Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2022/06, 36 pàg.
Working Paper 2022/06, 36 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2022/04, 36 pàg.
Working Paper 2022/04, 36 pag.

"Earthquake exposure and schooling: impacts and mechanisms"

Khalifany-Ash Shidiqi, Antonio Di Paolo and Álvaro Choi

UBIREA
Research Institute of Applied Economics

AQR
Regional Quantitative Analysis Research Group

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública
UNIVERSITAT DE BARCELONA

WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

AQR
Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Earthquake exposure and schooling: impacts and mechanisms

Abstract

Natural disasters are a significant threat to human development. In this paper, we analyze the effects of being exposed to a strong earthquake during school age on schooling outcomes. We merge geolocated data about the intensity of the shock at the district level with individual information from the Indonesia Family Life Survey. The identification strategy exploits variation in exposure to the natural shock by birth cohort and district of residence, considering as the treated group individuals who were residing in affected districts while they were in school age. Earthquake exposure reduces years of schooling by somewhat less than one year and negatively affects the probability of completing compulsory education but does not alter the chances of enrolling into post-compulsory education. Falsification analysis and several robustness checks corroborate the causal interpretation of our findings. The analysis of the potential mechanisms indicates that induced migration and casualties occurring at the family level as a consequence of the earthquake do not seem to play a relevant role. However, damages in educational infrastructures do represent a relevant channel through which natural disasters harm human capital formation. Part of the overall impact of the earthquake represents a delay in schooling progression, but a substantial share of its effect consists in a permanent loss of human capital among affected individuals.

JEL Classification: I25, I24, O15, Q54.

Keywords: Natural disasters, Earthquake, Schooling, Educational infrastructures.

Khalifany-Ash Shidiqi (corresponding author): Department of Economics, Universitas Muhammadiyah Yogyakarta, Faculty of Economics and Business, Indonesia & University of Barcelona, Faculty of Economics and Business, 690 Diagonal Av. 08034, Barcelona, Spain. Email: khalifany.ash@umy.ac.id

Antonio Di Paolo: AQR-IREA, University of Barcelona, Faculty of Economics and Business, 690 Diagonal Av. 08034, Barcelona, Spain. Email: antonio.dipaolo@ub.edu

Álvaro Choi: University of Barcelona, Faculty of Economics and Business, 690 Diagonal Av. 08034, Barcelona, Spain. Email: alvarochoi@ub.edu

Acknowledgements and funding

We are very grateful to André Groeger and Elisabetta Aurino for their valuable suggestions. Comments received at the LESE Workshop, SAEe, AEDE Meeting and at the IRSA International Conference are also welcome. Álvaro Choi acknowledges funding from the Project EDU206-76414-R, Ministry of Economy and Competitiveness. The usual disclaimers apply.

1. Introduction

Natural disasters are a major threat to human development. According to the United Nations Office for Disarmament Affairs (2020), up to 7,348 events were recorded during the first two decades of this century, claiming approximately 60,000 lives per annum, affecting more than 4 billion people, and with an economic cost of 2.97 trillion 2019 US$. Worryingly, although there have been improvements in disaster preparedness and response, which has reduced the loss of lives in single-hazard events, there has been an essential rise in climate-related disasters during the 2000-2019 period (CRED-UNDRR, 2020). Apart from the costs in lives and the immediate economic impact, natural disasters can affect a wide range of outcomes (Baez et al., 2010), including economic growth (Noy, 2009; Cavallo, Bank, et al., 2013; McDermott et al., 2014; Philipp Heger and Neumayer, 2019), poverty (Baez and Santos, 2008) labor market outcomes (Di Pietro and Mora, 2015; Kirchberger, 2017; Groen et al., 2019), electoral results (Gasper and Reeves, 2011; Masiero and Santarossa, 2020), crime (Hombrados, 2020), expenditure, spending behavior and income (Sulistyaningrum, 2015; Gignoux and Menéndez, 2016; Filipski et al., 2019), health (Cairo et al., 2010; Zhang et al., 2011; Bustelo et al., 2012), and religiosity (Belloc et al., 2016; Bentzen, 2019).

This paper investigates the effects of being exposed to natural disasters during school age on individuals' human capital formation, proxied by schooling attainments. The disruptive effects on education may operate through different channels in the form of negative income shocks and life losses at the household level, forced displacement of families, mental health and psychological effects, as well as the destruction of education facilities, among others (Kousky, 2016; O'Toole and Friesen, 2016; Esnard *et al.*, 2018). Given the close link between education and economic growth (Krueger and Lindahl, 2001), this may be one of the main channels through which natural disasters hinder the development of countries. Moreover, previous literature has shown that the negative impact on educational outcomes varies depending on the type of natural disaster (Nguyen and Pham, 2018) and the grade of development of the country, being greater in low-income countries (Toya and Skidmore, 2007; McDermott et al., 2014). However, there is still a lot to learn about the medium and long-term effects of these shocks on human capital formation and the channels that actually drive this relationship.

We analyze the impact of a strong earthquake that took place in 2006 in Yogyakarta, located in the Java Island of Indonesia. We provide novel evidence on the medium to long-term impact of a huge earthquake on schooling outcomes for a Southern Asian country that is frequently affected by this kind of natural shocks, but at the same time is making a great effort to improve the quality of education and promote human capital investment. The empirical analysis combines several data sources. On the one hand, we exploit geolocated information from the U.S. Geological Survey to capture geographical exposure to the earthquake and its intensity through the Modified Mercalli Intensity Index (MMI), measured at the district level. On the other hand, we use individual and family level information taken from the Indonesia Family Life Survey (IFLS). We mainly use the 2014 wave of the IFLS survey, which means that we measure education achievements eight years after the natural shock, although we also take advantage of previous waves for falsification analyses and other robustness checks. We identify the causal effect of earthquake exposure by exploiting variation by birth cohort and district of residence in 2006. That is, we compare completed education between individuals who were in school age in 2006 and who were living in affected and unaffected areas, taking older cohorts who were already out of school at that time as further control for idiosyncratic differences related to the district of residence. This identification strategy relies on the assumption that there are no district-specific cohort level unobservable determinants of education attainments. Our main outcome consists of years of schooling, although we also estimate the effect on the probability of completing compulsory education and on post-compulsory school enrollment.

Moreover, we explore the heterogeneous effects of earthquake exposure according to a battery of individual and family characteristics, ranging from age at exposure, gender, religion, ethnicity, parental education, number of siblings and birth order. This indeed represents a first contribution of our work to the literature, since none of the existing papers provided a heterogeneity analysis over so many dimensions. Most importantly, we carefully analyze several potential mechanisms behind the relationship between earthquake exposure and attained schooling, which represents a significant added value of our work. Specifically, first, using retrospective information about the entire migration history at the individual level, we are able to gauge the relevance of induced migration (i.e., post-earthquake) as a potential channel. Second, the availability of a specific set of variables contained in the 2007 wave of the IFLS regarding earthquake-related damages and injuries enables us to examine the role played by different possible issues occurred at the family level such as deaths, injuries, financial losses, etc. Third, and most importantly, using administrative information on school buildings at the district level, we provide for the first-time direct evidence about damages in educational infrastructures produced by the earthquake as a mechanism at work. Indeed, the shock on educational infrastructure is indeed a relevant albeit unexplored mechanism, especially in the light of the findings obtained by (Herrera-Almanza and Cas, 2021), indicating that school infrastructure recovery programmes may mitigate the detrimental effects of natural disasters on human capital accumulation. Finally, exploiting information about current school attendance, we also provide suggestive evidence about whether the overall impact of the earthquake represents a permanent, long-standing loss of human capital, or it (only) generates a certain transitory delay in schooling progression.

Our results show that earthquake exposure during school age generates a reduction of somewhat less than one year of schooling among affected individuals (0.74 years in our baseline estimation). Our findings further indicate that individuals exposed to the earthquake are approximately 10-11 percentage points less likely to complete primary and junior high school, respectively. Additionally, we do not find any statistical evidence of the impact on post-compulsory schooling enrollment rates. All the results from falsification exercises and sensitivity checks provide evidence in favor of the causal interpretation of our main findings, indicating that earthquake exposure during school-age harms human capital formation in a causal sense. We also find that the impact is greater for younger individuals who were still attending compulsory schooling when the earthquake took place. The detrimental effect of exposure is also more pronounced for individuals with low educated mothers, for those with fewer siblings and for first and second born individuals. Moreover, the analysis of potential mechanisms highlights that selective migration and household casualties are unlikely to be the main driver of the results. On the contrary, earthquake-related disruption of school infrastructure seems to be responsible for the loss in years of schooling experienced by younger cohorts affected by the natural disaster. Finally, we also show that part of the overall impact of earthquake exposure represents a (possibly) transitory delay in schooling progression, which is likely to be due to the afro mentioned disruption of schooling infrastructures. However, most of the overall negative effect indeed consists in a permanent loss of human capital among affected cohorts of individuals, who were in school age when the natural disaster occurred.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the literature on natural disasters and educational outcomes. Section 3 describes the data and descriptive statistics. Section 4 discusses the empirical models used in this study, and Section 5 presents the main findings. Section 6 concludes by discussing the implications of the empirical findings.

2. Earthquakes and educational outcomes

Natural disasters affect human capital accumulation through several channels (Baez et al., 2010; McDermott, 2012; Kousky, 2016; O'Toole and Friesen, 2016; Esnard et al., 2018; Rush, 2018), and analyzing their negative impacts on education is of crucial importance, especially for developing countries. The effects of natural disasters on educational outcomes depend on its type, the country's degree of development (Nguyen and Pham, 2018), and damage paths. In this paper, we focus on the impact of a specific type of natural disasters: a strong earthquake.

The understanding of the impact of earthquakes on educational outcomes at different ages has grown in the last decades. Caruso and Miller (2015) find that the exposure to the 1970 Ancash earthquake during early childhood or in utero reduces educational attainment. Similarly, Paudel and Ryu (2018) investigate the effects of the 1988 Nepal earthquake on human capital accumulation in infants exposed to disaster at a very young age. They find that infants born in areas severely affected by the earthquake achieved lower educational attainment and less school completion in middle and high school. Additionally, Gomez and Yoshikawa (2017) find that the 2010 Chilean earthquake decreased test scores in pre-literacy and early language assessments for preschool children.

Exposure to an earthquake at the primary school level age generates negative effects too. Wang *et al.* (2017) show that the 1976 Tangshan earthquake led to a reduction in schooling years of around 14% to 21% when exposed during primary school age. Bustelo *et al.* (2012) compare the outcomes of students aged 6 to 10 in 2005 in the most affected region –Quindío- to those from less-affected regions. Primary schooling enrolment was lower for children in the most affected areas –malnutrition at early stages in life and the lack of economic resources being two of the possible explanations. Moreover, Andrabi *et al.* (2021) found that children aged 3 to 11 at the time of the 2005 Northern Pakistan earthquake scored significantly worse on academic tests. Interestingly, they found that this was not the case for children whose mothers had completed at least the primary education level.

