects on the education system. Table 3 shows the results from the school census dataset, where we find that the earthquake resulted in a 1.5 percentage point surge (equivalent to approximately a 17 per cent increase) in the student dropout rate in 2007 — a trend sustained over time. Concurrently, the earthquake-induced an estimated 1 percentage point rise (equivalent to an approximately 20 per cent increase) in academic failures in the year of its occurrence. Together the findings hint at a long-term degradation of the schooling system, culminating in a persisting downturn in educational outcomes for students.

Table 3: Main School Census Results - Dropout and Failure Rates

Dependent Variables:	Dropout Rate	Failure Rate
Model:	(1)	(2)
	OLS	OLS
Variables		
Treatment × Year = 2004	8.08×10^{-5}	-0.0082
	(0.0020)	(0.0049)
Treatment × Year = 2005	0.0025	-0.0015
	(0.0019)	(0.0038)
Treatment × Year = 2007	0.0164***	0.0119**
	(0.0044)	(0.0044)
Treatment × Year = 2008	0.0097***	-0.0011
	(0.0025)	(0.0045)
Treatment × Year = 2009	0.0165***	0.0029
	(0.0026)	(0.0035)
Treatment × Year = 2010	0.0085**	0.0037
	(0.0030)	(0.0049)
Treatment × Year = 2011	0.0126***	0.0040
	(0.0033)	(0.0049)
Treatment × Year = 2012	0.0145***	-0.0014
	(0.0030)	(0.0036)
Treatment × Year = 2013	0.0153***	
	(0.0031)	
Treatment × Year = 2014	0.0099**	0.0045
	(0.0035)	(0.0040)
Treatment × Year = 2015	0.0112**	0.0114*
	(0.0036)	(0.0048)
After 2014 x Rural	-0.0284***	-0.0189***
	(0.0010)	(0.0011)
Treatment	Binary	Binary
Fixed-effects		
School	Yes	Yes
Year	Yes	Yes
Fit statistics		
Observations	459,797	417,956
R^2	0.51093	0.47583
Within R^2	0.01192	0.00368

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

6.1.1 Heterogeneity

Tables 10 and 11 of the Appendix underscore a significant disparity in the influence of earthquakes on dropout and failure rates across different school levels. Primary school pupils exhibit a pronounced sensitivity to the aftermath of an earthquake, leading to increased rates of academic failure and premature school leaving, compared to their counterparts in high school. One might expect the opposite, as high school students are more

inclined to seek external employment opportunities. However, considering the significant attrition rates in Peruvian high school cohorts, the earthquake may simply expedite the departure of students already poised to drop out in the near future. This is consistent with our findings on the schooling gap from the ENAHO data below.

Regarding gender disparity the evidence is weak, with only marginal differences in dropout and failure rates between male and female students.

6.2 ENAHO

In Table 4, we show the results from the ENAHO survey. We find evidence that the earthquake increased the schooling gap, which is consistent with our findings from the school census. When looking at the grouped sample of ages we see significant results after 2009 with the earthquake increasing the gap by approximately 0.1 years. This delayed impact suggests that the earthquake's impact could be greater among those who were younger at the time. We also find evidence that this educational loss translates into a fall in literacy rates, both at the time of the earthquake and in later years. However, we do not find evidence that there is a corresponding increase in the child's probability of working. Though this measure does not capture the possibility that children who help their family post earthquake and do not receive an income.

6.2.1 Heterogeneity

Assessing the change in the school gap on 12-year-olds is supportive of the theory that age is an important determinant on educational impacts. We find the earthquake increased the gap by 0.3 years in 2007, which is significant at the 0.1 per cent significance level. The impacts could have been especially large for this age group as it captures the transition from primary to secondary school. Furthermore, we find significant results in 2013 which would correspond to individuals who were five at the time of the earthquake and about to enter primary school. Additionally, the significant results for the final years of our sample correspond to the birth cohorts of 2005, 2006, and 2007 which aligns with the 1000-day hypothesis as discussed in the literature review.

Exploring the literacy impact on nine year olds, when most but not all are able to read, reveals delayed effects from the earthquake with literacy rates being approximately five percent lower. These effects are using a different cross sections of individuals to the those in the schooling gap at age 12, but once again those who were five at the time of the earthquake or born in 2005, 2006, and 2007 are most affected.

6.3 ECE Test Scores

Table 17 in the Appendix shows the results the impact on standardised test scores from the matching procedure used with the ECE dataset. Despite the impact on literacy rates in the ENAHO data document above, we do not find that average literacy score was significantly impacted for treated schools. We do however find that the average maths score was higher. This could be explained by a compositional change in the sample resulting from the increase in dropouts in treated regions.

Table 4: Main ENAHO Results - Binary Treatment

Dependent Variables:	Schooling Gap		Literacy		Child Labour
Sample Age:	5-17	12	5-17	9	5-17
Model:	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	OLS	OLS
Variables					
Treatment × Year = 2004	-0.0104	0.1120	0.0025	0.0025	0.0028
	(0.0346)	(0.0772)	(0.0145)	(0.0214)	(0.0044)
Treatment × Year = 2005	0.0308	0.1902	0.0026	0.0048	-0.0020
	(0.0348)	(0.0996)	(0.0138)	(0.0241)	(0.0050)
Treatment × Year = 2007	0.0315	0.3275***	-0.0322*	-0.0060	0.0020
	(0.0391)	(0.0968)	(0.0154)	(0.0191)	(0.0047)
Treatment × Year = 2008	0.0379	0.1103	-0.0167	-0.0150	0.0069
	(0.0387)	(0.0905)	(0.0157)	(0.0215)	(0.0049)
Treatment × Year = 2009	0.0732*	0.1757	-0.0178	-0.0116	0.0061
	(0.0372)	(0.0925)	(0.0150)	(0.0196)	(0.0046)
Treatment × Year = 2010	0.0821*	0.1943*	-0.0217	-0.0349	0.0035
	(0.0416)	(0.0922)	(0.0151)	(0.0198)	(0.0054)
Treatment × Year = 2011	0.1303***	0.2325*	-0.0469**	-0.0462*	-0.0026
	(0.0393)	(0.1063)	(0.0152)	(0.0223)	(0.0049)
Treatment × Year = 2012	0.1497***	0.1545	-0.0330*	-0.0288	-0.0057
	(0.0409)	(0.0856)	(0.0149)	(0.0177)	(0.0044)
Treatment × Year = 2013	0.1491***	0.3128***	-0.0336**	-0.0376*	-0.0032
	(0.0360)	(0.0855)	(0.0125)	(0.0166)	(0.0047)
Treatment × Year = 2014	0.0476	0.1844*	-0.0303*	-0.0465**	-0.0090*
	(0.0354)	(0.0828)	(0.0140)	(0.0179)	(0.0042)
Treatment × Year = 2015	0.0939*	0.1094	-0.0579***	-0.0598**	-0.0043
	(0.0370)	(0.0882)	(0.0149)	(0.0200)	(0.0048)
Treatment × Year = 2016	0.1323***	0.2542**	-0.0280*	-0.0233	-0.0050
	(0.0393)	(0.0846)	(0.0140)	(0.0177)	(0.0045)
Treatment × Year = 2017	0.1457***	0.2249**	-0.0267	-0.0153	-0.0053
	(0.0357)	(0.0841)	(0.0141)	(0.0167)	(0.0044)
Treatment × Year = 2018	0.1667***	0.3157***	-0.0470***	-0.0239	5.59×10^{-6}
	(0.0388)	(0.0869)	(0.0142)	(0.0185)	(0.0045)
Treatment × Year = 2019	0.1610***	0.3155***	-0.0601***	-0.0303	-0.0058
	(0.0355)	(0.0853)	(0.0139)	(0.0178)	(0.0044)
Age	0.1348***		0.1003***		0.0150***
	(0.0029)		(0.0007)		(0.0003)
Sex (male)	0.0525***	0.0646***	-0.0061***	-0.0032	0.0222***
	(0.0055)	(0.0120)	(0.0017)	(0.0028)	(0.0011)
State School	0.2368***	0.3397***	-0.0606***	-0.0074*	0.0145***
	(0.0106)	(0.0187)	(0.0036)	(0.0037)	(0.0012)
Rural	0.4813***	0.5316***	-0.0990***	-0.0864***	0.0214***
	(0.0267)	(0.0403)	(0.0062)	(0.0106)	(0.0018)
High Economic Dependence	0.1913***	0.4274***	-0.0418***	-0.0555***	-0.0349***
	(0.0201)	(0.0512)	(0.0060)	(0.0134)	(0.0012)
Households w/o Water	0.2979***	0.3137***	-0.0491***	-0.0457***	0.0068***
	(0.0128)	(0.0234)	(0.0031)	(0.0054)	(0.0010)
Rural × After 2014	-0.2847***	-0.3259***	0.0605***	0.0672***	0.0032
	(0.0194)	(0.0318)	(0.0048)	(0.0072)	(0.0016)
Fixed-effects					
District	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	443,686	37,217	244,929	34,717	443,686
R^2	0.22952	0.20701	0.32123	0.15172	0.08804
Within R^2	0.15957	0.04064	0.30354	0.01630	0.07653

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

6.4 Robustness Checks

For robustness, we use three alternative treatment definitions in our analysis. The first is based on an estimate of distance from the earthquake epicentre to the district centroid, as is commonly used in the literature, and the latter two employ the PGA values. To capture possible non-linearities between ground movement and building destruction we also used the squared value of the PGA measures.