There is also a certain amount of evidence on the effects of suffering an earthquake on secondary school attainment. For example, Cuaresma (2010) analyzes the impact of this type of geological disasters on secondary school enrollment in a cross-country framework. After averaging macro-level data, he concludes that geophysical disasters negatively affect secondary school enrollment rates between countries but not necessarily within countries. Rush (2018) confirms this finding by using the district level's secondary enrollment rate and focusing on different natural disasters (including earthquakes) occurred in a single country, Indonesia. He finds that the impact on secondary school enrollment depends on the paths of disaster damage. However, other studies using individual level data also point to detrimental effects on secondary school outcomes. For example, Paudel and Ryu (2018) assess the long-term effects of the Nepalese 1988 earthquake on the lower and upper secondary school completion rates. Their difference-in-differences model shows that children born in the affected areas showed lower completion rates in both levels (13.8% and 10% lower, respectively). Interestingly, they also demonstrate that this impact was heterogeneous across the population: while the negative impact was more acute for students from lower-caste households, it was null for students from higher caste households. Furthermore, Park et al. (2015) report that the household-level shocks due to the 2008 Sichuan earthquake worsened the child's psychosocial and family environment, reducing secondary school students' cognitive and non-cognitive skills.

Finally, other authors assess the effects at the higher education level. Di Pietro (2018) examines, using a difference-in-differences model, the immediate effect of the L'Aquila 2009 earthquake on the academic performance of the students from the local university. He finds that the earthquake significantly reduced the probability that a student would graduate on time and increased students' probability of dropping out during the academic year in which this natural disaster occurred. However, Cerqua and Di Pietro (2017) point out that the impact of that same earthquake on first-year enrolment at the University of L'Aquila was statistically not significant during the three years after the earthquake. They did, however, identify compositional changes in the first-year population.

This paper assesses the medium and long-term effects of the 2006 Yogyakarta earthquake at the compulsory and post-compulsory education levels. The literature assessing the impact of natural disasters on educational outcomes in Southeast Asia is scarce, and even scarcer for earthquakes. Nguyen and Pham (2018) analyzed the impact of climate disasters (i.e. drought, flood, frost, and hailstorms) on educational attainment from countries in three different continents, and South East Asia is one of them. Rush, (2018) combined climate and geological disasters (floods, strong winds, droughts, and landslides) and uses aggregated data at the district level to analyze the impact on enrollment rates in Indonesia. Evidence on the effects of the huge earthquake occurred in Yogyakarta in 2006 is even scarcer. As far as we know, the only study analyzing this earthquake's impact on students' educational outcomes is the paper by Sulistyaningrum (2017), who focused on test scores. Using a difference-in-differences framework, she found that: 1) the earthquake decreased the test scores of all children of age 11 or in their last year of primary school, the impact being larger for those who declared being in the affected area and in rural areas; 2) the negative impact slightly faded out one year after the earthquake; 3) there are no differences across gender, and 4) the negative impact is greater for children in the lowest quantile of test scores.

Our study takes the analysis four steps further: First, we analyze the medium to long-term impact of the earthquake on medium and long-term educational outcomes, considering years of schooling and education levels (enrollment and completion). Second, we use a more credible identification strategy that combines the use of the MMI and the residential history of citizens and exploits variation in exposure by birth cohort and district of residence at the time of the earthquake. Moreover, we present a battery of sensitivity checks and falsification exercises to validate the underlying hypothesis behind our identification strategy. Third, we allow for the heterogeneous effects of many individual and family characteristics. Finally and most importantly, we assess the relevance of different potential mechanisms through which the earthquake affected educational outcomes.

3. Data and Descriptive Statistics

Our empirical analysis focuses on the Java Island, the most populated island of Indonesia, and its capital, Jakarta, the country's most populated city, where the 2006 earthquake took place. Indonesia is a country that is prone to seismic upheaval due to its location on the so-called Pacific "Ring of Fire," volcanic arcs and fault lines surrounding the Pacific Basin. Between 2005-2015, there were more than 1,800 natural disasters occurred in Indonesia (Amri et al., 2018). The most destructive was the earthquake on the 27th of May 2006 at 05:55:03 local time with a magnitude of 5.9 on the Richter scale located in the southern part of Yogyakarta Province. It severely affected five districts in Yogyakarta and Central Java Province, respectively. According to Resosudarmo et al. (2012), up to 5,716 people lost their lives and it destroyed over 150,000 homes. The estimated cost was more than USD 3.1 billion in damage and losses (World Bank, 2007).

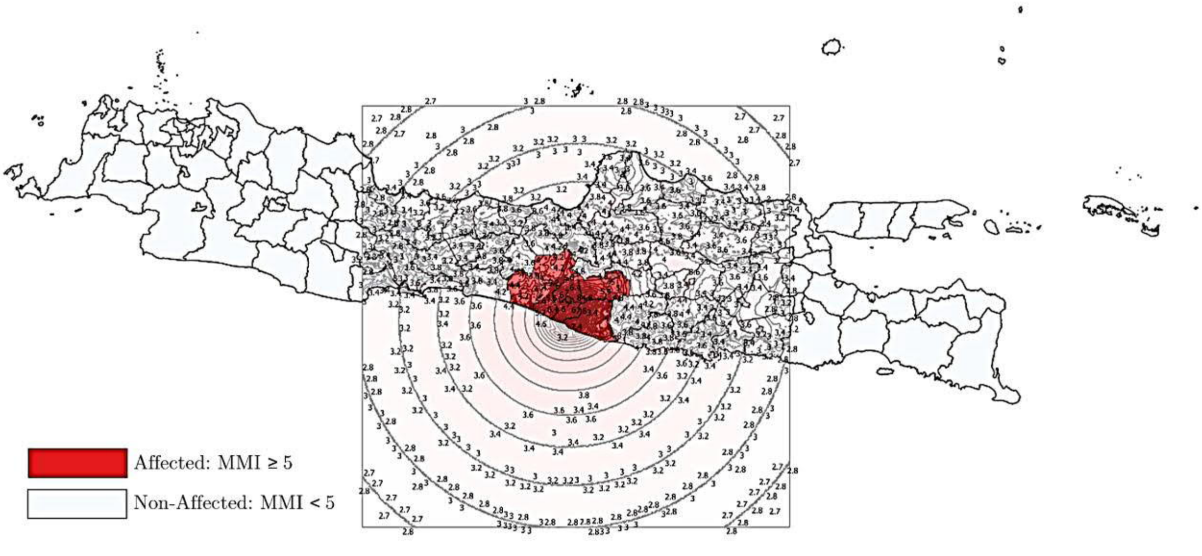
Furthermore, almost 3,000 educational facilities were damaged or destroyed. Bantul District, in Yogyakarta province, was one of the districts worst affected, with 917 -more than 90%- of its education buildings being damaged or destroyed. In Central Java, 558 buildings were damaged or destroyed, while the Klaten district experienced the highest level of damage in the province, with 298 buildings badly damaged, accounting for around 27% of all buildings (Bappenas, 2006). Bappenas (2006) joint team reports that the quality of school buildings was a significant factor in the high level of destruction. Many schools, especially in rural areas, were built in the 1970s without considering earthquake-resistant structures and other safety standards. Indeed, we carefully analyze the role of earthquake-related disruption of educational infrastructures, as explained below.

In this paper we exploit different data sources. First, to retrieve information about the geographical exposure to the earthquake and its intensity, we obtained a downloadable ShakeMap file provided by the U.S. Geological Survey, which contains information about the Modified Mercalli Intensity (MMI) measured at different locations. According to Worden and Wald (2016), the MMI data is an indicator based on Peak Ground Acceleration (PGA) and Peak Ground Velocity (PGV). Therefore, we rely on the recorded MMI to define affected and non-affected districts since it is plausibly the best measure of exposure to earthquake risk (Masiero and Santarossa, 2020).[1] We extract the ShakeMap file using the QGIS software to define the exposure to the earthquake and its intensity at the local (district) level. As illustrated in Figure 1, the ShakeMap file only covers the central area of the Java Island, meaning that the

[1]The exogeneity of earthquake-related deaths, injuries, and property damage across regions is debatable. The reported earthquake damage can be linked to a variety of unobservable district characteristics. As a result, using MMI to identify treatment and control districts is more precise.

relevant MMI value was only recorded for that square area. Districts that are not covered by the ShakeMap file are most likely to have had a very low-intensity value in which most people did not feel the tremor.

Figure 1: Java Island and MMI shapefile area for the 2006 Yogyakarta Earthquake



In order to exploit the information about local records of the MMI for the Yogyakarta earthquake, we follow the procedure adopted by Belloc *et al.* (2016) and Masiero and Santarossa (2020), among others.[2] Specifically, on the one hand, we classify districts with high earthquake intensity (hereafter "affected") if the highest registered MMI value is equal to or greater than 5[3], meaning that they were severely affected by the earthquake. On the other hand, districts with low seismic intensity (hereafter "unaffected") are those for which the highest registered MMI is less than 5. The range of variation in registered MMI for the Yogyakarta Earthquake is between 2.7 (the lowest) to 8.3 (the highest). However, we assign the MMI value equal to zero to districts outside of the area covered by the ShakeMap. Thus, as depicted in Figure 2, the areas colored red are the affected districts based on our definition.

The second database is the Indonesian Family Life Survey (IFLS) database[4], covering more than 80% of the Indonesian population within the survey area (Strauss et al., 2016). The IFLS is a longitudinal micro-level survey conducted in 1993, 1997, 2000, 2007, and 2014. The survey provides information about individuals' characteristics, educational attainment, and most importantly, the locations (province and district) of the respondents' birthplace, current residence, and entire migration history. As our main aim consists in analyzing long-term

[2]Similar approaches were also followed by Cipollone & Rosolia (2007), Kirchberger (2017), Paudel & Ryu (2018) and Hombrados (2020).
[3]According to the U.S. Geological Survey (2016), regions exposed to MMI greater than five are categorized as "strong". Below we also show that the share of destroyed schools relative to the pre-earthquake shock is strongly positive associated with registered MMI, only if it takes values equal or greater than 5 (Figure 4). Nevertheless, in the empirical analysis we check for the sensitivity of our results to different boundaries to define affected and unaffected districts.
[4]IFLS data can be obtained from https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html.

educational outcomes, we focus on the last wave of the IFLS survey (2014) to retrieve information about completed education, but we also exploit the information included in the 2007 and 2000 waves.[5] This procedure enables respondents' locations to be tracked at the district level, from the day they were born until the last wave of the survey (2014). Therefore, we (re)constructed the district of residence in the year of the earthquake, and we merged the information about MMI intensity at the district level accordingly, which enables us to group individuals according to whether they were residing in affected or unaffected areas when the earthquake struck (May 2006).