6.4.1 School Census

For the school census data we find that the use of PGA squared to define treatment leads to significant results consistent with our binary treatment measure. This is in line with the notion that the relationship between earthquake intensity and damage - and consequently educational outcomes - is non-linear. However, the inverse distance and PGA definition produces more concerning results with support for the parallel trends assumption not holding in 2004. These results can be seen in Tables 12 and 13 of the Appendix.

6.4.2 ENAHO

Our results utilising the ENAHO dataset are also sensitive to the treatment definition. Our estimates for the school gap for ages 5-17 are insignificant in all years when using the continuous treatment measures. This is also the case for the school gap when looking at children aged 12 except for the inverse distance measure which is significant in 2007, 2013, and 2019 as was the case in our main results. These results can be seen in Tables 6 and 7.

The literacy rate for those aged 5-17 produces consistent results when using both the Linear Probability and Logit estimator with binary treatment. However, the parallel trends assumption is not supported in 2004 for the other continuous treatment measures. When restricting the sample to those aged nine, the parallel trends assumption is supported but results are largely insignificant. These results can be seen in Tables 8 and 9.

7 Discussion

7.1 Policy Implications

It is clear from our results that younger children were disproportionately impacted by the earthquake. This is the case for young school age children (i.e those around the age of five), and those yet to enter school (including those in utero at the time of the earthquake). As such, we recommend that following future natural disasters policy makers should focus scarce resources on the youngest children and pregnant women.

As the earthquakes impact extends beyond school age children to those even younger at the time of the earthquake, it is important to emphasise that relevant government supports extend beyond educational provision. Policy makers should also ensure that young children and pregnant women have access to adequate nutrition and shelter.

It is also likely based on our paper's analysis that not enough was done to quickly ensure adequate educational access in impacted areas following the Perú 2007 earthquake. This is suggested by the lack of increased educational infrastructure spending in the most impacted districts when compared to those not impacted, as well as in the extremely modest and delayed increase in new schools in impacted compared to not impacted districts over time.

7.2 Limitations

7.2.1 Defining treatment

This paper identifies the weakness of using distance as a proxy for earthquake damage, leveraging Peak Ground Acceleration (PGA) as an effective alternative for economists to quantify the effects of earthquakes. However, adopting analogous methods or a synergistic blend, such as Peak Ground Velocity and the duration of shaking, could potentially enhance accuracy. Exploring the ideal regression model for PGA is also paramount, the correlation between PGA and damage is likely non-linear; small PGA levels typically inflict minimal harm, whereas the impact spectrum at high PGA levels exhibits a more uniform devastating effect. Accordingly, much of the variation in damage occurs at relatively moderate PGA levels, which we try to capture with our robustness check using the squared value of PGA. However, it is possible that other functional forms, such as a sigmoid transformation may better approximate the relationship.

7.2.2 Migration

A key concern addressed in this study is migration's impact on our analysis, as it has the potential to introduce selection bias. Individuals who migrate may possess different characteristics or experiences compared to those who remain in their original location. If relocation decisions are influenced by the earthquake or its consequences, it could confound the analysis and hinder the isolation of the true causal impact of the earthquake on human capital accumulation. To address this concern, we conducted extensive research on migration patterns in the Ica department, where most municipalities in our treated sample are located.

In a study by CEPAL et al. (2022), which analyzes migration patterns in Perú over the past century using data from the National Census conducted in 1993, 2007 (after the earthquake), and 2017, the authors calculate net migration and consider various geographical levels such as departments, provinces, and districts. The analysis reveals that Ica experienced low net out-migration migration between 1993 and 2007. In contrast, by 2017, it had changed towards low net inwards migration. Nevertheless, our study must determine the specific population that is migrating and whether this population is representative of our sample.

According to Aguilar (2015), that also uses the National Census, out of the total population of Ica in 2007 after the earthquake, 81.4% continued to reside in the district where they were born. Among the remaining 18.6% who migrated, 4% corresponded to natives of Ica who relocated within the department's provinces. A similar scenario was observed

in the 2017 census. The census data also provides information on migration over the past five years. It reveals that during 2002-2007 and 2012-2017, 94.7% and 90.4% of the population, respectively, remained in the same location. Although we lack specific data on the migration percentage during 2008-2011 and the particular characteristics of the migrants, these statistics offer valuable insights into Ica's migration patterns. They highlight its low migration rate and the tendency of most individuals born in Ica to stay and live there. Therefore migration is unlikely to cause selection bias in our study.

8 Conclusion

The 2007 earthquake in Perú had a significant and lasting impact on the country's education system. We utilise Ministry of Education data on national test scores and the school census, in addition to a long running household survey to estimate the impact on outcomes such as: dropouts, failures, retainment, literacy, test scores, and child labour. We employ both a dynamic difference and difference model as well as propensity score matching, estimating both the short to medium term impacts of the earthquake. We define treatment according to PGA values recorded by the US Geographic Survey, which improves upon traditional treatment measures in the literature.

We find that the earthquake resulted in an approximate 17 per cent increase in the student dropout rate, which was sustained over time. Concurrently, the earthquake-induced an approximately 20 per cent rise in academic failures in the year of its occurrence. These findings are corroborated in the household survey where we find an increase in the schooling gap and fall in literacy. These impacts are particularly pronounced for individuals who were born in the years just before the earthquake, as is consistent with the 1000-days hypothesis. Further heterogeneity analysis supports this finding, with dropouts and failure rates higher in primary schools. Though, some of our results are not robust to alternative continuous measures of treatment.

Overall, these findings underscore the profound and enduring impact of the earthquake on Perú's educational system. It highlights the critical need for robust disaster management and recovery strategies that not only address immediate needs but also consider the long term implications on vital sectors such as education.

References

Aguilar, A. (2015). Migraciones internas en el perú a nivel departamental.

Andrabi, T., Daniels, B., & Das, J. (2020). Human capital accumulation and disasters: Evidence from the pakistan earthquake of 2005. *Journal of Human Resources*.

Andrabi, T., & Das, J. (2017). In aid we trust : Hearts and minds and the pakistan earthquake of 2005. *The Review of Economics and Statistics*, *99*, 1–56.

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Baez, J. E., de la Fuente, A., & Santos, I. (2010). Do natural disasters affect human capital? an assessment based on existing empirical evidence. *Labor: Human Capital eJournal*.

Baez, J. E., Santos, I., Abadie, A., Afridi, F., Black, D., Engelhardt, G. V., Field, E. M., Jensen, R. L., Kubik, J. D., Liebman, J. C., & Zeckhauser, R. J. (2007). Children's vulnerability to weather shocks: A natural disaster as a natural experiment *.

Bobba, M., Ederer, T., León-Ciliotta, G., Neilson, C. A., & Nieddu, M. G. (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in peru. *SSRN Electronic Journal*.

Caruso, G. D., & Miller, S. (2015). Long run effects and intergenerational transmission of natural disasters: A case study on the 1970 ancash earthquake. *Journal of Development Economics*, *117*, 134–150.

CEPAL, N., et al. (2022). Perú: Migraciones internas y dinámica sociodemográfica de departamentos, provincias y distritos en las dos primeras décadas del siglo xxi.

Duflo, E. (2001). The medium run effects of education expansion: Evidence from a large school construction program in indonesia. *Labor: Human Capital*.

Elhawary, S., & Castillo, G. (2008). The role of the affected state: A case study on the peruvian earthquake response. *HGP Working Paper*.

Fritz, H. M., Kalligeris, N., Borrero, J. C., Broncano, P., & Ortega, E. (2008). The 15 august 2007 peru tsunami runup observations and modeling. *Geophysical Research Letters*, *35*.

Hanushek, E. A. (2021). Education and economic growth — oxford research encyclopedia of economics and finance.

Marcos, M., Vásquez, M. C., et al. (2018). Tipologıa y caracterización de las escuelas privadas en el perú.

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless : The nonpecuniary benefi ts of schooling.

Paudel, J., & Ryu, H. (2018). Natural disasters and human capital: The case of nepal's earthquake. *World Development*.

Shidiqi, K. A., Paolo, A. D., & Choi, Á. (2023). Earthquake exposure and schooling: Impacts and mechanisms. *Economics of Education Review*.

Tian, X., Gong, J., & Zhai, Z. (2022). Natural disasters and human capital accumulation: Evidence from the 1976 tangshan earthquake. *Economics of Education Review*.

U.S.G.S. (2023). Where do earthquakes occur?

A Appendix

A.1 Figures and Tables

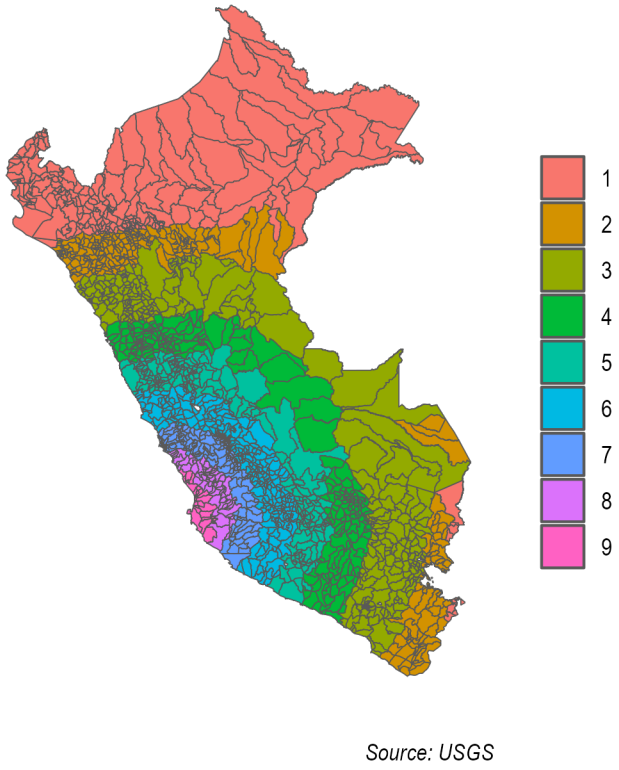


Figure 4: Districts in Peru by 2007 Earthquake Peak Ground Acceleration

Table 5: Peak Ground Acceleration and Treatment Rings

Treatment	Peak Ground Acceleration
1	0.1
2	0.2
3	0.5
4	1.0
5	2.0
6	5.0
7	10.0
8	20.0
9	50.0

A.2 Data

A.2.1 ENAHO - Description of Modules

Module 01: Household Characteristics . This module captures information about living conditions and poverty, including characteristics of housing, household expenses, social programs. The data from this module provides insight into the structural and material conditions of households in Peru, which is critical in understanding the socio-economic backdrop against which individuals and families live and make decisions.

Module 02: Individual Characteristics. Collects personal information on individual household members, which may include age, gender, marital status, and other demographic and personal characteristics. This data can provide a granular view of the population and facilitate analysis at an individual level.

Module 03: Education. Questions in this module are designed to gather information about educational attainment, school attendance, and possibly other related topics such as access to educational resources and quality of education.

Module 04: Health. The health section includes questions about health status, access to healthcare services, and potentially other health-related issues. This data is crucial for assessing the health conditions of the population and the reach of healthcare services.

Module 05: Labour. This module focuses on employment and income. The information collected includes employment status, type of employment, income levels, and financial inclusion. This data allows for an evaluation of the economic status of individuals and households, and the dynamics of the labour market in Peru.

Module 34: Household Summary. This module relates to calculated variables (sumaria), which includes the calculation of total household income, spending on particular categories and other aggregate or derived variables.

A.2.2 Schooling Census - Description of the Data Collection

The Educational Census is legally required according to Ministerial Resolution No. 627-2016-MINEDU, which approves the "Information Report from the Educational Institution to the Education System." However, it is important to note that the census is not a means of inspection. Each educational institution is responsible for conducting its own census, and the Director of the Educational Institution or the Coordinating Teacher is accountable for providing the necessary information.

The Educational Census consists of two modules, and data is collected through three types of forms that each educational institution must submit at the beginning and end of the school year. The first module is filled between May and June and has two forms: Form "A" for Enrollment, Teachers, Resources and Form "11" for Educational Facility. The second module is filled between December and February and has the Form "B" for Educational Exercise Results.

A.3 Results

A.3.1 ENAHO

Table 6: Robustness Checks for Schooling Gap

Dependent Variable:	Schooling Gap				
Sample Age:	5-17				
Model:	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Variables					
Treatment × Year = 2004	-0.0023	-0.0104	-3.808	-0.0010	-5.74×10^{-5}
	(0.0375)	(0.0346)	(4.659)	(0.0017)	(9.86×10^{-5})
Treatment × Year = 2005	-0.0088	0.0308	-4.534	-0.0020	-0.0001
	(0.0341)	(0.0348)	(4.018)	(0.0014)	(8.04×10^{-5})
Treatment × Year = 2007	0.0316	0.0315	-5.194	-0.0025	-0.0002
	(0.0433)	(0.0391)	(5.190)	(0.0019)	(0.0001)
Treatment × Year = 2008	0.0160	0.0379	-3.506	-0.0019	-0.0001
	(0.0431)	(0.0387)	(5.254)	(0.0021)	(0.0001)
Treatment × Year = 2009	0.0345	0.0732*	-1.380	-0.0009	-4.7×10^{-5}
	(0.0414)	(0.0372)	(5.055)	(0.0019)	(0.0001)
Treatment × Year = 2010	0.0643	0.0821*	-0.1170	0.0001	-3.26×10^{-6}
	(0.0457)	(0.0416)	(5.437)	(0.0021)	(0.0001)
Treatment × Year = 2011	0.1139**	0.1303***	4.317	0.0014	6.66×10^{-5}
	(0.0401)	(0.0393)	(5.664)	(0.0021)	(0.0001)
Treatment × Year = 2012	0.1262**	0.1497***	5.709	0.0020	7.04×10^{-5}
	(0.0415)	(0.0409)	(6.250)	(0.0023)	(0.0001)
Treatment × Year = 2013	0.1386***	0.1491***	6.018	0.0026	0.0001
	(0.0393)	(0.0360)	(4.857)	(0.0019)	(0.0001)
Treatment × Year = 2014	0.1312***	0.0476	-1.877	-0.0013	-9.33×10^{-5}
	(0.0379)	(0.0354)	(4.649)	(0.0017)	(9.62×10^{-5})
Treatment × Year = 2015	0.1670***	0.0939*	0.9403	0.0003	4.2×10^{-6}
	(0.0404)	(0.0370)	(4.641)	(0.0018)	(9.98×10^{-5})
Treatment × Year = 2016	0.2151***	0.1323***	6.628	0.0022	0.0001
	(0.0440)	(0.0393)	(5.112)	(0.0019)	(0.0001)
Treatment × Year = 2017	0.2204***	0.1457***	7.149	0.0028	0.0001
	(0.0422)	(0.0357)	(4.731)	(0.0017)	(9.69×10^{-5})
Treatment × Year = 2018	0.2462***	0.1667***	8.549	0.0027	0.0001
	(0.0430)	(0.0388)	(4.733)	(0.0017)	(9.3×10^{-5})
Treatment × Year = 2019	0.2622***	0.1610***	8.169	0.0027	0.0001
	(0.0410)	(0.0355)	(4.632)	(0.0016)	(8.79×10^{-5})
Age		0.1348***	0.1348***	0.1348***	0.1348***
		(0.0029)	(0.0029)	(0.0029)	(0.0029)
Sex (male)		0.0525***	0.0526***	0.0526***	0.0526***
		(0.0055)	(0.0055)	(0.0055)	(0.0055)
State School		0.2368***	0.2355***	0.2352***	0.2350***
		(0.0106)	(0.0106)	(0.0106)	(0.0106)
Rural		0.4813***	0.4836***	0.4846***	0.4853***
		(0.0267)	(0.0266)	(0.0266)	(0.0266)
High Economic Dependence		0.1913***	0.1915***	0.1915***	0.1915***
		(0.0201)	(0.0201)	(0.0201)	(0.0201)
Households w/o Water		0.2979***	0.2991***	0.2989***	0.2988***
		(0.0128)	(0.0128)	(0.0128)	(0.0128)
Rural × After 2014		-0.2847***	-0.2888***	-0.2895***	-0.2904***
		(0.0194)	(0.0192)	(0.0191)	(0.0190)
Treatment:	Binary	Binary	inverse_distance	PGA	PGA_squared
Fixed-effects					
District	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	468,023	443,686	443,686	443,686	443,686
R^2	0.07659	0.22952	0.22944	0.22944	0.22942
Within R^2	0.00057	0.15957	0.15948	0.15948	0.15946

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 7: Robustness Checks for Schooling Gap (Age 12)