Third, we also use pre-determined district level information from administrative registers, containing a set of local characteristics that we use to perform matching exercises that are aimed at retaining only unaffected districts that are comparable to affected districts. Specifically, we carry out a matching procedure, separately for each of the selected variables measured in 2005 (i.e., before the earthquake), considering a) total number of students, b) total number of teachers c) total number of schools, d) student to school ratio e) school density (i.e. number of schools per km^2), f) total population, and g) per capita gross regional domestic product at the district level.

In order to construct our estimation sample, we retain individuals in schooling age (6-19 years old) in 2006 and exclude individuals whose age is below 6. In the empirical analysis, we also consider individuals aged 16 to 19 because some individuals aged just above 15 might still be studying compulsory education due to previous grade repetition. Furthermore, there were more than 70% of post-compulsory school participation rates in Yogyakarta province in 2005.[6] In addition, our main estimation sample excludes individuals born before 1970 to avoid the inclusion of older cohorts, but we also exploit this information for falsification analysis.

To analyze the impact of the earthquake on schooling achievements, we exploit variation in exposure during schooling age by birth cohort and district of residence in 2006. Therefore, on the one hand, we consider as treated individuals those belonging to young birth cohorts, who were still in schooling age when the earthquake took place (i.e., those born between 1987 and 2000, young cohorts henceforth). Consequently, individuals from the control cohorts were born between 1970 and 1986 (old cohorts) and were already above schooling age in the year of the natural disaster analyzed in this work. Moreover, we consider individuals living in affected and unaffected areas, according to the registered value of the MMI scale for the district of residence in 2006.

The descriptive statistics for the main variables in the analysis are reported in Table 1.[7] The table provides sample means and standard deviations of the variables used in the empirical analysis (outcomes and controls) for young and old cohorts residing in affected and unaffected districts. As the main dependent variable, we use years of completed education.[8] We also

---

[5]We also take some individuals from the 2007 wave who are not in the schooling age but are not available in the 2014 wave and include them in the control group. Furthermore, we also use the 2000 wave along with the 2007 wave for falsification analysis.

[6]Detailed information is reported here: https://www.bps.go.id/indicator/28/301/6/school-participation-rate-s-p-r-.html.

[7]Aggregate districts' characteristics are reported in Table A1 of the Appendix.

[8]In the IFLS, there is information about the highest level of schooling attended and the highest grade ever completed by the respondents. Using both these data, we can calculate the years of completed education. For

estimate the effect on education level, namely the completion rates of primary and junior high school and the enrollment rates of post-compulsory schooling.[9]

Table 1: Summary Statistics

	OLD			YOUNG			
Variable	Not Affected	Affected	Diff	Not Affected	Affected	Diff	Diff-Diff
Years of Schooling	9.385	11.088	1.703***	9.328	10.422	1.094***	-0.609***
	(3.654)	(3.417)	(0.106)	(3.008)	(2.994)	(0.116)	(0.164)
Primary Education Completion	0.697	0.901	0.204***	0.808	0.925	0.117***	-0.087***
	(0.460)	(0.299)	(0.013)	(0.394)	(0.263)	(0.015)	(0.020)
Junior Secondary Education Completion	0.669	0.869	0.201***	0.642	0.746	0.104***	-0.097***
	(0.471)	(0.337)	(0.013)	(0.479)	(0.436)	(0.018)	(0.023)
Post Compulsory Education Enrollment	0.325	0.506	0.181***	0.347	0.504	0.157***	-0.024
	(0.469)	(0.500)	(0.014)	(0.476)	(0.500)	(0.018)	(0.023)
Currently Enrolled in Education	0.005	0.007	0.002	0.297	0.402	0.105***	0.104***
	(0.072)	(0.083)	(0.002)	(0.457)	(0.491)	(0.018)	(0.014)
Age in 2006	27.472	27.923	0.451***	12.597	12.626	0.029	-0.422*
	(4.680)	(4.634)	(0.137)	(4.105)	(4.179)	(0.158)	(0.216)
Male	0.506	0.500	-0.007	0.483	0.485	0.002	0.009
	(0.500)	(0.500)	(0.015)	(0.500)	(0.500)	(0.019)	(0.024)
Fathers' Education	6.796	8.599	1.803***	6.813	8.654	1.840***	-0.045
	(4.834)	(5.073)	(0.142)	(4.496)	(4.646)	(0.173)	(0.229)
Mothers' Education	6.738	8.500	1.762***	6.694	8.615	1.921***	0.149
	(4.752)	(5.016)	(0.140)	(4.364)	(4.630)	(0.169)	(0.224)
Moslems	0.966	0.903	-0.063***	0.981	0.917	-0.064***	-0.000
	(0.181)	(0.296)	(0.006)	(0.137)	(0.276)	(0.006)	(0.009)
Christians	0.029	0.095	0.066***	0.017	0.083	0.066***	-0.000
	(0.169)	(0.294)	(0.005)	(0.131)	(0.276)	(0.006)	(0.008)
Other Religions	0.004	0.002	-0.003	0.002	0.000	-0.002	0.001
	(0.067)	(0.039)	(0.002)	(0.043)	(0.000)	(0.002)	(0.003)
Javanese	0.574	0.972	0.399***	0.560	0.980	0.420***	0.022
	(0.495)	(0.164)	(0.014)	(0.496)	(0.140)	(0.018)	(0.023)
Sundanese	0.239	0.005	-0.234***	0.258	0.005	-0.253***	-0.019
	(0.426)	(0.068)	(0.012)	(0.438)	(0.073)	(0.016)	(0.020)
Other Ethnicities	0.187	0.023	-0.164***	0.182	0.015	-0.167***	-0,003
	(0.390)	(0.150)	(0.011)	(0.386)	(0.120)	(0.014)	(0.018)
Number of Siblings	3.378	2.996	-0.382***	3.439	3.138	-0.302***	0.081
	(2.555)	(2.257)	(0.074)	(2.358)	(2.214)	(0.090)	(0.119)
Birth Order	3.206	3.114	-0.092	3.811	3.583	-0.228**	-0.135
	(2.312)	(2.366)	(0.068)	(2.433)	(2.252)	(0.093)	(0.114)
Migrate between 2006-2014	0.173	0.141	-0.033***	0.090	0.094	0.004	0.036**
	(0.379)	(0.348)	(0.011)	(0.286)	(0.291)	(0.011)	(0.017)
Household Casualties	0.002	0.349	0.347***	0.003	0.350	0.347***	0.0003
	(0.045)	(0.477)	(0.005)	(0.053)	(0.477)	(0.006)	(0.01)
Observations	11,230	1,309	12,539	7,023	748	7,771	

According to descriptive statistics, individuals residing in affected districts have more years of education than those who were living in non-affected districts, regardless of their birth cohort. This is possibly due to the high number of schools and universities located in Yogyakarta province (Ramdhani et al., 2012). Moreover, and most importantly, the difference of the difference indicates that the younger cohort of individuals living in affected areas when

instance, if an individual's highest level of schooling is junior high school and his/her highest grade ever completed is 2, then his/her years of completed education is equal to 8 years.

[9]Indicators for enrollment and completion of education levels are constructed on the basis of completed years of schooling. For instance, an individual is considered to have completed his/her junior high school if he/she has years of completed education equal to 9 years or higher. Furthermore, he/she is considered to have enrolled in post-compulsory schooling if he/she has experienced at least a year in that level of education or years of completed education equal to 10 years or higher.

the earthquake struck cumulated relatively less human capital than other groups, which is likely to be due to earthquake exposure. Similar evidence is obtained for the (unconditional) probability of having completed primary or junior secondary education (but not for post-compulsory schooling). Moreover, we also detected that young individuals living in affected areas have a higher likelihood of being still a student at the time of the survey. This could be also a possible detrimental effect of the earthquake, which consists in a certain delay in schooling progression. Regarding control variables, we use only a parsimonious set of characteristics, namely gender, father's and mother's education, religion, ethnicity, number of siblings and individual's birth order.[10] Indeed, these control variables appear to be balanced, since although there are significant differences between individuals residing in affected and unaffected areas, these are similar between those belonging to the young and old cohorts. This is indeed a first piece of evidence that justifies our identification strategy for estimating the effect of the earthquake on human capital accumulation, which is described in the next section.

4. Identification Strategy

4.1 Baseline setup

The identification strategy that we adopt to estimate the causal effect of earthquake exposure on schooling attainments exploits two sources of variation, namely birth cohort and district of residence, in the same line as Caruso and Miller (2015), Paudel & Ryu (2018), and Hombrados (2020), who analyzed similar natural shocks. Specifically, on the one hand, we compare education achievements observed in 2014 of individuals who were in school age when the earthquake took place (i.e., those born between 1987 and 2000, who were between 6 to 19 in 2006), and were living in affected (MMI ≥ 5) and unaffected districts at that time. Therefore, our "treatment" group consists of young individuals who resided in districts that were severely affected by the 2006 earthquake, according to the measured MMI scale, and the "control" counterpart are those from the same birth cohort residing in unaffected areas. However, the difference in education achievements, even conditioning to a large set of observable characteristics, is not likely to be meaningful because individuals in the treatment and control groups might differ along many other dimensions besides having been exposed to the natural disaster while at school, i.e., unobservable local and school inputs of human capital formation. Therefore, we use as additional control older cohorts of individuals who were beyond school age in 2006 (born in 1986-1970) and who were living in the two areas of the Java Island. The baseline regression that we estimate is,

$$Y_{id} = \alpha + \beta I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \qquad (1)$$

[10] The table also report descriptive information for two additional variables that we use to analyses potential channels, specifically, the probability of having changed place of residence after the earthquake and the probability of having suffered earthquake-related casualties at the household level.

where Y_{id} corresponds to the measure of schooling achievements of individual i (either years of attained schooling, dummies for completed levels of compulsory schooling, or post-compulsory enrollment) residing in district d (in 2006), θ_{yb} and δ_d represent, respectively, year of birth (yb) and district (d) fixed effects, while X_i is a set of individual controls.[11] Our interest relies on the coefficient (β) attached to the interaction between the indicator for living in an affected district in 2006 ($I(MMI_d \geq 5)$) and the one for individuals born between 1987 and 2000. This captures the difference in schooling for individuals belonging to the young cohort, who were living in affected and unaffected areas, in excess with respect to the difference observed among individuals living in the same districts but belonging to the older cohorts, which are those who were already out of (pre-university) education at the time of the earthquake. Indeed, this resembles a difference-in-difference approach, with the main difference that instead of using data from affected and unaffected areas obtained before and after the shock, we rely on cohort variation to capture exposure during school age. This is appealing since it is impossible to anticipate the timing of an earthquake's exogenous shock (Cavallo, Galiani, et al., 2013; Hombrados, 2020). However, two main underlying identifying assumptions need to be satisfied to interpret the estimated β coefficient as the causal effect of having been exposed to the earthquake during school age on completed education. First, older cohorts are assumed to be a valid counterfactual to capture unobservable differences between districts; that is, unobserved heterogeneity at the local level is the same for individuals belonging to different birth cohorts and are thus absorbed by the year of birth fixed effects (θ_{yb}). Second, differences by cohort in the unobservable heterogeneity are the same for individuals living in affected and unaffected districts and are captured by district fixed effects (δ_d). As detailed below, we perform several robustness checks and falsification exercises to provide evidence regarding the validity of these two main assumptions, as well as assessing other potential issues that could invalidate our empirical setup. In addition, we cluster the standard errors of equation (1) at the district level, which is the level of variation of exposure to the earthquake.