Dependent Variable:	Schooling Gap				
Sample Age:	12				
Model:	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Variables					
Treatment × Year = 2004	0.0893	0.1120	10.63	0.0033	0.0001
	(0.0861)	(0.0772)	(9.850)	(0.0037)	(0.0002)
Treatment × Year = 2005	0.1446	0.1902	10.36	0.0032	0.0001
	(0.1024)	(0.0996)	(10.92)	(0.0040)	(0.0002)
Treatment × Year = 2007	0.2797*	0.3275***	28.38**	0.0074	0.0003
	(0.1167)	(0.0968)	(10.44)	(0.0040)	(0.0002)
Treatment × Year = 2008	0.0545	0.1103	7.383	0.0034	0.0002
	(0.1083)	(0.0905)	(11.74)	(0.0052)	(0.0003)
Treatment × Year = 2009	0.1331	0.1757	5.110	0.0017	2.02×10^{-5}
	(0.1050)	(0.0925)	(10.70)	(0.0042)	(0.0002)
Treatment × Year = 2010	0.1142	0.1943*	9.940	0.0038	0.0001
	(0.0980)	(0.0922)	(11.06)	(0.0045)	(0.0003)
Treatment × Year = 2011	0.2082	0.2325*	21.09	0.0078	0.0004
	(0.1201)	(0.1063)	(15.84)	(0.0062)	(0.0004)
Treatment × Year = 2012	0.0736	0.1545	9.317	0.0040	0.0002
	(0.0974)	(0.0856)	(9.789)	(0.0049)	(0.0003)
Treatment × Year = 2013	0.2609*	0.3128***	23.47*	0.0075	0.0003
	(0.1057)	(0.0855)	(10.44)	(0.0044)	(0.0003)
Treatment × Year = 2014	0.2461*	0.1844*	14.96	0.0044	0.0002
	(0.0954)	(0.0828)	(11.43)	(0.0039)	(0.0002)
Treatment × Year = 2015	0.2036	0.1094	4.144	0.0020	7.39×10^{-5}
	(0.1061)	(0.0882)	(10.44)	(0.0044)	(0.0003)
Treatment × Year = 2016	0.3100**	0.2542**	14.82	0.0044	0.0002
	(0.0999)	(0.0846)	(10.75)	(0.0041)	(0.0002)
Treatment × Year = 2017	0.3059**	0.2249**	15.77	0.0051	0.0002
	(0.0974)	(0.0841)	(9.712)	(0.0040)	(0.0002)
Treatment × Year = 2018	0.3674***	0.3157***	19.41	0.0060	0.0002
	(0.0991)	(0.0869)	(9.964)	(0.0037)	(0.0002)
Treatment × Year = 2019	0.3906***	0.3155***	27.76**	0.0102*	0.0005*
	(0.1023)	(0.0853)	(10.71)	(0.0042)	(0.0002)
Sex (male)		0.0646***	0.0648***	0.0649***	0.0651***
		(0.0120)	(0.0120)	(0.0120)	(0.0121)
State School		0.3397***	0.3389***	0.3389***	0.3387***
		(0.0187)	(0.0187)	(0.0187)	(0.0187)
Rural		0.5316***	0.5354***	0.5364***	0.5371***
		(0.0403)	(0.0401)	(0.0400)	(0.0400)
High Economic Dependence		0.4274***	0.4278***	0.4276***	0.4275***
		(0.0512)	(0.0512)	(0.0512)	(0.0512)
Households w/o Water		0.3137***	0.3152***	0.3147***	0.3146***
		(0.0234)	(0.0234)	(0.0234)	(0.0234)
Rural × After 2014		-0.3259***	-0.3323***	-0.3330***	-0.3341***
		(0.0318)	(0.0312)	(0.0311)	(0.0308)
Treatment:	Binary	Binary	inverse_distance	PGA	PGA_squared
Fixed-effects					
District	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	38,170	37,217	37,217	37,217	37,217
R^2	0.17112	0.20701	0.20683	0.20665	0.20662
Within R^2	0.00119	0.04064	0.04041	0.04020	0.04016

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 8: Robustness Checks for Literacy

Dependent Variable:	Literacy								
Sample Age:	5-17								
Model:	(1) OLS	(2) OLS	(3) Logit	(4) OLS	(5) Logit	(6) OLS	(7) Logit	(8) OLS	(9) Logit
Variables									
Treatment × Year = 2004	-0.0074	0.0025	-0.0117	3.877*	31.85*	0.0018***	0.0154***	0.0001***	0.0009***
	(0.0150)	(0.0145)	(0.1238)	(1.660)	(13.64)	(0.0005)	(0.0040)	(2.77×10^{-5})	(0.0002)
Treatment × Year = 2005	0.0163	0.0026	-0.0419	2.274	14.76	0.0008	0.0047	4.65×10^{-5}	0.0003
	(0.0130)	(0.0138)	(0.1187)	(1.381)	(11.30)	(0.0006)	(0.0048)	(3.28×10^{-5})	(0.0003)
Treatment × Year = 2007	-0.0225	-0.0322*	-0.3278*	1.008	0.4814	0.0011	0.0061	8.12×10^{-5}*	0.0005
	(0.0144)	(0.0154)	(0.1345)	(1.859)	(14.96)	(0.0006)	(0.0052)	(3.45×10^{-5})	(0.0003)
Treatment × Year = 2008	-0.0071	-0.0167	-0.1991	2.186	13.08	0.0014	0.0097	8.63×10^{-5}*	0.0006*
	(0.0163)	(0.0157)	(0.1317)	(1.683)	(13.91)	(0.0007)	(0.0056)	(4.05×10^{-5})	(0.0003)
Treatment × Year = 2009	0.0007	-0.0178	-0.1520	0.7487	5.193	0.0007	0.0057	4.64×10^{-5}	0.0004
	(0.0135)	(0.0150)	(0.1303)	(1.698)	(13.24)	(0.0007)	(0.0057)	(4.14×10^{-5})	(0.0003)
Treatment × Year = 2010	-0.0082	-0.0217	-0.1833	0.0164	-0.5385	0.0005	0.0044	4.52×10^{-5}	0.0004
	(0.0161)	(0.0151)	(0.1258)	(1.783)	(14.06)	(0.0007)	(0.0056)	(4.14×10^{-5})	(0.0003)
Treatment × Year = 2011	-0.0263	-0.0469**	-0.3715**	-1.649	-12.21	-0.0005	-0.0031	-2.06×10^{-5}	-0.0001
	(0.0151)	(0.0152)	(0.1284)	(1.678)	(12.61)	(0.0007)	(0.0051)	(3.69×10^{-5})	(0.0003)
Treatment × Year = 2012	-0.0164	-0.0330*	-0.2713*	-0.7384	-3.241	-0.0002	-0.0005	-7.16×10^{-6}	-6.32×10^{-6}
	(0.0154)	(0.0149)	(0.1278)	(1.566)	(12.24)	(0.0007)	(0.0052)	(3.75×10^{-5})	(0.0003)
Treatment × Year = 2013	-0.0130	-0.0336**	-0.2728*	-0.8635	-5.971	-0.0003	-0.0011	-6.85×10^{-6}	2.93×10^{-5}
	(0.0122)	(0.0125)	(0.1096)	(1.377)	(11.21)	(0.0005)	(0.0045)	(3.03×10^{-5})	(0.0002)
Treatment × Year = 2014	-0.0349*	-0.0303*	-0.2495*	-1.067	-7.335	3.65×10^{-5}	0.0013	1.88×10^{-5}	0.0002
	(0.0137)	(0.0140)	(0.1172)	(1.595)	(12.32)	(0.0006)	(0.0045)	(3.12×10^{-5})	(0.0002)
Treatment × Year = 2015	-0.0550***	-0.0579***	-0.4764***	-2.847	-23.39	-0.0005	-0.0031	-2.78×10^{-6}	1.42×10^{-5}
	(0.0161)	(0.0149)	(0.1223)	(1.755)	(13.29)	(0.0006)	(0.0048)	(3.37×10^{-5})	(0.0003)
Treatment × Year = 2016	-0.0253	-0.0280*	-0.2159	-0.1084	-1.546	0.0005	0.0040	4.45×10^{-5}	0.0004
	(0.0137)	(0.0140)	(0.1173)	(1.641)	(12.93)	(0.0006)	(0.0048)	(3.22×10^{-5})	(0.0003)
Treatment × Year = 2017	-0.0217	-0.0267	-0.1933	0.2109	2.990	0.0004	0.0034	3.33×10^{-5}	0.0003
	(0.0139)	(0.0141)	(0.1158)	(1.688)	(11.95)	(0.0006)	(0.0045)	(3.32×10^{-5})	(0.0002)
Treatment × Year = 2018	-0.0491***	-0.0470***	-0.3564**	-1.530	-11.81	-7.54×10^{-5}	-0.0002	1.65×10^{-5}	0.0001
	(0.0144)	(0.0142)	(0.1212)	(1.735)	(13.30)	(0.0007)	(0.0057)	(3.96×10^{-5})	(0.0003)
Treatment × Year = 2019	-0.0591***	-0.0601***	-0.4401***	-3.041	-20.59	-0.0007	-0.0038	-1.65×10^{-5}	-6.28×10^{-5}
	(0.0136)	(0.0139)	(0.1162)	(1.641)	(12.76)	(0.0006)	(0.0045)	(2.99×10^{-5})	(0.0002)
Age		0.1003***	0.9257***	0.1003***	0.9256***	0.1003***	0.9256***	0.1003***	0.9256***
		(0.0007)	(0.0140)	(0.0007)	(0.0140)	(0.0007)	(0.0140)	(0.0007)	(0.0140)
Sex (male)		-0.0061***	-0.0475***	-0.0061***	-0.0477***	-0.0060***	-0.0476***	-0.0060***	-0.0475***
		(0.0017)	(0.0125)	(0.0017)	(0.0124)	(0.0017)	(0.0124)	(0.0017)	(0.0124)
State School		-0.0606***	-0.4097***	-0.0603***	-0.4080***	-0.0602***	-0.4072***	-0.0601***	-0.4066***
		(0.0036)	(0.0255)	(0.0036)	(0.0254)	(0.0036)	(0.0253)	(0.0036)	(0.0253)
Rural		-0.0990***	-0.7762***	-0.0999***	-0.7815***	-0.1005***	-0.7864***	-0.1010***	-0.7897***
		(0.0062)	(0.0486)	(0.0062)	(0.0490)	(0.0062)	(0.0490)	(0.0062)	(0.0491)
High Economic Dependence		-0.0418***	-0.3662***	-0.0418***	-0.3652***	-0.0417***	-0.3650***	-0.0418***	-0.3654***
		(0.0060)	(0.0475)	(0.0060)	(0.0476)	(0.0060)	(0.0476)	(0.0060)	(0.0476)
Households w/o Water		-0.0491***	-0.3580***	-0.0494***	-0.3599***	-0.0493***	-0.3593***	-0.0493***	-0.3591***
		(0.0031)	(0.0240)	(0.0031)	(0.0241)	(0.0031)	(0.0241)	(0.0031)	(0.0241)
Rural × After 2014		0.0605***	0.4986***	0.0622***	0.5079***	0.0630***	0.5140***	0.0637***	0.5190***
		(0.0048)	(0.0379)	(0.0047)	(0.0372)	(0.0047)	(0.0371)	(0.0047)	(0.0370)
Treatment:	Binary	Binary	Binary	inverse_distance	inverse_distance	PGA	PGA	PGA_squared	PGA_squared
Fixed-effects									
District	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics									
Observations	265,758	244,929	244,727	244,929	244,727	244,929	244,727	244,929	244,727
Squared Correlation	0.02916	0.32123	0.47828	0.32121	0.47822	0.32121	0.47819	0.32119	0.47819
Pseudo R^2	0.02319	0.32846	0.36691	0.32843	0.36688	0.32843	0.36686	0.32840	0.36685
BIC	351,214.1	214,003.7	195,685.5	214,011.9	195,693.6	214,011.8	195,697.6	214,018.6	195,702.5

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 9: Robustness Checks for Literacy (Age 9)

Dependent Variable:	Literacy								
Sample Age:	9								
Model:	(1) OLS	(2) OLS	(3) Logit	(4) OLS	(5) Logit	(6) OLS	(7) Logit	(8) OLS	(9) Logit
Variables									
Treatment × Year = 2004	-0.0013	0.0025	-0.1699	1.513	7.262	0.0012	0.0148	6.81×10^{-5}	0.0009
	(0.0231)	(0.0214)	(0.3458)	(2.004)	(36.40)	(0.0007)	(0.0123)	(4.03×10^{-5})	(0.0007)
Treatment × Year = 2005	0.0153	0.0048	-0.0309	0.9626	16.63	0.0009	0.0149	4.24×10^{-5}	0.0007
	(0.0237)	(0.0241)	(0.4325)	(2.322)	(49.53)	(0.0011)	(0.0218)	(6.31×10^{-5})	(0.0012)
Treatment × Year = 2007	-0.0014	-0.0060	0.0348	-0.5574	7.582	0.0003	0.0107	1.08×10^{-5}	0.0004
	(0.0197)	(0.0191)	(0.3991)	(2.216)	(46.39)	(0.0010)	(0.0191)	(5.62×10^{-5})	(0.0010)
Treatment × Year = 2008	-0.0085	-0.0150	-0.1376	0.3743	33.89	0.0008	0.0246	5.38×10^{-5}	0.0017
	(0.0223)	(0.0215)	(0.4075)	(1.973)	(46.80)	(0.0009)	(0.0194)	(5.24×10^{-5})	(0.0012)
Treatment × Year = 2009	-0.0103	-0.0116	-0.0516	-2.307	-34.21	-0.0002	-0.0032	-7.23×10^{-6}	-0.0001
	(0.0207)	(0.0196)	(0.4083)	(2.535)	(42.72)	(0.0012)	(0.0197)	(6.78×10^{-5})	(0.0011)
Treatment × Year = 2010	-0.0322	-0.0349	-0.4621	-2.475	-39.11	-0.0002	-0.0012	3.19×10^{-6}	0.0002
	(0.0221)	(0.0198)	(0.3860)	(1.770)	(40.50)	(0.0009)	(0.0189)	(5.15×10^{-5})	(0.0011)
Treatment × Year = 2011	-0.0475*	-0.0462*	-0.6691	-3.025	-37.72	-0.0002	0.0004	3.07×10^{-6}	0.0004
	(0.0240)	(0.0223)	(0.4171)	(1.982)	(35.29)	(0.0009)	(0.0150)	(4.73×10^{-5})	(0.0009)
Treatment × Year = 2012	-0.0262	-0.0288	0.3134	-1.147	91.51	-5.06×10^{-5}	0.0390	-5.49×10^{-6}	0.0014
	(0.0197)	(0.0177)	(0.5248)	(2.054)	(107.1)	(0.0009)	(0.0466)	(5.17×10^{-5})	(0.0020)
Treatment × Year = 2013	-0.0351	-0.0376*	0.3395	-1.845	154.1*	-0.0004	0.0390	-1.72×10^{-5}	0.0020
	(0.0183)	(0.0166)	(0.4687)	(1.575)	(71.28)	(0.0008)	(0.0257)	(4.28×10^{-5})	(0.0014)
Treatment × Year = 2014	-0.0771***	-0.0465**	-0.7395	-2.795	-20.26	-0.0004	0.0050	-8.98×10^{-6}	0.0005
	(0.0198)	(0.0179)	(0.4145)	(1.787)	(44.35)	(0.0008)	(0.0176)	(4.6×10^{-5})	(0.0010)
Treatment × Year = 2015	-0.0905***	-0.0598**	-1.138*	-4.056*	-52.25	-0.0009	-0.0088	-3.57×10^{-5}	-0.0003
	(0.0217)	(0.0200)	(0.4526)	(1.757)	(35.58)	(0.0008)	(0.0155)	(4.64×10^{-5})	(0.0009)
Treatment × Year = 2016	-0.0487*	-0.0233	0.0915	-1.120	56.96	0.0002	0.0459*	1.88×10^{-5}	0.0031*
	(0.0199)	(0.0177)	(0.5339)	(1.511)	(44.69)	(0.0008)	(0.0200)	(4.32×10^{-5})	(0.0013)
Treatment × Year = 2017	-0.0407*	-0.0153	1.115	-1.644	63.17	-4.52×10^{-5}	0.0397	1.93×10^{-7}	0.0016
	(0.0188)	(0.0167)	(0.7152)	(1.638)	(100.7)	(0.0008)	(0.0399)	(4.47×10^{-5})	(0.0017)
Treatment × Year = 2018	-0.0514**	-0.0239	0.1751	-2.011	5.162	-7.52×10^{-5}	0.0234	1.85×10^{-6}	0.0013
	(0.0199)	(0.0185)	(0.5812)	(1.591)	(50.75)	(0.0008)	(0.0236)	(4.51×10^{-5})	(0.0013)
Treatment × Year = 2019	-0.0526**	-0.0303	0.0218	-2.285	21.14	-0.0003	0.0200	-5.57×10^{-6}	0.0014
	(0.0194)	(0.0178)	(0.5483)	(1.603)	(45.11)	(0.0008)	(0.0232)	(4.58×10^{-5})	(0.0015)
Sex (male)		-0.0032	-0.0615	-0.0032	-0.0613	-0.0032	-0.0617	-0.0032	-0.0628
		(0.0028)	(0.0493)	(0.0028)	(0.0491)	(0.0028)	(0.0490)	(0.0028)	(0.0491)
State School		-0.0074*	-0.5176**	-0.0071	-0.5108**	-0.0071	-0.5112**	-0.0070	-0.5087**
		(0.0037)	(0.1582)	(0.0037)	(0.1584)	(0.0037)	(0.1593)	(0.0037)	(0.1598)
Rural		-0.0864***	-1.056***	-0.0870***	-1.068***	-0.0874***	-1.074***	-0.0877***	-1.073***
		(0.0106)	(0.1208)	(0.0106)	(0.1220)	(0.0106)	(0.1218)	(0.0106)	(0.1216)
High Economic Dependence		-0.0555***	-0.5718***	-0.0552***	-0.5663***	-0.0552***	-0.5667***	-0.0553***	-0.5679***
		(0.0134)	(0.1216)	(0.0134)	(0.1213)	(0.0134)	(0.1214)	(0.0134)	(0.1213)
Households w/o Water		-0.0457***	-0.5016***	-0.0461***	-0.5019***	-0.0461***	-0.5015***	-0.0460***	-0.5021***
		(0.0054)	(0.0605)	(0.0054)	(0.0605)	(0.0054)	(0.0604)	(0.0054)	(0.0605)
Rural × After 2014		0.0672***	0.4290***	0.0683***	0.4515***	0.0687***	0.4685***	0.0692***	0.4697***
		(0.0072)	(0.1227)	(0.0072)	(0.1232)	(0.0072)	(0.1230)	(0.0072)	(0.1233)
Treatment:	Binary	Binary	Binary	inverse_distance	inverse_distance	PGA	PGA	PGA_squared	PGA_squared
Fixed-effects									
District	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics									
Observations	35,720	34,717	27,901	34,717	27,901	34,717	27,901	34,717	27,901
Squared Correlation	0.14408	0.15172	0.15886	0.15150	0.15883	0.15144	0.15856	0.15136	0.15840
Pseudo R^2	0.63254	1.1016	0.19145	1.0998	0.19103	1.0994	0.19072	1.0988	0.19057
BIC	19,341.6	15,542.4	21,586.0	15,551.6	21,593.3	15,554.0	21,598.4	15,557.1	21,601.0