4.2 Robustness and falsification checks

As the first set of robustness checks of our baseline specification, we check for the sensitivity of the results to the MMI threshold used to define affected and unaffected districts. More specifically, rather than using a single indicator per district with a registered MMI greater than or equal to 5, we consider dummies for segments of the observed MMI range[12] and estimate the following equation:

$$Y_{id} = \alpha + \sum_j \beta_j I(1987 \leq yb_i \leq 2000) \times I(MMI_d \in k) + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \qquad (2)$$

[11]As mentioned in section 3, we consider only a parsimonious set of pre-determined controls, namely gender, father's and mother's education, religion, ethnicity, the number of sibling and birth order. Most of the estimates reported in this work are obtained without conditioning to any observable, but we also show the main results provided by models with controls for robustness (which are indeed very stable).

[12]That is, k = 1 if MMI <3.5, k = 2 if 3.5 ≤ MMI < 5, k = 3 if 5 ≤ MMI < 7.5, k = 4 if MMI ≥ 7.5.

This last equation clarifies whether the effect of earthquake exposure increases with its intensity and whether the baseline results are sensitive to the choice of the MMI threshold selected to define affected and unaffected areas (i.e., equal to or above five). The second battery of sensitivity analyses we perform is also related to the definition of affected and unaffected districts but considering distance with respect to the "core" of the affected area. That is, we replicate our main estimation (equation (1)) by excluding observations of individuals who were residing far away from the part of the island that was most strongly shaken by the earthquake. This enables us to analyze whether the results are robust to the exclusion of districts that are likely to be different with respect to the affected ones. We do this in two different ways: a) excluding districts that are not covered in the MMI shapefile (see Figure 1) and b) excluding districts located more than 200 or even 100 kilometers away from the closest district with MMI ≥ 5.[13] Related to that, we adopt a matching approach based on the method used in Redding and Sturm (2008), which enables using only unaffected districts that are similar to affected districts along several local characteristics (separately for each exercise) measured in 2005. Specifically, as mentioned below, we apply a matching algorithm that retains selected unaffected districts by minimizing the squared difference in terms of pre-earthquake a) total number of students, b) total number of teachers c) total number of schools, d) student to school ratio e) school density (i.e., number of schools per km^2), f) total population, g) per capita gross regional domestic product at the district level.

Additionally, we carry out a falsification exercise aimed at discarding the possibility that the coefficient of interest is blurred by spurious differences across districts. Our approach is based on a permutation test, similar to the one applied by Kuka et al. (2020). Specifically, the test involves the random assignment of an indicator for exposure to a fake earthquake to locations that were not affected by the natural disaster of 2006. To ensure the affected locations are not selected, we exclude districts located in Yogyakarta and Central Java provinces. We replicate this exercise 10,000 times and estimate equation (1) with observations from unaffected districts and obtain the resulting distribution of the placebo beta coefficient. Obtaining fake betas that are distributed around zero would be reassuring for the validity of our identification strategy.

Subsequently, to understand whether our identification strategy is invalidated by potential trends across heterogeneous cohorts between affected and unaffected locations, we implement three different falsification exercises based on older cohorts of individuals who were already out of school in 2006. Using 2014 data from IFLS 5 (as in our baseline), we consider a cohort of older individuals, initially excluded from our estimation sample, born between 1969 and 1956, and we treat them as a fake control cohort. Therefore, we use our original control cohort of individuals born between 1986 and 1973 (6 to 19 in 1992) as a fake treated cohort and individuals born between 1972 and 1956 (20-36 years old in 1992) as a fake control cohort. We then estimate a placebo regression "as if" the earthquake occurred in 1992 rather than in 2006 but maintaining the division between affected and unaffected districts based on individuals' place of residence in 2006 (i.e., keeping the real distribution of MMI across

[13] We perform vector analysis for this robustness check by extracting geometry attributes that produce latitude and longitude information for all districts. We then create straight lines between the centroids of non-affected districts and the nearest affected districts.

districts). Similarly, we use 2007 data from IFLS 4 and retain the same cohorts of individuals as in the previous falsification exercise, that is 1973-1986 for the fake treated group and 1972-1956 for the fake control group, neither having ever been affected by the natural disaster. Hence, we repeat the same placebo regression, again considering the place of residence in 2006, but using completed education observed in 2007 (i.e., one year after the real earthquake) as outcome. Finally, we use 2000 data from IFLS 3 and select only individuals who, at the time of the interview (2000) and at the time of the placebo earthquake (1992), were in the same age range as our baseline sample (14-44 and 6-36 respectively). However, this time, we impute the observed values of MMI by district according to their place of residence in 1992. For the three possibilities, finding placebo coefficients that are different from zero would indicate potential spurious heterogeneous trends across the cohorts, preventing a causal interpretation of the results. On the contrary, obtaining not significant estimates close to zero would constitute supporting evidence in favor of the validity of our approach.

4.3 Heterogeneity analysis and mechanisms

The last step of our empirical analysis consists in exploring any heterogeneous effects and potential mechanisms that could drive the obtained findings. First of all, we examine whether being exposed to the earthquake has a differential effect on schooling outcomes based on age at exposure, considering boundaries (*j*) defined according to whether individuals were in primary education, junior secondary (10-14) or upper secondary education (15-19) when the natural disaster occurred.[14] The corresponding equation takes the form:

$$Y_{id} = \alpha + \sum_j \beta_j I(yb_i \in j) \times I(MMI_d \geq 5) + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \quad (3)$$

Moreover, we analyze whether the effect of the 2006 earthquake affected differently education achievements of individuals according to other predetermined individual and family characteristics. Specifically, we include interaction terms to allow for heterogeneous β coefficients by gender, religion (Moslem versus others), ethnicity (Javanese versus others), father's and mother's education (compulsory versus post-compulsory education), number of siblings and birth order.

Regarding the potential mechanisms at work, we examine whether a) endogenous migration, b) earthquake-related casualties at the family level and c) damages in local education infrastructures are, to some extent, the driving forces behind the (negative) relationship between earthquake exposure and schooling achievement. To the best of our knowledge, these are the candidates for being channels that can be explored with the available data. Regarding the first potential mechanism, we track back the history of residential movements that occurred between 2006 and 2014. Therefore, we estimate an equation in which the dependent variable is an indicator that takes the value 1 if the individual changed district of residence during this period, using the same specification as for equation (1). This clarifies whether affected individuals (i.e. in school age and residing in affected districts in 2006) are more likely to have

[14] Specifically, $j = 1$ for those born between 2000 and 1997, whose age was 6 to 9 in 2006, $j = 2$ for those born between 1996 and 1992 (10-14) and $j = 3$ for those born between 1991 and 1987 (15-19), respectively.

changed place of residence after the earthquake. Moreover, we also estimate another equation for schooling outcomes that includes a triple interaction with the aforementioned indicator for being a mover (plus the corresponding base effects and double interactions). This alternative model shows whether movers and stayers were differently affected by the earthquakes in terms of attained schooling. Theoretically, the sign of this triple interaction is ambiguous since, on the one hand, migration can be a way to escape from the damages produced by the natural shock, but on the other hand it can represent an obstacle in the schooling process due to the need to adapt to another environment. In any case, finding a positive effect of earthquake exposure on the probability of being a mover together with a differential impact of the earthquake of education outcomes would indicate that (endogenous) migration behaviors could be one of the channels through which the natural disaster affected human capital formation at the individual level.

Second, in a similar vein, we also constructed an indicator for whether the family suffered death or major injuries of a household member, direct financial loss to the household, or relocation of the household member in the last five years because of the earthquake, using ad-hoc information included in IFLS 4.[15] Therefore, on the one hand, we also estimate equation (1) using as outcome the dummy for having suffered some kind of earthquake-related casualty at the family level. On the other hand, we allow for a triple interaction with the casualties' indicator in the schooling outcome's equation, as done for post-earthquake migration. Again, finding a positive effect of earthquake exposure on the likelihood of having experienced any kind of casualties together with a differential effect of the earthquake on schooling according to whether the individual's family was directly affected in some aspect (i.e. death of family members, injuries, financial losses or relocation) by the earthquake would point to a relevant role of this potential channel in explaining the link between the natural disaster and education achievements.

Third, to analyze the unexplored channel of damages on educational infrastructures, we retrieved administrative data regarding the number of education infrastructures destroyed or damaged due to 2006's natural disaster by district (expressed in percentage of 2005, pre-earthquake, stock). First, we check whether this measure of the destruction of schools correlates with registered MMI at the district level and, second, we estimate the following equation in which we substitute the indicator for living in affected districts in 2006 with the sum of damaged/destroyed schools in the district (*dsch*$_d$) over the pre-earthquake (2005) stock of school buildings (sch_d^{2005}), that is:

$$Y_{id} = \alpha + \beta I(1987 \leq yb_i \leq 2000) \times \frac{\sum dsch_d}{sch_d^{2005}} + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \qquad (4)$$

This alternative estimation is already suggestive of whether the impact of the earthquake on school infrastructures represents one of the channels through which this natural disaster had a detrimental effect on school achievements among affected individuals. Moreover, we also adopt a triple difference model that includes the interaction between the indicators for being in

[15] We use IFLS 4 because the questionnaire asks about the natural disaster that occurred in the last five years, and choose earthquake for the type of natural disaster and 2006 as the year of occurrence. We match that information with our sample in the main estimation according to the place of residence in 2006.

the affected cohorts, living in affected districts and another one that captures affected districts in which a certain proportion of schools were destroyed or damaged. Specifically, we consider the differential effect of exposure to the natural disasters according to whether at least some schools were disrupted by the earthquake. Lastly, we also analyze the impact of living in districts in which the majority of existing schools (i.e. more than 75%) suffered some kind of damage due to the earthquake. Overall, these additional estimations would reveal whether the disruption of educational infrastructures represents one of the possible mechanisms at work.

To conclude, and with the aim of shedding light about whether the overall impact of earthquake exposure obtained from our empirical setup (and the available data) consists in a transitory delay of schooling progression, or indeed represent a long-term negative effect on human capital accumulation, we exploit information on current school attendance. This is because in the survey's year (2014) some residents of the affected districts belonging to the cohorts who were in school age when the earthquake struck might be still at school. That is, it is possible that, at some point, they would catch up their counterparts who belong to the same cohorts, but were living in unaffected districts, in terms of completed years of schooling over the medium-long run.[16] Specifically, first we estimate equation (1) using the baseline sample but considering as dependent variable the indicator for being still enrolled in education in 2014. This would provide a first indication about whether the earthquake generated a certain delay in schooling progression. Moreover, we re-estimate the same equation again using years of schooling as outcome but excluding individuals who were still students at the time of the survey, which would provide evidence about the long-standing effect of the earthquake on education attainment.