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

A.3.2 School Census

Table 10: Main School Census Results - Heterogeneity Analysis

Dependent Variables:	Dropout Rate		Dropout Rate	
Sample Type:	Male	Female	Primary School	Secondary School
Model:	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Variables				
Treatment × Year = 2004	0.0013	-0.0034	0.0014	-0.0004
	(0.0017)	(0.0023)	(0.0023)	(0.0028)
Treatment × Year = 2005	0.0017	0.0010	0.0027	0.0025
	(0.0016)	(0.0020)	(0.0017)	(0.0030)
Treatment × Year = 2007	0.0173***	0.0175***	0.0205***	0.0100
	(0.0042)	(0.0051)	(0.0046)	(0.0055)
Treatment × Year = 2008	0.0085***	0.0090***	0.0161***	-0.0008
	(0.0020)	(0.0026)	(0.0027)	(0.0039)
Treatment × Year = 2009	0.0159***	0.0152***	0.0182***	0.0125**
	(0.0020)	(0.0028)	(0.0023)	(0.0041)
Treatment × Year = 2010	0.0087***	0.0094***	0.0143***	-0.0012
	(0.0023)	(0.0024)	(0.0026)	(0.0052)
Treatment × Year = 2011	0.0131***	0.0143***	0.0175***	0.0040
	(0.0031)	(0.0035)	(0.0033)	(0.0043)
Treatment × Year = 2012	0.0137***	0.0167***	0.0188***	0.0065
	(0.0028)	(0.0033)	(0.0031)	(0.0039)
Treatment × Year = 2013	0.0154***	0.0176***	0.0203***	0.0065
	(0.0028)	(0.0037)	(0.0032)	(0.0038)
Treatment × Year = 2014	0.0101**	0.0125**	0.0147***	0.0020
	(0.0032)	(0.0041)	(0.0036)	(0.0042)
Treatment × Year = 2015	0.0112***	0.0145***	0.0162***	0.0031
	(0.0032)	(0.0043)	(0.0034)	(0.0047)
After 2014 x Rural	-0.0269***	-0.0334***	-0.0277***	-0.0264***
	(0.0009)	(0.0013)	(0.0010)	(0.0013)
Fixed-effects				
School	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	454,216	455,739	355,255	104,542
R^2	0.43590	0.44855	0.50590	0.52459
Within R^2	0.00831	0.00822	0.01277	0.00815

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 11: Main School Census Results - Heterogeneity Analysis

Dependent Variable:	Failure Rate			
Sample Type:	Male	Female	Primary School	Secondary School
Model:	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Variables				
Treatment × Year = 2004	-0.0130	-0.0033	-0.0022	-0.0161
	(0.0069)	(0.0034)	(0.0027)	(0.0121)
Treatment × Year = 2005	-0.0084	0.0055	-0.0006	-0.0023
	(0.0058)	(0.0039)	(0.0021)	(0.0089)
Treatment × Year = 2007	0.0066	0.0176***	0.0195***	0.0014
	(0.0063)	(0.0038)	(0.0043)	(0.0078)
Treatment × Year = 2008	-0.0055	0.0041	0.0051*	-0.0108
	(0.0061)	(0.0037)	(0.0025)	(0.0100)
Treatment × Year = 2009	-0.0017	0.0079*	0.0107***	-0.0089
	(0.0057)	(0.0032)	(0.0019)	(0.0084)
Treatment × Year = 2010	-0.0026	0.0105**	0.0108***	-0.0082
	(0.0070)	(0.0037)	(0.0022)	(0.0102)
Treatment × Year = 2011	-0.0020	0.0104*	0.0118***	-0.0089
	(0.0065)	(0.0044)	(0.0022)	(0.0103)
Treatment × Year = 2012	-0.0084	0.0061*	0.0091***	-0.0181*
	(0.0058)	(0.0029)	(0.0019)	(0.0082)
Treatment × Year = 2014	-0.0027	0.0120**	0.0166***	-0.0133
	(0.0057)	(0.0037)	(0.0024)	(0.0085)
Treatment × Year = 2015	0.0045	0.0185***	0.0212***	-0.0033
	(0.0074)	(0.0036)	(0.0022)	(0.0100)
After 2014 x Rural	-0.0169***	-0.0212***	-0.0263***	0.0067***
	(0.0011)	(0.0011)	(0.0011)	(0.0018)
Fixed-effects				
School	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	412,902	414,342	323,573	94,383
R^2	0.44094	0.43926	0.55276	0.37959
Within R^2	0.00227	0.00405	0.01035	0.00070

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 12: Main School Census Results - Robustness Checks

Dependent Variable:	Dropout Rate			
Model:	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Variables				
Treatment × Year = 2004	8.08×10^{-5}	-44.35***	-0.0001*	-3.55×10^{-6}
	(0.0020)	(11.94)	(5.65×10^{-5})	(2.73×10^{-6})
Treatment × Year = 2005	0.0025	6.082	4.82×10^{-5}	3.3×10^{-6}
	(0.0019)	(7.953)	(4.11×10^{-5})	(2.34×10^{-6})
Treatment × Year = 2007	0.0164***	80.50***	0.0004***	2.44×10^{-5}***
	(0.0044)	(17.19)	(7.31×10^{-5})	(4.1×10^{-6})
Treatment × Year = 2008	0.0097***	54.00***	0.0002***	1.4×10^{-5}***
	(0.0025)	(14.82)	(6.03×10^{-5})	(3.58×10^{-6})
Treatment × Year = 2009	0.0165***	107.5***	0.0005***	2.44×10^{-5}***
	(0.0026)	(19.56)	(6.28×10^{-5})	(3.31×10^{-6})
Treatment × Year = 2010	0.0085**	84.87***	0.0004***	1.74×10^{-5}***
	(0.0030)	(23.37)	(6.45×10^{-5})	(3.28×10^{-6})
Treatment × Year = 2011	0.0126***	142.0***	0.0006***	2.64×10^{-5}***
	(0.0033)	(27.78)	(8.77×10^{-5})	(3.97×10^{-6})
Treatment × Year = 2012	0.0145***	135.7***	0.0006***	2.85×10^{-5}***
	(0.0030)	(24.16)	(7.52×10^{-5})	(3.44×10^{-6})
Treatment × Year = 2013	0.0153***	160.3***	0.0006***	3.1×10^{-5}***
	(0.0031)	(27.43)	(8.56×10^{-5})	(3.77×10^{-6})
Treatment × Year = 2014	0.0099**	105.3***	0.0004***	2.11×10^{-5}***
	(0.0035)	(22.40)	(6.66×10^{-5})	(3.18×10^{-6})
Treatment × Year = 2015	0.0112**	133.7***	0.0005***	2.44×10^{-5}***
	(0.0036)	(26.70)	(7.9×10^{-5})	(3.58×10^{-6})
After 2014 x Rural	-0.0284***	-0.0299***	-0.0301***	-0.0305***
	(0.0010)	(0.0009)	(0.0009)	(0.0009)
Treatment	Binary	Inverse Distance	PGA	PGA Squared
Fixed-effects				
School	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	459,797	560,302	560,302	560,302
R^2	0.51093	0.52259	0.52227	0.52180
Within R^2	0.01192	0.01593	0.01527	0.01430