5 Results

5.1 The impact of the 2006 earthquake

We begin by presenting the impact of the 2006 earthquake on years of education, which are displayed in Table 2. We present two specifications, one without control variables (column (1)) and another with individual and family characteristics included as controls (column (2)), both including fixed effects for year of birth and district of residence in 2006. The estimate of interest is unaffected by the inclusion of controls and indicates that being affected by the earthquake during school age reduces years of schooling by 0.74 years, which corresponds to around 0.22% of one standard deviation point of years of schooling for the whole sample (mean 9.5, s.d. 3.43).

[16]Notice that this additional analysis would have been easily done with a new, more recent version of IFLS. Unfortunately, the IFLS survey stopped in 2014 and right now it is very unlikely that a new wave of the survey will be implemented in the near future.

Table 2: Impact of the 2006 earthquake on years of education

	(1)	(2)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-0.742***	-0.744***
	(0.162)	(0.173)
R-squared	0.183	0.272
District & Year of Birth Fixed Effects	Yes	Yes
Controls	No	Yes
Number of Observations	20,304	20,304

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Table 3 shows the effects on completed and enrolled education levels. Individuals exposed to the earthquake during school age are 11.6 percentage points less likely to complete primary school than those in non-affected areas. The impact is slightly smaller in panel B[17], indicating a reduction of around 10.6 percentage points in the probability of completing junior high school. Moreover, the estimate reported in panel C indicated that the effect on enrollment into post-compulsory education is virtually zero and not significant. Also, for education levels, the results are unaffected by the inclusion of controls, which is consistent with descriptive evidence regarding the balancing test of individual and family characteristics and speaks in favor of the exogeneity of the shock.

Table 3: Impact of the 2006 earthquake on school completion and enrollment

	Panel A: Primary School Completion		Panel B: Junior High School Completion		Panel C: Post Compulsory Enrollment	
	(1)	(2)	(3)	(4)	(5)	(6)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-0.116***	-0.117***	-0.106***	-0.106***	-0.004	-0.004
	(0.019)	(0.019)	(0.025)	(0.024)	(0.023)	(0.022)
R-squared	0.160	0.203	0.168	0.208	0.120	0.178
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District & Year of Birth Fixed Effects	No	Yes	No	Yes	No	Yes
Number of Observations	20,304	20,304	19,689	19,689	19,120	19,120

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

5.2 Robustness and falsification checks

Focusing on years of schooling as outcome, as a first sensitivity check we analyze whether the results are robust to the MMI threshold we adopted to define affected and unaffected districts (i.e., MMI $\geq$ 5). Therefore, we define categorical dummies for different values of the registered MMI, which leads to the estimation of equation (2). The results are shown in Table 4 and suggest that the detrimental effect of exposure to the natural shock occurs when the MMI takes values equal to or greater than 5 (but not lower). Moreover, there is

[17]Notice that the estimations in columns (3) to (4) are based on a smaller sample, since we exclude individuals who could still be in junior high school in 2014.

virtually no difference in the estimates for different segments of the MMI distribution above the cut-off we used in the baseline estimations.

Table 4: Sensitivity to MMI thresholds

Variables	(1)	(2)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d = 0)$	*reference category*	
$I(1987 \leq yb_i \leq 2000) \times I(0 < MMI_d < 3.5)$	0.387*	0.575**
	(0.209)	(0.224)
$I(1987 \leq yb_i \leq 2000) \times I(3.5 \leq MMI_d < 5)$	0.496**	0.642***
	(0.196)	(0.203)
$I(1987 \leq yb_i \leq 2000) \times I(5 \leq MMI_d < 7.5)$	-0.536**	-0.555**
	(0.230)	(0.266)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 7.5)$	-0.568**	-0.432*
	(0.220)	(0.224)
R-squared	0.184	0.274
District & Year of Birth Fixed Effects	Yes	Yes
Controls	No	Yes
Number of Observations	20,304	20,304

Notes: OLS estimation, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

The following set of robustness checks involves the definition of unaffected districts. Specifically, instead of using the pool of districts of the Java Island that were either outside the MMI's shape file or had a registered MMI for the 2006 earthquake below 5, in the first exercise we retain only districts that were not excessively distant from affected districts (100 and 200 kilometers away using a straight line between districts' centroids). Second, we keep only districts that appear in the MMI map. As shown in columns (1) to (6) of Table 5, the results obtained using this restricted group of unaffected districts are qualitatively similar to the main results. The point estimates are somewhat higher than the baseline and highlight a reduction in years of schooling by around 1 year or slightly more for having been exposed to the earthquake while in school age.

Table 5: Sensitivity to the choice of unaffected districts

Districts selection:	100 KM	100 KM	200 KM	200 KM	MMI Map	MMI Map
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-1.077***	-1.217***	-1.009***	-1.140***	-1.031***	-1.157***
	(0.241)	(0.239)	(0.166)	(0.174)	(0.164)	(0.170)
R-squared	0.213	0.317	0.192	0.295	0.194	0.297
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	4,637	4,637	8,742	8,742	9,478	9,478

Notes: OLS estimation, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

In a similar vein, with the aim of showing that the resulting evidence is not driven by the choice of unaffected districts, Table 6 displays the results obtained after repeating the estimation of equation (1) after implementing the matching procedure, which was carried out separately for each district's characteristics. As can be observed, the number of observations is reduced drastically since few unaffected districts can be matched with affected districts according to the selected pre-earthquake variables (even less than for the previous check). However, the main results remain qualitatively similar and very close, in terms of point estimates, to those obtained after restricting the number of unaffected districts based on geographical criteria.

Table 6: Matching results

Matching based on:	Total No. Students		Total No. Teachers		Total No. Schools	
	(1)	(2)	(3)	(4)	(5)	(6)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-1.084**	-1.087**	-0.711***	-0.942***	-1.058**	-1.109**
	(0.403)	(0.398)	(0.246)	(0.267)	(0.419)	(0.408)
R-squared	0.219	0.313	0.189	0.306	0.232	0.321
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	3,429	3,429	3,678	3,678	3,344	3,344

Notes: OLS estimation, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

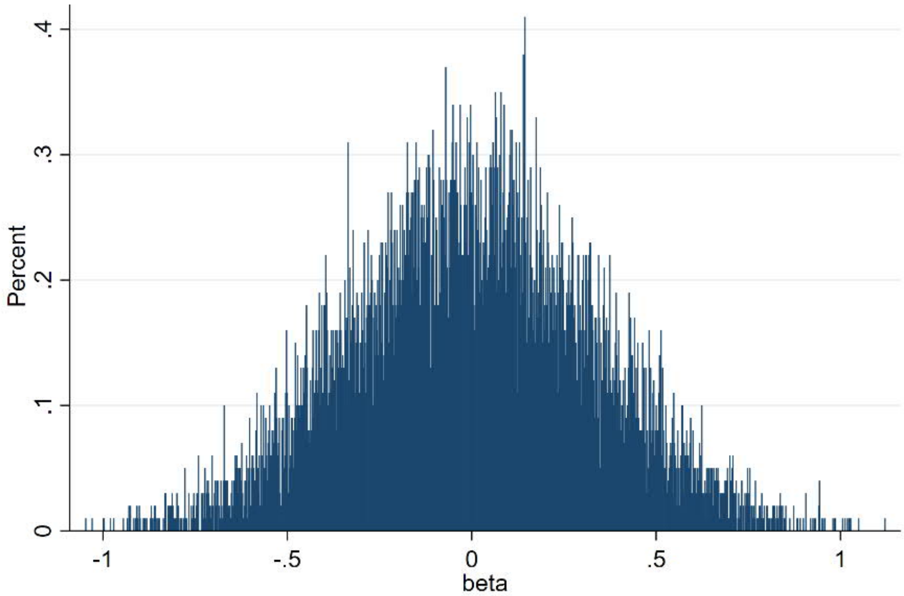
Table 6 (continued): Matching results

Matching based on:	Student-School Ratio		School Density		Total Population		GRDP per Capita	
	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-0.701***	-0.803***	-0.680*	-0.713*	-1.152***	-1.004**	-1.157***	-1.357***
	(0.218)	(0.252)	(0.345)	(0.411)	(0.339)	(0.426)	(0.225)	(0.198)
R-squared	0.162	0.272	0.244	0.337	0.168	0.281	0.215	0.326
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Number of Observations	3,435	3,435	4,926	4,926	2,450	2,450	4,806	4,806

Notes: OLS estimation, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Next, we show the evidence from different falsification tests. First, we perform a permutation test involving the random assignment of an indicator for exposure to a fake earthquake to locations not affected by the natural disaster of 2006. We exclude districts located in Yogyakarta and Central Java provinces to ensure that the affected area is not selected during randomization. After running 10,000 replications, we find that the estimates of the fake exposure coefficient follow a bell-shaped distribution centered around zero. This evidence indicates that our main results are unlikely to be driven by spurious differences across districts that distort the coefficient of interest.

Figure 2: Permutation tests of the fake earthquake locations



The second falsification exercise entails creating a fake earthquake year. As explained in section 4, we turn the old cohort into a fake young cohort and use a very old cohort, which was not in our main sample, to be a fake control cohort. We then estimate a placebo regression "as if" the earthquake occurred in 1992 rather than in 2006 using IFLS waves 5 (2014) and 4 (2007). In IFLS 3 (2000), we select only individuals who, at the time of the interview (2000) and at the time of the placebo earthquake (1992), were in the same age range as our baseline sample (14-44 and 6-36 respectively). This time, we rely on the place of residence in 1992 and impute the MMI values based on their residence in 1992. Table 8 column (1) to (6) shows that the results of these placebo estimations are substantially smaller in size, generally positive (except for the falsification exercise using IFLS 3) and not statistically different from zero, implying no indication of potential spurious heterogeneous trends across the cohorts, which further strengthens the causal interpretation of our main findings.

Table 7: Impact of the 1992 fake-year earthquake on years of education (Using old and very old cohorts)

Variables	IFLS 5 (1)	IFLS 5 (2)	IFLS 4 (3)	IFLS 4 (4)	IFLS 3 (5)	IFLS 3 (6)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	0.111	0.193	0.089	0.197	-0.077	-0.157
	(0.306)	(0.297)	(0.199)	(0.140)	(0.315)	(0.319)
R-squared	0.213	0.340	0.220	0.436	0.214	0.355
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	18,063	18,063	15,868	15,868	14,308	14,308

Notes: OLS estimation, dependent variable = years of schooling Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

5.3 Heterogeneous Effects

The findings reported so far point towards the existence of a negative causal effect of earthquake exposure on schooling outcomes, and the evidence from all the robustness checks and falsification exercises indicate that our identification strategy has a reasonable degree of internal validity. In this subsection, we allow for heterogeneous effects of exposure by different individual and family characteristics, which is also informative of the external validity of the results reported in this paper. First, we analyze whether the earthquake had a differential effect according to age at exposure, which corresponds to equation (3). As can be appreciated in Table 8, we find that the effect of the natural disaster decreases with age at exposure, and is significantly stronger among very young individuals who were still in compulsory education when the earthquake struck. Specifically, individuals born between 1997 and 2000, who were still in primary school at the time of the earthquake, are much more severely affected (coefficient equal to -1.74 without controls, s.e. 0.281). There is still a significant and negative effect for those born between 1992 and 1996, who were in junior high school, but substantially smaller than for the younger cohort. However, the earthquake did not significantly affect years of schooling of individuals born between 1987 and 1991.