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 13: Main School Census Results - Robustness Checks

Dependent Variable:	Failure Rate			
Model:	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Variables				
Treatment × Year = 2004	-0.0082	-66.70*	-0.0002	-1.17×10^{-5}
	(0.0049)	(28.70)	(0.0001)	(6.52×10^{-6})
Treatment × Year = 2005	-0.0015	-19.77	-7.2×10^{-5}	-4.33×10^{-6}
	(0.0038)	(22.95)	(8.32×10^{-5})	(4.75×10^{-6})
Treatment × Year = 2007	0.0119**	35.85	0.0003**	1.62×10^{-5}**
	(0.0044)	(26.95)	(9.52×10^{-5})	(5.56×10^{-6})
Treatment × Year = 2008	-0.0011	-12.23	-2.58×10^{-5}	-1.96×10^{-6}
	(0.0045)	(25.91)	(0.0001)	(6.05×10^{-6})
Treatment × Year = 2009	0.0029	35.48	0.0002	6.71×10^{-6}
	(0.0035)	(26.98)	(9.13×10^{-5})	(4.92×10^{-6})
Treatment × Year = 2010	0.0037	38.58	0.0002	7.27×10^{-6}
	(0.0049)	(36.34)	(0.0001)	(6.52×10^{-6})
Treatment × Year = 2011	0.0040	78.75	0.0003*	1.12×10^{-5}
	(0.0049)	(42.01)	(0.0001)	(6.84×10^{-6})
Treatment × Year = 2012	-0.0014	31.67	0.0001	2.09×10^{-6}
	(0.0036)	(30.34)	(9.82×10^{-5})	(5.14×10^{-6})
Treatment × Year = 2014	0.0045	22.66	0.0001	6.81×10^{-6}
	(0.0040)	(28.73)	(9.95×10^{-5})	(5.49×10^{-6})
Treatment × Year = 2015	0.0114*	84.89*	0.0004***	1.87×10^{-5}**
	(0.0048)	(37.23)	(0.0001)	(6.27×10^{-6})
After 2014 x Rural	-0.0189***	-0.0209***	-0.0209***	-0.0211***
	(0.0011)	(0.0010)	(0.0010)	(0.0009)
Treatment	Binary	Inverse Distance	PGA	PGA Squared
Fixed-effects				
School	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	417,956	509,221	509,221	509,221
R^2	0.47583	0.50124	0.50120	0.50111
Within R^2	0.00368	0.00448	0.00440	0.00422

Clustered (District) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

A.3.3 Test Scores

Table 14: Balance Pre-Matching

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
Propensity Score	6.270602e-01	0.1340085	1.7552740	2.1833989	0.4118109	0.6694293	NA
Mean Literacy Score	5.391098e+02	467.9103312	1.3153809	0.4187984	0.2517475	0.4369472	NA
Mean Mathematics Score	5.180326e+02	483.9085981	0.6187434	0.4707293	0.1230756	0.2549806	NA
Variance Mathematics	5.459987e+03	5638.2146015	-0.0343015	0.7486986	0.0227113	0.0483831	NA
Variance Literacy	5.356465e+03	4638.3179787	0.1617830	0.8728349	0.0844455	0.1445361	NA
Skewness Mathematics	5.000190e-02	-0.0795838	0.2143734	0.9025354	0.0546107	0.0856107	NA
Skewness Literacy	1.421941e-01	-0.0113095	0.2625002	0.8423065	0.0673955	0.1121231	NA
Number of Studen ts	1.315301e+01	11.4325833	0.1210658	1.0007442	0.0168741	0.0772727	NA
Proportion Male	5.116718e-01	0.5109456	0.0032637	0.9786765	0.0078282	0.0195587	NA
Rural Area	8.733520e-02	0.6335645	-1.9347491	NA	0.5462293	0.5462293	NA
State School	3.495763e-01	0.8398748	-1.0282328	NA	0.4902985	0.4902985	NA
Average Income in School District (2006)	3.409458e+04	13962.6536701	1.2867891	3.4404993	0.3647373	0.6787693	NA

Table 15: Balance Post-Matching

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
Propensity Score	6.270602e-01	0.3247392	1.0762689	1.8224054	0.1561335	0.4983522	1.0762697
Mean Literacy Score	5.391098e+02	517.1673375	0.4053779	0.6334311	0.0824913	0.1624294	1.1652476
Mean Mathematics Score	5.180326e+02	507.9631024	0.1825816	0.6916509	0.0404921	0.1014595	1.1891140
Variance Mathematics	5.459987e+03	5625.6232316	-0.0318782	0.9992770	0.0150295	0.0313089	0.8820197
Variance Literacy	5.356465e+03	5169.3713681	0.0421481	1.1631814	0.0129099	0.0287194	0.9138989
Skewness Mathematics	5.000190e-02	-0.0062988	0.0931382	0.9026404	0.0234510	0.0435499	1.1557244
Skewness Literacy	1.421941e-01	0.1258670	0.0279203	0.9116621	0.0071664	0.0240113	1.1331741
Number of Studen ts	1.315301e+01	15.7057439	-0.1796344	0.5746511	0.0192864	0.0574388	1.0488500
Proportion Male	5.116718e-01	0.5103820	0.0057964	0.9716048	0.0049165	0.0143597	1.1102848
Rural Area	8.733520e-02	0.2135122	-0.4469201	NA	0.1261770	0.1261770	0.7154058
State School	3.495763e-01	0.5974576	-0.5198460	NA	0.2478814	0.2478814	0.9098539
Average Income in School District (2006)	3.409458e+04	22894.4249053	0.7158897	4.8283690	0.0847213	0.4865819	0.7372041

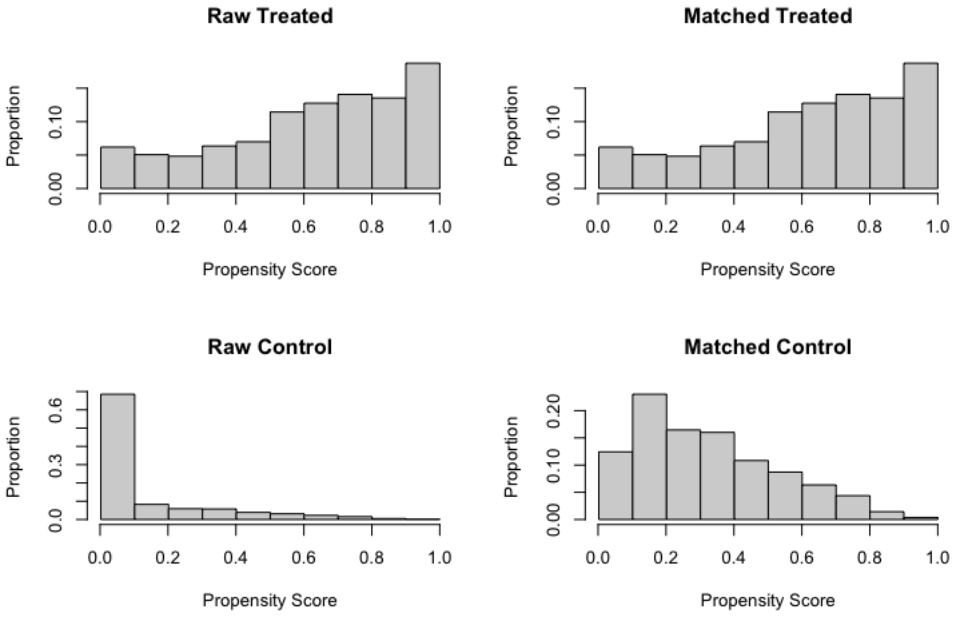


Figure 5: Distribution of Propensity Score

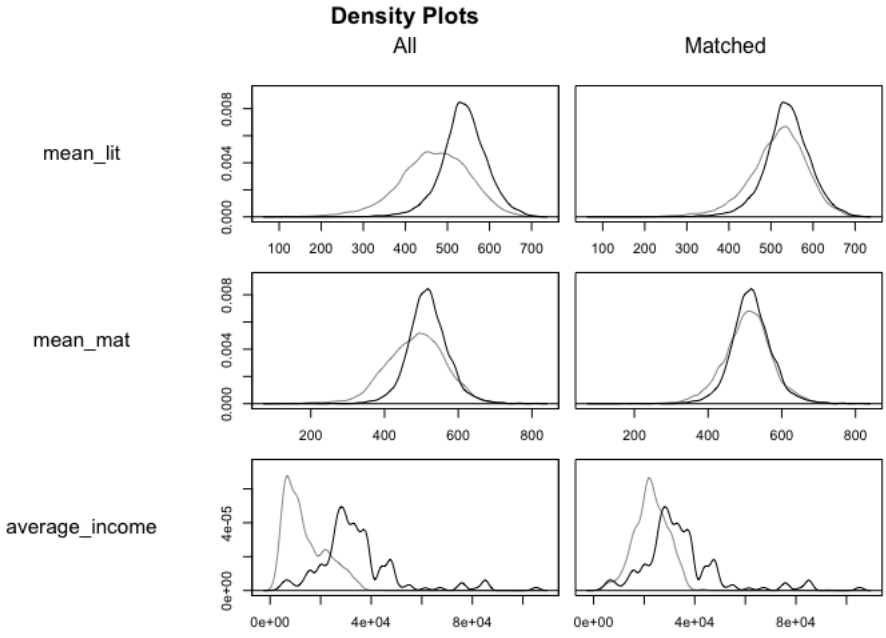


Figure 6: Distribution of selected matching covariates

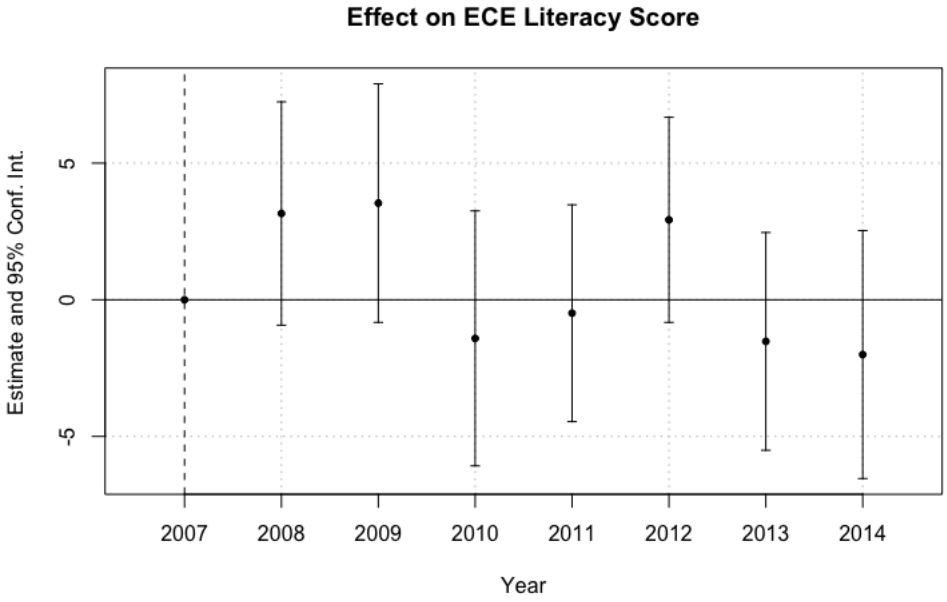


Figure 7

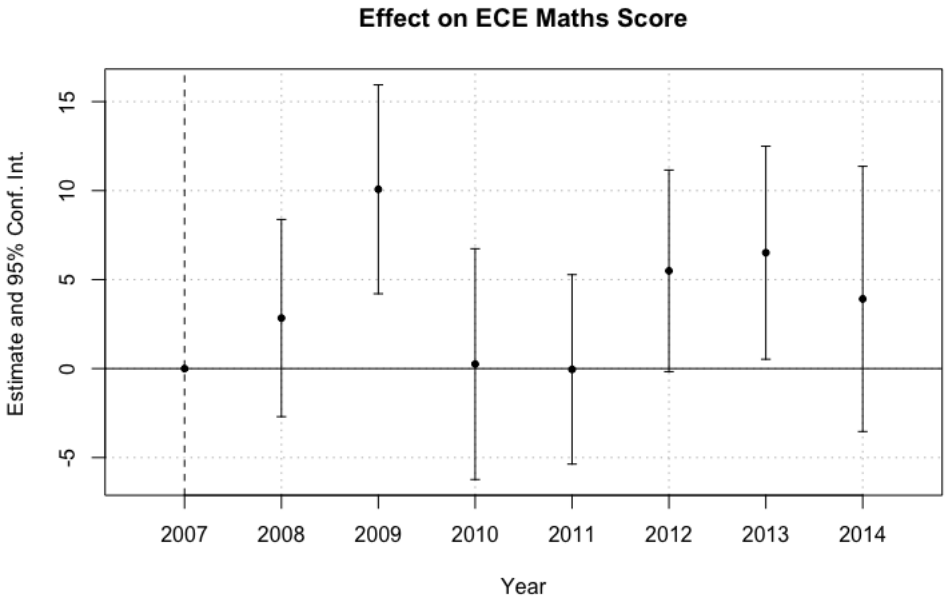


Figure 8

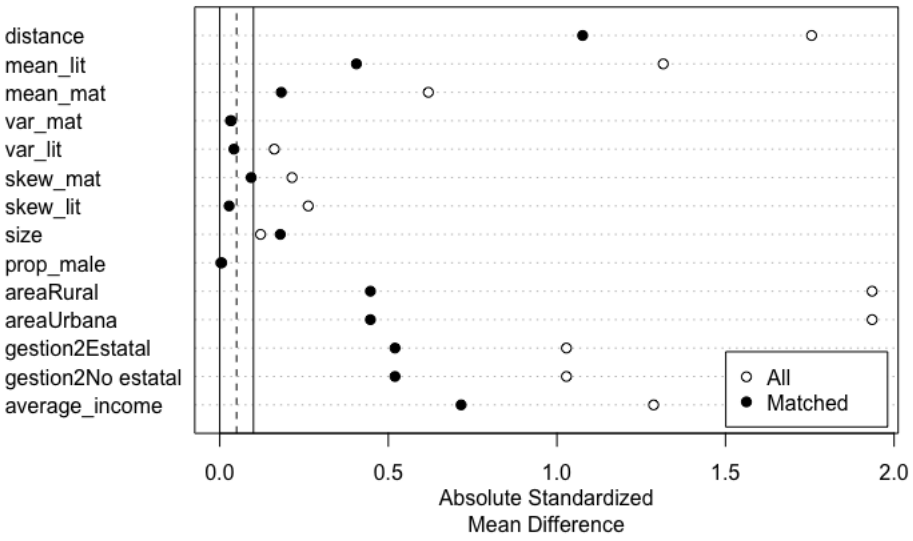


Figure 9

Table 16: ECE Standardized Test Effects

Dependent Variables: Model:	Median ECE Literacy (1)	Median ECE Maths (2)
Variables		
Treatment × Year = 2008	1.777	0.9686
	(2.064)	(2.837)
Treatment × Year = 2009	2.214	8.392**
	(2.200)	(3.007)
Treatment × Year = 2010	-2.171	0.0086
	(2.350)	(3.300)
Treatment × Year = 2011	-1.474	-0.4583
	(2.001)	(2.689)
Treatment × Year = 2012	2.146	4.827
	(1.872)	(2.823)
Treatment × Year = 2013	-2.309	5.835
	(1.974)	(2.983)
Treatment × Year = 2014	-2.736	3.332
	(2.249)	(3.696)
Treatment × Year = 2015	-6.122**	-2.042
	(2.281)	(3.487)
Treatment × Year = 2016	-9.545***	-9.595*
	(2.382)	(3.922)
Size of School (Students)	0.0389**	0.2530***
	(0.0147)	(0.0237)
% Male	-10.18***	5.485
	(2.443)	(3.432)
	Binary	Binary
Fixed-effects		
School	Yes	Yes
Year	Yes	Yes
Fit statistics		
Observations	26,084	26,084
R^2	0.80285	0.67504
Within R^2	0.00425	0.00827

Clustered (School) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 17: ECE Standardized Test Effects

Dependent Variables:	Median ECE Literacy	Median ECE Maths
Model:	(1)	(2)
Variables		
Treatment × Year = 2008	3.158	2.839
	(2.085)	(2.826)
Treatment × Year = 2009	3.537	10.07***
	(2.227)	(2.994)
Treatment × Year = 2010	-1.412	0.2502
	(2.380)	(3.308)
Treatment × Year = 2011	-0.4908	-0.0437
	(2.024)	(2.716)
Treatment × Year = 2012	2.925	5.491
	(1.915)	(2.890)
Treatment × Year = 2013	-1.522	6.509*
	(2.033)	(3.053)
Treatment × Year = 2014	-2.005	3.914
	(2.314)	(3.802)
Size of School (Students)	-0.0001	0.1861***
	(0.0161)	(0.0259)
% Male	-9.497***	6.838
	(2.658)	(3.704)
	Binary	Binary
Fixed-effects		
School	Yes	Yes
Year	Yes	Yes
Fit statistics		
Observations	22,089	22,089
R^2	0.81261	0.67603
Within R^2	0.00242	0.00512

Clustered (School) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*