Table 8: Heterogeneous effects by age at exposure

	(1)	(2)
$I(1987 \leq yb_i \leq 1991) \times I(MMI_d \geq 5)$	-0.183	-0.144
	(0.245)	(0.233)
$I(1992 \leq yb_i \leq 1996) \times I(MMI_d \geq 5)$	-0.475**	-0.454*
	(0.183)	(0.233)
$I(1997 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-1.740***	-1.818***
	(0.281)	(0.325)
R-squared	0.184	0.273
District & Year of Birth Fixed Effects	Yes	Yes
Controls	No	Yes
Number of Observations	20,304	20,304

Notes: OLS estimation, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

This result is consistent with the evidence reported in Table 3, indicating that the effect on years of schooling is mostly driven by a reduction in the probability of completing compulsory education, but also with additional evidence (which we will analyze further below) regarding the disruption of educational infrastructures. Indeed, most of the stock of primary school buildings available in 2006 were constructed under the primary school expansion program analyzed, among others, by Duflo (2001), which was implemented during the nineteen-seventies. This school construction policy was effective in shaping schooling opportunities and increasing education attainments. However, other sources report that the quality of school buildings was poor due to a low enforcement of development regulations. The Government opted for maximizing the number of newly constructed schools over compliance with earthquake-resistant building standards and other safety standards (Bappenas, 2006).

In a subsequent step, we estimate heterogeneous effects for several covariates that were used as controls, namely gender, religion, fathers' and mothers' educational background, ethnicities, number of siblings, and birth order. The results are presented in Table 9. The model with heterogeneous effects by gender provides a larger point estimate of the coefficient of interest for males but is less precisely estimated. Indeed, the test for the equality of the coefficients for males and females does not provide sufficient evidence to reject the corresponding null hypothesis. Similar evidence is obtained for differences by religion, ethnicity and paternal education, since in none of these cases we can reject the null hypothesis of equal coefficients. However, the impact of earthquake exposure is significantly stronger for individuals whose mothers have completed at most compulsory education. Moreover, we also detect stronger effects for individuals with fewer brothers and sisters than those who have three or more siblings, as well for those who are the first and second born children in the family.

Table 9: Heterogeneous effects of the earthquake

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(female)$	-0.555***			
	(0.173)			
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(male)$	-0.942***			
	(0.267)			
$I(1987 \leq yb_i \leq 2000) \times I(MMId \geq 5) \times I(non\text{-}moslem)$		-0.205		
		(0.615)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(moslem)$		-0.681***		
		(0.152)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(non\text{-}javanese)$			0.186	
			(0.760)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(javanese)$			-0.890***	
			(0.165)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(dad's\ educ \leq 9)$				-0.602***
				(0.208)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(dad's\ educ > 9)$				-0.608**
				(0.257)
Test for coefficients' equality, p-value	0,168	0,413	0,169	0,987
R-squared	0.272	0.274	0.271	0.276
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Number of Observations	20,304	20,304	20,304	20,304

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Table 9 (continued): Heterogeneous effects of the earthquake

	(5)	(6)	(7)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{mom's educ} \leq 9)$	-0.828***		
	(0.202)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{mom's educ} > 9)$	-0.414**		
	(0.162)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{siblings} = 0)$		-0.978**	
		(0.401)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{siblings} = 1)$		-1.487***	
		(0.455)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{siblings} = 2)$		-1.325***	
		(0.276)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{siblings} = 3)$		-0.386	
		(0.534)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{siblings} \geq 4)$		-0.306	
		(0.245)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{birth-order} = 1)$			-0.947***
			(0.280)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{birth-order} = 2)$			-1.304***
			(0.214)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\text{birth-order} \geq 3)$			-0.440**
			(0.204)
Test for coefficients' equality, p-value	0,033	0,018	0.002
R-squared	0.271	0.275	0.275
District & Year of Birth Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Number of Observations	20,304	20,304	20,304

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

5.4 Potential Mechanisms

In order to provide evidence regarding the relevance of migration as one of the potential mechanisms behind the negative relationship between earthquake exposure and human capital accumulation, we start by analyzing whether individuals who, in 2006, were in school age and residing in affected districts are more likely to have changed district of residence between the 2006 and 2014. As can be appreciated in columns (1) and (2) of Table 10, the natural disaster did not affect the probability of migrating, regardless of whether we include or not control variables. However, the estimates from the model that includes a triple interaction between birth cohort, living in an affected district and the indicator for being a mover (columns (3) and (4)) show that the reduction in years of schooling is significantly lower for individuals who moved to another district after the earthquake, according to the positive (although only marginally significant) interaction coefficient. Our interpretation of these results is that although migration does not seem to be a relevant mechanism, since it was not induced by earthquake exposure, changing place of residence (possibly due to other household decisions) could be a way to mitigate the detrimental effects of natural disasters on schooling outcomes.[18]

Table 10: Migration as potential mechanism

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	0.028	0.031	-0.846***	-0.888***
	(0.017)	(0.021)	(0.172)	(0.177)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(mover_i)$			0.724*	0.690*
			(0.423)	(0.350)
R-squared	0.065	0.174	0.186	0.302
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Number of Observations	20,304	20,304	20,304	20,304

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order. The dependent variable for columns (1) and (2) is the indicator for having changed place of residence between 2006 and 2014, $I(mover_i)$. The dependent variable for columns (3) and (4) is years of schooling; the models also include the base effect of being a mover and the corresponding double interactions.

In order to assess the role of earthquake-related family casualties as a potential channel, we examine whether the loss of life or any injuries, financial losses and relocation suffered by the household members due to the 2006 earthquake played some role in explaining our main findings. Similarly, to the analysis of migration, in the first two columns of Table 11 we first show the direct effect of earthquake exposure on the probability of having experienced some kind of casualties. As expected, the coefficient of the interaction between the indicators for being in school age in 2006 and residing in affected districts is positive and significant, indicating that the likelihood of having suffered household-level issues as a consequence of the earthquake increases by 2.5 percentage points. However, the coefficient of the triple interaction displayed in Columns (3) and (4) of Table 11 are imprecisely estimated and not statistically significant, which means that we do not detect any evidence in favor of the hypothesis that family casualties do not represent a relevant mechanism that links exposure to the natural disaster and completed education.

[18]We also tried to re-estimate the main model after excluding individuals who changed place of residence during the period 2000-2014, which provided very similar results (available upon request).

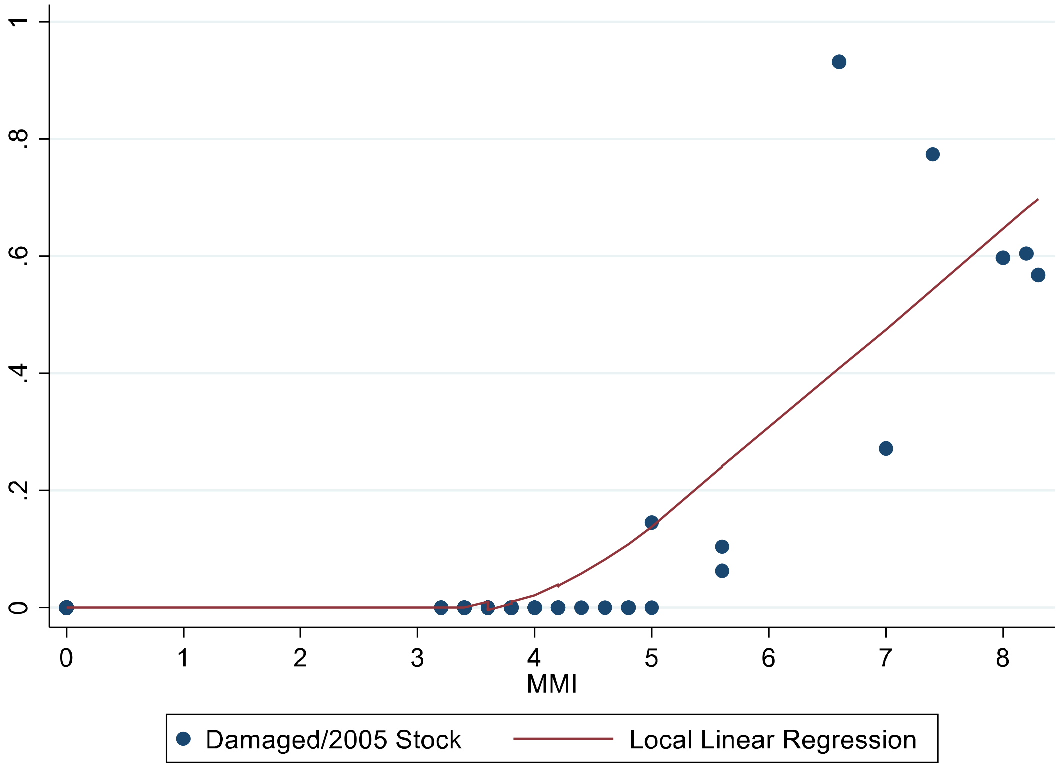
Table 11: Household casualties as potential mechanism

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	0.025**	0.025**	-0.614***	-0.717***
	(0.010)	(0.010)	(0.166)	(0.186)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(casualties_i)$			0.449	0.441
			(0.727)	(0.812)
R-squared	0.552	0.554	0.184	0.273
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Number of Observations	20,304	20,304	20,304	20,304

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order. The dependent variable for columns (1) and (2) is the indicator for having suffered earthquake-related casualties in the household, $I(casualties_i$ The dependent variable for columns (3) and (4) is years of schooling; the models also include the base effect of having experienced some kind of casualties at the household level and the corresponding double interactions.

For the last potential mechanism, we analyze whether the 2006 earthquake caused damages or destruction of educational infrastructures and, subsequently, if individuals in school age at the time of the earthquake who were living in districts with a higher school disruption rate were most severely affected in terms of human capital formation. Specifically, using district level data, we investigate whether the intensity of the earthquake is associated with a higher level of disruption of educational infrastructures. We express damage and destruction of school buildings as a percentage of the pre-earthquake (2005) stock to control for the size of the district in terms of number of schools and, indirectly, to the school age population. Figure 4 reports the scatter plot of the share of disrupted schools as a function of registered MMI at the district level, together with a local linear regression fit. The figure clearly indicates a positive and strong relationship between the intensity of the natural shock and the fraction of affected schools. Moreover, it also provides evidence regarding the adequateness of our MMI threshold to define affected and unaffected districts, since no schools were damaged or destroyed in districts where the registered MMI was below five.

Figure 4: MMI and damaged/destroyed schools by district (over the 2005 stock)



In columns (1) and (2) of Table 12 we report the estimate(s) of the coefficient of interest from equation 4, in which we substituted the indicator for living in a district with a registered MMI equal to or greater than five with the share of disrupted schools relative to the 2005 stock. The results highlight that school age individuals living in districts with a higher fraction of affected schools obtain significantly less education, in a similar vein to that of our baseline estimates. These two pieces of evidence are already suggestive of the relevance of the disruption of educational infrastructures as a mechanism at work. Third, in order to further examine the importance of this channel, in the subsequent columns we show the results obtained from a triple interaction model that includes an additional indicator for living in districts with a) at least some school damaged or destroyed (columns (3) and (4)) and b) at least 75% of available schools affected by the earthquake. In the first case, it is possible to see that although even individuals residing in the few affected districts with no disrupted school were negatively affected by earthquake exposure, the reduction in schooling achievements is more pronounced for those residing in places with at least some disrupted schools. Indeed, given the strong coincidence between the MMI cutoff and the risk of school disruption, the overall effect (base coefficient and interaction) is virtually identical to our baseline estimate. Moreover, when we allow for a differential effect of living in districts where most of the schools were damaged or destroyed by the earthquake, the estimate(s) indicates that the detrimental effect of earthquake exposure is even stronger when accompanied by a significant disruption of school infrastructures. Overall, these last findings highlight that earthquake-induced disruption of educational facilities indeed represents a relevant and significant mechanism through which earthquakes, and possibly natural disasters in general, tend to dampen human capital formation.

Table 12: Distruption of educational infrastructures as potential mechanism

	(1)	(2)	(3)	(4)	(5)	(6)
$I(1987 \leq yb_i \leq 2000) \times \sum dsch_d / sch_d^{2005}$	-0.986**	-1.000***				
	(0.167)	(0.233)				
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$			-0.489***	-0.414***	-0.554***	-0.563***
			(0.116)	(0.127)	(0.175)	(0.191)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\sum dsch_d / sch_d^{2005} > 0)$			-0.276**	-0.359***		
			(0.130)	(0.136)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(\sum dsch_d / sch_d^{2005} > 0.75)$					-0.413***	-0.399*
					(0.145)	(0.223)
R-squared	0.183	0.272	0.183	0.272	0.183	0.272
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	20,304	20,304	20,304	20,304	20,304	20,304

Notes: OLS estimation. Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

The last piece of evidence that we report is regarding current school attendance of individuals in our sample. This enables understanding whether the overall negative impact of exposure to the earthquake represents a transitory shock, which generates a certain delay in schooling progression, or a long-standing effect that implies a lower endowment of human capital among individuals affected by the natural disasters during school age. Indeed, the evidence reported above is consistent with a potential transitory effect. This is because, on the one hand, we detected a stronger effect among individuals who were still in compulsory schooling age when the earthquake occurred and could be still enrolled in education in 2014. On the other hand, the relevance of educational infrastructures' disruption as channel could also imply that students living in affected areas were prevented to attend school until the reconstruction process was completed.[19] Also, the youngest might have experienced a delay in the access to the education system. To provide suggesting evidence about this point, in Table 13 we show the results obtained from the estimation of the baseline equation (1) but using as dependent variable a dummy that captures current school attendance (columns (1) and (2)). This additional estimation highlights that actually young individuals who were living in affected areas at the time of the earthquake are more likely to be still students in 2014 (+ 10 percentage points), which is indicative of a certain delay in schooling progression induced by the natural disaster. However, this is just part of the overall effect, since re-estimating the main model for years of schooling using the subsample of individuals who are not currently enrolled in education provides an estimate that is just somewhat lower than the baseline (coefficient equal to -0.494, s.e. 0.146, relative to -0.742 from the baseline model without controls). This indicate that around 67% of the overall effect detected from the main specification actually represents a long-standing impact of the natural disasters, which reduced the accumulation of human capital for affected individuals.

[19]Unfortunately, to the best of our knowledge, detailed information about school reconstruction after the 2006 earthquake is not available.

Table 13: Additional evidence regarding current school attendance

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	0.103***	0.103***	-0.494***	-0.368**
	(0.0189)	(0.0188)	(0.146)	(0.167)
R-squared	0.597	0.599	0.174	0.279
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Number of Observations	20,304	20,304	17,849	17,849

Notes: OLS estimation, Colu Standard errors, in parentheses, are clustered at the district level. *** significant at 1% , ** significant at 5% , * significant 10%. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order. The dependent variable for columns (1) and (2) is the indicator for being currently enrolled in education. The dependent variable for columns (3) and (4) is years of schooling the estimations are obtained after excluding individuals who are currently students from the sample.

6 Conclusion

We analyzed the impact of natural disasters on human capital formation, considering as a natural experiment a disastrous earthquake that occurred in 2006 in the Java Island of Indonesia. Drawing on combined individual-level and aggregate datasets and focusing on the effect of suffering an earthquake during school age, we adopted an identification strategy that exploits variation in exposure to the earthquake by birth cohort and district of residence at the time of the natural disaster. The main results indicate that exposure to the earthquake during school age has a significant and negative impact on the accumulation of human capital, proxied by years of schooling, as well as on enrollment and completion of compulsory and post-compulsory education levels. Specifically, the baseline estimates highlight a reduction of somewhat less than one year of schooling because of the earthquake (-0.74 years, although other estimates indicate a slightly stronger effect) and a lower probability of completing compulsory education of around 10-11 percentage points. However, no effect was detected for the chances of enrolling in post-compulsory education levels.

The results are robust to several sensitivity analyses and, most importantly, the findings from falsification and matching exercises point towards the internal validity of our identification strategy and validate the causal interpretation of the results. Therefore, the evidence reported in this paper is consistent with previous results from the existing literature, which indicate that natural disasters are worrisome events not only for their direct impacts in terms of human lives and economic damage, but also because of their detrimental effects on the endowment of the human capital of affected countries. This is indeed especially relevant for emerging countries, since education represents one of the main factors through which they can foster economic growth and achieve the desirable level of economic and social development. The evidence from the analysis of heterogeneous effects also indicates that the impact of exposure to the shock appears to be stronger for younger individuals who were still in compulsory school when the earthquake struck. Moreover, the effect was more pronounced for children of low educated mothers, pointing towards the protective effect of material human capital but also to the fact that governments and policymakers should consider tailoring recovery interventions at the individual/family level, in order to be more generous to those with a less advantaged educational and social background. Moreover, the evidence from the potential mechanism at work, according to data availability, suggests that earthquake-related casualties at the family level do not seem to play a relevant role. Endogenous migration responses do not appear to be relevant channels either, although the results indicate that migration could be a way to reduce the negative effect of natural disasters. This is indeed consistent with the results reported by Park et al. (2015), who found that forced migration policies of students affected by an earthquake helped to mitigate earthquake-related mental health problems such as depression, as well as to enhance self-esteem and the test scores of affected children. Most importantly, the analysis of the unexplored mechanism of the disruption of educational infrastructures shows that this is indeed a relevant issue since it represents a channel through which natural disasters harm human capital formation. Finally, we also reported additional evidence regarding whether the impact of earthquake exposure, which appears to be stronger for younger cohorts of affected individuals and mediated by the disruption of educational infrastructures, represents a transitory shock that generates a delay in schooling progression, or a permanent loss of human capital. The results suggest that both effects are present, although the latter one seems to be more prominent, since a substantial fraction of the overall impact of the natural disaster induced lower educational attainment among affected individuals who stopped studying before their unaffected counterparts.

Altogether, we are confident that our results are also characterized by a high degree of external validity, which means that the evidence reported in our work can be reasonably extrapolated to other realities (especially for developing countries). Therefore, a direct policy implication of the results reported in this work is that policymakers should focus their efforts on improving the quality of school buildings and complying with modern ant seismic regulation and technical recommendations to withstand the disruptive effects of earthquakes and other natural disasters. Indeed, Herrera-Almanza and Cas (2021) show that Typhoon-resistant school construction policies implemented in 1989 in the Philippines almost entirely offset the harmful impact of typhoons on educational attainment. Therefore, governments of countries that are often subject to earthquakes and other harmful natural shocks, which cannot be accurately forecasted nor eradicated with public interventions (as they are an intrinsic feature of our world), should consider devoting more resources to improving the quality of school facilities. Most importantly, policymakers and administrators of educational facilities should try to double their efforts to immediately allocate a certain (and sufficient) amount of recovery funds to school reconstruction, as well as to provide temporary schooling infrastructures to prevent young individuals from interrupting their schooling process due to the occurrence of natural disasters. In fact, a private interview with the head of the education department of one of the most affected districts of the Java Island highlighted that, in the aftermath of the earthquake, students enrolled in disrupted schools were temporarily dismissed from their learning activities. They were not relocated in other schools or in temporary infrastructures. The reconstruction process prioritized the rebuilding of destroyed private houses and then focused on public infrastructures such as schools, hospitals, roads and bridges, only several months later. The funds for the reconstruction came from the central and regional governments, accompanied by international donors. Also, the reconstruction rate was not homogeneous across affected districts and villages, ranging between one and two and a half years.

Overall, there is still a lot of work to be done in emerging countries which, like Indonesia, substantially expanded the supply of educational infrastructures at different levels to provide education opportunities, but sometimes at the expense of the quality of infrastructures. This appears to be a sensible route to follow, not only because it would prevent the future cost of natural disasters in terms of human lives and reconstruction expenditure, but also because having earthquake-resistant school buildings would mitigate the detrimental effects of natural disasters on human capital accumulation, and in turn on economic growth and development.

References

Amri, M.R., Yulianti, G., Yunus, R., Wiguna, S., W. Adi, A., Ichwana, A.N. and Randongkir, Roling Evans Septian, R.T. (2018) *RBI (Risiko Bencana Indonesia), BNPB Direktorat Pengurangan Risiko Bencana*.

Andrabi, T., Daniels, B. and Das, J. (2021) 'Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005', *Journal of Human Resources*, 128182, pp. 0520-10887R1.

Baez, J., Fuente, A. De and Santos, I. (2010) *Do Natural Disasters Affect Human Capital? An Assessment Based on Existing Empirical Evidence*, *Discussion Paper Series No. 5164*. IZA. Bonn.

Baez, J. and Santos, I. (2008) *On shaky ground: The effects of earthquakes on household income and poverty*, *United Nations Development Programme Regional Bureau for Latin America and the Caribbean*. New York. Available at: http://www.preventionweb.net/preventionweb-files/english/hyogo/gar/2011/en/bgdocs/GAR-2009/background_papers/Chap3/LAC-overview/El-Salvador/El-Salvador-Earthquakes.pdf.

Bappenas (2006) *Preliminary Damageand Loss Assessment Yogyakarta and Central Java Natural Disaster*. Jakarta, Indonesia. Available at: https://recovery.preventionweb.net/publication/preliminary-damage-and-loss-assessment-yogyakarta-and-central-java-natural-disaster.

Belloc, M., Drago, F. and Galbiati, R. (2016) 'Earthquake, religion, and transition to self-government in Italian cities', *The Quarterly Journal of Economics*, pp. 1875–1926.

Bentzen, J.S. (2019) 'Acts of God? Religiosity and Natural Disasters Across Subnational World Districts', *The Economic Journal*, 129, pp. 2295–2321.

Bustelo, M., Arends-Kuenning, M. and Lucchetti, L. (2012) 'Persistent impact of natural disasters on child nutrition and schooling: Evidence from the 1999 Colombian earthquake', *IZA Discussion Paper* [Preprint], (6354).

Cairo, J.B., Dutta, S., Nawaz, H., Hashmi, S., Kasl, S. and Bellido, ; Edgar (2010) 'The prevalence of post-traumatic stress disorder among landslide victims', *Disaster Medicine and Public Healrh Preparedness*, 4, pp. 39–46.

Caruso, G. and 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, pp. 134–150.

Cavallo, E., Bank, I.D., Galiani, S. and Pantano, J. (2013) 'Catastrophic Natural Disasters and Economic Growth', *The Review of Economics and Statistics*, 95(5), pp. 1549–1561.

Cavallo, E., Galiani, S., Noy, I. and Pantano, J. (2013) 'Catastrophic natural disasters and economic growth', *Review of Economics and Statistics*, 95(5), pp. 1549–1561.

Cerqua, A. and Di Pietro, G. (2017) 'Natural disasters and university enrolment: evidence from L'Aquila earthquake', *Applied Economics*, 49(14), pp. 1440–1457.

Cipollone, P. and Rosolia, A. (2007) 'Social interactions in high school: Lessons from an earthquake', *American Economic Review*, 97(3), pp. 948–965.

CRED-UNDRR (2020) *The human cost of disasters: an overview of the last 20 years (2000-2019), Human Cost of Disasters: An overview of the last 20 years*. Geneva. Available at: https://www.undrr.org/media/48008/download.

Cuaresma, J.C. (2010) 'Natural disasters and human capital accumulation', *World Bank Economic Review*, 24(2), pp. 280–302.

Duflo, E. (2001) 'Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment', *American Economic Review*, 91(4), pp. 795–813.

Esnard, A.M., Lai, B.S., Wyczalkowski, C., Malmin, N. and Shah, H.J. (2018) 'School vulnerability to disaster: examination of school closure, demographic, and exposure factors in Hurricane Ike's wind swath', *Natural Hazards*, 90(2), pp. 513–535.

Filipski, M., Jin, L., Zhang, X. and Chen, K.Z. (2019) 'Living like there's no tomorrow: The psychological effects of an earthquake on savings and spending behavior', *European Economic Review*, 116, pp. 107–128.

Gasper, J.T. and Reeves, A. (2011) 'Make It Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters', *American Journal of Political Science*, 55(2), pp. 340–355.

Gignoux, J. and Menéndez, M. (2016) 'Benefit in the wake of disaster: Long-run effects of earthquakes on welfare in rural Indonesia', *Journal of Development Economics*, 118, pp. 26–44.

Gomez, C.J. and Yoshikawa, H. (2017) 'Earthquake effects: Estimating the relationship between exposure to the 2010 Chilean earthquake and preschool children's early cognitive and executive function skills', *Early Childhood Research Quarterly*, 38, pp. 127–136.

Groen, J.A., Kutzbach, M.J. and Polivka, A.E. (2019) 'Storms and jobs: The effect of hurricanes on individuals' employment and earnings over the long term', *Journal of Labor Economics*, 38(3), pp. 653–685.

Herrera-Almanza, C. and Cas, A. (2021) 'Mitigation of Long-Term Human Capital Losses from Natural Disasters: Evidence from the Philippines', *The World Bank Economic Review*, 35(2), pp. 436–460.

Hombrados, J.G. (2020) 'The lasting effects of natural disasters on property crime : Evidence from the 2010 Chilean earthquake', *Journal of Economic Behavior and Organization*, 175, pp. 114–154.

Kirchberger, M. (2017) 'Natural disasters and labor markets', *Journal of Development Economics*, 125, pp. 40–58.

Kousky, C. (2016) 'Impacts of natural disasters on children', *Future of Children*, 26(1), pp. 73–92.

Krueger, A.B. and Lindahl, M. (2001) 'Education for Growth: Why and For Whom?', *Journal* *of Economic Literature*, 39(4), pp. 1101–1136.

Kuka, E., Shenhav, N. and Shih, K. (2020) ‘Do Human Capital Decisions Respond to the Returns to Education? Evidence from DACA’, *American Economic Journal: Economic Policy*, 12(1), pp. 293–324.

Masiero, G. and Santarossa, M. (2020) ‘Natural disasters and electoral outcomes’, *European Journal of Political Economy*, p. 101983.

McDermott, T.K.J. (2012) *The Effects of Natural Disasters on Human Capital Accumulation, IIIS Discussion Paper*. 391. Dublin.

McDermott, T.K.J., Barry, F. and Tol, R.S.J. (2014) ‘Disasters and development: Natural disasters, credit constraints, and economic growth’, *Oxford Economic Papers*, 66(3), pp. 750–773.

Nguyen, C.V. and Pham, N.M. (2018) ‘The impact of natural disasters on children ’ s education : Comparative evidence from Ethiopia ,’ *Review of Development Economics*, 22(4), pp. 1561–1589.

Noy, I. (2009) ‘The macroeconomic consequences of disasters’, *Journal of Development Economics*, 88(2), pp. 221–231.

O’Toole, V.M. and Friesen, M.D. (2016) ‘Teachers as first responders in tragedy: The role of emotion in teacher adjustment eighteen months post-earthquake’, *Teaching and Teacher Education*, 59, pp. 57–67.

Park, A., Sawada, Y., Wang, H. and Wang, S. (2015) ‘Natural Disaster and Human Capital Accumulation: The Case of the Great Sichuan Earthquake in China’, in Sawada, Y. and Oum, S. (eds) *Disaster Risks, Social Preferences, and Policy Effects: Field Experiments in Selected ASEAN and East Asian Countries*. Jakarta: ERIA Research Project Report FY2013, pp. 201–232.

Paudel, J. and Ryu, H. (2018) ‘Natural disasters and human capital: The case of Nepal’s earthquake’, *World Development*, 111, pp. 1–12.

Philipp Heger, M. and Neumayer, E. (2019) ‘The impact of the Indian Ocean tsunami on Aceh’s long-term economic growth’, *Journal of Development Economics*, 141.

Di Pietro, G. (2018) ‘The academic impact of natural disasters: evidence from L’Aquila earthquake’, *Education Economics*, 26(1), pp. 62–77.

Di Pietro, G. and Mora, T. (2015) ‘The effect of the L’Aquila earthquake on labour market outcomes’, *Environment and Planning C: Government and Policy*, 33(2), pp. 239–255.

Ramdhani, S., Istiqomah, E.N. and Ardiyanti, G.K. (2012) ‘The History of Yogyakarta, an Education City’, in *International Proceedings of Economics Development and Research*, pp. 21–24.

Redding, S.J. and Sturm, D.M. (2008) ‘The Costs of Remoteness: Evidence from German Division and Reunification Structure of the Technical Appendix’, *American Economic Review*, 98(5), pp. 1766–1797.

Resosudarmo, B.P., Sugiyanto, C. and Kuncoro, A. (2012) ‘Livelihood Recovery after Natural Disasters and the Role of Aid: The Case of the 2006 Yogyakarta Earthquake’, *Asian Economic Journal*, 26(3), pp. 233–259.

Rush, J. V (2018) ‘The Impact of Natural Disasters on Education in Indonesia’, *Economics of Disasters and Climate Change*, 2, pp. 137–158.

Strauss, J., Witoelar, F. and Sikoki, B. (2016) *The Fifth Wave of the Indonesia Family Life Survey: Overview and Field Report: Volume 1, WR-1143/1-NIA/NICHD*. RAND Working Papers.

Sulistyaningrum, E. (2015) 'Household expenditure in response to natural disasters', *Journal of Indonesian Economy and Business*, 30(3), pp. 257–272.

Sulistyaningrum, E. (2017) 'the Impact of Earthquake on Child Test Score', *Journal of Indonesian Economy and Business*, 32(2), pp. 104–120.

Toya, H. and Skidmore, M. (2007) 'Economic development and the impacts of natural disasters', *Economics Letters*, 94(1), pp. 20–25.

U.S. Geological Survey (2016) *The Severity of an Earthquake*. Available at: https://pubs.usgs.gov/gip/earthq4/severitygip.html.

United Nations Office for Disarmament Affairs (2020) *Human Cost of Disasters: An overview of the last 20 years 2000-2019*, *Human Cost of Disasters: An overvie of the last 20 years*. United Nations.

Wang, J., Yang, J. and Li, B. (2017) 'Pain of disasters: The educational cost of exogenous shocks evidence from Tangshan Earthquake in 1976', *China Economic Review*, 46(July), pp. 27–49.

Worden, C.B. and Wald, D.J. (2016) *ShakeMap 4 Manual — ShakeMap Documentation documentation*, *U.S. Geological Survey*. Available at: http://usgs.github.io/shakemap/.

World Bank (2007) *One year after the java earthquake and tsunami: reconstruction achievements and the results of the java reconstruction fund*, *Java reconstruction fund progress report*. Washington, D.C. Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/647621468338510405/one-year-after-the-java-earthquake-and-tsunami-reconstruction-achievements-and-the-results-of-the-java-reconstruction-fund.

Zhang, Z., Shi, Z., Wang, L. and Liu, M. (2011) 'One year later: Mental health problems among survivors in hard-hit areas of the Wenchuan earthquake', *Public Health*, 125(5), pp. 293–300.

Appendix

Table A1: Summary Statistics (mean and s.d.) of Districts' Characteristics

Variable	Unaffected Districts	Affected Districts	Diff
Total Number of Students	227,133.41	139,661.41	-87,472.01**
	(137,479.66)	(42,213.89)	(43,826.28)
Total Number of Teachers	12,236.54	10,410.20	-1,826.34
	(6,038.10)	(3,113.01)	(1,938.97)
Total Number of Schools	1,022.58	758.300	-264.28
	(520.63)	(277.98)	(167.32)
Student to School Ratio	231.18	196.368	-34.816
	(67.05)	(61.371)	(22.045)
School Density	2.739	3.131	0.392
	(5.500)	(4.103)	(1.788)
Population	1,170,445	762,506.12	-407,938.88*
	(732,475.37)	(259,372.39)	(233,812.83)
GRDP/1000	10,721.09	6,692.98	-4,028.12
	(16,307.47)	(3,572.72)	(5,188.22)
Total Number of Districts	105	10	115

Note: GRDP stands for Gross Regional (district-level) Domestic Product.

UBIREA

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics

WEBSITE: www.ub-irea.com • **CONTACT:** irea@ub.edu

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona