UNIVERSITAT DE BARCELONA

Essays on Economics of Education

Khalifany Ash Shidiqi

PhD in Economics

Essays on Economics of Education

Khalifany Ash Shidiqi

UNIVERSITAT DE BARCELONA

2024

PhD in Economics | Khalifany Ash Shidiqi

UNIVERSITAT DE BARCELONA

PhD in Economics

Thesis title:

Essays on Economics of Education

PhD candidate:

Khalifany Ash Shidiqi

Advisors:

Antonio Di Paolo

Álvaro Choi

Date:

March 2024

UNIVERSITAT DE BARCELONA

Acknowledgements

I express my deepest gratitude to الله سبحانه وتعالى for the strength, smoothness, and ease that have been so valuable to me, crucial in overcoming the challenges faced during my PhD journey. My experience was immensely enriched by a robust support system and the privilege of being surrounded by remarkable individuals. I am profoundly appreciative of each person who contributed to making these years not only a period of significant learning but also of great joy.

First and foremost, I am grateful to my supervisors, Antonio Di Paolo and Álvaro Choi, for their ongoing support, patience, and guidance. Looking back on the start of my PhD path, I'm astounded at how much I've learned from both of you. Thank you for always being there for me, even during your busiest moments, showing me the bright side, teaching me how to discover it, and assisting me in finding solutions to difficulties that initially were impossible to address. Your ongoing advice, both in research and in life, in exploring new views has been essential. My advancement as a researcher is due to both of you, my role models, who taught me that a successful academic career can coexist with a satisfying life.

My sincere thanks go to Bernat Mallén Alberdi, Georgios Tsiachtsiras, and Hamza Benazzi, who have been my discussion buddies, especially during my first year in the PhD program. I am also extremely grateful to my colleagues and friends from Indonesia, Romi Hartarto and Wahyu tri Wibowo, who provided valuable insights into my research; Taufik Akbar, Muhammad Miftahun Nadzir, and Susilo Darsono, who have become online hangout friends, Dr. Caesar Marga Putri and her family who have been the closest thing to the family during my stay in Barcelona, and all my Indonesian hangout friends who live in Barcelona.

Special thanks to Jordi Roca, whose unwavering support and dedication to streamlining every administrative procedure have been instrumental from the outset of my PhD journey. His efforts have significantly eased the complexities associated with my academic endeavors, ensuring a smoother path until the end.

I am grateful to Universitas Muhammadiyah Yogyakarta for their financial support during my PhD studies, particularly Prof. Gunawan Budiyanto

(Rector), Prof. Nano Prawoto (Deputy Chancellor for Human Resources), Prof. Achmad Nurmandi (Vice Chancellor for Cooperation and International Affairs), Prof. Eko Priyo Purnomo, Prof. Rizal Yaya (Dean of Faculty of Economics and Business), Dr. Diah Setyawati Dewanti (Deputy Dean for Academic, Student Affairs and AIK), Dr. Meika Kurnia Puji (Deputy Dean for Human Resources and Finance), Prof. Endah Saptutyningsih (Heads of Economics Department), Dr. Dyah Titis Kusuma Wardani (Secretary of the Economics Department), Dr. Rini Juni Astuti (Former Head for Human Resources), Dr. Adhianty Nurjanah (Head for Human Resources), Dr. SN Nurul Makiyah (Head of Lecturer & Educational Careers Division), Mr. Suhariyanto (Known as Pak Alek, staff of Lecturer Career), and all staffs in economics department.

To my beloved family, your unfailing support has been my rock through every stage of life. Mama and Papa, no words can completely explain how grateful I am for everything you have done for me. Thank you to my brothers, Muhammad Saddam Sofyandi and Muhammad Ilman Nafi'a, for bringing fraternity and joy into my life. My heartfelt gratitude also goes to my father, mother, sister, and brother-in-law, whose encouragement has been the foundation of my path. Your faith in me has given me unending strength and gratitude.

Finally, and most importantly, to my cherished wife, Andika Ridha Ayu Perdana, and our beautiful and precious children, Muhammad Sheva Aqeil Shidiqi and Shaneen Muhayya Shidiqi, for your steadfast support and encouragement. Your faith in me, even in difficult times, has motivated me to strive for the stars. Sharing every step of this trip with you has been a great delight and privilege.

Contents

List of Tables

List of Figures

1. Introduction

This dissertation focuses on the determinants and the effects of human capital formation in Indonesia and contribute to the extensive literature on the economics of education. Since the seminal works by Mincer (1958), Goode (1959), Schultz (1961), and Becker (1975), a large amount of research has been conducted within this field, highlighting how educational investment boosts job performance and economic productivity. Such advancements are attributed to the acquisition of knowledge, skills, and personal traits through education, as discussed by Nelson et al. (1966) and further developed by Lucas (1988). Human capabilities, when combined with physical capital in the production function, play a crucial role in economic growth, as demonstrated by the empirical research of Barro (1991), Benhabib and Spiegel (1994), Barro and Sala-i-Martin (1995), and Griliches (1996). Within the context of globalization, the quantity and quality of human capital play a growing role in determining the competitive advantage of emerging countries (Awan et al., 2011; Cremin and Nakabugo, 2012; Montenegro and Patrinos, 2014). Indeed, the economic rise of East Asian economies was clearly driven by their strategic investments in human capital building, which led to significant progress in economic development, and an important reduction in poverty and income inequality (World Bank, 1993).

At the individual level, the benefits of investment in education can be framed in monetary and non-monetary terms. From the financial perspective, the primary return to human capital investment is in the form of higher lifetime earnings, which has been largely studied (Angrist and Krueger, 1991; Card, 1999; Blundell et al., 2004; Cascio and Lewis, 2006; Oreopoulos, 2006a; Aydemir and Kirdar, 2017; Hampf, 2019). Psacharopoulos and Patrinos (2018) conducted a comprehensive comparison among countries, highlighting the latest trends and patterns in returns to schooling based on a database of 139 countries for the 1950–2014 period. Their analysis reveals universally positive returns to schooling, though the magnitude varies by country classification and region.

On the other hand, extends its impact beyond monetary aspects, influencing non-economic facets such as criminal behavior (Sabates and Feinstein, 2007; Lochner, 2011) and health outcomes (Lleras-Muney, 2005; Silles, 2008;

Powdthavee, 2010; Lochner, 2011), which hold significant importance in developing countries where crime rates remain elevated (Harrendorf et al., 2010; Natarajan, 2016), and health indicators lag behind those of developed nations (Ortiz-Ospina and Roser, 2020). Additionally, education reinforces political participation (Dee, 2004; Hoskins et al., 2008; Wantchekon et al., 2015; Croke et al., 2016; Larreguy and Liu, 2023), a critical factor for enhancing democratic institutions and governance in developing countries with nascent democratic frameworks. Moreover, by enhancing a greater understanding of the ideal timing for marriage, promoting efficient family planning, and managing fertility, education directly mitigates the issue of early marriage, teen pregnancy, and the spread of sexually transmitted infections (Duflo et al., 2015; Marchetta and Sahn, 2016), empowering individuals in the marriage market with the self-awareness and necessary tools for informed decision-making.

Economists have extensively analyzed the multitude of factors shaping educational outcomes and human capital accumulation, as well as their impact. During the last decades, the focus rely on results from empirical strategies that enables obtaining causal estimates, such as randomized controlled trials (Banerjee et al., 2007; Duflo et al., 2011; Fryer, 2011; Chetty et al., 2014). For instance, Kremer et al. (2013) provide a comprehensive overview on the impact of educational interventions in developing countries, focusing on enhancing school participation and learning outcomes through randomized evaluations, and exploring access, quality improvement, and technology's role in education. However, when randomization is not feasible, researchers often turn to quasi-experimental designs to establish causality (Angrist and Pischke, 2009). Such methods includes differences in differences, instrumental variables, and regression discontinuity designs which have been widely used in the area of economics of education (Hanushek et al., 2006a, 2006b, 2011a, 2011b, 2016). For instance, numerous studies have been produced by employing quasi-experimental designs on the government interventions aimed at increasing educational attainment, time in school, enrollment, cognitive performance, as well as reducing drop-out rates through an extension of the compulsory schooling law (Angrist and Krueger, 1991; Harmon and Walker, 1995, 1999; Oreopoulos, 2006a; Grenet, 2013; Eble and Hu, 2019). Large literature has also examined the impact of unexpected external shocks, such as natural disasters, on individuals and families, utilizing quasi-experimental methods to assess how these events can disrupt the effectiveness and implementation of educational policies. (Jensen, 2000; Baez et al., 2010; Wilkinson et al., 2013; Caruso and Miller, 2015; Kousky, 2016; Cerqua and Di Pietro, 2017; Di Pietro, 2017; Paudel and Ryu, 2018). Considering the contemporary challenges such as conflict, epidemics, natural disasters, and rapid technological changes, it is paramount to explore how these global transitions might alter the drivers of educational outcomes, human capital accumulation, and the broader benefits of pursuing education.

In this dissertation, I study the effect of education policies and natural disasters on human capital formation, as well as its effects on socioeconomic outcomes, with a focus on Indonesia. This country is selected due to a unique context characterized by its vast archipelagic geography, diverse ethnic composition, and varying levels of economic development across regions. However, it's important to clarify that my analysis predominantly concentrates on Java, which, despite being a singular part of Indonesia, reflects broader national trends due to its large population and its role as economic hub. Furthermore, Indonesia's varied educational policies and the frequent occurrence of natural disasters offer a rich context to explore the impact on educational outcomes and the ensuing effects on social structures, such as interethnic marriage patterns. This setting allows for a detailed study of how external shocks and policy intervention in the educational sector contribute to societal change and human capital accumulation in the emerging economies, with Java serving as a focal point within Indonesia's complex and evolving educational landscape.

This dissertation is composed of three independent articles. In the first article (chapter 2)[1], together with both my advisors Antonio Di Paolo and Álvaro Choi, we examine the medium to long-term impact of earthquake exposure during school age on educational attainment, utilizing data from the Indonesia Family Life Survey and geolocated information on earthquake intensity by district in Indonesia. We employ a difference-in-difference approach by exploiting variations in exposure across birth cohorts and districts, and we reveal a notable reduction in years of schooling by nearly one year and a decreased likelihood of completing mandatory education.

[1] The chapter has been published in the Economics of Education Review and is available online at the following link: https://www.sciencedirect.com/science/article/pii/S0272775723000444?via%3Dihub

However, it does not significantly impact enrollment in higher education. Most importantly, we identify the disruption of educational infrastructure as a critical mechanism in how natural disasters negatively affect human capital accumulation. We also detected that young individuals living in affected areas are more likely to be still a student at the time of the survey, pointing to potential delays in educational progression as another adverse consequence of earthquake exposure.

In the second article (chapter 3), together with my advisor, Álvaro Choi, we evaluate the medium to long-term impacts of the extension of compulsory education from six to nine years on educational outcomes. We adopt a sharp regression discontinuity strategy using panel data from the Indonesia Family Life Survey (IFLS). Our findings suggest that the reform successfully increased junior secondary completion, enrollment in senior secondary schooling, completion of 12 years of schooling, and overall years of education, especially among females and individuals from less-educated families. Nevertheless, it did not significantly affect university attendance. Furthermore, the analysis of heterogeneous effects emphasized the reform's significant advantages in rural areas and for females, demonstrating its impact on decreasing educational inequalities and improving gender equality. Overall, our findings provide policymakers with valuable guidance on customizing educational reforms to meet the specific needs of diverse demographic groups, improving their effectiveness on educational outcomes.

In the third and last paper (chapter 4), together with my advisor Antonio Di Paolo, we analyze the effect of educational attainments on interethnic marriages within Indonesia's diverse and developing context. Utilizing data from Java Island derived from the 2014 Indonesian Family Life Survey, combined with administrative information on the location and year of establishment of Higher Education Institutions (HEI), we apply an instrumental variable/two-stage least squares (IV/TSLS) approach to discern causal effects. This approach leverages on the variability in HEI accessibility based on the individual's year of birth and district, using the proximity of HEIs—within a 10-kilometer radius of a district's centroid at age 18—as an instrumental variable for educational attainment. By conducting our analysis at the individual level, with separate regressions for males and females, we find that higher education attainments significantly increase the probability of exogamy, that is, having a partner from a different ethnic group. Notably, this effect is more pronounced for females. Our findings, consistent across numerous robustness checks, suggest a credible causal relationship. Interestingly, the influence of education on interethnic marriage does not vary with parental education levels or mixed ethnicity, though it is notably lesser among Javanese individuals relative to other ethnic groups. Further exploration into the mechanisms behind these findings points to migration patterns and shifts in social norms as relevant channels behind observed link between higher education expansion, educational achievements and interethnic marriage, suggesting that fostering human capital formation may play a crucial role in mitigating ethnic segregation.

Finally, chapter 5 concludes the dissertation, highlighting the main results, policy implications, and potential future research.

2. Earthquake exposure and schooling: impacts and mechanisms

2.1. 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 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, 2021), crime (García and 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). 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.

Most of the existing papers focus on a specific type of natural disasters: earthquakes. 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, Tian et al.

(2022) also find evidence of negative effects of in utero exposure to the 1976 Tangshan earthquake on educational attainment. Interestingly, they suggest maternal psychological stress as one of the main mechanisms behind the negative effect of earthquakes on educational outcomes. 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. This paper focuses, however, on exposure during schooling age.

Indeed, 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 investigates the effects and the underlying mechanisms of exposure to a strong earthquake during school age on human capital formation, proxied by individual 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. Indeed, 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, being 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 very limited. 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, who were in their last year of primary school, 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, combining the use of the MMI and the residential history of citizens, 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.

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 the 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, considering both demand and supply-side factors, which might drive the relationship between earthquake exposure and attained schooling. This indeed represents a significant value added of our work. Specifically, as for the demand side factors, 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, that is, a supply-side channel. 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 programs 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 aforementioned 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 2.2 describes the data and descriptive statistics. Section 2.3 discusses the empirical models used in this study, and Section 2.4 presents the main findings. Section 2.5 concludes by discussing the implications of the empirical findings.

2.2. Data and descriptive statistics

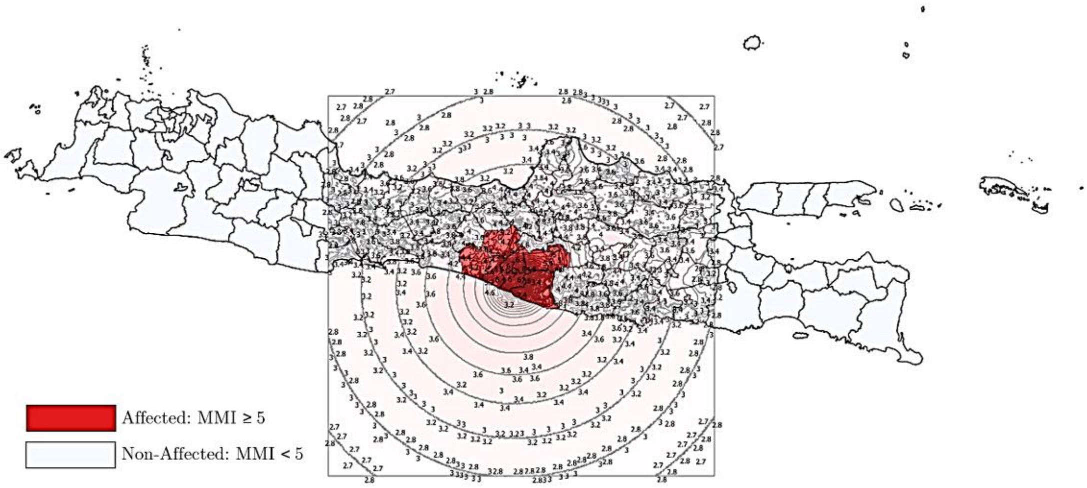
Our empirical analysis focuses on the Java Island, the most populated island of Indonesia, and its capital, Jakarta, the country's most populated city, where the 2006 earthquake took place. Indonesia is a country that is prone to seismic upheaval due to its location on the so-called Pacific "Ring of 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, 2021).[2] 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 2.1, the ShakeMap file only covers the central area of the Java Island, meaning that the 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 2.1: 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.*

---

[2] 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. Moreover, no other measures to capture the strength of the earthquake are available at the district level for the earthquake we investigate in this paper. Nevertheless, there is a clear (although approximated) relationship between the MMI and other indicators, such as the Richter Scale.

(2016) and Masiero and Santarossa (2020), among others.[3] 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[4], 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 2.1, the areas colored red are the affected districts based on our definition.

The second database is the Indonesian Family Life Survey (IFLS) database[5], 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 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.[6] 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

---

[3] Similar approaches were also followed by Cipollone & Rosolia (2007), Kirchberger (2017), Paudel & Ryu (2018) and Hombrados (2020).
[4] According to the U.S. Geological Survey (2016), regions exposed to MMI greater than five are categorized as "strong" (which approximatively corresponds to a value equal or above 4 of the Richter scale). 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 (Table A.2.1 in the Appendix). Nevertheless, in the empirical analysis we check for the sensitivity of our results to different boundaries to define affected and unaffected districts.
[5] IFLS data can be obtained from https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html.
[6] We also use the data from the 2007 wave to retrieve information about pre-determined variables with missing values in the 2014 wave, with the aim of preventing the loss of observations. Furthermore, we exploit the 2000 wave along with the 2007 wave for falsification analysis.

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.[7] 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.[8] Moreover, we consider individuals living in affected and

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

[8] Notice also that the individuals in our sample, residing in the Java Island, were not affected by any other relevant and dramatic natural disaster, comparable to the Yogyakarta earthquake (BMKG, 2018).

unaffected areas, according to the registered value of the MMI scale for the district of residence in 2006.

Table 2.1: 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***	-0.610***
	(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	0.037**
	(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	

The descriptive statistics for the main variables in the analysis are reported in Table 2.1.[9] 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.[10] We also

[9] Aggregate districts' characteristics are reported in Table A.2.1. of the Appendix.
[10] 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

estimate the effect on education level, namely the completion rates of primary and junior high school and the enrollment rates of post-compulsory schooling.[11]

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 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.[12] Indeed, these control variables appear to be balanced, since although there are significant differences between individuals residing in affected and unaffected areas,

---

the years of completed education. For 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.

[11] Indicators for enrollment and completion of education levels are constructed on the basis of completed years of schooling (without considering grade repetition). 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.

[12] 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. In order to construct the indicator for earthquake-related family casualties, we exploited a specific set of questions included in IFLS 4, which allow a) identifying households that were affected by the Yogyakarta's earthquake of 2006 and b) selecting families that answered yes to the question "Did any of the disaster was severe enough to cause death or major injuries of a household member, cause direct financial loss to the household, or cause household member to relocate?".

these are similar between those belonging to the young and old cohorts. This is indeed the 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.

2.3. Identification strategy

2.3.1. 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 1970-1986) and who were living in the two areas of the Java Island. The baseline regression that we estimate is,

$$\begin{aligned} Y_{id} = \alpha + \beta I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \\ +\theta_{yb} + \delta_d + X_i + \varepsilon_{id} \end{aligned} \tag{2.1}$$

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.[13] 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 et al., 2013; García and 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

[13] 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).

errors of equation (2.1) at the district level, which is the level of variation of exposure to the earthquake.

2.3.2. 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[14] and estimate the following equation:

$$Y_{id} = \alpha + \sum_{j} \beta_j I(1987 \leq yb_i \leq 2000) \times I(MMI_d \in k) + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \quad (2.2)$$

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 (2.1)) 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 2.1) and b) excluding districts located more than 200 or even 100 kilometers away from the closest district with MMI ≥ 5.[15] Related to that, we adopt a matching approach based on the method used in Redding

[14] 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.

[15] 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.

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. We replicate this exercise 10,000 times and estimate equation (2.1) 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.[16]

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 1956 and 1972, 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 districts). Similarly, we use 2007 data from IFLS 4 and retain the same cohorts of individuals as in the previous

[16] We also repeated the same exercise after excluding districts located in Yogyakarta and Central Java provinces (i.e. the areas that contain the affected districts), which provided similar evidence (available upon request).

falsification exercise, that is 1973-1986 for the fake treated group and 1956-1972 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.

2.3.3. 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.[17] The corresponding equation takes the form:

$$Y_{id} = \alpha + \sum_{j} \beta_j I(yb_i \in j) \times I(MMI_d \geq 5) + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \tag{2.3}$$

[17] 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.

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 (2.1). This clarifies whether affected individuals (i.e. in school age and residing in affected districts in 2006) are more likely to have 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 or damages 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.[18] Therefore, on the one hand, we also estimate equation (2.1) 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:

$$Y_{id} = \alpha + \beta I(1987 \leq yb_i \leq 2000) \times \frac{\sum dsch_d}{\sum sch_d^{2005}} + \theta_{yb} + \delta_d + X_i + \varepsilon_{id} \tag{2.4}$$

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

[18] 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.

among affected individuals. Moreover, we also adopt a triple difference model that includes the interaction between the indicators for being in 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.[19] Specifically, first we estimate equation (2.1) 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.

[19] 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.

2.4. Results

2.4.1. 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 2.2. 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).

Table 2.2: 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***	-0.752***
	(0.162)	(0.176)
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%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

We begin by presenting the impact of the 2006 earthquake on years of education, which are displayed in Table 2.2. 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).

Table 2.3 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[20], 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 2.3: 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***	-0.103***	-0.004	-0.005
	(0.019)	(0.019)	(0.025)	(0.021)	(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%. Primary School Completion is an indicator that individuals have completed their education for at least primary school (years of education ≥ 6). Junior High School Completion is an indicator that individuals have completed their education for at least junior secondary schooling (years of education ≥ 9). Post Compulsory Enrollment is an indicator that individuals have ever attended for at least one year in senior secondary schooling (educational attainment ≥ 10 years). The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

The magnitude of these effects is large, compared to those obtained by studies focusing on in utero exposure. For example, Tian et al. (2022) estimated that individuals whose mothers lived in areas that were intensely affected by the 1976 Tangshan earthquake while they were pregnant completed 0.18 fewer years of schooling. Moreover, in utero exposure to the earthquake reduced the probability of completing middle school, completing

[20] 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.

high school, or attending college by 1.5, 2.1 and 1.2 percentage points, respectively.

2.4.2. 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 ≥ 5). Therefore, we define categorical dummies for different values of the registered MMI, which leads to the estimation of equation (2.2). The results are shown in Table A.2.2 of the Appendix and indicate 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).[21] Moreover, there is virtually no difference in the estimates for different segments of the MMI distribution above the cut-off we used in the baseline estimations.

Table 2.4: Sensitivity to the choice of unaffected districts

Districts selection:	100 KM	100 KM	200 KM	200 KM	MMI Map	MMI Map
	(1)	(2)	(3)	(4)	(5)	(6)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-1.077***	-1.230***	-1.009***	-1.152***	-1.031***	-1.169***
	(0.241)	(0.241)	(0.166)	(0.176)	(0.164)	(0.172)
R-squared	0.213	0.317	0.192	0.295	0.194	0.296
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%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order. District selection based on distance from the affected areas (columns (1)-(4)) or on the MMI map of Figure 2.1 (columns (5) and (6)).

The following set of robustness checks involves the definition of unaffected districts. Specifically, instead of using the pool of districts of the Java Island

[21] Notice that the positive coefficients for values of the MMI higher than 0 and lower than 5 a are due to the fact that young cohorts residing in these districts achieved more schooling than older cohorts, relatively to what happened to individuals residing in districts outside the MMI map (for which we imputed an MMI equal to 0). This evidence is reported in Table A.2.3 of the Appendix, which is also related to the higher coefficients displayed in Table 2.4.

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 2.4, 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.

In a similar vein, with the aim of showing that the resulting evidence is not driven by the choice of unaffected districts, Table 2.5 displays the results obtained after repeating the estimation of equation (2.1) 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 2.5: 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**	-1.093**	-0.711***	-0.954***	-1.058**	-1.113**
	(0.403)	(0.400)	(0.246)	(0.270)	(0.419)	(0.411)
R-squared	0.219	0.316	0.189	0.306	0.232	0.322
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%. The term "yb_i" stands for year of birth of individual i. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Table 2.5 (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***	-0.807***	-0.680*	-0.715*	-1.152***	-1.030**	-1.157***	-1.351***
	(0.218)	(0.259)	(0.345)	(0.413)	(0.339)	(0.436)	(0.225)	(0.200)
R-squared	0.162	0.273	0.244	0.339	0.168	0.285	0.215	0.324
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%. The term "yb_i" stands for year of birth of individual *i*. 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. After running 10,000 replications (Figure A.2.1 in the appendix), 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.

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 in 2000 and 6-36 in 1992 respectively). This time, we rely on the place of residence in 1992 and impute the MMI values based on their residence in 1992. Table 2.6 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 2.6: Impact of the 1992 fake-year earthquake on years of education (using old and very old cohorts)

Wave:	IFLS 5	IFLS 5	IFLS 4	IFLS 4	IFLS 3	IFLS 3
	(1)	(2)	(3)	(4)	(5)	(6)
$I(1972 \leq ybi \leq 1986) \times I(MMI_d \geq 5)$	0.111	0.196	0.089	0.217	-0.077	-0.158
	(0.306)	(0.298)	(0.199)	(0.134)	(0.315)	(0.324)
R-squared	0.213	0.339	0.220	0.433	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%. The term "yb_i" stands for year of birth of individual *i*. Control group: individuals born between 1956 and 1972. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

2.4.3. Heterogenous 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. The analysis of heterogeneous effects is useful for policymaking, since it can help to design policies that are specifically targeted to the most affected (and possibly vulnerable) subgroups of individuals.

First, we analyze whether the earthquake had a differential effect according to age at exposure, which corresponds to equation (2.3). As can be appreciated in Table 2.7, 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 the years of schooling of individuals born between 1987 and 1991.

Table 2.7: Heterogenous effects by age at exposure

	(1)	(2)
$I(1987 \leq yb_i \leq 1991) \times I(MMI_d \geq 5)$	-0.183	-0.170
	(0.245)	(0.237)
$I(1992 \leq yb_i \leq 1996) \times I(MMI_d \geq 5)$	-0.475**	-0.449*
	(0.183)	(0.232)
$I(1997 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-1.740***	-1.817***
	(0.281)	(0.327)
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%. The term "yb_i" stands for year of birth of individual *i*. 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 2.3, 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 2.8.[22] The model with heterogeneous effects by

[22] Following the suggestion of an anonymous referee, we also tried to implement the Romano-Wolf adjustment for multiple hypothesis testing. Overall, the results are

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[23], 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. This evidence points to a protective effect of maternal education, which possibly reduces the risk of dropping out from school even after the troublesome consequences of a natural disaster. Moreover, we also detect stronger effects for individuals with fewer brothers and sisters than those who have three or more siblings, as well as for those who are the first and second born children in the family. Indeed, these groups of individuals are likely to be those who are more reactive to the shock produced by the earthquake in terms of education attendance and progression.[24]

---

consistent, although the degree of significance of some coefficient for the heterogeneous effects of the earthquake is slightly reduced (complete results are not reported but are available upon request).

[23] Similar results are obtained by using the highest level of education between father and mother (available upon request), which is possibly due to the fact that for 82% of individuals in our sample father's education is greater or equal than mother's education.

[24] That is, they are more prone to be the "compliers" (adopting the IV-LATE terminology) to the natural disaster and would have studied more in the counterfactual state of the world in which the earthquake never occurred.

Table 2.8: Heterogenous effects of the earthquake

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(female)$	-0.554***			
	(0.177)			
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(male)$	-0.959***			
	(0.267)			
$I(1987 \leq yb_i \leq 2000) \times I(MMId \geq 5) \times I(non\text{-}moslem)$		-0.146		
		(0.629)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(moslem)$		-0.693***		
		(0.155)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(non\text{-}javanese)$			0.222	
			(0.748)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(javanese)$			-0.905***	
			(0.168)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(dad's\ educ \leq 9)$				-0.654***
				(0.196)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(dad's\ educ > 9)$				-0.579***
				(0.208)
Test for coefficients' equality, p-value	0.147	0.358	0.146	0.772
R-squared	0.272	0.274	0.271	0.275
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, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1%, ** significant at 5%, * significant 10%. The term "yb_i" stands for year of birth of individual i. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Table 2.8 (continued): Heterogenous effects of the earthquake

	(5)	(6)	(7)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(mom's\ educ \leq 9)$	-0.823***		
	(0.207)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(mom's\ educ > 9)$	-0.437***		
	(0.167)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(siblings = 0)$		-1.010**	
		(0.396)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(siblings = 1)$		-1.499***	
		(0.465)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(siblings = 2)$		-1.351***	
		(0.278)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(siblings = 3)$		-0.424	
		(0.548)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(siblings \geq 4)$		-0.278	
		(0.246)	
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(birth\text{-}order = 1)$			-0.935***
			(0.276)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(birth\text{-}order = 2)$			-1.354***
			(0.217)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(birth\text{-}order \geq 3)$			-0.437**
			(0.207)
Test for coefficients' equality, p-value	0.047	0.008	0.002
R-squared	0.270	0.274	0.274
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, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1%, ** significant at 5%, * significant 10%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

2.4.4. Potential mechanisms

We separately assess two demand-side (internal migration and household casualties) and one supply-side mechanism, which may explain the negative relationship between earthquake exposure and human capital accumulation. Thus, 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 2.9, 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. Similar evidence is obtained when we only consider "permanent movers", that is, individuals who changed place of residence after 2006 and did not migrate back to the same district where they were living between 2006 and 2014 (columns (5) - (6) and (7) – (8), respectively). 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.[25]

Table 2.9: Migration as potential mechanism

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	0.028	0.032	-0.846***	-0.900***	0.027	0.031	-0.847***	-0.823***
	(0.017)	(0.021)	(0.172)	(0.180)	(0.018)	(0.199)	(0.163)	(0.179)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(mover_i)$			0.724*	0.699**				
			(0.423)	(0.346)				
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(permanent\ mover_i)$							0.884**	0.699
							(0.429)	(0.427)
R-squared	0.065	0.174	0.186	0.302	0.059	0.185	0.186	0.302
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	20,304	20,304	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%. The term "yb_i" stands for year of birth of individual *i*. 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. The dependent variable for columns (5) and (6) is the indicator for having permanently changed place of residence after 2006, $I(mover_i)$. The dependent variable for columns (7) and (8) is years of schooling; the models also include the base effect of being a permanent mover and the corresponding double interactions.

[25] 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).

In order to investigate 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 2.10 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 2.10 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.

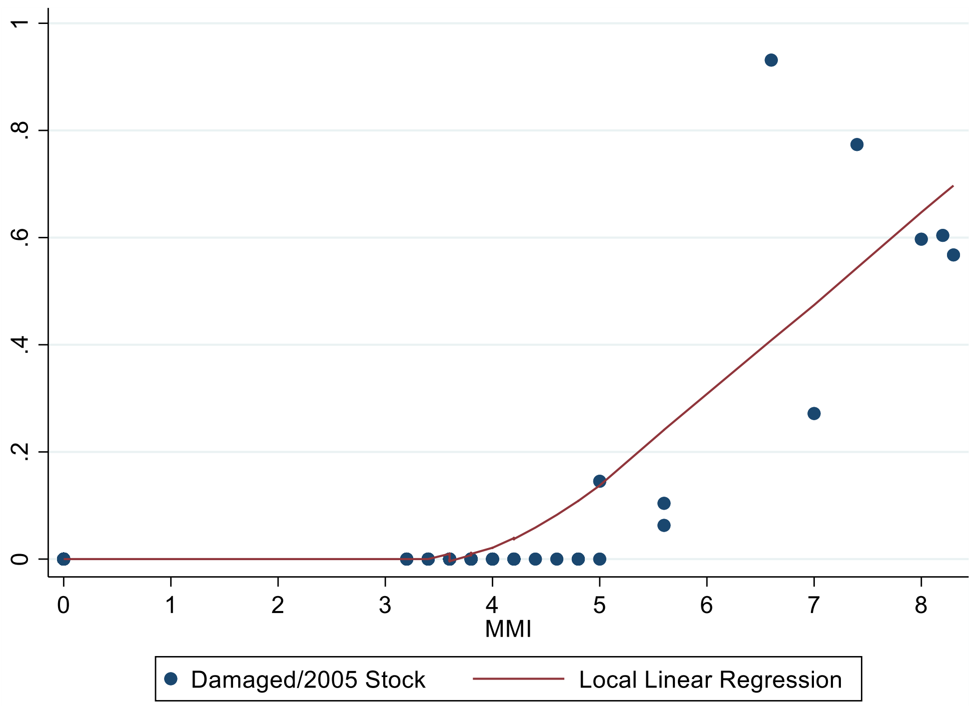
Table 2.10: Household casualties as potential mechanism

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	0.025**	0.025**	-0.614***	-0.726***
	(0.010)	(0.010)	(0.166)	(0.188)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5) \times I(casualties_i)$			0.449	0.495
			(0.727)	(0.818)
R-squared	0.552	0.554	0.184	0.272
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%. The term "yb_i" stands for year of birth of individual *i*. 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.

Finally, we analyze the role of a supply-side factor as a potential mechanism: the disruption of educational infrastructures. That is, we investigate 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 2.2 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 2.2: MMI and damaged/destroyed schools by district (over the 2005 stock)



In columns (1) and (2) of Table 2.11 we report the estimate(s) of the coefficient of interest from equation 2.4, 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. Moreover, we also interacted the share of damaged/destroyed schools with dummies for age at exposure (using the same age ranges than in the analysis of heterogeneous effects). The results, reported in Table A.2.4 of the Appendix, indicate that the impact of the disruption of educational infrastructure mostly affected individuals who were in primary school when the earthquake took place. 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.

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.[26] Also, the youngest might have experienced a delay in the access to the education system. To provide suggesting evidence about this point, in Table 2.12 we show the results obtained from the estimation of the baseline equation (2.1) 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 indicates 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.

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

Table 2.11: Disruption 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**	-1.022***				
	(0.167)	(0.235)				
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$			-0.489***	-0.344***	-0.554***	-0.563***
			(0.116)	(0.125)	(0.175)	(0.196)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$						
$\times I(\sum dsch_d / sch_d^{2005} > 0)$			-0.276**	-0.444***		
			(0.130)	(0.137)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$						
$\times I(\sum dsch_d / sch_d^{2005} > 0.75)$					-0.413***	-0.416*
					(0.145)	(0.226)
R-squared	0.183	0.271	0.183	0.271	0.183	0.271
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, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1%, ** significant at 5%, * significant 10%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Table 2.12: Additional evidence regarding current school attendance

	(1)	(2)	(3)	(4)
$I(young) \times I(MMI_d \geq 5)$	0.103***	0.103***	-0.494***	-0.374**
	(0.019)	(0.019)	(0.146)	(0.169)
R-squared	0.597	0.599	0.174	0.278
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. Standard errors, in parentheses, are clustered at the district level. *** significant at 1%, ** significant at 5%, * significant 10%. The term "*yb*i" stands for year of birth of individual *i*. 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.

2.5. 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 maternal 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. Additionally, 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 anti 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. In addition, 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 that, 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.

Appendix

Table A.2.1: 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	

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

Table A.2.2: 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*	0.572**
	(0.209)	(0.223)
$I(1987 \leq yb_i \leq 2000) \times I(3.5 \leq MMI_d < 5)$	0.496**	0.648***
	(0.196)	(0.202)
$I(1987 \leq yb_i \leq 2000) \times I(5 \leq MMI_d < 7.5)$	-0.536**	-0.554**
	(0.230)	(0.269)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 7.5)$	-0.568**	-0.445*
	(0.220)	(0.229)
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%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Table A.2.3: Additional sensitivity checks for the choice of the control group

	baseline				excluding districts with MMI = 0 (i.e. inside the MMI map)				excluding districts with MMI > 0 & MMI < 5			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 5)$	-0.742***	-0.752***			-1.031***	-1.169***			-0.542***	-0.491**		
	(0.162)	(0.176)			(0.164)	(0.172)			(0.190)	(0.204)		
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d = 0)$			*reference category*								*reference category*	
$I(1987 \leq yb_i \leq 2000) \times I(0 < MMI_d < 3.5)$			0.387*	0.572**			*reference category*				--	--
			(0.209)	(0.223)							--	--
$I(1987 \leq yb_i \leq 2000) \times I(3.5 \leq MMI_d < 5)$			0.496**	0.648***			0.115	0.053			--	--
			(0.196)	(0.202)			(0.199)	-0.208			--	--
$I(1987 \leq yb_i \leq 2000) \times I(5 \leq MMI_d < 7.5)$			-0.536**	-0.554**			-0.932***	-1.222***			-0.528**	-0.537**
			(0.230)	(0.269)			(0.234)	(0.285)			(0.232)	(0.264)
$I(1987 \leq yb_i \leq 2000) \times I(MMI_d \geq 7.5)$			-0.568**	-0.445*			-0.967***	-1.052***			-0.556**	-0.449**
			(0.220)	(0.229)			(0.229)	(0.233)			(0.212)	(0.222)
R-squared	0.183	0.272	0.184	0.273	0.194	0.296	0.1936	0.296	0.185	0.2695	0.1853	0.2695
District & Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of Observations	20,304	20,304	20,304	20,304	9,478	9,478	9,478	9,478	12,883	12,883	12,883	12,883

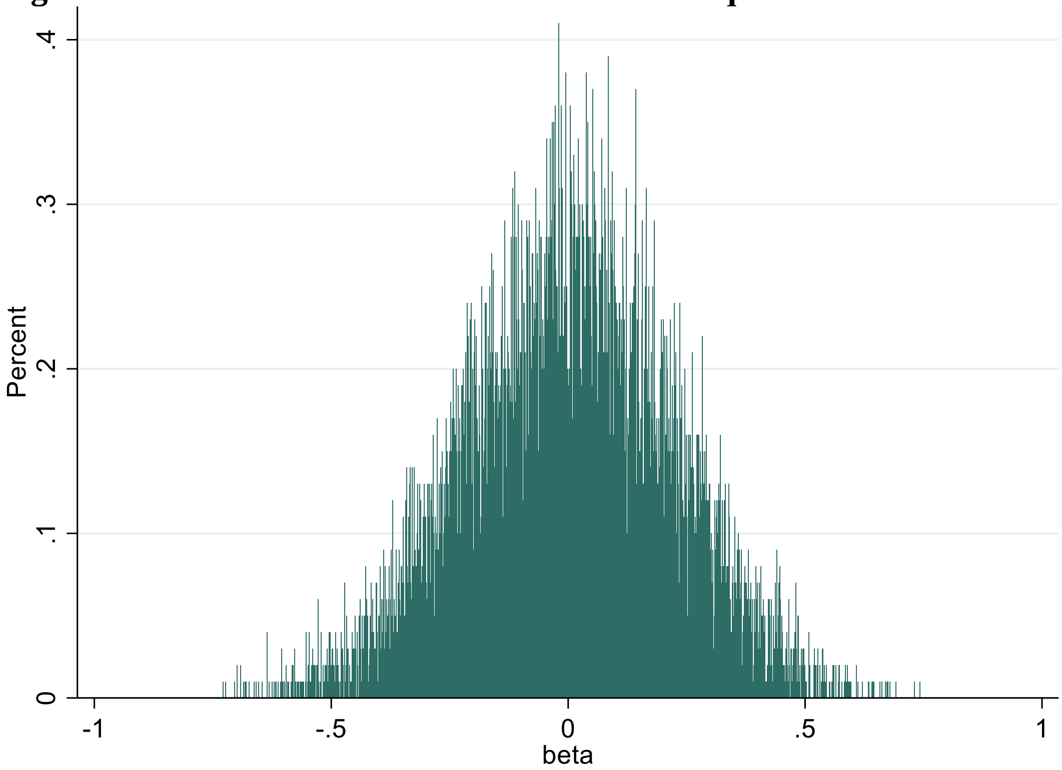
Notes: OLS estimation, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1%, ** significant at 5%, * significant 10%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity,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

Table A.2.4: Heterogeneous effect of the disruption of educational infrastructures by age at exposure

	(1)	(2)	(3)	(4)
$I(1987 \leq yb_i \leq 2000) \times \sum dsch_d / sch_d^{2005}$	-0.986**	-1.022***		
	(0.167)	(0.235)		
$I(1987 \leq yb_i \leq 1991) \times \sum dsch_d / sch_d^{2005}$			-0.262	-0.220
			(0.300)	(0.261)
$I(1992 \leq yb_i \leq 1996) \times \sum dsch_d / sch_d^{2005}$			-0.562**	-0.526
			(0.253)	(0.374)
$I(1997 \leq yb_i \leq 2000) \times \sum dsch_d / sch_d^{2005}$			-2.372***	-2.587***
			(0.361)	(0.419)
R-squared	0.183	0.271	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, dependent variable = years of schooling. Standard errors, in parentheses, are clustered at the district level. *** significant at 1%, ** significant at 5%, * significant 10%. The term "yb_i" stands for year of birth of individual *i*. Control variables are gender, religion, ethnicity, father's and mother's education, number of siblings, and birth order.

Figure A.2.1: Permutation tests of the fake earthquake locations



3. The impact of compulsory schooling reform on educational outcomes: the case of Indonesia

3.1. Introduction

Compulsory education laws aim to enhance access to education by substituting the decision-making capacity of individuals in the level of consumption of educational services to correct information failures, maximize social positive externalities, and feed the economic system with skilled workers. The starting age and extension of compulsory education varies widely, the former being in most developed countries around ages 5-6 and, the latter, between 9 and 12 years (*UIS*, 2022). In developing countries, compulsory education laws have also been used as tools for expanding access to higher levels in a sequential way (Choi, 2009).

Existing evidence on the impact of mandatory education laws on educational outcomes in developed countries usually finds that compulsory education laws lead to significant gains in the accumulation of human capital. For the US, Angrist and Krueger (1991) and Acemoglu and Angrist (1999), using the 1960 - 1980 censuses and taking advantage of variation in the timing of law changes across states over time, showed that compulsory schooling laws were effective in compelling students to stay in school until they reach the legal dropout age. Using a one-percent sample of the 1960 Census too, Lleras-Muney (2002) found that the Compulsory Attendance Law and the Child Labor Laws increased educational attainment. More recent research such as Stephens and Yang (2014), Lleras-Muney and Shertzer (2015), Clay et al. (2021) or Shanan (2021) reach similar conclusions. Findings for the UK by Goldin and Katz (2011) or Harmon and Walker (1995) show the positive effects of compulsory education laws on secondary school enrollment rate and years of completed schooling. Cross-country analyses by Oreopoulos (2006a, 2006b) concluded that minimum school leaving age reforms increased educational attainment in the US, Canada and the UK.

Evidence for emerging countries is however scarcer and its results less clear. Pischke and Von Wachtser (2008) utilized the changes in Germany's obligatory schooling law following World War II and discovered that the law only boosted the average number of years of education by a small amount. Moving to Africa, Elsayed (2019) used a natural experiment to analyze the additional one-year extension to primary education in 1999 and found that the reform had a large positive effect on educational attainment. The extension of compulsory education to lower secondary education significantly increased compulsory education completion and improved the average number of post-compulsory graduates in Senegal (Momo et al., 2021).

Evidence for Asia is mixed. In Turkey, the 1997 school compulsory reform increased the likelihood of completing eight-grade and high school and effectively enhances years of education (Kirdar et al., 2016; Dayioglu and Kirdar, 2022). Furthermore, this policy substantially increased female junior high school completion rates who had to discontinue their formal education due to parental opposition (Dursun et al., 2022). In 2012, the law was once again revised and compulsory education expanded from eight to twelve years. Utilizing the latter, Erten and Keskin (2019) found that the reform increased high school attendance. Spohr (2003) and Tsai et al. (2009) discovered that the Taiwan's 1968 compulsory education expansion from six to nine years increased the average years of schooling. Similar evidence was found by Korwatanasakul (2019) when assessing a 1978 reform in Thailand. Remarkably, the increase in the number of years of schooling almost doubled the additional schooling required by the law. Similarly, Fang et al. (2012a) analyzed the 1968 nine years of compulsory education reform in China and found that this policy increased years of completed education by almost one year. However, in 2006 to 2007, the 2006-2007 free compulsory education reform enforcing nine years of compulsory schooling had no significant impact on school enrollment (Tang et al., 2020). Thus, the relationship between compulsory education laws and educational outcomes in developing countries is far from being straightforward.

Moreover, the analysis of heterogeneous effects of compulsory education laws has been seldom assessed. In terms of gender, some studies have concluded that the policy change has a stronger effect on females than on males (Lleras-Muney, 2002; Tsai et al., 2009; Fang et al., 2012; Dayioglu and Kirdar, 2022) but also vice versa (Spohr, 2003; Goldin and Katz, 2011; Elsayed, 2019). In terms of area of residence, the law has a larger impact on individuals who lived in rural areas in China, Egypt and Turkey (Fang et al., 2012; Elsayed, 2019; Dayioglu and Kirdar, 2022, respectively). Finally, in terms of parental education, the 1997 mandatory schooling reform in Turkey affected individuals whose parents had completed over five years of education more than those whose parents had less. However, the effect became insignificant for those whose parents had completed education at least ten years of schooling (Dayioglu and Kirdar, 2022).

Evidence for Indonesia is even scarcer. A natural experiment which consisted in the extension of the academic year due to the change of the start of the academic year from January to July in 1978 was assessed by Parinduri (2014 and 2017). According to our knowledge, the only study that has evaluated the causal impact of compulsory schooling policies in Indonesia is Lewis and Nguyen (2020), who found that the 1994 nine-year compulsory schooling initiative had no discernible impact on child educational attainment. This would be explained by shortcomings in the implementation of the program. Lewis and Nguyen (2020) acknowledge potential weaknesses[27] in their analysis, particularly as regards lack of data on possible socioeconomic determinants of school participation. In this paper we assess this same reform but applying an enhanced identification strategy -as well as adding these missing socioeconomic characteristics and overcoming other shortcomings[28]. This allows us to estimate strikingly different results. We contribute to the literature by evaluating the medium to long-term impacts of the extension of compulsory education from six to nine years on a set of educational outcomes in which previous studies focus on the immediate effects. In addition to assessing the average effect, we also conduct the heterogeneous effects

[27] Lewis and Nguyen (2020) did not have access to data on socioeconomic characteristics (i.e., parents' education, household income, household size, the number and gender of child's siblings, child's birth order or school costs) that may influence the positive effects of government's compulsory schooling.

[28] Lewis and Nguyen (2020) assume that all children start school at the age of seven, with the cut-off date for determining school age being August 31st. This contrasts with the Indonesian context, where children can begin schooling at six, introducing a challenge in accurately determining their assignment to either the treatment or control group. Therefore, in our identification strategy, we employ a 'donut hole' RDD approach following Cattaneo et al. (2020) by excluding individuals who are exactly at the cut-off point in which we detail extensively in section 3.

according to gender, parental education, and area (urban/rural) of residence of a law passed in Indonesia in the mid-nineties.

We follow a sharp regression discontinuity design (RDD) using panel data from the Indonesia Family Life Survey (IFLS). Our findings suggest that the reform successfully increased junior secondary completion, enrollment on senior secondary schooling, completion on 12 years of schooling, and overall years of education, especially among females and individuals from less-educated families. However, it did not significantly affect university attendance. Furthermore, the analysis of heterogeneous effects emphasized the reform's significant advantages in rural areas and for females, demonstrating its impact on decreasing educational inequalities and improving gender equality. Our robustness tests, including falsification exercises and sensitivity analyses, support the causal link between the reform and the observed educational improvements.

The remainder of the paper is organized as follows. Section 3.2 overviews Indonesia's 1994 compulsory education law. Section 3.3 describes the data, descriptive statistics and empirical models used in this study, and Section 3.4 presents the main findings. Section 3.5 concludes.

3.2. The 1994 national compulsory education law

The education system in Indonesia follows a 6+3+3+4 framework, where students spend 6 years in elementary school, 3 years in junior high, 3 years in senior high, and up to 4 years in college. This format has remained unchanged since the implementation of the initial curriculum in 1946 (Al-samarrai and Lewis, 2021). According to World Bank (2019), the education system is primarily dominated by public institutions, particularly at the elementary and junior secondary levels. Approximately 98% of students are enrolled in public primary schools, while around 75% are enrolled in public junior secondary schools.

After years of a gradual expansion at the primary education level[29], the government of Indonesia turned to boosting secondary graduates during the

---

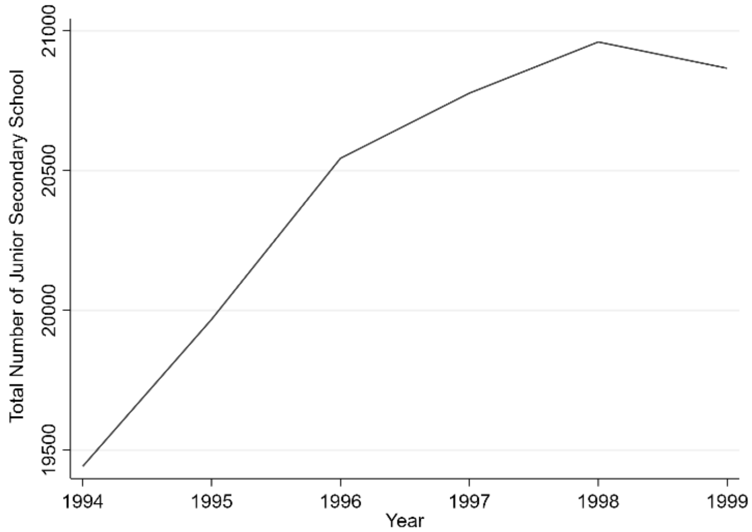
[29] One policy that was phenomenal in making primary education universal was the *Sekolah Dasar* INPRES program. From 1973 to 1978, the Indonesian government undertook one of the most extensive school construction initiatives under this program (Duflo, 2001).

late 1980s. By that time, the six-year compulsory schooling policy (ages 7 to 12)[30], in force since 1988, had succeeded in making primary school almost universal, with completion rates exceeding 95% (World Bank, 2022a; Suryadarma et al., 2006)). However, by 1990, junior secondary school graduation rates were still below 50% (World Bank, 2022a). Since then, the government, private institutions and international donor agencies responded by implementing several programs, including training for teachers, providing textbooks for students and teachers, supplying and distributing science equipment to schools, and establishing more buildings for secondary schools (Yeom et al., 2002). Those early 1990s programs[31], intended to support the imminent move to a nine-years compulsory education system. As reported by Indonesia Statistics (Badan Pusat Statistik, BPS), the number of junior secondary schools and teachers in junior secondary schools increased by around 7% (Figure 3.1) and 12% (Figure 3.2), respectively, within five years of the 1994 reform. These trends underline that the support of educational infrastructure development accompanied the extension of compulsory schooling.

[30] Although, by law the compulsory age for children to start elementary school is at age seven, some begin a year earlier or a year later (Barakat and Bengtsson, 2018).

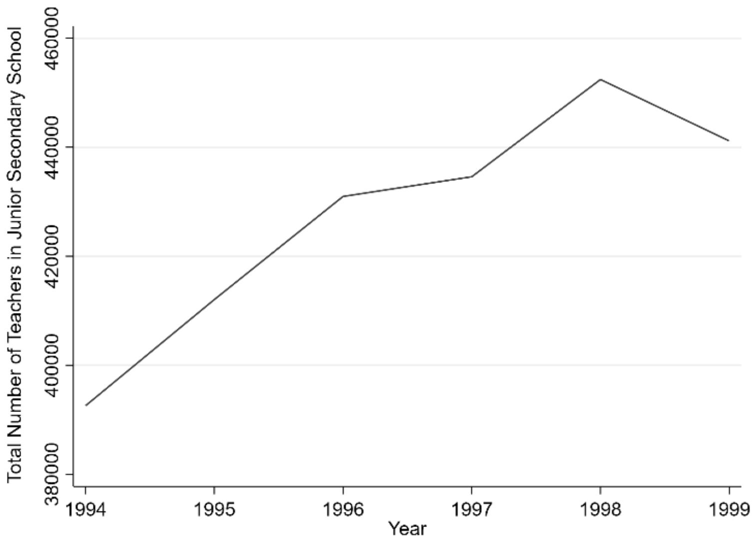
[31] Based on Law Number 2 of 1989 concerning the National Education System, article 14, paragraph 2, states that citizens aged seven must attend primary school or equivalent education until graduation. However, the implementation of nine-years compulsory education was only effective nationally in the academic year of 1994/1995 after the issue of Presidential Instruction Number 1 of 1994.

Figure 3.1: Number of junior secondary school



Source: Indonesian Statistics

Figure 3.2: Number of teachers in junior secondary school



Source: Indonesian Statistics

Indeed, in April 1994, the President of Indonesia issued Presidential Instruction Number 1, which extended compulsory schooling from 6 to 9 years. The reform was effectively implemented at the beginning of the 1994/95 academic year, that is, in September 1994. Students who had not graduated from junior secondary school by the end of the 1993/1994 academic year were bound to complete nine years of education. This extension aimed to reduce child labor and keep kids in school, especially those who could not afford to pursue higher education (Yeom et al., 2002).

3.3. Data, empirical methodology, and preliminary checks

3.3.1. Data

We exploit data from Indonesia Family Life Survey (IFLS)[32], covering more than 80% of the Indonesian population within the survey area (Strauss et al., 2016). The survey provides information about individuals' characteristics, educational attainment, and the locations (province and district) of the respondents' birthplace, current residence and entire migration history. As our main aim is to analyze the medium-long term impact of the 1994's nine years compulsory schooling reform on the completion of nine years of compulsory schooling, enrollment in senior secondary schooling, completion of 12 years of schooling, university attendance, and years of education, we exploit the third, fourth, and fifth wave of the IFLS survey (2000, 2007, and 2014, respectively). Anyway, we also exploit information from previous waves for specific purposes.

We create a set of dummy variables for measuring the different outcomes (completion rate of nine years of compulsory schooling, enrollment in senior secondary schooling, completion of 12 years of schooling, and university attendance). Additional variables are created to capture whether an individual was enrolled for at least one year in senior secondary schooling, completed at least senior secondary schooling or been enrolled for at least one year at the university level, respectively. As for years of education, we exploit information about the highest level of schooling attended and the highest grade ever completed by the respondents. Using these data we can calculate the years of completed education.[33] We distinguish between individuals who were subjected to the compulsory schooling policies (treated) and those who were not (control) by using month and year of birth. In accordance with the

[32] IFLS data can be obtained from https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html.

[33] For 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 year of completed education is equal to 8 years. This calculation already considers the grade repetition.

directives of the President, the educational reforms were initiated during the 1994/1995 academic year, compelling students who had not attained completion of their junior high school to pursue their studies until attaining graduation.

Table 3.1 illustrates summary statistics for individuals that belong to treatment (born in and after September 1978) and control (born before September 1978) groups. The table provides sample means and standard deviations of the variables used in the empirical analysis (outcome and controls). Regarding control variables, we only use a set of characteristics which are relevant factors in the education production function (Hanushek, 2020): gender, fathers' and mothers' education, number of siblings, birth order, and religions. The treated and non-treated samples seem unbalanced at both sides of the threshold. However, the significant differences in those predetermined characteristics are the overall differences between treated and control groups. These differences are no longer there when we assess these variables around the cutoff point (refer to section 4.3) which make us feel confident about the comparability of both subgroups as suggested in De La Cuesta and Imai (2016).

Table 3.1: Summary statistics

Variable	Control (1)	Treatment (2)	Diff (3)
Junior Secondary Completion	0.637	0.655	0.017**
	(0.481)	(0.475)	(0.008)
Senior Secondary Enrollment	0.316	0.346	0.031***
	(0.465)	(0.476)	(0.008)
Senior Secondary Completion	0.304	0.281	-0.023***
	(0.460)	(0.449)	(0.007)
University Attendance	0.136	0.118	-0.019***
	(0.343)	(0.322)	(0.005)
Years of Education	9.167	9.262	0.095*
	(3.573)	(3.226)	(0.055)
Male	0.516	0.492	-0.024***
	(0.500)	(0.500)	(0.008)
Fathers' Educ >= 9 Years	0.286	0.283	-0.002
	(0.452)	(0.451)	(0.007)
Mothers' Educ >= 9 Years	0.244	0.222	-0.022***
	(0.430)	(0.416)	(0.007)
Number of Siblings	3.840	3.904	0.065
	(2.970)	(2.546)	(0.045)
Birth Order	4.680	5.186	0.506***
	(3.138)	(2.638)	(0.047)
Islam	0.904	0.893	-0.011**
	(0.295)	(0.309)	(0.005)
Christian	0.049	0.061	0.012***
	(0.216)	(0.240)	(0.004)
Other Religion	0.047	0.046	-0.001
	(0.212)	(0.209)	(0.003)
Individual Live in Urban at age 12	0.386	0.424	0.038***
	(0.487)	(0.494)	(0.008)
Individual Live in Java Island at age 12	0.431	0.455	0.024***
	(0.495)	(0.498)	(0.008)
Observations	7,074	8,176	15,250

Note: *** significant at 1%, ** significant at 5%, * significant 10%.
Source: Indonesian Family Life Survey Wave 2000, 2007, and 2014.

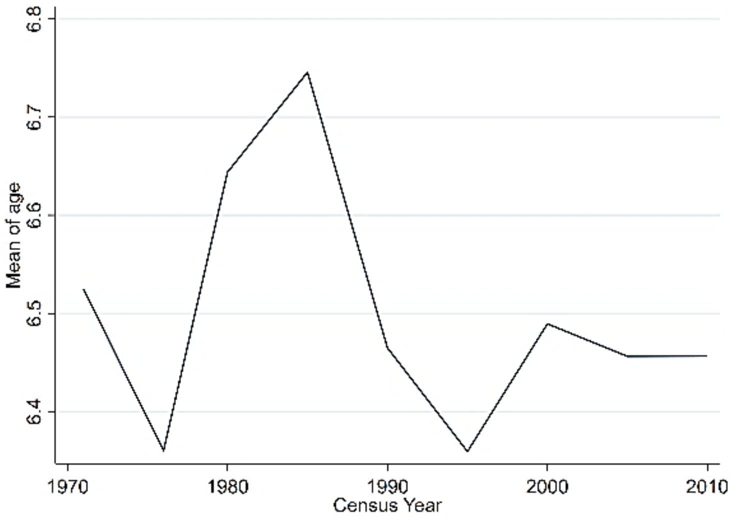
3.3.2. Identification strategy

The compulsory education reform implemented in 1994 stipulates that individuals born in September 1978 and subsequent months were mandated to complete a minimum of nine years of formal education. In contrast, those born prior to this cut-off date had the provision to discontinue their education earlier (before completing 9 years of education). We exploit the individual's month and year of birth to create the cut-off point and running variable in a sharp RDD for estimating the impact of the reform on a set of educational outcomes. Individuals born on and after September 1978 were mainly in grade eight at the end of the 1993/1994 academic year and exposed to the reform. Thus, in our RDD those born on and after September 1978 represent the treatment group. However, under Indonesian Law No. 2 of 1989, Article 14, children who are six years old have the right to start basic education, while at the age of seven, enrollment in basic education becomes compulsory. This regulation introduces possible confusion in determining whether a child should be placed in the treatment or control group for our identification strategy -Lewis and Nguyen (2020) did not take this into account. As we can see from Figure 3.3, it depicts the trend of the average age at which children began their first year of primary school in Indonesia from 1970 to 2010. [34] The data shows that the mean enrollment age has hovered around 6.5 years, suggesting a mix of students starting school at ages six and seven which in line with the finding in Barakat and Bengtsson (2018). Therefore, to identify precisely, we exclude individuals precisely at the cut-off point.[35] This procedure is known as "donut hole" approach as discussed in Cattaneo et al. (2020) and have been used in a number of studies, namely for performing robustness checks (see, e.g., Almond and Doyle, 2011; Fukushima et al., 2016; Hoxby and Bulman, 2016; Kirdar et al., 2018).

[34] The data was obtained from a 10% sample of the Indonesian census records for the years 1970, 1980, 1990, 2000, and 2010. Detailed information about these samples can be found at https://international.ipums.org/international-action/sample_details/country/id

[35] Lewis and Nguyen (2020) attempted to address this issue by employing a fuzzy regression discontinuity (RD) design. However, their findings indicated that the results were not statistically significant. Similarly, we have also conducted our analysis without excluding individuals at the cutoff point and we obtained a lack of statistical significance as well.

Figure 3.3: Average age at first enrollment in first grade



Source: IPUMS (10% of the Indonesian census records for the years 1970, 1980, 1990, 2000, and 2010)

Following Imbens and Lemieux (2008) and Cattaneo et al. (2020) we use local linear regressions in our RDD estimations and implement the optimal bandwidth selection using the Calonico et al. (2014) procedure to minimize bias and maximize precision. Our main RDD estimate is as follow:

$$Y_i = \alpha + \beta treat_i + f(x_i) + \varepsilon_i$$
$$\forall x_i \in (c - h, c + h)$$

where Y_i is the dependent variable, $treat_i$ is the treatment status, x_i is the forcing variable, h is the bandwidth around the cut-off point c, and ε_i is the error term. Following Lee and Card (2008), we cluster standard errors at the month-year of birth level to accommodate for specification errors in the forcing variable. We conduct our analysis by both excluding and including individual covariates as control variables. These controls include a set of dummy variables indicating gender, whether the individual's father or mother has completed nine years of schooling, number of siblings, the individual's birth order, religion, district of residence at age 12 fixed effects, month-of-birth fixed effects, and year of survey fixed effects.

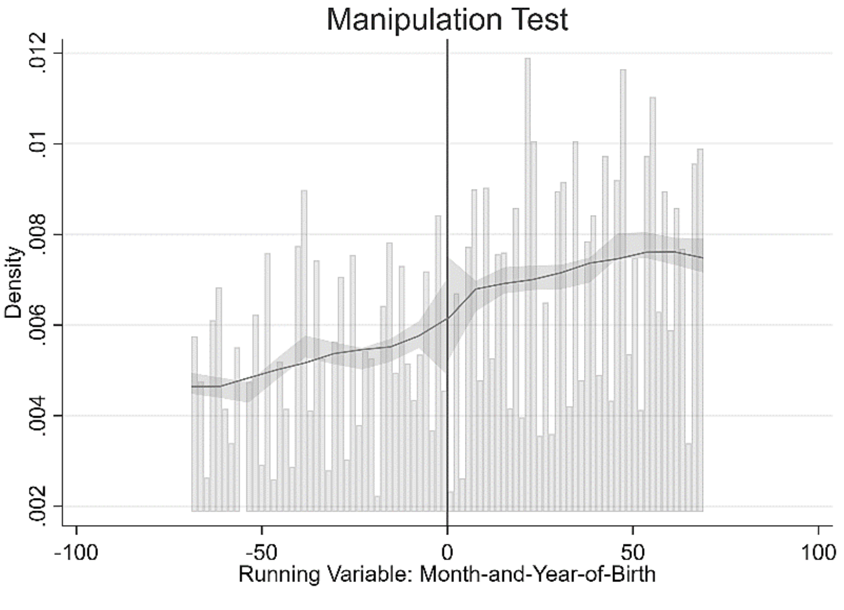
Before conducting our main analysis, we first perform essential preliminary assessments. The initial check involves a density test of the running variable to check for manipulation and ensure the validity of our analysis (McCrary, 2008). If the density of the running variable changes sharply at the cutoff point would challenge the validity of our RDD. The second test is performing estimation on the predetermined characteristics of the individuals to verify that there is a balance in the characteristics on both sides around the cutoff, treating these characteristics as dependent variables and assessing them against the running variable (Cattaneo et al., 2020). Similarly, a significant difference in any predetermined characteristics would invalidate its causal inferences in our RDD.

Furthermore, we follow Erten and Keskin (2019) to perform sensitivity analyses using 1.5, 2, and 4.5 times the size of the optimal bandwidth selection, and static bandwidth for the heterogenous analysis. We also conduct a falsification exercise, in which we treat a group of older people who are belong to the control group in our baseline model (1971–1977) as if they were treated (fake treatment) and much older individuals, who were born between 1964 and 1970, as if they were in the control group (fake control). Then, we set the fake policy reform as if it happened in 1987 by estimating a placebo regression instead of the actual one in 1994.

3.3.3. Preliminary checks

We begin by conducting validity checks for the RDD, based on the assumption that the treatment assignment around the cutoff is essentially random. This idea relies on the fact that individuals cannot manipulate their birth dates to be on a specific side of the cutoff. Since month-year of birth is determined are fixed well before reform's announcement, such manipulation is implausible in our context. Nevertheless, to ensure the integrity of our design, we perform a manipulation check on our running variable, as recommended by McCrary (2008). Following the procedure described in Cattaneo et al. (2018), we do not find substantial evidence of a discernible discontinuity in the density of our running variable. This suggests that the underlying distribution of the data remains consistent across the spectrum of the running variable, thereby upholding the validity of the RDD (Figure 3.4 and Table 3.2).

Figure 3.4: Local density plots of month-and-year of birth around the cutoff point



Another important check that we need to validate is whether the predetermined individual's characteristics that we control for in our estimation are continuous at the discontinuity (Imbens and Lemieux, 2008). A significant difference in any individuals' predetermined characteristics would question the validity of its causal inferences in our RDD. Therefore, we check this assumption by estimating individuals' predetermined characteristics against the running variable, allowing for a discontinuity at the threshold. The predetermined characteristics that we plot are a set of dummy variables indicating gender, whether individuals' fathers and mothers have completed at least nine years of schooling, number of siblings, and religion. The results in Table 3.3 indicate that there are no significant jumps at the cutoff point for any predetermined individuals' characteristic, indicating that the changes we see in the main estimation are most likely due to the impact of the reform rather than individuals' characteristics.[36]

[36] The graphical analysis of the local averages of the individual's predetermined characteristics against the running variable are depicted in Figure A.3.1. It indicates that there are no significant jumps around the cutoff point for predetermined characteristics in each wave.

Table 3.2: Manipulation test of running variable

Method:	
T	0.479
P>|T|	0.632
Bandwidth Left	23.000
Bandwidth Right	23.000
Observations	15250

Note: This table presents the findings from the manipulation test of the running variable following the procedure of Cattaneo et al. (2018). The running variable is individuals' month and year of birth.

Table 3.3: RD estimate for pre-determined characteristics

Male	-0.037
	(0.042)
Father's Education	0.0028
	(0.035)
Mother's Education	0.028
	(0.041)
No. Siblings	-0.203
	(0.152)
Birth Order	-0.186
	(0.204)
Islam	0.006
	(0.025)
Christian	0.009
	(0.019)
Other Religions	-0.015
	(0.019)
Observations	15250

Note: This table presents the results of preliminary checks using a Regression Discontinuity Design (RDD) to assess the balance of pre-determined characteristics across the treatment threshold. Each row in the table represents an estimated treatment effect on a pre-determined characteristic, with separate models run for each characteristic. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

3.4. Results

3.4.1. The average impact of the 1994 compulsory schooling reform

We initiate our analytical work by quantifying the estimated average impact of the policy shift on the educational outcomes of individuals. For each respective outcome, we introduce two specifications, without and with control variables, defined in accordance with the optimal bandwidth selection described in section 3.3. Table 3.4 reports the results from our RDD, showing the effects of the intervention on various educational achievements: fulfilling nine years of obligatory schooling (Panel A); enrollment in senior secondary education (Panel B); completion of a 12-year schooling period (Panel C); matriculation into university (Panel D); and cumulative years of education (Panel E).[37]

Our findings suggest that the obligatory schooling reform had a fruitful impact in fostering an increase in the accomplishment of nine years of education (8.1 percentage points), the enrolment in senior secondary education (8.3 percentage points), the completion of a 12-year educational cycle (7.9 percentage points) and completed years of education (0.55 years). However, the influence on the probability of university enrollment is statistically insignificant. The results displayed in Table 4 are in line with those documented in different countries. For instance, Momo et al. (2021) observed that the impact of the reform exhibits greater magnitude in Senegal, with an increased likelihood of completing grade 10 by seven percentage points. This effect size is comparatively smaller when compared to the outcomes observed in Thailand, where a similar reform initiative resulted in an approximate extension of four additional years of schooling, nearly twice the duration mandated by the corresponding legislation (Korwatanasakul, 2019). Furthermore, the findings align closely with those observed in other developing nations, where the range of outcomes varied from 0.8 to 2.14 years (Tsai et al., 2009; Fang et al., 2012; Elsayed, 2019).

[37] Graphical analyses are reported in Figure A.3.2.

Table 3.4: Treatment effect on educational outcomes (optimal bandwidth)

	Panel A: Completion of nine years of compulsory schooling		Panel B: Enrollment in senior secondary schooling		Panel C: Completion of 12 years of schooling		Panel D: University attendance		Panel E: Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RD Estimate	0.080**	0.081**	0.087**	0.083***	0.082**	0.079***	0.032	0.028	0.568**	0.551***
	(0.036)	(0.034)	(0.035)	(0.024)	(0.035)	(0.026)	(0.028)	(0.020)	(0.273)	(0.176)
Opt. bw ($\hat{h}$)	24	20	22	23	22	19	26	31	24	21
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	15250	15250	15250	15250	15250	15250	15250	15250	15250	15250

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and exploits the optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

3.4.2. Robustness checks and falsification test

We then perform a set of robustness checks to ensure the reliability of our findings in the baseline estimation. Specifically, we examine the sensitivity of our results to two factors: the adjustment of bandwidth selections and a falsification analysis using a fake placebo reform. To begin, we implemented bandwidths of 1.5, 2, and 4.5 times the size of the optimal bandwidth selection. Table A.3.1 presents these results, demonstrating that the estimated effects remain stable under different bandwidths.

Finally, we performed a falsification exercise to further strengthen the validity of our main findings. In this exercise, we considered a fictitious compulsory schooling reform and replicated our baseline analysis using never-treated individuals. We take a cohort of older individuals, born between 1964 and 1970 who were excluded from our main analysis, as a fake control cohort, and we use our original control cohort of individuals born between 1971 and 1977 as a fake treated cohort. We then estimate a placebo regression by assuming that the compulsory reform took place in 1987 rather than the actual year of 1994. The RDD estimates of these placebo estimations are presented in Table A.3.2 (columns 1 to 10). Results are not statistically different from zero for all educational outcomes.[38] These null results of the placebo estimations reinforce the causal interpretation of our main findings which further bolsters the robustness of our conclusions and emphasizes the credibility of the estimated relationship between the compulsory schooling reform and the educational outcomes.

3.4.3. The heterogeneous effects of the 1994 compulsory education law

This section provides heterogenous analysis based on gender, parental education levels, and geographical location of residency when individuals reached the age of 12. Panels A, B C, D and E in Table A.3.3 provide the set of educational outcomes as in the baseline estimation. Overall, the reform appears to have a variety of impacts on educational outcomes, with the most consistent and significant effects observed in the completion of nine year of compulsory schooling, senior secondary enrollment and completion, and years of education, particularly for females and individuals with less educated parents (fathers and mothers). This finding indicates that the reform especially helped female children to complete nine years of compulsory schooling, which also increasing their propensity to enroll and complete senior secondary schooling. On the other hand, the insignificant impact of the reform on boys as well as individuals with highly educated parents (fathers and mothers) suggests that most of them would pursue their education regardless of the reform. This result is in line with the findings in Lleras-Muney (2002), Tsai et al. (2009), Fang et al. (2012), and Dayioglu and Kirdar (2022) in which the effects of the reform are stronger on females. The policy did not affect university attendance, regardless of gender.

In a subsequent stage of our analysis, we explore a heterogenous analysis rooted in the urban or rural areas when individuals reach the age of 12 (Table A.3.4). Our findings reveal that the reform positively affects both individuals who lived in urban and rural areas to complete mandatory schooling.

[38] The graphical illustrations are displayed in Figure A.3.3.

Nonetheless, the most pronounced benefits, including enrollment and completion of senior secondary education, university attendance, and the total years of education were exclusive to individuals living in rural areas. However, in urban areas, progressing beyond compulsory schooling has become the norm, even in the absence of this reform. These results actually in line with the findings of Tsai et al. (2009), Fang et al. (2012), Elsayed (2019), and Dayioglu and Kirdar (2022), where the observed effects were more pronounced for rural children compared to urban children.

We also carry out a robustness test of our heterogenous analysis, by applying a fixed bandwidth around the cutoff point. The fixed bandwidth corresponds to the optimal bandwidth selection for the entire sample for each educational outcome that is used in the baseline estimation. The results displayed in Table A.3.5 and A.3.6 are very similar to those in Table A.3.3 and A.3.4, implying that the estimates are robust to alternative specifications with different bandwidths.

Furthermore, we conduct analysis by combining individuals' gender with parental education and urban and rural contexts to gain a deeper understanding of the heterogeneous effect observed in Table A.3.3 and A.3.4. This additional task aims to determine which gender would derive the greatest advantage from this reform in different scenarios, considering parental education and childhood residential location. Understanding which gender benefits more is important for several reasons. First, it empowers policymakers to tailor affirmative educational reforms to specifically target the distinct demands and obstacles encountered by various populations. This precision can result in more effective strategies aimed at reducing educational inequalities. Second, closing gender gaps in education may translate into long term productivity gains which may reflect in labor market outcomes, economic development, and social equity (Ghosh and Ramanayake, 2020). This issue is especially relevant in Indonesia, where there are noticeable disparities between genders in education (Afkar et al., 2020). One noteworthy gap is the difference in enrollment rates between males and females in secondary education, as shown in Figure A.3.4 and also highlighted in Ridwan (2017).

As illustrated in Table A.3.7 and A.3.8, the reform benefits females with lower parental education (father and mother), especially in completing mandatory schooling and overall years of education. However, the positive impact is stronger for females with highly educated mothers, especially for senior secondary enrollment and completion rates. This difference can be explained by two main factors. Firstly, data from the IFLS for 1993 and 1997 demonstrate that females with highly educated mothers had significantly higher increases in senior secondary enrollment and completion rates than those with less educated mothers. Specifically, the enrollment rate for females with less educated mothers rose from 7.41% in 1993 to 9.5% in 1997, while for those with highly educated mothers, it increased from 32.1% to 35.31%. Similarly, the completion rate for females with less educated mothers increased from 6% to 8.5%, and for those with highly educated mothers, it went from 26.3% to 31.8%. Secondly, the greater impact on females with highly educated mothers is likely due to these mothers having higher educational aspirations for their daughters and being more involved in their education, leading to a prioritization of their daughters' schooling as highlighted in Sathar et al. (2013) and Khalid (2023).

The varying result by parental education of the effect of the reform introduces an intricate discussion which is inherently linked to the intergenerational transmission of education. Research in this area has produced varied results, influenced by factors such as parental involvement, sociocultural backgrounds, and the gender of the children. For instance, Chevalier (2004), Black et al. (2005), Holmlund et al. (2011) and Stella (2013) conclude that the importance of maternal education appears to supersede that of paternal education. On the other hand, Behrman and Rosenzweig (2005), Pronzato (2012), and Amin et al. (2015) argued that the significance of paternal education appears to outweigh that of maternal. In our case, the impact of the reform on female who have either a low-educated father or a low-educated mother appears to be similar. Therefore, we further tested for coefficient equality between daughters of low-educated fathers and those of low-educated mothers. Our tests found statistically significant for all outcomes indicating that the reform has a slightly higher effect on females with less educated fathers than those with less educated mothers.[39]

Similarly, Table A.3.9 presents a detailed heterogeneous analysis that intersects gender with urban-rural backgrounds, demonstrating that the reform significantly advantages females across both urban and rural settings.

---

[39] Results are available upon request.

Importantly, it highlights the reform's role in motivating females from rural backgrounds to pursue higher education, suggesting that the reform offers substantial long-term benefits, particularly for females in rural areas. This underscores the reform's effectiveness in bridging educational gaps and fostering greater educational equity.

3.5. Conclusion

This paper analyzed the medium-long-term effect of 1994's compulsory schooling reform on educational outcomes in Indonesia. Beginning in the school years of 1994/1995, the Indonesian government expanded the mandatory education system from six to nine years. Using panel data from IFLS, the reform significantly enhanced educational attainment, particularly increasing the completion rates of nine years of compulsory schooling, enrollment in senior secondary education, and the overall years of education. However, it showed no statistically significant effect on university enrollment. These outcomes demonstrate the reform's success in expanding access from primary to secondary education, although its influence at the tertiary level remains uncertain.

Our results rely on the identification assumption that individuals at the two sides of the thresholds are identical in terms of observable and unobservable characteristics and cannot manipulate their location with respect to the cutoff. We tested this by conducting a manipulation test and the individual's predetermined characteristics against our running variable. The results suggest that we did not find evidence of a significant break in density in our running variable and no significant jumps at the cutoff point for each individual's predetermined characteristics. Furthermore, the results are robust to sensitivity analysis and falsification exercises that point towards the internal of our identification strategy and validate the causal interpretation of the results. Therefore, this paper lines up with the strand in literature which identifies compulsory schooling reforms as effective instruments for expanding access to higher levels of education. This finding is especially relevant for an emerging country such as Indonesia since education represents one of the necessary conditions through which they can foster economic growth and achieve economic and social development.

The heterogeneous analysis also reveals that the reform's benefits were especially noticeable among females and individuals from less-educated family backgrounds, emphasizing its contribution in reducing educational inequalities. Moreover, the reform had a greater impact on individuals from rural areas, indicating its effectiveness in addressing educational gaps between rural and urban areas too. Overall, the result from heterogenous analyses suggests that compulsory schooling reforms do not only increase educational attainment -efficiency gains- but can also play a crucial role in promoting enhanced educational opportunities for underprivileged groups.

All in all, the results displayed along this paper show that compulsory education laws have the potential for expanding the human capital endowment for the population. The insights from Indonesia suggest that previous investments in educational facilities and sufficient enforcement capacity may help to the success of the law. Our findings also emphasize the necessity for supplemental policies specifically designed for individuals in rural areas, females, and those from households with lower socioeconomic status, as these groups are more likely to abandon school and join the labor market at very young ages, a common challenge in developing countries (Psacharopoulos and Patrinos, 2004; Hannum and Buchmann, 2005; Ray and Lancaster, 2005; Klasen and Lamanna, 2009).

Finally, complementary interventions may be considered in order to boost the egalitarian effects of compulsory educational reforms. Scholarships and mentoring programs targeted at girls, as well as campaigns aimed at shifting societal perceptions regarding female education, have shown significant positive outcomes in terms of educational opportunities by gender. For instance, Unterhalter et al. (2014) provide comprehensive reviews of these interventions, underscoring their effectiveness in enhancing girls' educational achievements. Additionally, Duflo (2012) and Kazianga et al. (2013) found that measures like providing uniforms and sanitary pads can notably increase school attendance and decrease dropout rates among girls. The same applies for policymakers aiming to tackle socioeconomic educational gaps. The implementation of measures such as conditional cash transfers has shown its potential to increase enrollment and attendance among children from economically disadvantaged households. Progresa/Oportunidades in México (Attanasio et al., 2012) and Bolsa Familia in Brazil (de Brauw et al., 2015) are excellent examples in that sense. Similarly, findings by Barrera-Osorio et al. (2011 and 2019) in Colombia, Gazeaud and Ricard (2024) in Morocco, and Dustan (2020) in Mexico further illustrate the potential of conditional cash transfers in improving educational outcomes. Lastly, a wide range of measures have been shown to be effective in order to reduce the disparity in education between urban and rural areas. Two good examples are the elimination of transportation fees to school, as highlighted by Adukia et al. (2020), and the supply of school meals, as demonstrated by Aurino et al. (2023). Additionally, interventions focused on improving school infrastructure and teacher training in rural areas, as highlighted by Glewwe et al. (2011) for Ghana and Glewwe et al. (2004) for Vietnam, have demonstrated promising results in addressing the challenges faced by students in these communities.

Appendix

Table A.3.1: Robustness check – Main estimation using various bandwidths

Bandwidth	Optimal ($\hat{h}$)		1.5*($\hat{h}$)		2*($\hat{h}$)		4.5*($\hat{h}$)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Completion on Nine Years of Compulsory Schooling								
RD Estimate	0.080**	0.081**	0.075**	0.072**	0.051**	0.057**	0.081***	0.074***
	(0.036)	(0.034)	(0.034)	(0.030)	(0.030)	(0.026)	(0.022)	(0.019)
($\hat{h}$)	24	20	36	30	48	30	108	90
Panel B: Enrollment on Senior Secondary Schooling								
RD Estimate	0.087**	0.083***	0.091***	0.084***	0.070**	0.065***	0.071***	0.071***
	(0.035)	(0.024)	(0.031)	(0.023)	(0.028)	(0.022)	(0.022)	(0.018)
($\hat{h}$)	22	23	33	35	44	46	99	104
Panel C: Completion on 12 Years of Schooling								
RD Estimate	0.082**	0.079***	0.089***	0.085***	0.074***	0.078***	0.072***	0.070***
	(0.035)	(0.026)	(0.032)	(0.024)	(0.029)	(0.023)	(0.021)	(0.017)
($\hat{h}$)	22	19	33	29	44	38	99	86
Panel D: University Attendance								
RD Estimate	0.032	0.028	0.035	0.026	0.028	0.019	0.013	0.009
	(0.028)	(0.020)	(0.021)	(0.017)	(0.018)	(0.015)	(0.015)	(0.012)
($\hat{h}$)	26	31	39	47	52	62	117	140
Panel E: Years of Education								
RD Estimate	0.568**	0.552***	0.505**	0.480***	0.391*	0.351**	0.298*	0.266**
	(0.273)	(0.176)	(0.250)	(0.171)	(0.228)	(0.162)	(0.173)	(0.133)
($\hat{h}$)	24	21	36	32	48	42	108	95
Observations	15250	15250	15250	15250	15250	15250	15250	15250
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff. Columns (1) and (2) report the baseline model, employing an optimal bandwidth selection as per the methodology outlined in Calonico et al. (2014). Columns (3) and (4) explore local RD regressions with a linear control function, utilizing one-half the optimal bandwidth, while columns (5) and (6) double this bandwidth, and columns (7) and (8) employ a bandwidth five times the optimal size. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.2: Falsification test for educational outcomes using old and very old cohorts

	Panel A: Completion of nine years of		Panel B: Enrollment in senior secondary		Panel C: Completion of 12 years of		Panel D: University attendance		Panel E: Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RD Estimate	0.049	0.043	0.007	0.008	0.024	0.024	-0.007	-0.010	0.454	0.420
	(0.057)	(0.049)	(0.032)	(0.026)	(0.034)	(0.027)	(0.020)	(0.017)	(0.418)	(0.341)
Opt. bw ($\hat{h}$)	23	25	19	20	19	19	24	25	21	20
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	10988	10988	10988	10988	10988	10988	10988	10988	10988	10988

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents falsification analysis of the regression discontinuity (RD). We consider a cohort of older individuals, initially excluded from our estimation sample born between 1964 and 1970 as a fake control cohort and we use our original control cohort of individuals born between 1971 and 1977 as a fake treated cohort. We set the fake policy reform as if it happened in 1987. All results are estimated using the optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.3: Analysis of heterogenous effect of RD treatment (individual covariates)

	Male		Female		Dad Educ ≥ 9		Dad Educ < 9		Mom Educ ≥ 9		Mom Educ < 9	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Completion on Nine Years of Compulsory Schooling												
RD Estimate	0.054	0.061	0.118*	0.124**	0.019	0.018	0.106**	0.101**	0.020	0.008	0.091**	0.106***
	(0.041)	(0.038)	(0.071)	(0.059)	(0.038)	(0.038)	(0.043)	(0.041)	(0.043)	(0.041)	(0.040)	(0.038)
Optimal Bandwidth ($\hat{h}$)	22	28	19	18	23	23	22	21	23	23	23	22
Panel B: Senior Secondary Enrollment												
RD Estimate	0.050	0.046	0.133**	0.126**	0.072	0.075	0.097***	0.089***	0.112	0.111	0.067*	0.076**
	(0.053)	(0.045)	(0.059)	(0.051)	(0.056)	(0.052)	(0.030)	(0.027)	(0.073)	(0.070)	(0.035)	(0.034)
Optimal Bandwidth ($\hat{h}$)	25	26	26	28	24	27	23	23	22	21	22	22
Panel C: Completion of 12 years of schooling												
RD Estimate	0.044	0.010	0.139**	0.139**	0.064	0.061	0.104***	0.100***	0.108	0.108	0.088***	0.081**
	(0.046)	(0.048)	(0.067)	(0.060)	(0.059)	(0.058)	(0.033)	(0.031)	(0.086)	(0.079)	(0.031)	(0.032)
Optimal Bandwidth ($\hat{h}$)	30	22	22	20	25	25	20	19	19	18	28	23
Panel D: University Attendance												
RD Estimate	0.005	0.009	0.044	0.040	0.047	0.045	0.028	0.024	0.007	0.011	0.032	0.033
	(0.034)	(0.026)	(0.044)	(0.037)	(0.059)	(0.055)	(0.025)	(0.022)	(0.084)	(0.093)	(0.032)	(0.029)
Optimal Bandwidth ($\hat{h}$)	25	32	29	31	32	32	31	32	27	23	24	25
Panel E: Years of Education												
RD Estimate	0.194	0.163	0.890*	0.805**	0.257	0.220	0.757***	0.771***	0.107	0.015	0.635**	0.728***
	(0.251)	(0.216)	(0.467)	(0.369)	(0.413)	(0.371)	(0.271)	(0.235)	(0.396)	(0.405)	(0.283)	(0.257)
Optimal Bandwidth ($\hat{h}$)	24	21	26	27	26	28	21	19	28	24	23	22
Observations	7675	7675	7575	7575	4337	4337	10913	10913	3541	3541	11709	11709
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.4: Analysis of heterogenous effect of RD treatment (place of residence)

	Urban		Rural	
	(1)	(2)	(3)	(4)
Panel A: Completion on Nine Years of Compulsory Schooling				
RD Estimate	0.080*	0.095**	0.070*	0.062*
	(0.043)	(0.040)	(0.039)	(0.036)
Optimal Bandwidth ($\hat{h}$)	24	22	23	23
Panel B: Senior Secondary Enrollment				
RD Estimate	0.086	0.108**	0.087***	0.075**
	(0.056)	(0.050)	(0.030)	(0.031)
Optimal Bandwidth ($\hat{h}$)	26	21	25	24
Panel C: Completion of 12 years of schooling				
RD Estimate	0.081	0.098*	0.085***	0.087***
	(0.061)	(0.055)	(0.029)	(0.028)
Optimal Bandwidth ($\hat{h}$)	21	19	24	25
Panel D: University Attendance				
RD Estimate	-0.002	0.002	0.052**	0.047**
	(0.062)	(0.047)	(0.023)	(0.021)
Optimal Bandwidth ($\hat{h}$)	25	31	26	27
Panel E: Years of Education				
RD Estimate	0.431	0.432	0.658**	0.603**
	(0.374)	(0.271)	(0.273)	(0.242)
Optimal Bandwidth ($\hat{h}$)	21	24	23	23
Observations	6196	6196	9054	9054
Controls	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.5: Robustness check - Analysis of heterogenous effect of RD treatment (individual covariates)

	Male		Female		Dad Educ ≥ 9		Dad Educ < 9		Mom Educ ≥ 9		Mom Educ < 9	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Completion on Nine Years of Compulsory Schooling												
RD Estimate	0.050	0.051	0.119*	0.109*	0.022	0.022	0.114***	0.104**	0.022	0.007	0.096**	0.115***
	(0.043)	(0.046)	(0.071)	(0.057)	(0.040)	(0.040)	(0.044)	(0.042)	(0.048)	(0.045)	(0.042)	(0.039)
Optimal Bandwidth ($\hat{h}$)	20	20	20	20	20	20	20	20	20	20	20	20
Panel B: Senior Secondary Enrollment												
RD Estimate	0.037	0.031	0.135**	0.129**	0.065	0.060	0.097***	0.089***	0.111	0.103	0.069*	0.077**
	(0.060)	(0.052)	(0.064)	(0.054)	(0.060)	(0.056)	(0.034)	(0.030)	(0.072)	(0.068)	(0.035)	(0.034)
Optimal Bandwidth ($\hat{h}$)	23	23	23	23	23	23	23	23	23	23	23	23
Panel C: Completion of 12 years of schooling												
RD Estimate	0.019	0.011	0.151**	0.148**	0.050	0.029	0.107***	0.097***	0.106	0.099	0.070*	0.079**
	(0.067)	(0.057)	(0.071)	(0.059)	(0.071)	(0.064)	(0.037)	(0.034)	(0.088)	(0.081)	(0.038)	(0.037)
Optimal Bandwidth ($\hat{h}$)	19	19	19	19	19	19	19	19	19	19	19	19
Panel D: University Attendance												
RD Estimate	0.016	0.008	0.046	0.040	0.045	0.040	0.029	0.022	0.017	0.003	0.029	0.032
	(0.032)	(0.026)	(0.042)	(0.037)	(0.062)	(0.058)	(0.025)	(0.022)	(0.077)	(0.075)	(0.028)	(0.026)
Optimal Bandwidth ($\hat{h}$)	31	31	31	31	31	31	31	31	31	31	31	31
Panel E: Years of Education												
RD Estimate	0.173	0.144	1.004*	0.949**	0.307	0.182	0.759***	0.677***	0.130	0.047	0.658**	0.750***
	(0.294)	(0.227)	(0.524)	(0.394)	(0.484)	(0.419)	(0.287)	(0.241)	(0.517)	(0.468)	(0.294)	(0.262)
Optimal Bandwidth ($\hat{h}$)	21	21	21	21	21	21	21	21	21	21	21	21
Observations	7675	7675	7575	7575	4337	4337	10913	10913	3541	3541	11709	11709
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.6: Robustness check - Analysis of heterogenous effect of RD treatment (place of residence)

	Urban		Rural	
	(1)	(2)	(3)	(4)
Panel A: Completion on Nine Years of Compulsory Schooling				
RD Estimate	0.094**	0.104***	0.070*	0.066*
	(0.043)	(0.040)	(0.041)	(0.039)
Optimal Bandwidth ($\hat{h}$)	20	20	20	20
Panel B: Senior Secondary Enrollment				
RD Estimate	0.088	0.099*	0.076**	0.073**
	(0.062)	(0.052)	(0.031)	(0.030)
Optimal Bandwidth ($\hat{h}$)	23	23	23	23
Panel C: Completion of 12 years of schooling				
RD Estimate	0.084	0.094	0.081***	0.080***
	(0.067)	(0.058)	(0.030)	(0.029)
Optimal Bandwidth ($\hat{h}$)	19	19	19	19
Panel D: University Attendance				
RD Estimate	-0.004	-0.003	0.054**	0.050***
	(0.056)	(0.048)	(0.021)	(0.019)
Optimal Bandwidth ($\hat{h}$)	31	31	31	31
Panel E: Years of Education				
RD Estimate	0.394	0.404	0.673**	0.631**
	(0.391)	(0.286)	(0.290)	(0.261)
Optimal Bandwidth ($\hat{h}$)	21	21	21	21
Observations	6196	6196	9054	9054
Controls	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.7: Cross heterogenous analysis of RD treatment (individual covariates) - Male

	Male*Dad Educ ≥ 9		Male*Dad Educ < 9		Male*Mom Educ ≥ 9		Male*Mom Educ < 9	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Completion on Nine Years of Compulsory Schooling								
RD Estimate	-0.053	-0.049	0.073	0.073	-0.018	-0.017	0.065	0.079
	(0.073)	(0.071)	(0.066)	(0.059)	(0.085)	(0.081)	(0.065)	(0.059)
Optimal Bandwidth ($\hat{h}$)	26	26	33	36	32	35	38	39
Panel B: Senior Secondary Enrollment								
RD Estimate	-0.067	-0.076	0.086	0.079	-0.029	-0.034	0.050	0.049
	(0.097)	(0.096)	(0.087)	(0.079)	(0.116)	(0.109)	(0.095)	(0.094)
Optimal Bandwidth ($\hat{h}$)	23	22	39	40	26	27	33	31
Panel C: Completion of 12 years of schooling								
RD Estimate	-0.119	-0.116	0.051	0.040	-0.070	-0.091	0.048	0.042
	(0.111)	(0.093)	(0.100)	(0.088)	(0.118)	(0.101)	(0.082)	(0.082)
Optimal Bandwidth ($\hat{h}$)	24	23	35	36	28	33	38	36
Panel D: University Attendance								
RD Estimate	0.052	0.066	-0.019	-0.020	0.015	0.021	-0.015	-0.013
	(0.132)	(0.118)	(0.062)	(0.056)	(0.134)	(0.124)	(0.045)	(0.043)
Optimal Bandwidth ($\hat{h}$)	30	30	31	30	32	28	34	33
Panel E: Years of Education								
RD Estimate	-0.617	-0.585	0.474	0.384	-0.379	-0.449	0.234	0.280
	(0.832)	(0.734)	(0.555)	(0.465)	(0.748)	(0.584)	(0.406)	(0.352)
Optimal Bandwidth ($\hat{h}$)	24	24	30	30	27	29	35	36
Observations	2052	2052	5622	5622	1695	1695	5979	5979
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.8: Cross heterogenous analysis of RD treatment (individual covariates) - Female

	Female*Dad Educ ≥ 9		Female*Dad Educ < 9		Female*Mom Educ ≥ 9		Female*Mom Educ < 9	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Completion on Nine Years of Compulsory Schooling								
RD Estimate	0.087	0.072	0.148*	0.158**	0.058	0.048	0.130**	0.189***
	(0.074)	(0.068)	(0.082)	(0.064)	(0.063)	(0.062)	(0.065)	(0.063)
Optimal Bandwidth ($\hat{h}$)	19	20	19	17	20	20	20	17
Panel B: Senior Secondary Enrollment								
RD Estimate	0.109	0.101	0.141**	0.134**	0.282**	0.267***	0.144**	0.137**
	(0.082)	(0.075)	(0.062)	(0.056)	(0.114)	(0.097)	(0.064)	(0.057)
Optimal Bandwidth ($\hat{h}$)	29	28	26	26	15	15	24	25
Panel C: Completion of 12 years of schooling								
RD Estimate	0.129	0.136	0.131**	0.139**	0.358***	0.339***	0.119**	0.116**
	(0.111)	(0.088)	(0.064)	(0.062)	(0.130)	(0.115)	(0.060)	(0.058)
Optimal Bandwidth ($\hat{h}$)	16	21	21	22	25	25	28	29
Panel D: University Attendance								
RD Estimate	0.036	0.035	0.048	0.040	-0.004	-0.002	0.068	0.080
	(0.092)	(0.090)	(0.043)	(0.040)	(0.110)	(0.117)	(0.052)	(0.049)
Optimal Bandwidth ($\hat{h}$)	35	34	29	27	25	23	22	21
Panel E: Years of Education								
RD Estimate	0.295	0.256	1.034**	1.004**	0.301	0.356	0.998*	1.184**
	(0.614)	(0.589)	(0.479)	(0.427)	(0.522)	(0.546)	(0.531)	(0.488)
Optimal Bandwidth ($\hat{h}$)	33	35	26	24	26	23	21	21
Observations	2285	2285	5290	5290	1846	1846	5729	5729
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Table A.3.9: Cross heterogenous analysis of RD treatment (place of residence) – Male and Female

	Male*Urban		Male*Rural		Female*Urban		Female*Rural	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Completion on Nine Years of Compulsory Schooling								
RD Estimate	0.075**	0.074*	0.038	0.019	0.308***	0.226**	0.130*	0.137*
	(0.034)	(0.041)	(0.059)	(0.057)	(0.106)	(0.100)	(0.078)	(0.080)
Optimal Bandwidth ($\hat{h}$)	23	25	25	24	23	25	29	29
Panel B: Senior Secondary Enrollment								
RD Estimate	0.101	0.112	-0.009	-0.014	0.286***	0.221**	0.123*	0.140*
	(0.069)	(0.082)	(0.052)	(0.055)	(0.098)	(0.091)	(0.066)	(0.072)
Optimal Bandwidth ($\hat{h}$)	26	21	19	18	23	26	26	26
Panel C: Completion of 12 years of schooling								
RD Estimate	0.053	0.060	0.015	-0.015	0.238**	0.158*	0.167**	0.183**
	(0.078)	(0.091)	(0.036)	(0.039)	(0.101)	(0.083)	(0.071)	(0.075)
Optimal Bandwidth ($\hat{h}$)	24	21	27	24	25	28	31	32
Panel D: University Attendance								
RD Estimate	0.025	0.018	-0.006	-0.008	0.110	0.082	0.128**	0.139**
	(0.062)	(0.060)	(0.036)	(0.029)	(0.113)	(0.111)	(0.053)	(0.056)
Optimal Bandwidth ($\hat{h}$)	28	23	19	18	23	25	34	32
Panel E: Years of Education								
RD Estimate	0.305	0.402	0.100	0.062	2.151***	1.654***	1.401**	1.437***
	(0.366)	(0.397)	(0.268)	(0.228)	(0.621)	(0.627)	(0.549)	(0.534)
Optimal Bandwidth ($\hat{h}$)	24	21	28	28	21	23	35	35
Observations	3094	3094	4581	4581	3102	3102	4473	4473
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The estimation uses panel data from the Indonesian Family Life Survey (IFLS) for the waves 2000, 2007, and 2014. Each column in the table presents the regression discontinuity (RD) treatment effects for individuals born on or after September 1978, applying a linear control function based on month and year of birth on either side of the cutoff and uses optimal bandwidth selection based on Calonico et al. (2014) algorithm. The dependent variable in: Panel A is a dummy variable equal to one if the individual completed junior secondary school; Panel B is a dummy variable equal to one if the individual enrolled to senior secondary school; Panel C is a dummy variable equal to one if the individual completed senior secondary school; Panel D is a dummy variable equal to one if the individual enrolled to university; and Panel E is the years of education attained. Control variables are gender, father's and mother's education, number of siblings, birth order, and religions. Standard errors, in parentheses, are clustered at the month-year-of-birth. *** significant at 1%, ** significant at 5%, * significant 10%.

Figure A.3.1: predetermined individuals' characteristics

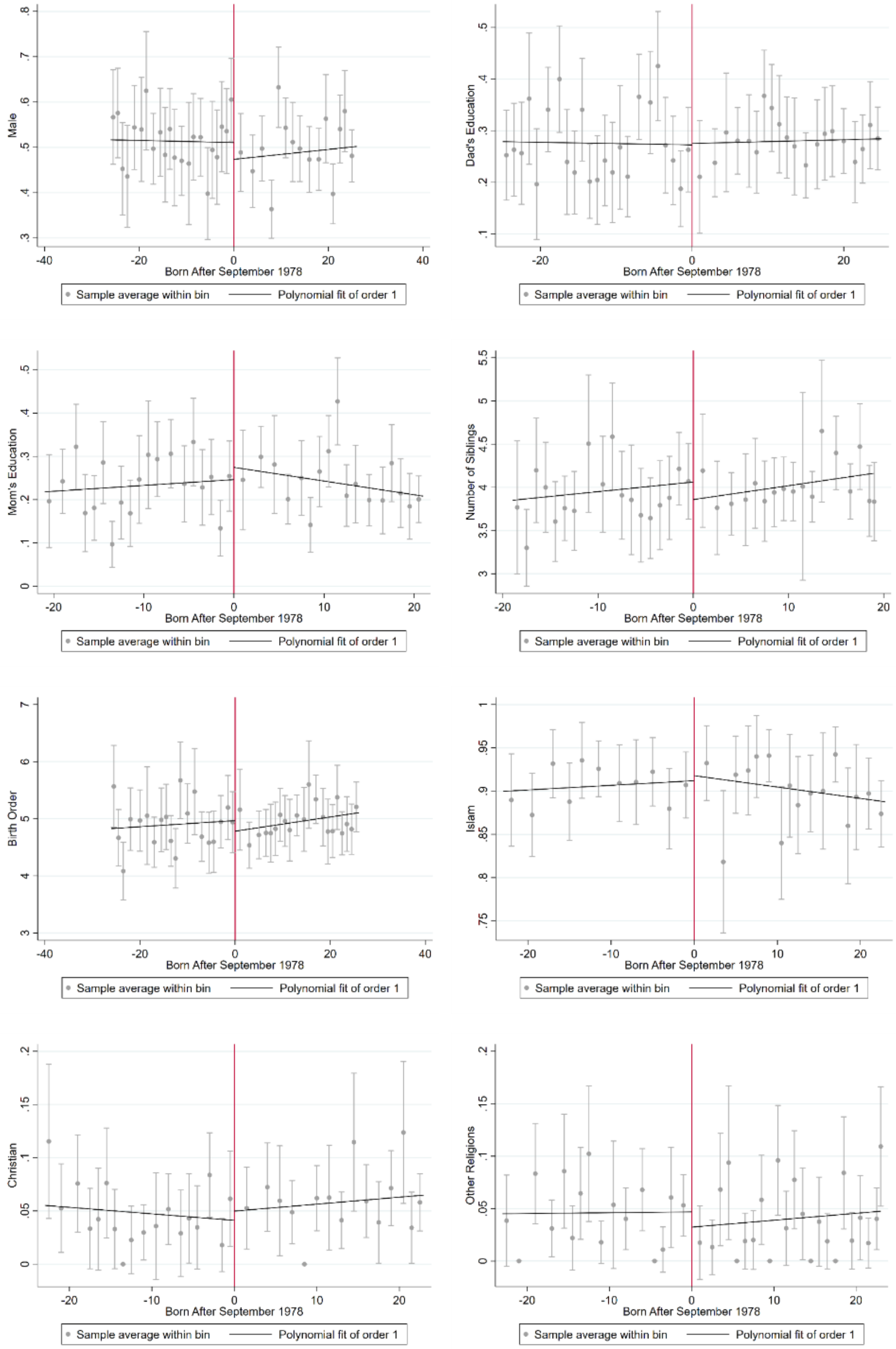


Figure A.3.2: The impact of compulsory schooling on educational outcomes

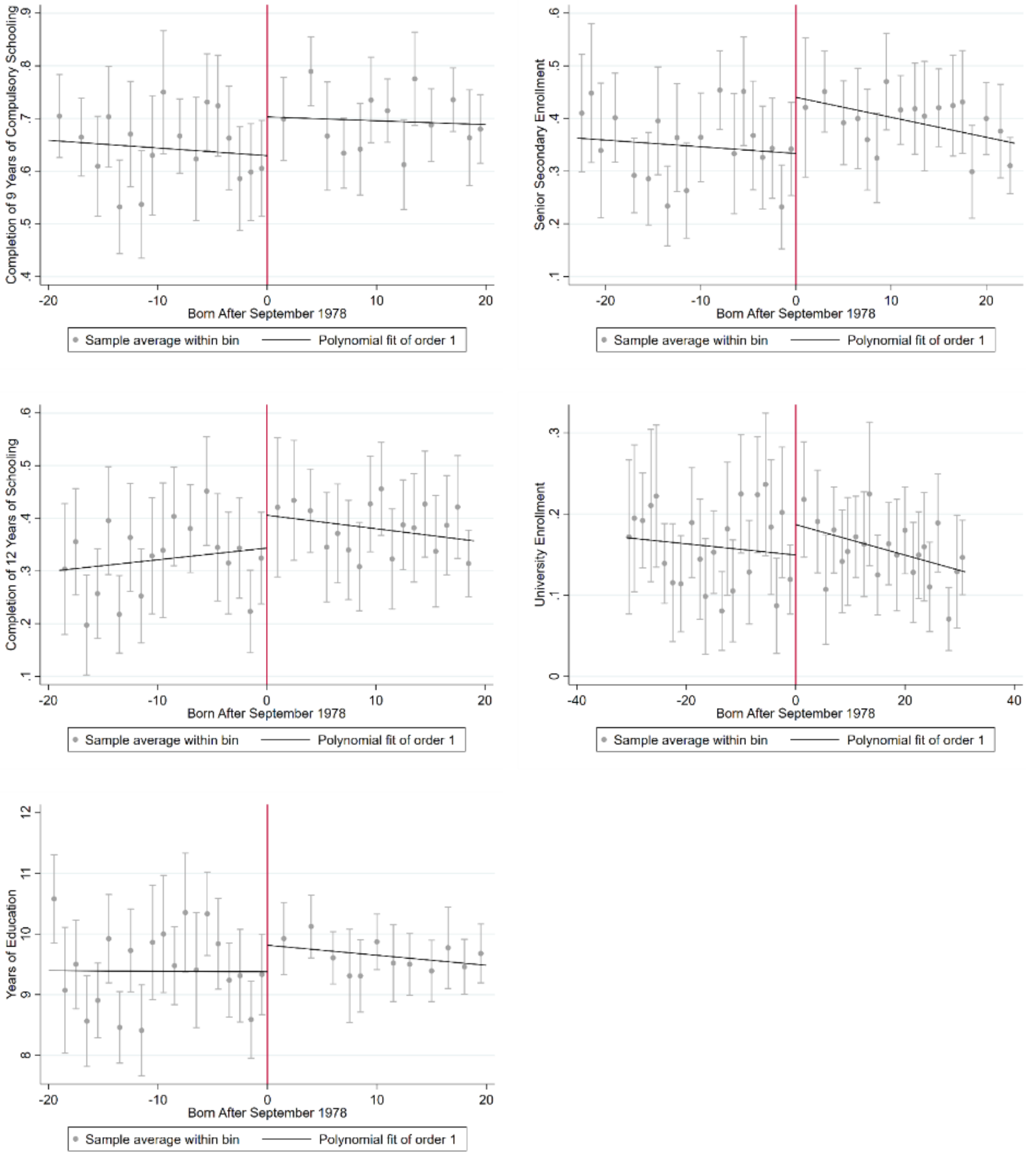


Figure A.3.3: Falsification analysis using fake compulsory schooling reform in 1987

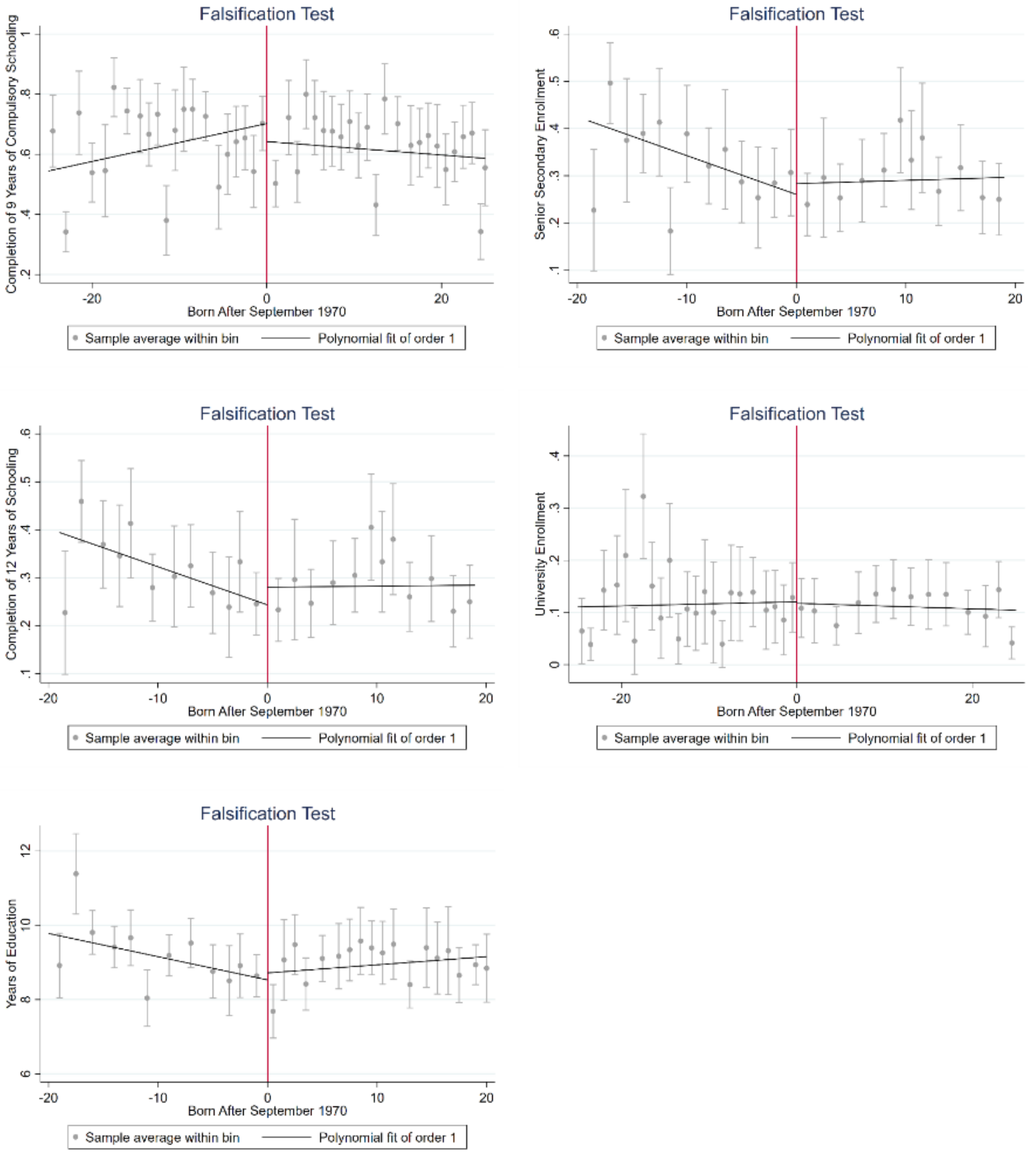
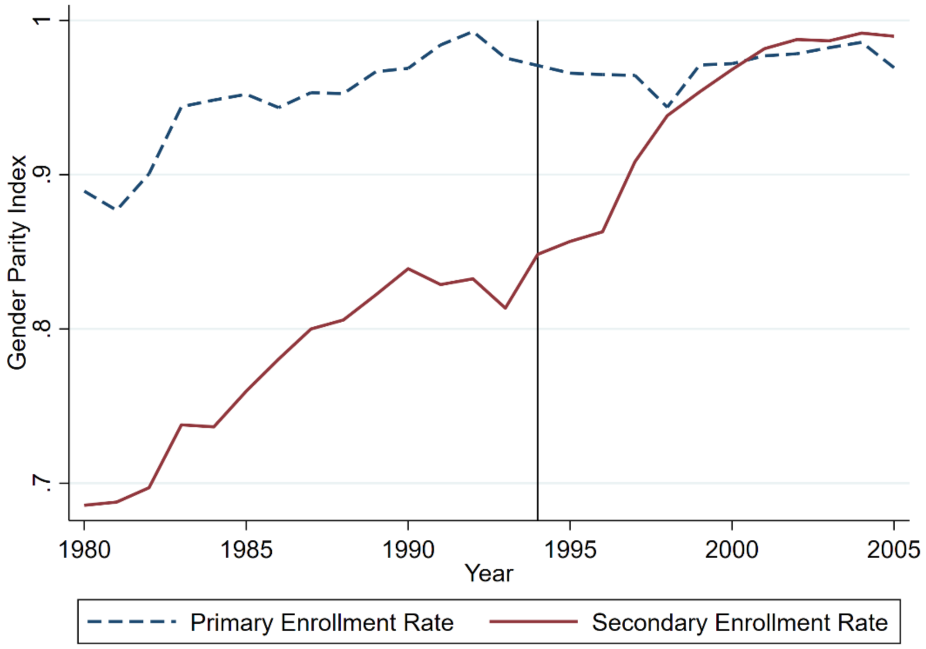


Figure A.3.4: Enrollment in primary and secondary education



4. Education and Ethnic Intermarriage: Evidence from Higher Education Expansion in Indonesia

4.1. Introduction

Education generates several positive effects both at the individual and aggregate levels. The increase in human capital endowments is especially important for developing countries since it shapes economic growth and development (Barro, 2001). Indeed, governments of several developing countries have undertaken diverse policies to enhance human capital formation during the last decades. These policies typically encompass large-scale interventions such as the extension of compulsory schooling and the expansion of educational infrastructures at the primary, secondary, and, more recently, tertiary education levels, following the patterns that developed countries have experienced. Indeed, fostering education through the expansion of Higher Education Institutions (HEI) represents an effective policy to enhance economic growth (Valero and Van Reenen, 2019). In this paper, we focus on a specific impact of the increase in educational attainments induced by the expansion of HEI: ethnic intermarriages in a multiethnic developing country (Indonesia).

Understanding whether and to what extent higher education attainments increase (in a causal sense) the likelihood of interethnic marriages (i.e. exogamy) is relevant in ethnically mixed societies for several reasons. On the one hand, the ethnic intermarriage rate is a clear indicator of ethnic attachment, which is strongly related to ethnic fractionalization and ethnically related socioeconomic segregation (Bazzi et al., 2019; Kukić, 2023). On the other hand, lower levels of ethnic fractionalization and segregation can mitigate civil conflicts, which in turn would favor economic development (Esteban et al., 2012; Corvalan and Vargas, 2015; Sanjaya et al., 2022). Indeed, these potential impacts could be relevant channels through which education is likely to a) reduce conflict (Rohner and Saia, 2019) and

b) increase interethnic tolerance and diversity in general (Roth and Sumarto, 2015). Therefore, analyzing the effect induced by HEI expansion would provide evidence regarding whether this policy represents an effective tool to achieve the aforementioned goals.

From the theoretical point of view, there are several possible justifications for the existence of a positive causal relationship between educational attainments and interethnic marriages. First, the (Indonesian) education system promotes a shared national identity and the adoption of a single language (Bahasa Indonesia, also known as standard Indonesian) and a unitary culture. Indeed, this is in line with the existing papers about the role of education on identity formation (e.g. Bandiera et al., 2019; Alesina et al., 2021). Second, education might change cultural and social norms, mitigating the degree of attachment to traditional (and possibly ethnically segregated) values, thus favoring interethnic tolerance (Roth and Sumarto, 2015). Third, education increases earnings potential and, therefore, fosters financial autonomy, thus limiting the dependency on the family, which could be especially important for women living in matrilocal enclaves.[40] Finally, more educated individuals have a higher propensity to migrate, possibly to larger agglomerations characterized by a higher degree of ethnic diversity, which could affect the likelihood of finding a partner from a different ethnic background.

There is a large body of literature regarding the determinants and the socioeconomic effects of ethnic/racial intermarriages in developed countries (mostly the U.S.), mainly focused on first- and second-generation migrants (for a review, see Furtado and Song, 2022). However, despite the relevance of the topic, there is a clear lack of evidence regarding the causal relationship between education and interethnic marriages in multiethnic developing countries. Some recent works focused on the determinants of intermarriages (not exclusively on education) in developing countries. For example, Ray et al. (2020) analyzed the association between inter-caste marriages and husband’s, wife’s and parents’ education in India. The paper by Allendorf

[40] Matrilocality is a social system in which the couple lives in the neighborhood of his wife or wife's family after marriage. This is different from patrilocality, where the wife moves to her husband's neighborhood or husband's family. Matrilocality is often associated with matriarchal societies, where women have a central role in social structure and family decisions.

and Thornton (2015) examines the determinants of inter-caste marriages in Nepal, including education as an explanatory variable. Crespin-Boucaud (2020) and Bandyopadhyay and Green (2021) studied the determinants of interethnic marriages in Sub-Saharan countries. Nevertheless, none of these works provide causal estimates.

There are also a few papers on the case of Indonesia. The most relevant work is the one by Bazzi et al. (2019), in which the authors exploit a large-scale population resettlement program in Indonesia during the '80 (the so-called Transmigration Program) to investigate the causal effect of intergroup contact on national integration. Although educational attainments are not the focus of the paper, the authors consider interethnic marriages as one of the proxies for national integration and show that the interethnic marriage rate is negatively affected by ethnic polarization. There are also other descriptive papers about ethnic intermarriages in Indonesia (Utomo and McDonald, 2016, 2020; Utomo, 2019), which also consider the association with education, but none of these papers addresses causality.

In this paper, we analyze the causal effect of educational attainment on the probability of being engaged in an interethnic marriage in Indonesia. As such, this is the first work that provides plausibly causal evidence on this topic, representing our work's main contribution to the existing literature. To achieve identification, we leverage the geographical expansion of Higher Education Institutions in Indonesia, especially on the Island of Java (where we focus on), since the second half of the XX century. Therefore, we also contribute to the evidence regarding the effects of investment on educational infrastructures (Duflo, 2001 and related papers), as well as to the growing literature about the local effect of college expansion (Cottini et al., 2019; Jagnani and Khanna, 2020; Carneiro et al., 2023, among others[41]), with an additional piece of evidence for an emerging country. Moreover, we also provide suggestive evidence regarding potential mechanisms that could be at play in the causal chain between HEI expansion, educational attainments and

[41] We are not the first in using college expansion as an instrumental variable to address the endogeneity of educational attainment. Starting from the paper by Currie and Moretti (2003), this approach has been used in several recent works (Kyui, 2016; Kamhöfer et al., 2019; Belskaya et al., 2020; Bratti et al., 2022; Westphal et al., 2022). In the empirical methodology section, we carefully describe similarities and differences between our identification strategy and the framework adopted in previous papers.

interethnic marriages, representing an additional value added to this paper. More generally, we contribute to the body of evidence highlighting the role of education as a tool to reduce ethnic-related segregation in multiethnic developing countries.

The empirical analysis integrates various data sources. First, we employ administrative data regarding the year of establishment and the exact location of all higher education institutions that provide undergraduate education on Java Island, the most populated island in Indonesia. A notable aspect of this data is its disaggregation at the campus level, considering the possibility of multiple locations for each institution. Second, we draw on data from the Indonesian Family Life Survey (IFLS). Our primary focus is on the latest available wave in 2014, supplemented by relevant information from preceding waves for specific analytical purposes. Based on information about individual ethnicity and households' identifiers, we can create an indicator for exogamy; that is, the ethnicity of one member of the couple differs from that of the other, representing the main outcome variable of our analysis. Moreover, ILFS data includes details about the district of residence and provides a comprehensive residential history dating back to the year of birth. Therefore, we can impute the geographical exposure to available HEI during different stages of adolescence based on the individual's district of residence. This serves as the basis for constructing our Instrumental Variable (IV) employment to address the endogeneity of educational attainments in the exogamy equation. More specifically, we instrument education with the number of HEI present in a radius of 10km from the district of residence of the individuals at age 18. We leverage on variation in geographical exposure to HEI across cohorts and locations, exploiting the expansion of HEI that took place over time in Java Island. The model explaining the probability of engaging in an interethnic marriage is separately estimated for males and females. We test for the robustness of the results to the definition of the instrument, particularly with respect to age and radii of exposure. Most importantly, we perform several sensitivity checks to discard the possibility that the instrument captures spurious correlations driven by either unobserved time-varying local factors – related to the demand for higher education – that could be correlated with the propensity for interethnic marriages or by issues of endogenous residential sorting. Furthermore, we test for possible heterogeneous effects of educational attainments on the probability of being married to someone from a different ethnic background.

Finally, we provide suggestive evidence regarding the role of possible mechanisms behind the causal chain between HEI exposure, educational attainments and interethnic marriages. Specifically, we examine the potential relevance of migration/residential locations and social norms related to ethnicity. This analysis of mechanisms indeed constitutes another significant contribution of our work to the existing literature.

The results indicate that higher educational attainments, induced by the expansion of HEI, positively impact the likelihood of being in an interethnic marriage. Following Currie and Moretti (2003) and Jagnani and Khanna (2020), among others, we consider different proxies for educational attainments: years of schooling, university enrolment and university completion. The positive effect on interethnic marriages is observed for the three outcomes and is somewhat higher for females than males. The results are very robust to all the sensitivity checks, pointing to the validity of the underlying assumption behind our IV approach. The analysis of heterogeneous effects highlights that the impact of education on the probability of having a partner from a different ethnic background is the same regardless of parental education and having parents with mixed ethnicities. However, increased educational attainments induced by HEI expansion exert a lower effect on exogamy for individuals with Javanese ethnicity than their counterparts from other ethnic backgrounds. This evidence indeed suggests that education could be a tool to mitigate segregation of ethnic minorities. Finally, the evidence regarding potential mechanism highlights the relevance of both dimensions. On the one hand, more educated individuals are more likely to migrate and reside in large cities, with a higher degree of ethnic fractionalization, thereby increasing the likelihood of exogamy. On the other hand, the increase in educational attainments induced by the expansion of HEI fosters trust towards individuals from different ethnic backgrounds (our proxy for social norms), which could lead to a higher propensity to form an ethnically mixed couple.

Overall, the results presented in this paper highlight the relevance of education and the expansion of higher education as tools for promoting the social integration of individuals from diverse ethnic backgrounds. Therefore, the beneficial effects on human capital formation induced by the establishment of new HEI could not only materialize into positive impacts in terms of earnings and other labor market outcomes but can also enhance other social outcomes and, more in general, can mitigate ethnic-related segregation in multiethnic countries and foster social cohesion.

The rest of the paper proceeds as follows: Section 4.2 summarizes the institutional background regarding ethnicities and marriages in Indonesia and its education system. Section 4.3 contains a description of the data used in the empirical analysis and presents some descriptive evidence. Section 4.4 illustrates the empirical strategy, and Section 4.5 reports the results. Finally, Section 4.6 concludes.

4.2. Institutional background

4.2.1. Ethnicity and interethnic marriages in Indonesia

With a population of over 240 million, Indonesia is one of the world's most populous countries. It is also an extremely rich and diverse country from a cultural point of view. Its major religion is Islam, although several other religions coexist. Moreover, Indonesian inhabitants belong to a wide and diverse range of ethnic groups, each with its cultural norms and traditions. In Indonesia, ethnicity is largely assigned based on language (Rademakers and van Hoorn, 2021), with minimal variations in terms of physical appearance in the majority of instances. Moreover, the ethnic diversity in Indonesia offers a fascinating chance to explore the interaction between ethnicity, culture, and family dynamics, specifically regarding choices for marriage and family formation.

Every ethnic group deeply values marriage as it represents the union of two individuals and their families. These ceremonies celebrate and maintain diverse cultural heritage and ethnic identities by following specific ethnic traditions (Buttenheim and Nobles, 2009). Meanwhile, the practice of interethnic marriage encounters notable challenges. For instance, Parker et al. (2014) explored how ethnic and religious groups in Indonesia interact, from socializing to marriage. They observed strong resistance to interreligious relationships, impacting even casual dating, largely due to strict religious teachings. While Indonesian society increasingly accepts interethnic relationships, endogamy remains the most common practice.

Java Island, the focus of our study and the most densely populated island in Indonesia, is largely inhabited by the Javanese, who comprise more than 55%

of its population. They predominantly reside in Central Java, D.I. Yogyakarta, and East Java Province. The Sundanese, constituting around 25% of the population, primarily reside in West Java. The Betawi and Madurese, with approximately 5% of the populace each, are primarily concentrated in Jakarta and Madura Island, situated immediately north of East Java, respectively. The remaining portion of the population, approximately 10%, comprises various minority ethnic groups (Statistics Indonesia, 2010). Considering this demographic context, Utomo and McDonald (2016) found notable disparities in marriage trends between Jakarta, the primary urban and economic hub, which displayed the lowest propensity for endogamous marriages at 67%, and regions heavily influenced by Javanese culture, where this rate surpasses 95%. According to Utomo (2019), Jakarta has lower rates of endogamy since it serves as a hub for migrants and represents a place where different cultures mix together. The city's heterogeneous population favors interethnic partnerships and marriages, particularly in its higher education institutions that attract students from across the country. Utomo (2019) also highlights that individuals do not engage in random marriage pairings but consider ethnicity a significant aspect of their decision-making process. In general, the primary challenge in interethnic unions often lies in adapting to the spouse's customs, traditions, culture, and strict customary. These strict traditional norms often lead to a preference for marrying within the same ethnicity (Ida Bagus, 2008; Parker et al., 2014).

4.2.2. The education system and higher education in Indonesia

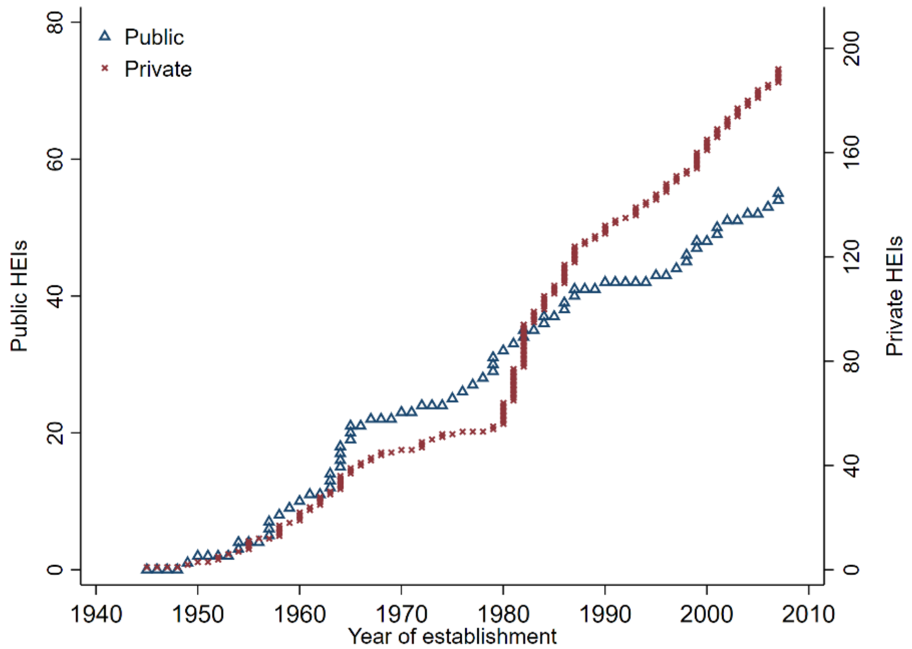
Indonesia's education system follows the 6-3-3-4 model, which includes six years of elementary school, three years of junior high, three years of senior high, and up to four years of higher education (Mukminin et al., 2019). The higher education system comprises vocational degrees, with a duration of one to four years, and undergraduate degrees, which typically consist of four-year programs. After completing their undergraduate studies, graduates can pursue either a two-year master's degree or a doctoral program, which typically lasts three to five years.

Indonesia's Higher Education Institutions (HEI) include universities, institutes, colleges, polytechnics, and academies, which can be either public or private. Public institutions are funded through public subsidies and tuition fees. Funding of private institutions primarily relies on tuition fees and other financing sources. Additionally, public HEIs are under the authority of government-appointed administration and adhere to stringent regulations. In contrast, private HEIs have greater independence in their governance and management, although they may encounter varying degrees of government influence that impact their funding, governance, and regulatory supervision (Welch, 2007; Ngo and Meek, 2019). In general, for both types of HEI, the enrolment cost paid by students varies according to the institution (and its quality) and the field of study. However, when enrolling in private institutions, students also have to pay an entry fee, which is not fixed and is specific to each institution and study program. On average, the overall cost paid by students is generally higher in private institutions, although there could be specific undergraduate degrees that are more expensive in prestigious public institutions than in less renowned private centers.

From a historical perspective, the expansion of HEIs in Indonesia began immediately after the country achieved independence in 1945. Just between 1945 and 1950, national student enrolment in higher education degrees increased from 1,600 to 5,200 between 1945 and 1950 (Buchori and Malik, 2004). The Higher Education Act of 1961 was one of the first substantial advances the newly independent nation made. DGHE (2003) outlined that this legislation established the foundation for future HE advancements and brought about significant improvements. Following this new law, HEI adopted an ordered framework with a precise division of faculties. The legislation defined the requirements for establishing universities, colleges, academies, and other HEI, along with the procedures for creating faculties.

HEI are established through different processes, depending on whether they are public or private. Public institutions are opened through a public procedure (and inaugurated directly by the President of the Republic of Indonesia), while private institutions are typically initiated by private corporations or foundations, which are obliged to inform the Ministry of Education of their intent (Welch, 2007). This notification requires the submission of a notarial deed confirming the legal entity governing the HEI, its articles of association, assets, expected sources of funding for its operation, curricular plans, and a complete description of each faculty member's credentials and teaching positions. The government supervises and guides private HEIs to ensure quality and compliance with standards through an agency called the Private Higher Education Coordinator (KOPERTIS). This agency, led by the Minister of Education, is present in all Indonesian provinces (Buchori and Malik, 2004). In terms of admission to undergraduate degrees, initially, the only requirement was a senior high school diploma. To unify standards, the government and the major public HEI in Java Island implemented a general admissions test in 1976 (SKALU). The admission system changed in 1989 (UMPTN), mostly because specialized exams based on the chosen major were introduced. On the contrary, private HEIs have maintained independent admission processes at the college level without a unified testing system.

Figure 4.1: Year of establishment (public and private HEI)



The number and variety of Indonesian HEI have grown significantly since the HE Act was enacted in 1961. According to Pannen (2018), there were 450 HEI formed in 1970, with a student population of 237 thousand. However, by 1990, the number had risen dramatically to 900 schools, serving nearly 1.5 million students. Figure 4.1 depicts the number of public and private HEI offering undergraduate degrees in Java Island by year of establishment. From 1945 to the mid-1960s, public and private HEI development was relatively moderate and steady. Around the mid-1960s, there was a pronounced increase in the establishment of public HEI, which continued to grow steadily during the following decades. Private HEI, however, experienced a constant rise during the ’70s, but a sharp increase during the ’80. The increase in the presence of private HEI was more moderated, although still very pronounced, during the following decades. At the end of the XX century, private HEI more than doubled public HEI in Java Island. Buchori and Malik (2004) argued that the rapid growth of private HEIs in the 1980s was driven by the increasing demand for HE that emerged in the 1970s. During this period, the state's budget was insufficient to satisfy this demand (Ngo and Meek, 2019). Notably, private foundations or organizations responded by creating schools such as universities, institutes, colleges, polytechnics, and academies, which provide a variety of programs and degrees.

Figure 4.2: The geographical location of HEI on Java Island over time

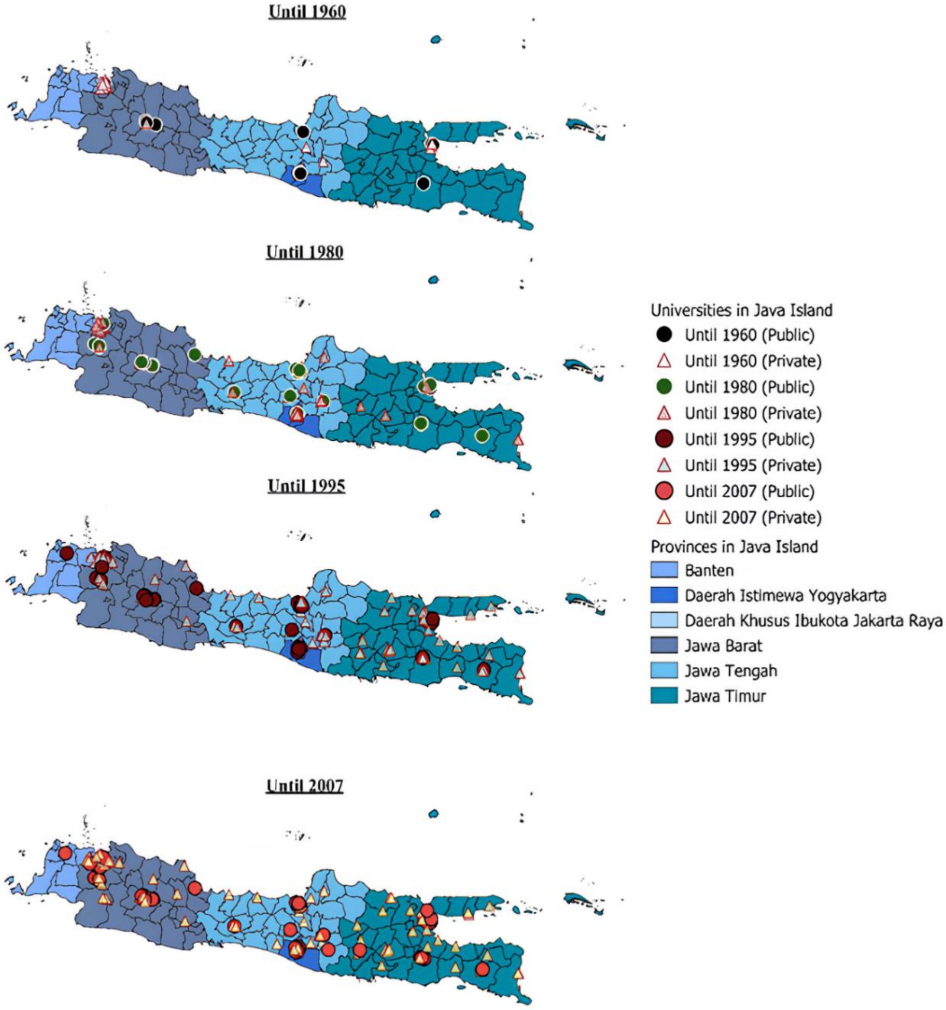


Figure 4.2 display the temporal evolution of the geographical location of public and private HEI, again focussing on institutions that offer undergraduate programs. In 1960, the few existing HEI were concentrated in major urban centers, notably Jakarta, Bandung, Semarang, Yogyakarta, and Surabaya. From 1980 to 1995, the higher education sector expanded considerably, with public institutions increasingly concentrated medium and large agglomerations. Nevertheless, throughout this period, many private institutions emerged, both in urban centers and in small towns. At the end of the relevant period (2007[42]), the presence of HEI was more widespread at the geographical level, providing generalized coverage of all Java's provinces, especially thanks to the extensive expansion of private institutions.

4.3. Data and descriptive statistics

Our empirical analysis focuses on Java Island, the most populated island of Indonesia and its capital and most populated city – Jakarta – is located. Moreover, most Higher Education Institutions are located on Java Island (PDSP Kemdikbud, 2013). We combine different data sources. First, we employ data regarding all HEI obtained from the National Accreditation Body for Higher Education (BAN-PT). This dataset includes information about the exact location of each campus for both public and private HEI, the year of establishment, as well as details on the type of higher education offered by each institution and their accreditation status. For the empirical analysis, we retain only institutions offering undergraduate education degrees that achieved a minimum accreditation score.[43] The site of HEI campuses has been geolocated using their detailed address (see Figure 4.2).

---

[42] We consider 2007 as the end of the relevant period because, as explained in what follows, we mainly consider exposure to HEI at age 18, and the youngest individual in our estimation sample turned 18 in that year.

[43] Based on the BAN-PT (National Accreditation Board for Higher Education) Regulation No. 2 of 2017, which details the mechanisms for accreditation, Higher Education Institutions in Indonesia are evaluated and classified into three categories of accreditation: A (excellent compliance with the standards), B (good compliance with the standards), and C (represents the minimum fulfilment of national standards).

Second, we use individual and family-level data from the Indonesian Family Life Survey (IFLS) database[44], which is representative of more than 80% of the Indonesian population within the survey area (Strauss et al., 2016). We mostly use data from the last wave of 2014, although we also exploit information from previous waves for specific purposes. The survey provides information about several individual and parental characteristics, including detailed information about educational attainments. Most importantly, the last two waves (2014 and 2007) of the IFLS database contain information about the respondents' ethnicity, as well as the ethnicity of his/her parents. The questionnaire includes 29 different ethnicities, representing the large majority of ethnic groups in terms of the country's population. Thanks to household identifiers, we are able to construct our outcome variable, exogamy, which is an indicator of having a partner from a different ethnic background. We consider several measures of educational attainments. Specifically, we use explanatory variables of interest, such as years of schooling, college attendance, or college completion. These variables have been constructed by combining information about the highest grade attended and the highest completed grade.[45]

Moreover, the IFLS database also includes information about the place of birth and the current residence, defined according to two main administrative geographical units – provinces and districts – and the entire migration history. Given the lack of information about the precise place of residence of households within the districts, we combine the two data sources based on the centroids of the districts. Specifically, as better explained in the next section, we construct different measures of geographical exposure to HEI during adolescence. These are defined according to the number of HEI located within a certain radius of distance from the districts' centroid,

---

[44] IFLS data are freely available from this link: https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html.

[45] That is, if an individual's highest level of education is junior high school and his/her highest grade ever completed is 2, then we impute 8 years of schooling. Furthermore, the indicator for college attendance is equal to one if an individual attended at least one year of college, while the indicator for college completion takes the value 1 if the individual attended and completed college.

covering the period from the year of birth until the year in which the individual turned 18 years old.[46]

The estimation sample has been obtained by retaining married individuals aged between 25 and 65. In this way, we avoid including individuals who could still be studying, and limit selection issues related to the age at marriage.[47] Moreover, we also exclude older individuals due to potential issues of selective mortality. We include only individuals who were born in Java and lived on the island for the entire relevant period. Finally, we exclude observations with missing values in the variables of interest. After applying these conditions, we obtain a sample of 6352 males and 6181 females.[48]

Table 4.1 reports descriptive information about ethnicity and exogamy for the estimation sample by gender. The largest ethnic group is Javanese (64-65%), followed by Sundanese (20%).

Table 4.1: Endogamy and exogamy by ethnicity

		Males			Females	
Variable	% sample	Endogamy	Exogamy	% sample	Endogamy	Exogamy
Javanese	0.640	0.931	0.069	0.648	0.928	0.072
Sundanese	0.209	0.833	0.167	0.213	0.817	0.183
Madurese	0.049	0.877	0.123	0.049	0.904	0.096
Betawi	0.067	0.611	0.389	0.062	0.652	0.348
Other Ethnicities	0.035	0.413	0.587	0.028	0.489	0.511
Total	1	0.868	0.132	1	0.874	0.126
Observations	6352	5548	843	6181	5403	778

Madurese and Betawi ethnicities are significantly less common (we grouped other minority ethnic groups due to the low number of observations, although

[46] Actually, in order to perform a robustness check for our Empirical framework, we also consider exposure to HEI at age 25.

[47] According to the World Bank (2023), the average age at marriage in Indonesia is 27.1 and 22.4 for male and female respectively. Notice that, using information about the year of marriage, we also perform a robustness check in which we only retain individuals who got married after completing education.

[48] The estimation sample contains a slightly higher number of males than females, since there are cases in which the wife is younger than 25 and, therefore, does not satisfy the 25-65 age range criteria.

all ethnic groups are used for the construction of the exogamy indicator). Overall, around 13% of individuals in the sample are engaged in an interethnic marriage, with this proportion being significantly lower for individuals from the Javanese ethnicity, the largest ethnic group in Java. Table 4.2 also displays the proportion of interethnic marriages according to education level. The probability of having a partner from a different ethnic background increases with educational attainments. More concretely, among individuals with less than compulsory education (junior high school), the exogamy rate is 7.8% for males and 9.2% for females. However, this proportion increases to around 20% for university-educated individuals. Of course, this change in the likelihood of exogamy associated with educational attainments cannot be interpreted in causal terms, since there could be differences in observed and unobserved characteristics that affect both education and the propensity for interethnic marriages.

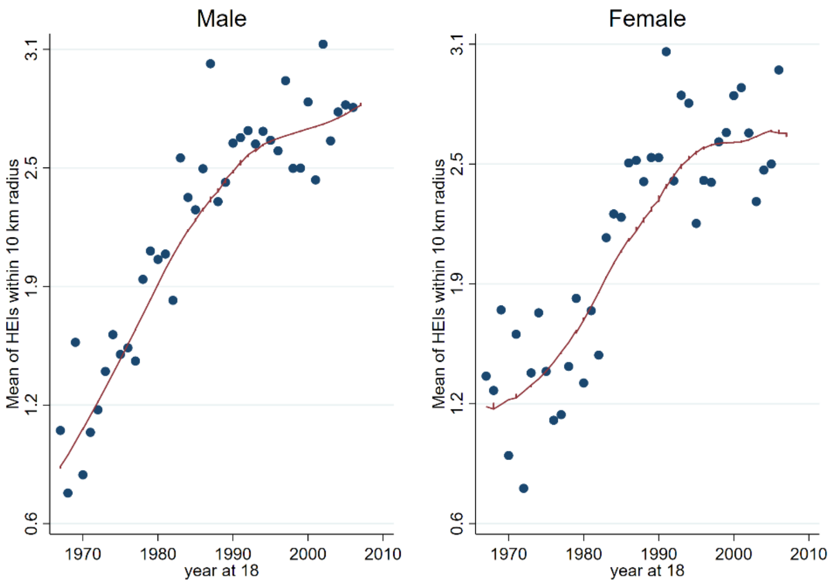
Table 4.2: Endogamy and exogamy by level of education

	Males			Females		
Variable	% sample	Endogamy	Exogamy	% sample	Endogamy	Exogamy
Less than Compulsory Education	0.399	0.922	0.078	0.459	0.908	0.092
Post Compulsory Education	0.601	0.832	0.168	0.541	0.845	0.155
No University Attendance	0.863	0.880	0.120	0.870	0.884	0.116
University Attendance	0.137	0.791	0.209	0.130	0.805	0.195
No University Completion	0.883	0.878	0.122	0.880	0.884	0.116
University Completion	0.116	0.792	0.208	0.120	0.802	0.198
Total		0.868	0.132		0.874	0.126
Observations	6352	5548	843	6181	5403	778

Table 4.3 displays basic summary statistics for all the variables used in the empirical analysis for males and females. Besides exogamy and the three measures of educational attainments, we also report descriptive information about the number of available HEI from a certain radius of the district of residence at age 18 (exposure at other ages in not reported for space reasons). As expected, exposure increases with the radius. Moreover, exposure is higher for private than for public HEI, which is in line with the figures reported in section 2. To provide suggestive information about the changes across the cohort in exposure to HEI, driven by the expansion process, in Figure 4.3, we display a scatter plot and a lowess fit of the average number of HEI surrounding the district of residence at age 18 by year. We observe a

pronounced positive trend for both genders, indicating that exposure to HEI increases across the cohorts.

Figure 4.3: Average HEIs within a 10 km radius by year at age 18



As control variables, we use own ethnicity and religion and family background. Specifically, we consider the number of siblings, a dummy for having low-educated parents, and an indicator of mixed parental ethnicity (i.e., the father's ethnicity is different from the mother's). Moreover, we also employ additional variables that are used for the analysis of potential mechanisms. Using information about residential history, we construct an indicator for having changed the residence district between the year the individual turned 18 and 2014. Moreover, combining this information with the district of residence in 2014, we constructed a dummy that is equal to 1 if the individual resided in a large city: Jakarta, Bandung, Semarang, Surabaya, Surakarta, and Yogyakarta, the largest urban areas in the Java Island.

Table 4.3: Descriptive statistics

Variable	Males		Females	
	Mean	*S.D.*	Mean	*S.D.*
Exogamy	0.132	*0.338*	0.126	*0.332*
Years of Schooling	8.687	*4.178*	8.176	*4.291*
University Attendance	0.137	*0.344*	0.130	*0.337*
University Completion	0.116	*0.320*	0.120	*0.324*
HEI within 5 Km radius at age 18	1.264	*2.495*	1.290	*2.597*
HEI within 10 Km radius at age 18	3.146	*5.851*	3.158	*6.009*
HEI within 15 Km radius at age 18	5.292	*9.195*	5.331	*9.372*
HEI within 20 Km radius at age 18	6.762	*11.108*	6.855	*11.379*
HEI within 25 Km radius at age 18	8.197	*12.568*	8.361	*12.929*
Public HEI within 10 Km radius at age 18	0.685	*1.504*	0.690	*1.536*
Private HEI within 10 Km radius at age 18	2.479	*4.806*	2.493	*4.941*
Javanese	0.640	*0.480*	0.648	*0.477*
Sundanese	0.209	*0.406*	0.213	*0.409*
Madurese	0.049	*0.217*	0.049	*0.216*
Betawi	0.067	*0.250*	0.062	*0.241*
Other Ethnicities	0.035	*0.184*	0.028	*0.165*
Moslems	0.971	*0.168*	0.969	*0.174*
Christians	0.027	*0.161*	0.029	*0.169*
Hindus	0.001	*0.028*	0.001	*0.031*
Other Religions	0.002	*0.040*	0.001	*0.031*
Number of Siblings	3.069	*2.221*	3.363	*2.500*
Low Parental Education	0.123	*0.328*	0.136	*0.342*
Ethnically-Mixed Parents	0.070	*0.256*	0.064	*0.244*
Change District of Residence (18 - 2014)	0.138	*0.345*	0.124	*0.330*
Move to Large Cities (18 - 2018)	0.035	*0.183*	0.039	*0.193*
Fractionalization	0.427	*0.495*	0.423	*0.494*
Being a Minority in 2014	0.216	*0.411*	0.207	*0.405*
Trust Own Ethnicity	0.636	*0.481*	0.687	*0.464*
Observations	6391		6181	

Note: Low parental education = 1 if parents did not complete primary education. Fractionalization has been defined according to ethnicity, based on district-level information from the 10% of the 2010 Census. Being a minority in 2014 = 1 if the individual's ethnicity is different than the most prevalent ethnicity in the district of residence in 2014. Trust own ethnicity = 1 if the individual declares he/she completely agrees or agrees with the sentence "I trust individuals from my own ethnic group more than others". This last variable is available only for 4515 males and 4872 females (i.e. is missing for 25% of the estimation sample). The corresponding descriptive statistics have been obtained only with valid observations.

The indicator for being a minority group in the place of residence is directly obtained from the IFLS data, combining the information about the largest ethnic group in the community of residence[49] in 2014 and own ethnicity. We also imputed ethnic fractionalization in the district of residence in 2010. In order to do this, we computed the fractionalization index at the district level using information on individual ethnicity from the 2010 Census (10% sample), following Bazzi et al. (2019). Finally, we constructed a proxy for social norms based on the question regarding trust in individuals from the same ethnic group relative to individuals from different ethnic backgrounds. Specifically, the question asks whether the individual: 1) strongly agrees, 2) agrees, 3) disagrees, or 4) strongly disagrees with the statement that they trust more individuals from the same ethnic group than others. Therefore, we use an indicator that takes the value of one if the individual agrees or strongly agrees with the above statement. Unfortunately, this variable is missing for 25% of the estimation sample.

4.4. Empirical strategy

Our objective is to estimate the (causal) impact of education on the likelihood of being in a relationship with a partner from a different ethnic background (exogamy). The equation of interest takes the following form:

$$EXO_i = \alpha + \delta EDUC_i + \beta' X_i + \theta_{tp(i)} + \varepsilon_i \tag{4.1}$$

Here, EXO_i represents the indicator for having a partner with a different ethnicity, while $EDUC_i$ encompasses different proxies for educational attainment, namely i) years of schooling, ii) college attendance, and iii) college completion, which represent our main explanatory variables of interest. The model also includes a set of control variables (X_i), which comprise dummy variables for one's own ethnicity and religion, the number of siblings, an indicator for having low-educated parents, and another dummy for having ethnically mixed parents. We also control for year of birth (t) ×

[49] This information proceeds from “Community-Facility Survey” of IFLS and is reported by the official village/township leader.

province of residence[50] (p) fixed effects, which capture province-cohort specific trends in local time-varying factors that might affect the outcome. Throughout the whole empirical analysis, we estimate the model separately for males and females.

We start with the OLS estimation of equation (4.1). However, the causal interpretation of the OLS estimate of the δ parameter is challenging, mostly because of the likely relevance of unobserved factors that correlate both with educational attainments and with the propensity to form an ethnically mixed couple. To deal with this omitted variable issue and obtain a plausibly causal estimate of the effect of education on exogamy, we employ and Instrumental Variable (IV) approach that leverages the presence of Higher Education Institutions (HEI) in the place of residence during adolescence, exploiting the massive geographical expansion of HEI that took place in the Java Island over time. More specifically, our instrument ($HEI^{r}_{d(i)\tau(i)}$) consists in the number of existing HEI (at the relevant age, τ) in a certain radius (r) from the centroid of the district of residence (d).[51] In our preferred specification, we define the instrument based on the district of residence at age 18, which is the typical university entrance age in Indonesia. Similarly, we consider a radius of 10km to compute the number of available HEI surrounding the district of residence. For both dimensions of the instrument, we select the option that maximizes the instrument's strength. However, we also conduct robustness tests using alternative reference ages and different radii. Equation (4.2) represents the corresponding first-stage equation:

$$EDUC_i = \mu + \rho HEI^{r}_{d(i)\tau(i)} + \gamma' X_i + \omega_{tp(i)} + u_i \qquad (4.2)$$

Therefore, we use within birth cohort and province variation in the geographical exposure to HIE as an exogenous source of variation in educational attainments. This approach is valid under the assumption that the

[50] We primarily focus on the province of residence at age 18 due to reasons related to our identification strategy.

[51] We adopt clustered standard errors at the district level, which represents the primary level of variation for the instrument. Additionally, we experimented with two-way clusters at the district-year of birth level, yielding similar results (available upon request).

presence of HEI at the local level is a strong predictor of educational attainments while not being directly related to ethnic exogamy. The IV counterpart of equation (4.1) is thus represented by the following equation:

$$EXO_i = \alpha + \delta_{IV}\widehat{EDUC}_i + \beta' X_i + \theta_{tp(i)} + \varepsilon_i \tag{4.3}$$

Under the validity of the underlying assumptions, the coefficient associated with educational attainments (δ_{IV}) can be interpreted as the causal effect of education on interethnic marriages among individuals induced into higher educational attainments due to the geographical expansion of higher education (in a LATE framework).

This IV approach resembles the one employed in the seminar paper by (Currie and Moretti, 2003), and its variants that have been adopted by other authors in more recent papers (Kyui, 2016; Kamhöfer et al., 2019; Belskaya et al., 2020; Bratti et al., 2022; Westphal et al., 2022, among others). Nevertheless, there are certain notable differences in our setting that warrant further discussion. On the one hand, an advantage of our dataset is that it provides retrospective information about the district of residence since birth, year by year. Hence, we are able construct our instrument based on the district of residence at age 18, a pivotal year when individuals typically enroll in university in Indonesia (although we also explore previous ages for robustness, as elaborated below). Indeed, data about the place of residence during adolescence is not always available and several works rely on information about residence at birth. On the other hand, unfortunately, to the best of our knowledge information about the size of the cohort of individuals in the age range to attend college is not available for the case of Indonesia, neither at the district nor at the province level. This constitutes a data limitation for our identification strategy. In fact, as noticed by Currie and Moretti (2003), the geographical variation in the number of HEI across cohorts could be capturing both the demand and supply for university education. While the supply-side can be reasonably taken as exogenous, demand-side factors can (directly) correlate with other local-level variables that could associated with to the decision to form an ethnically mixed couple. Despite controlling for cohort × province of residence specific fixed effects should account for local-level confounders varying across birth cohorts, questions may still arise regarding the exogeneity of the instrument. That is, there could be unobserved local factors correlated with both the demand for higher education and the propensity for exogamy, influencing individuals born in a given cohort in different ways within their province of residence. An additional, but related, potential concern that might undermine the validity of the instrument is the endogenous residential sorting of families and/or individuals. This is because decisions regarding residential locations could be influenced by unobserved factors that are linked to both the inclination for interethnic marriages and demand-side elements related to the presence of universities. Nevertheless, we conduct several robustness checks that are aimed at providing evidence in favor of the validity of the instrument and the causal interpretation of the corresponding estimate of the parameter of interest (δ_{IV}).

4.4.1. Alternative specifications and robustness checks

To validate our IV approach and the general empirical framework, we perform a battery of sensitivity tests. First, we test for the robustness of the results with respect to the two main dimensions along which we construct the instrument: the radius (r) and age at exposure (τ). Regarding the former element, we compute the number of universities surrounding the individual's district of residence using buffers of a certain radius from the district's centroid. We adopt this strategy to define the availability of HEI because the IFLS data contain information on two main geographical identifiers: the province, which is possibly too broad to define the relevant area of influence, and the district, which is likely to be too narrow.[52] Of course, the choice of the radius is, by definition, subject to some degree of arbitrariness. We therefore computed the instrument based on different radii of exposure: 5km, 10km, 15km, 20km and 25km. Moreover, we also adopt a similar approach than in Kamhöfer et al. (2019) and Westphal et al. (2022), which consists in considering data on the location of all university campuses in the Java Island and compute the number of available colleges weighted by their distance from the centroid of the district of residence using Gaussian Kernel weights (using the Silverman's rule for bandwidth selection). To determine the best

[52] Authors of existing papers focused on the number of universities within administrative geographical units that are in between provinces and districts such as US counties (Currie and Moretti, 2003) and municipalities (Kyui, 2016; Bratti et al., 2022).

specification, we select the option that maximizes the strength of the instrument, i.e. maximizes the first stage F-statistic. Second, we also check for the sensitivity to the choice of the relevant age at exposure. Although the natural choice consists in selecting the typical age at which people enroll into college (18 in the case of Indonesia), as done in other papers, to some extent this is also an arbitrary choice. Moreover, using age 18 could also be related to the issue of endogenous residential sorting, because individuals and families might decide to relocate to areas in which not only college accessibility is higher, but there is also a more favorable environment for the formation of ethnically mixed couples. Therefore, we defined the instrument based on the district of residence at ages 18, 15, 12, 6, and at birth. Subsequently, we selected the option that yields a higher F-statistic in the first stage.[53] Additionally, we also repeat the estimation while retaining only individuals who did not change their district of residence either between the year of birth and the year in which they turned 18.

After determining the preferred specification of the instrumental variable ($HEI^{r}_{d(i)\tau(i)}$), we implement other checks that are aimed at validating its exogeneity, especially regarding the concern that the number of available HEI could be capturing time-varying demand-side local factors that directly affect the outcome. For these checks we also focus on the reduced-form equation, which corresponds to:

$$EXO_i = \alpha + \lambda_{RF} HEI^{r}_{d(i)\tau(i)} + \beta' X_i + \theta_{tp(i)} + \varepsilon_i \qquad (4.4)$$

First, we compare the estimate of the reduced-form coefficient (λ_{RF}) from equation (4.4) to the coefficient obtained from an alternative specification in which we also include the presence of HEI surrounding the district of residence at birth ($\tau = 0$) as additional control. This additional variable should capture long-standing unobsevables at the local level that could correlate with both the demand for higher education and interethnic marriages. If these factors are actually relevant, the reduced-form coefficient of our instrument should be significantly lower, which would suggest that the exogeneity

[53] In conducting this exercise, we also change the year of birth × province of residence accordingly, considering the province of residence at the corresponding age.

assumption is not satisfied. In a similar vein, we re-estimate the model while conditioning to the presence of at least one HEI at birth around the district of residence. This implies considering only individuals who were born in districts that should be generally similar in terms of local level characteristics. Finding similar results than our baseline estimation would provide supporting evidence for the validity of the underlying assumption of our IV approach. Second, we aim to account for potential recent changes in local demand-related factors by including an additional control for the presence of "new" HEI established between the individual's birth year and the year they turned 18, located in proximity to the district. If what really matters in both the reduced-form and first-stage equations is the number of newly established HEI and not the overall stock, this is probably indicative of the higher relevance of (potentially endogenous) demand-side factors rather than supply-side elements. Third, borrowing from Currie and Moretti (2003), we include as additional control variable the number of available universities at age 25. In the hypothetical case in which our instrument is capturing spurious correlation with local level unobservables, we would observe a higher estimated coefficient for the number of HEI at 25 than at 18, and a significant reduction in the coefficient of the instrument relative to the baseline estimation. Fourth, also following Currie and Moretti (2003), we compute the exposure to public and private HEI separately and re-estimate the model with each of these two instruments. As the establishment of private universities is more likely to be related to (potentially endogenous) geographical characteristics such as the price of soil, but also to the expected demand. Therefore, finding larger effects of the presence of private HEI than public HEI would be indicative of the lack of exogeneity of the instrument.[54] Finally, we conduct a falsification exercise based on a permutation test, in which we randomly assign the district of residence. This process is repeated 10,000 times, and we estimate the reduced-form equation for each replication, generating a distribution of fake reduced-form coefficients. If these placebo estimates are not symmetrically distributed around 0, it would

[54] Currie and Moretti (2003) also refer to potential issue related to the prices of tuition fees between public and private institutions. This concern is less relevant for the case of Indonesia. As explained in the institutional background section, differences in prices between public and private Higher Education Institutions (HEI) are not very pronounced, though they are indeed field- and university-specific. The primary distinction in cost lies in the entry fee for private colleges.

be evidence that the real instrument could be capturing some kind of spurious correlation. Moreover, we also estimate an overidentified model in which we use dummies for the presence of HEI at the local level and present the results of the Hansen J-test for overidentification.

Besides this battery of sensitivity checks regarding the definition and the validity of our instrumental variable, we also perform two additional checks to provide further evidence about the internal validity of our estimations. On the one hand, the causal chain that we hypothesized is that the expansion of HEI shaped educational attainments, and this in turns increased the propensity to find a couple from a different ethnicity. However, although rare, there could be cases where marriage occurs before completing education. To address this, we re-estimate the model after excluding individuals who married before leaving the education system. On the other hand, we observe the ethnicity of both members of the couple in 2014, which is after marriage. Many existing papers on ethnicity assume this to be a predetermined and immutable feature. However, Rademakers and van Hoorn (2021) provide evidence of the likelihood of changing ethnicity in Indonesia, noting that this pattern is more prevalent among members of interethnic marriages. In IFLS ethnicity is reported from the last two waves (2014 and 2007). Therefore, we also repeat our estimations considering only individuals who i) are interviewed in both waves and ii) report the same ethnicity in 2014 as in 2007.

4.4.2. Analysis of heterogenous effects and potential mechanisms

The additional evidence that we report in this paper concerns the analysis of heterogeneous effects of education on exogamy, as well as potential mechanisms that lie behind the link between HEI, educational attainments, and interethnic marriages.

As for heterogeneous effects, we consider whether the impact of education differs along three main features: own ethnicity, parental education and having parents from a mixed ethnic background. In doing that, we use interactions rather than splitting the sample, with the aim of avoiding small sample issues. Therefore, for each of these three variables in a separate fashion, we estimate the model that includes interactions with the instrument as additional exclusion restriction, as well as interaction with educational attainments as additional endogenous regressor.

In terms of potential mechanisms, while there are several factors that could be relevant in this setting, we are limited by data availability. Consequently, we focus on two main elements: migration/residential location and social norms. Regarding the former, the hypothesis is that the expansion of higher education leads individuals to attain higher educational levels, influencing their propensity to migrate and, possibly, settle in larger and more ethnically fractionalized cities. This, in turn, could increase their likelihood of marrying someone from a different ethnic background. Therefore, we consider alternative outcomes related to these factors: i) an indicator for having changed place of residence between age 18 and 2014, ii) an indicator for currently residing in large cities, iii) being a minority in the place of residence in 2014 and iv) and ethnic fractionalization in the district of residence. As for social norms, the idea is to employ a proxy for tolerance and openness toward different ethnic groups. This, in turn, could be fostered by increased educational attainment and consequently affecting the propensity to match with a partner from a different ethnicity. Based on available data, we rely on the variable capturing whether the individual trusts more others from the same ethnicity or not, which has been described in the data section. Therefore, we use the indicator for trusting more in individuals from the same ethnicity than others as proxy for social norms.

Because justifying the adoption of our IV approach while using these alternative variables as outcomes, we focus on the reduced-form equation in which they are directly regressed against the presence of HEI at the local level. However, it is important to note that all the variables that we consider in the analysis of are observed possibly several years after marriage (i.e. in 2010 for fractionalization and in 2014, the survey year, for other variables). Therefore, the results should be taken with caution because these variables could actually reflect "consequences" of interethnic marriages rather than pure mechanisms (i.e. an individual who is married with someone from a different ethnicity could develop more trust toward others from a different ethnic background). While acknowledging this limitation, we remain convinced that analyzing the impact of exposure to HEI on these proxies for migration/residential choices and social norms provides suggestive evidence about the relevance of these factors in the underlying causal effect between educational attainment and the formation of interethnic marriages.

4.5. Results

Table 4.4 displays the main results from the OLS estimation of equation (4.1), for each of the three measures of educational attainments (complete results are reported in Table A.4.1 of the Appendix). The estimates are separately obtained for males and females. We estimate the model without control variables (i.e. only including fixed effects for year of birth × province of residence at 18), as well as controlling for own ethnicity, own religion, number of siblings, parental education, and mixed parental ethnicity. In general, education is positively and significantly associated with the probability of being engaged in an ethnically mixed marriage. Each additional year of increase in years of schooling is associated with an increase in the probability of exogamy of 0.7 and 0.6 percentage points (p.p.) for males and females, respectively. Having attended or completed university is associated with a higher propensity of interethnic marriages as well (around 6-7p.p.).

The inclusion of control variables leads to a certain reduction in the coefficients of all measures of educational attainments, more pronounced for females, although their significance remains unchanged. The results regarding control variables are of independent interest and warrant further discussion. As for own ethnicity, people from the Sundanese ethnicity are not more likely to engage in mixed marriages than those from the Javanese ethnicity (the largest ethnic group in Java). However, those from other ethnicities are generally more likely to be married to a partner from other ethnic groups, except for Betawi females. Religion does not seem to play an important role while other variables are controlled for. Specifically, only Hindu males exhibit a lower likelihood of engaging in interethnic marriages compared to their Javanese counterparts. Having low-educated parents is associated with a slightly lower probability of exogamy. Moreover, as expected, parental exogamy is an important predictor of own exogamy, indicating a certain intergenerational pattern in interethnic marriages.

Table 4.4: OLS estimations – Dependent variable: Exogamy

	Males		Females	
	(1)	(2)	(3)	(4)
Panel A:				
Years of Schooling	0.007***	0.006***	0.006***	0.004***
	(0.001)	(0.001)	(0.002)	(0.002)
R-squared	0.200	0.260	0.204	0.238
Panel B:				
University Attendance	0.075***	0.058***	0.064***	0.047***
	(0.014)	(0.012)	(0.018)	(0.017)
R-squared	0.199	0.259	0.204	0.239
Panel C:				
University Completion	0.072***	0.057***	0.067***	0.049***
	(0.015)	(0.013)	(0.019)	(0.018)
R-squared	0.198	0.258	0.204	0.239
Controls	No	Yes	No	Yes
Observations	6391	6391	6181	6181

Notes: OLS estimations with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Main regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Additional control variables: ethnicity, religion, number of siblings, having parents with low education and having parents with different ethnicities.

However, due to the potential endogeneity of educational attainment in the exogamy equation, the previous results cannot be interpreted in causal terms. Therefore, to obtain plausibly causable estimates, we employ our measure of geographical exposure to HEI as instrument for educational attainment. We start with exposure to HEI defined according to the district of residence at 18, considering the number of available institutions within a 10km radius from the district's centroid. Table 4.5 reports the results (with and without controls) for the three educational outcomes. The first-stage coefficients are generally positive and highly significant, highlighting the strength of the exposure to HEI as predictor of years of schooling and university attendance/completion. The IV/TSLS estimates of equation (4.3) confirm that education exerts a positive effect on the probability of exogamy.

Table 4.5: IV/2SLS estimations – Dependent variable: Exogamy

	Males		Females	
	(1)	(2)	(3)	(4)
Panel A: Years of Schooling				
First Stage	0.235***	0.195***	0.241***	0.184***
	(0.037)	(0.031)	(0.041)	(0.033)
Second Stage	0.035***	0.032***	0.046***	0.048***
	(0.011)	(0.011)	(0.015)	(0.018)
First-Stage F-statistic	39.635	39.451	34.813	31.203
P-Value	0.000	0.000	0.000	0.002
Panel B: University Attendance				
First Stage	0.019***	0.017***	0.015***	0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Second Stage	0.441***	0.372***	0.727***	0.696***
	(0.147)	(0.139)	(0.223)	(0.249)
First-Stage F-statistic	36.631	30.146	23.906	19.110
P-Value	0.000	0.000	0.000	0.009
Panel C: University Completion				
First Stage	0.016***	0.014***	0.015***	0.013***
	(0.002)	(0.002)	(0.003)	(0.003)
Second Stage	0.516***	0.435***	0.709***	0.668***
	(0.164)	(0.155)	(0.212)	(0.235)
First-Stage F-statistic	40.821	33.390	19.873	16.572
P-Value	0.000	0.000	0.000	0.000
Controls	No	Yes	No	Yes
Observations	6391	6391	6181	6181

Notes: 2SLS estimation with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Endogenous regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Instrumental variable: number of Higher Education Institutions (HEI) within a 10 km radius from the centroid of the district of residence at age 18. Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Additional control variables: ethnicity, religion, number of siblings, having parents with low education and having parents with different ethnicities.

Generally, the coefficients are higher than those obtained from OLS, which is consistent with a LATE interpretation of the results. Specifically, these coefficients represent the (causal) impact of education on the likelihood of interethnic marriages among those who are induced into higher educational attainments due to the presence of HEI surrounding their place of residence at 18 (i.e. the compliers). The results from the model without control indicate that each additional year of schooling increases the propensity for exogamy by 3.5 p.p. among males and 4.6 p.p. for females. University education rises the probability of having a partner from a different ethnic background by around 44-52 p.p. for males and 71-73 p.p. for females. The model with control variables provides similar evidence, generally with slightly lower second-stage coefficients. Finding similar results from the model that includes controls is a first indication in favor of the internal validity of the results.

4.5.1. Analysis of heterogenous effects and potential mechanisms

To validate our findings, we report the evidence from several sensitivity checks. For simplicity, we report these results for years of schooling only.[55] First, we show the results obtained by adopting different definitions of the radius of exposure for calculating the number of available HEI, which are displayed in Tables 4.6 (males) and 4.7 (females). In general, the results are virtually identical across all alternatives, including when employing the Kernel Density Weighting based on the distance from the district's centroid and the location of HEI. However, using a radius of 10km yields the highest F-statistic for the first stage, and thus represents our preferred option.

Table 4.6: Robustness check – Using different radii - Males

	(1)	(2)	(3)	(4)	(5)	(6)
Radii of exposure:	5 km	10 km	15 km	20 km	25 km	Kernel
Panel A: First Stage - Dependent Variable: Years of Schooling						
HEI within X radius at age 18	0.261***	0.235***	0.233***	0.212***	0.195***	0.408***
	(0.066)	(0.037)	(0.038)	(0.038)	(0.042)	(0.080)
First-Stage F-statistic	15.845	39.635	38.384	31.281	21.605	26.228
Panel B: Second Stage - Dependent Variable: Exogamy						
Years of Schooling	0.037***	0.035***	0.036***	0.035***	0.032***	0.038***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)
Observations	6391	6391	6391	6391	6391	6391

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

[55] The results of robustness checks for other educational attainments are available upon request.

Table 4.7: Robustness check – Using different radii - Females

	(1)	(2)	(3)	(4)	(5)	(6)
Radii of exposure:	5 km	10 km	15 km	20 km	25 km	Kernel
Panel A: First Stage - Dependent Variable: Years of Schooling						
HEI within X radius at age 18	0.295***	0.241***	0.228***	0.220***	0.195***	0.359***
	(0.064)	(0.041)	(0.044)	(0.041)	(0.043)	(0.085)
First-Stage F-statistic	21.422	34.813	27.447	28.505	20.102	17.794
Panel B: Second Stage - Dependent Variable: Exogamy						
Years of Schooling	0.051***	0.046***	0.047***	0.042***	0.033**	0.043**
	(0.016)	(0.015)	(0.013)	(0.013)	(0.013)	(0.017)
Observations	6181	6181	6181	6181	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Second, we consider different relevant ages at exposure (Tables 4.8 and 4.9). As it can be appreciated, the results are not affected by the choice of age at exposure. The first-stage coefficients remain positive and significant for both males and females, even when defining the number of available HEI within a 10km radius based on the district of residence at birth—though slightly reduced. Indeed, this is also an indication that the instrument is not blurred by endogenous residential sorting. Using exposure at 18 years old provides the largest F-statistic for males, although employing age 12 as reference to compute exposure seems to be the best option for females. Nevertheless, given the overall stability of the result, we retain 18 as reference age as baseline for both genders. To further discard the possibility that the results are affected by endogenous residential sorting, we also replicate the estimations after retaining only individuals who never changed district of residence from their birth year until they turned 18 (see Table A.4.2 of the Appendix). Again, the results are virtually the same as for the original estimation sample.

Table 4.8: Robustness checks – Using different age at exposure - Males

	(1)	(2)	(3)	(4)	(5)
age at exposure:	18	15	12	6	0
Panel A: First Stage - Dependent Variable: Years of Schooling					
HEI within 10 km radius	0.235***	0.216***	0.215***	0.204***	0.187***
	(0.037)	(0.036)	(0.036)	(0.039)	(0.041)
First-Stage F-statistic	39.635	36.148	34.863	27.898	20.540
Panel B: Second Stage - Dependent Variable: Exogamy					
Years of Schooling	0.035***	0.038***	0.037***	0.039***	0.038***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.013)
Observations	6391	6391	6391	6391	6391

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table 4.9: Robustness checks – Using different age at exposure - Females

age at exposure:	(1)	(2)	(3)	(4)	(5)
	18	15	12	6	0
Panel A: First Stage - Dependent Variable: Years of Schooling					
HEI within 10 km radius	0.241***	0.230***	0.218***	0.210***	0.193***
	(0.041)	(0.038)	(0.035)	(0.036)	(0.037)
First-Stage F-statistic	34.813	36.518	39.075	34.681	27.709
Panel B: Second Stage - Dependent Variable: Exogamy					
Years of Schooling	0.046***	0.048***	0.048***	0.051***	0.054***
	(0.015)	(0.015)	(0.015)	(0.017)	(0.019)
Observations	6181	6181	6181	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Subsequently, we present the sensitivity checks that are aimed at dispelling doubts about the possibility that the number of available HEI is capturing (potentially endogenous) demand-side factors. The results are reported in Tables 4.10 and 4.11. Here, we mainly focus on reduced-form equations, although we also display the results for the first-stage and the second-stage for comparison. Column (1) contains the results from the reduced-form equation (4.4) obtained from the baseline specification of the instrument. As expected, geographical exposure to HEI at age 18 exerts a positive and significant effect on the likelihood of exogamy for both genders, which is in line with the previous IV/TSLS results. In column (2), we repeat the estimations after controlling for the number of HEI at birth, which would capture for potential long-standing trends in the demand for higher education at the local level. Indeed, this additional control has a very small and insignificant point estimate in the reduced form equation, as well as in the second-stage equation. Moreover, the main results remain virtually unchanged. We obtain similar evidence when restricting the sample to individuals born in districts with at least one Higher Education Institution nearby. This restriction implies comparing districts that were generally similar in terms of pre-existing factors related to the demand for higher education. In column (4) we seek to control for potential recent changes in the demand for higher education across birth cohorts, by controlling for the number of newly established HEI (i.e., those created since the individual's birth year and the year in which he/she turned 18). Also in this case, the corresponding coefficient is virtually zero and insignificant in the reduced form equation and in the second stage, while the coefficients of years of schooling remain qualitatively unchanged.

Finally, as in Currie and Moretti (2003), we control for the number of HEI surrounding the district of residence at age 25, which does not alter the overall results. We also obtain reassuring evidence regarding the validity of the instrument from the falsification based on the random assignment of the district of residence at 18 and the replication of 10000 estimations of the reduced form equation using fake exposure to HEI (permutation test). As displayed in Figure A.4.1 of the Appendix, the distribution of fake reduced form coefficient is centered around zero and the real reduced-form coefficients are clearly outside its mass. Moreover, we report the results of the overidentified model that includes as instruments dummies for exposure to HEI (Table A.4.3 of the Appendix). Although both the first-stage F-statistic and the second-stage coefficient of years of schooling are slightly lower than in the baseline, the results are qualitatively the same. Most importantly, the Hansen J-test for overidentification provides evidence in favor of the null hypothesis that the instruments can be excluded from the second-stage, indicating that geographical exposure to HEI seems not to be directly related to exogamy.

Table 4.10: Robustness check for demand-related factors - Males

	(1)	(2)	(3)	(4)	(5)
Panel A: Reduced Form - Dependent Variable: Exogamy					
HEI within 10km radius at age 18	0.008***	0.013**	0.008**	0.007*	0.009***
	(0.003)	(0.006)	(0.003)	(0.004)	(0.003)
HEI within 10km radius at age 0		-0.007			
		(0.006)			
new HEI (0-18) in 10km radius				0.001	
				(0.005)	
HEI within 10km radius at age 25					-0.001
					(0.001)
Panel B: First Stage - Dependent Variable: Years of Schooling					
HEI within 10km radius at age 18	0.235***	0.365***	0.152***	0.166***	0.296***
	(0.037)	(0.064)	(0.049)	(0.052)	(0.042)
HEI within 10km radius at age 0		-0.184***			
		(0.064)			
new HEI (0-18) in 10km radius				0.099**	
				(0.045)	
HEI within 10km radius at age 25					-0.072***
					(0.023)
First-Stage F-statistic	39.635	32.246	9.833	10.203	49.430
Panel C: Second Stage - Dependent Variable: Exogamy					
Years of Schooling	0.035***	0.036**	0.053**	0.044*	0.031***
	(0.011)	(0.017)	(0.024)	(0.025)	(0.009)
HEI within 10km radius at age 0		-0.001			
		(0.004)			
new HEI (0-18) in 10km radius				-0.003	
				(0.007)	
HEI within 10km radius at age 25					0.001
					(0.001)
Observations	6391	6391	2709	6391	6391

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimations in column (3) are obtained after retaining only individuals who were born in districts with at least one HEI within a radius of 10km.

Table 4.11: Robustness check for demand-related factors - Females

	(1)	(2)	(3)	(4)	(5)
Panel A: Reduced Form - Dependent Variable: Exogamy					
HEI within 10km radius at age 18	0.011***	0.013**	0.013***	0.009**	0.012***
	(0.003)	(0.006)	(0.004)	(0.004)	(0.003)
HEI within 10km radius at age 0		-0.004			
		(0.007)			
new HEI (0-18) in 10km radius				0.003	
				(0.003)	
HEI within 10km radius at age 25					-0.001
					(0.001)
Panel B: First Stage - Dependent Variable: Years of Schooling					
HEI within 10km radius at age 18	0.241***	0.363***	0.152***	0.159***	0.276***
	(0.041)	(0.071)	(0.050)	(0.053)	(0.044)
HEI within 10km radius at age 0		-0.171**			
		(0.073)			
new HEI (0-18) in 10km radius				0.117***	
				(0.038)	
HEI within 10km radius at age 25					-0.041**
					(0.020)
First-Stage F-statistic	34.813	26.425	9.323	8.841	39.975
Panel C: Second Stage - Dependent Variable: Exogamy					
Years of Schooling	0.046***	0.037**	0.086***	0.054*	0.043***
	(0.015)	(0.018)	(0.024)	(0.030)	(0.013)
HEI within 10km radius at age 0		0.003			
		(0.005)			
new HEI (0-18) in 10km radius				-0.003	
				(0.007)	
HEI within 10km radius at age 25					0.001
					(0.001)
Observations	6181	6181	2675	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimations in column (3) are obtained after retaining only individuals who were born in districts with at least one HEI within a radius of 10km.

Finally, again following Currie and Moretti (2003) we estimate the model considering two different instruments, which are based on exposure to public and private HEI respectively. As shown in Table 4.12, the overall results are very similar when considering exposure to the two types of institutions. The first-stage coefficients are somewhat lower for private HEI, while the second stage coefficients are slightly higher. However, the stability of the results is again reassuring and suggest that the presence of HEI is not capturing anticipated changes in the demand for higher education, or other local-level unobserved factors that could be directly related to the propensity to form interethnic marriages. Overall, these results suggest that our instrument is not capturing spurious effect that are due to changing trends in local demand for higher education, supporting the underlying assumption of its exogeneity. As final robustness checks, we also replicate the estimations after excluding individuals who got married before completing education (Table A.4.4 of the Appendix), as well as while retaining in the estimation sample only individuals who report the same ethnicity in 2014 (IFLS 5) than in 2007 (IFLS 4) and appear in both waves of the survey. For both robustness checks, the results are virtually identical with respect to the baseline.

Table 4.12: Separate exposure to public and private HEI

	Males		Females	
	Public HEI	Private HEI	Public HEI	Private HEI
Panel A: First Stage - Dependent Variable: Years of Schooling				
HEI within 10km radius at age 18	0.299***	0.250***	0.285***	0.240***
	(0.066)	(0.040)	(0.073)	(0.046)
First-Stage F-statistic	20.730	38.482	15.339	27.744
Panel B: Second Stage - Dependent Variable: Exogamy				
Years of Schooling	0.028***	0.040***	0.043**	0.052***
	(0.011)	(0.012)	(0.019)	(0.017)
Observations	6391	6391	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at age 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

4.5.2. Evidence about heterogenous effects and potential mechanisms

The evidence obtained so far indicates that higher educational attainments increase the likelihood of having a partner from a different ethnic background. Moreover, the set of sensitivity checks point out the strong stability of the results, and that they can be plausibly interpreted as causal evidence. The next step consists in understanding whether the effect of education on exogamy is heterogeneous according to individual’s and parental characteristics, and what could be the potential mechanisms that underlie the causal chain between HEI expansion, education and interethnic marriages. As for the first objective, Table 4.13 displays the results of the estimation of IV/TSLS with heterogeneous coefficients, in which we interacted years of schooling (and the instrument) with i) own ethnicity[56], ii) the dummy for parental education and iii) the dummy for having parents with

[56] Here we grouped Sundanese, Madurese, Betawi and other ethnicities due to the low number of observations and used a dummy for belonging to the Javanese ethnicity.

mixed ethnic background. The results indicate that the effect of education on exogamy is not significantly different according to parental education and having ethnically-mixed parents. However, the impact of schooling on the likelihood of having a partner from a different ethnic background is lower for individuals with Javanese ethnicity (the largest ethnic group in Java) than for those belonging to other ethnic groups, for whom we detect a larger effect of education on the propensity to interethnic marriage. This result points out that increased educational attainments induced by HEI expansion can reduce segregation of ethnic minorities.

Table 4.13: IV/TSLS with Heterogeneous Effects

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Schooling	0.064**	0.038***	0.033***	0.072***	0.052***	0.042***
	(0.026)	(0.012)	(0.010)	(0.023)	(0.017)	(0.014)
Years of Schooling X Javanese	-0.039*			-0.036*		
	(0.023)			(0.022)		
Years of Schooling X Low Parental Education		-0.006			-0.005	
		(0.032)			(0.020)	
Years of Schooling X Ethnically-Mixed Parents			-0.032			-0.017
			(0.024)			(0.028)
First-Stage F-statistic	17.015	20.796	18.930	19.317	15.159	16.713
Observations	6391	6391	6391	6181	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at age 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. The regression reported in columns (1) and (4) include as control a dummy for being Javanese (versus other ethnicities). The regression reported in columns (2) and (5) include as control a dummy for having low-educated parents. The regression reported in columns (3) and (6) include as control a dummy for having ethnically-mixed parents.

Concerning the analysis of potential mechanisms, we focus on reduced-form estimations that directly relate exposure to HEI surrounding the district of residence at 18 and the different variables that we consider, given data availability. Although we acknowledge that these variables are not ideal for this purpose, because they are observed possibly several years after marriage, we are still convinced that they deserve a certain attention and could highlight interesting patterns regarding potentially relevant channels. The results, reported in Table 4.14, indicate that exposure to HEI has a positive impact on the probability of changing place of residence between age 18 and 2014 (column (1)). Moreover, it also exerts a positive on the probability of moving to a large city (column (2)), where several ethnicities are more likely to coexist. Consistently, being exposed to more HEI at age 18 also increases expected ethnic fractionalization in the district of residence in 2010 (column (3)), although there is no impact on the probability of being an ethnic minority in the community of residence at the time of the survey (column (4)). This evidence indeed suggests that the relevant channel could be migration towards larger agglomerations, where the chances of matching with a person from a different ethnicity are higher, rather than constraints in the marriage market due to residing in enclaves with a very limited number of inhabitants from one's own ethnic group. Finally, we also obtain suggestive evidence regarding the role of changes in social norms. Specifically, individuals exposed to a higher number of HEI during their adolescence are less likely to trust (relatively) more others from the same ethnic group than their counterparts with a different ethnic background. This result highlights the relevance of higher education opportunities in shaping tolerance and trust towards other ethnicities, which could be one of the possible channels through which educational attainments favor the formation of interethnic marriages.

Table 4.14: Potential mechanisms

Dependent Variable:	Migrated (18 - 2014)	Migrated to Large Cities (18 - 2014)	Fractionalization (2010)	Being a minority (2014)	Trust Own Ethnicity (2014)
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
HEI within 10km radius at age 18	0.020***	0.006***	0.045***	0.006	-0.019***
	(0.004)	(0.002)	(0.014)	(0.005)	(0.003)
Observations	6391	6391	6391	6391	4515
Panel B: Females					
HEI within 10 radius at age 18	0.019***	0.007***	0.046***	0.007	-0.024***
	(0.003)	(0.002)	(0.014)	(0.004)	(0.003)
Observations	6181	6181	6181	6181	4872

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

4.6. Conclusion

We investigated the effect of educational attainments on the formation of interethnic marriages in Indonesia, exploiting the expansion of Higher Education Institutions that took place in the country from the last half of the 20th century. We focused on Java Island, the most populated island of the country, where its capital (Jakarta) is located. The empirical analysis was carried out using data from the 2014 wave of the Indonesian Family Life Survey, combined with administrative information about the year of establishment and the exact location of HEI that offer undergraduate degrees across the Java Island. The main outcome consists of the probability of having a partner from a different ethnic background than one's own ethnicity, i.e., exogamy. As for educational attainments, we considered three main measures: years of completed schooling, college attendance, and college completion. To address the issue of endogeneity of education, we exploited variation by year of birth and district of residence at age 18 in geographical exposure to HEI in an Instrumental Variable framework.

The results indicate that education has a positive impact on the propensity to form an ethnically-mixed couple, with somewhat stronger effects observed for females compared to males. Specifically, each additional years of schooling increases the likelihood of exogamy by 3.5 p.p. for males and 4.6 p.p. for females, while the effects of college attendance/completion range between 44-52 p.p. and-71-73 p.p. for males and females, respectively (considering the baseline model without control variables). These results remain largely unchanged across different specifications and are robust to various sensitivity checks, providing supporting evidence for the validity of the Instrumental Variable approach and its underlying assumptions. We do not find evidence of heterogeneous effects of schooling on the propensity to form an interethnic marriage according to parental education or mixed parental ethnicity. However, the effect of education on exogamy is lower for individuals belonging to the largest ethnic group (Javanese) than their counterparts with other ethnic background. This evidence highlights the relevance of education as a tool to reduce segregation of ethnic minorities. Finally, the analysis of potential mechanisms reveals that migration/residential choices and changes in social norms are likely channels through which the expansion of higher education could foster the likelihood of interethnic marriage. Specifically, geographical exposure to HEI rises the propensity to migrate and reside in large cities, characterized by a higher degree of ethnic fractionalization, where ethnically-mixed marriages are more likely. Moreover, individuals exposed to a higher number of HEI during their adolescence are more prone to trust on others from a different ethnic background. This result highlights the potential role of higher education opportunities on changing social norms and favoring interethnic tolerance and social integration.

From the policy perspective, the results reported in this paper suggest that fostering human capital formation through the increase in higher education opportunities driven by the expansion of college education infrastructure is likely to beneficial for several reasons. This is because a wider presence of HEI across the territory not only could lead to higher educational attainments, which could generate positive impacts at the individual level in terms of earning potential and labor market outcomes, health status, and other socioeconomic outcomes. Indeed, the increase in education driven by the expansion of HEI can foster changes in social norms that are likely to break existing ethnic-related barriers, promote a sense of unity, and reduce ethnic segregation in multi-ethnic societies.

Appendix

Table A.4.1a: Complete OLS results - Males

	Years of Education		HE Attendance		HE Completion	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	0.007***	0.006***	0.076***	0.058***	0.074***	0.057***
	(0.001)	(0.001)	(0.015)	(0.013)	(0.016)	(0.014)
Ethnicity						
I(Javanese)			*Reference Category*			
I(Sundanese)		0.010		0.010		0.010
		(0.034)		(0.034)		(0.034)
I(Maduranese)		0.082***		0.077**		0.077**
		(0.030)		(0.030)		(0.030)
I(Betawi)		0.102**		0.103**		0.103**
		(0.048)		(0.049)		(0.049)
I(Other Ethnicities)		0.329***		0.330***		0.332***
		(0.050)		(0.050)		(0.050)
Religion						
I(Islam)			*Reference Category*			
I(Christian)		-0.023		-0.021		-0.019
		(0.028)		(0.027)		(0.028)
I(Hindu)		-0.263*		-0.265*		-0.255*
		(0.147)		(0.144)		(0.136)
I(Other Religions)		0.182		0.164		0.162
		(0.137)		(0.135)		(0.135)
Number of Siblings		-0.003		-0.002		-0.002
		(0.002)		(0.002)		(0.002)
Low Parental Education		-0.016		-0.037***		0.037***
		(0.012)		(0.012)		(0.012)
Ethnically-Mixed Parents		0.210***		0.210***		0.210***
		(0.019)		(0.019)		(0.019)
R-squared	0.166	0.226	0.164	0.225	0.163	0.225
Controls	No	Yes	No	Yes	No	Yes
Observations	6391	6391	6391	6391	6391	6391

Notes: OLS estimations with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Main regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A.4.1b: Complete OLS results - Females

	Years of Education		HE Attendance		HE Completion	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	0.006***	0.004***	0.064***	0.049***	0.066***	0.050***
	(0.002)	(0.002)	(0.019)	(0.018)	(0.020)	(0.019)
Ethnicity						
I(Javanese)			*Reference Category*			
I(Sundanese)		0.021		0.021		0.021
		(0.031)		(0.032)		(0.032)
I(Maduranese)		0.056*		0.049*		0.049*
		(0.030)		(0.029)		(0.029)
I(Betawi)		0.044		0.045		0.045
		(0.046)		(0.047)		(0.046)
I(Other Ethnicities)		0.262***		0.262***		0.261***
		(0.048)		(0.048)		(0.048)
Religion						
I(Islam)			*Reference Category*			
I(Christian)		0.020		0.021		0.021
		(0.025)		(0.025)		(0.025)
I(Hindu)		-0.097		-0.101		-0.102
		(0.222)		(0.221)		(0.222)
I(Other Religions)		-0.038		-0.045		-0.045
		(0.176)		(0.174)		(0.174)
Number of Siblings		0.001		0.002		0.002
		(0.002)		(0.002)		(0.002)
Low Parental Education		-0.021**		-0.035***		-0.035***
		(0.010)		(0.009)		(0.009)
Ethnically-Mixed Parents		0.189***		0.190***		0.190***
		(0.023)		(0.023)		(0.023)
R-squared	0.171	0.207	0.170	0.207	0.170	0.207
Controls	No	Yes	No	Yes	No	Yes
Observations	6181	6181	6181	6181	6181	6181

Notes: OLS estimations with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Main regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A.4.2: Robustness check - excluding individuals who changed district of residence (0-18)

	Males		Females	
	Baseline	Never Move (0 - 18)	Baseline	Never Move (0 - 18)
Panel A: First Stage - Dependent Variable: Years of Schooling				
HEI within 10km radius at age 18	0.235***	0.231***	0.241***	0.246***
	(0.037)	(0.039)	(0.041)	(0.040)
First-Stage F-statistic	39.635	35.143	34.813	37.693
Panel B: Second Stage - Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.034***	0.046***	0.041***
	(0.011)	(0.011)	(0.015)	(0.014)
Observations	6391	6257	6181	6066

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A.4.3: Overidentified IV/TSLS with dummies for the number of HEI

	Males		Females	
	(1)	(2)	(3)	(4)
Panel A: First Stage - Dependent Variable: Years of Schooling				
HEI within 10 km radius	0.235***		0.241***	
	(0.037)		(0.041)	
I(HEI within 10 km radius at age 18 = 0)		*reference category*		*reference category*
I(HEI within 10 km radius at age 18 = 1)		1.817***		1.690***
		(0.355)		(0.482)
I(HEI within 10 km radius at age 18 = 2)		0.433		0.512
		(0.568)		(0.601)
I(HEI within 10 km radius at age 18 = 3)		0.769		1.192
		(0.527)		(0.828)
I(HEI within 10 km radius at age 18 = 4)		2.017***		1.966***
		(0.396)		(0.419)
I(HEI within 10 km radius at age 18 = 5)		3.522***		1.871**
		(0.487)		(0.858)
I(HEI within 10 km radius at age 18 = 6)		1.349*		1.346
		(0.697)		(0.888)
I(HEI within 10 km radius at age 18 = 7)		2.472***		1.910***
		(0.372)		(0.364)
I(HEI within 10 km radius at age 18 = 8)		2.176***		2.666***
		(0.554)		(0.694)
I(HEI within 10 km radius at age 18 = 9)		1.329***		1.930***
		(0.397)		(0.460)
I(HEI within 10 km radius at age 18 ≥ 10)		2.194***		2.312***
		(0.315)		(0.324)
First-Stage F-statistic	39.635	13.199	34.813	7.996
P-Value(1st-Stage F-statistic)	0.000	0.000	0.000	0.000
Panel B: Second Stage - Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.027***	0.046***	0.032***
	(0.011)	(0.008)	(0.015)	(0.011)
P-Value(Hansen J statistic)		0.417		0.387
Observations	6391	6391	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A.4.4: Robustness check – Removing individuals who married before completing education

	Males		Females	
	Baseline	Married after completing education	Baseline	Married after completing education
	(1)	(2)	(3)	(4)
Panel A: First Stage - Dependent Variable: Years of Schooling				
HEI within 10 radius at age 18	0.235***	0.238***	0.241***	0.253***
	(0.037)	(0.037)	(0.041)	(0.040)
First-Stage F-statistic	39.635	40.832	34.813	39.498
Panel B: Second Stage - Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.035***	0.046***	0.042***
	(0.011)	(0.011)	(0.015)	(0.014)
Observations	6391	6355	6181	5982

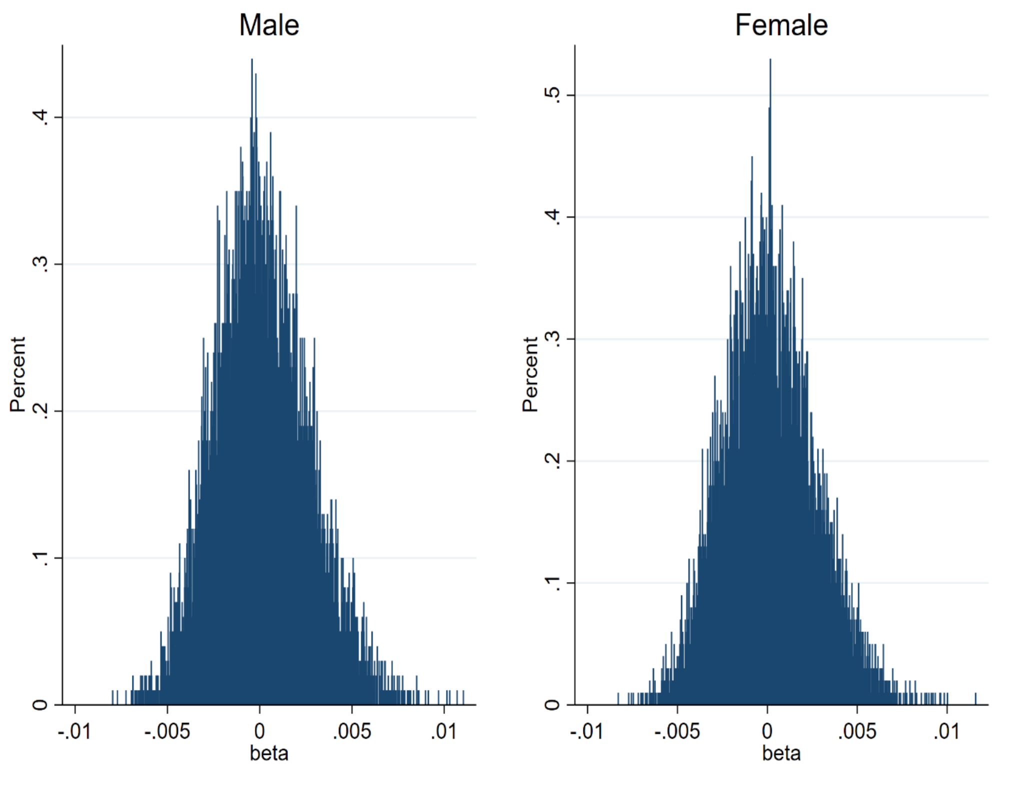
Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimates in columns (2) and (4) are obtained after excluding individuals who married before the year in which they completed education (= year of birth + 6 + years of schooling).

Table A.4.5: Robustness check – Removing individuals who changed ethnicity between 2007 and 2014

	Male		Female	
	Baseline	Same Ethnicity	Baseline	Same Ethnicity
	(1)	(2)	(3)	(4)
Panel A: First Stage - Dependent Variable: Years of Schooling				
HEI within 10 radius at age 18	0.235***	0.225***	0.241***	0.206***
	(0.037)	(0.035)	(0.041)	(0.038)
First-Stage F-statistic	39.635	41.187	34.813	29.593
Panel B: Second Stage - Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.036***	0.046***	0.046***
	(0.011)	(0.012)	(0.015)	(0.015)
Observations	6391	4461	6181	4563

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimates reported in columns (2) and (4) are obtained after retaining only individuals who report the same ethnicity in 2014 than in 2007 and are interviewed in both waves of IFLS.

Figure A.4.1: **Fake reduced form coefficient – permutation test with random assignment of districts of residence at 18**



5. Conclusion

This dissertation provides evidence about the impact of natural disasters on human capital formation, the effects of education policies on educational attainments, and their impacts on the marriage market and ethnic segregation in an emerging country, Indonesia. It examines how natural disasters disrupt the education landscape, potentially hindering human capital accumulation by affecting access to and the quality of education. Simultaneously, it evaluates the effectiveness of educational policy aimed at improving educational outcomes. Beyond these direct effects, the research further explores the broader social implications, particularly how education can bridge ethnic segregation through interethnic marriage. The three empirical chapters provide a comprehensive analysis of these dynamics that aims to offer important policy implications and contribute valuable insights for future research.

Chapter 2 analyses the impact of natural disasters on human capital formation, using a catastrophic earthquake in 2006 on Java Island as a natural experiment. We document that experiencing a powerful earthquake during school age produces medium- and long-term negative impacts on educational attainments. By using 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 negatively affects human capital accumulation, measured by years of schooling as well as enrollment and completion of compulsory and post-compulsory education levels. However, no effect was detected on the likelihood of enrolling in post-compulsory education levels. The analysis of heterogeneous effects highlights that the impact appears stronger for younger individuals who were still in compulsory school when the earthquake struck. Moreover, the effect was also more pronounced for children whose mothers had lower levels of education, pointing towards the protective effect of maternal human capital, which also suggests that governments and policymakers should consider tailoring recovery interventions at the individual/family level, particularly focusing on providing more support to those from disadvantaged educational and social backgrounds. Additionally, the evidence from the potential mechanism suggests that earthquake-related casualties at the family level do not seem to play a relevant role. Internal migration responses do not seem to be significant pathways either, although the results indicate that migration could serve as a method to mitigate the adverse effects of natural disasters. Furthermore, the analysis of the unexplored mechanism of the disruption of educational infrastructures indicates that the loss in years of schooling among younger cohorts is primarily due to the damages suffered by school infrastructure as a consequence of the natural disaster. 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 delays schooling progression or a permanent loss of human capital. The results suggest that both effects are present, although the latter seems 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.

This chapter mainly contributes to the large economics literature on the detrimental effects of natural disasters on human capital formation by providing the medium to long-term impact of the earthquake on educational outcomes. Furthermore, we also provide analysis of heterogenous effects based on individual and family characteristics, ranging from age at exposure, gender, religion, ethnicity, parental education, number of siblings, and birth order, in which no existing papers have provided such a heterogenous analysis across multiple dimensions. Most importantly, we thoroughly analyze several potential mechanisms, considering both demand and supply-side factors, which might drive the connection between exposure to earthquake and educational attainment. A natural extension of this work consists of understanding the effects of exposure to the earthquake beyond educational attainment, for example, on labor market outcomes. More concretely, we plan to analyze the impact of the earthquake on the labor supply decision of women and children within affected households. Exploring the labor supply reactions of women and children after natural catastrophes is relevant for several reasons. First, it can provide insights into how households adjust their labor and income strategies in the aftermath of a shock, which is vital for understanding the broader economic consequences of natural disasters. Second, focusing on women and children can reveal gender and age-specific vulnerabilities and resilience strategies, informing more targeted and effective policy interventions. Lastly, understanding these labor supply dynamics can help predict long-term impacts on human capital development and economic growth, as shifts in labor supply, especially involving children, can have profound implications for education and future earning potential. Investigating this topic can thus contribute to designing better disaster response and recovery programs that mitigate negative outcomes and support sustainable development.

Exploring educational policy, Chapter 3 analyzes the medium to long-term impacts of extending compulsory education in Indonesia on various educational outcomes. In 1994/1995, the Indonesian government extended the mandatory education program from six to nine years. By using three waves of individual panel data, the reform has a notable impact on increasing educational attainment. There were improvements in nine-year schooling completion rates, increased enrollment in senior secondary education, and a rise in overall years of schooling. Despite these improvements, the policy did not significantly impact university attendance rates. Our findings emphasize the effectiveness of the policy in connecting the educational path from primary to secondary levels, although its impact on higher education is still uncertain. It is worth mentioning that the advantages of this reform were particularly significant for girls and individuals from less-educated households, highlighting its impact on reducing educational inequalities. Moreover, the reform had a greater effect in rural areas, demonstrating its ability to reduce educational disparities between different locations. The thorough examination highlights the crucial importance of mandatory education reforms in increasing educational levels, promoting gender equality, and broadening access for socially disadvantaged groups.

This chapter adds valuable insights to the economics literature by examining the medium to long-term impacts of mandatory schooling reform on a set of educational outcomes such as completion rate in junior secondary schooling, enrollment and completion rate in senior secondary education, and university attendance. This is an important contribution to literature because most of the previous papers on this topic only focus on the immediate effects. Furthermore, the analysis of heterogeneous effects reveals the diverse impacts of the reform across different groups, highlighting the reform’s effectiveness in reducing educational inequalities among gender, parental educational background, and urban and rural areas. This evidence is especially valuable for policymakers in tailoring educational policies to meet the specific needs of diverse demographic groups, improving their effectiveness on educational outcomes. Looking ahead, this chapter lays the groundwork for future research into the long-term socioeconomic impacts of the reform, particularly its implications on labor market outcomes, earnings, and social mobility. Such exploration is essential for capturing the full scope of benefits associated with extending compulsory education and contributes to a wider discussion on how educational policies can foster human capital development. Importantly, our finding suggests that this policy is especially beneficial for traditionally underserved groups, including those from less-educated family backgrounds, females, and rural areas, emphasizing its contribution to reducing educational inequalities. This leads to a deeper examination of the gender-specific outcomes of educational reforms in a country marked by inequality. Understanding these dynamics is crucial for designing policies that not only elevate educational standards but also address gender disparities and support the advancement of women and other marginalized groups.

Finally, Chapter 4 shifts the focus from the determinants of education to its socioeconomic impacts, specifically its effect on the likelihood of engaging in an interethnic marriage in Indonesia. We investigate the role of educational attainments, including years of schooling, university enrollment, and completion, in facilitating interethnic marriage by utilizing the geographical expansion of HEI across Java Island. Our findings reveal that higher educational attainments, induced by the expansion of HEI, significantly increase the likelihood of being in an interethnic marriage. The positive effect on interethnic marriages is somewhat more pronounced among females than males. The analysis of heterogeneous effects shows that the impact of education on engaging in a marriage with a partner from a different ethnic background remains the same, regardless of parental education and having parents with mixed ethnicities. However, increased educational attainments induced by HEI expansion exert a lower effect on exogamy for individuals of Javanese ethnicity than those from other ethnic backgrounds. This evidence suggests that education can serve as a critical means to reduce ethnic segregation. Additionally, we identify potential mechanisms, representing an additional value added to this paper, behind these trends:

more educated individuals are more likely to move and settle in larger cities with a higher degree of ethnic fractionalization, thereby increasing the probability of exogamy. Moreover, the educational attainments prompted by HEI expansion enhance trust towards people from different ethnic backgrounds, highlighting its role in fostering more integrated and inclusive societies.

This chapter not only deepens our understanding about investments in educational infrastructure and the local impact of college expansion but also enriches the economics literature by introducing the first study to offer plausible causal evidence that links educational attainment with the likelihood of being engaged in an interethnic marriage. This causal relationship is crucial because interethnic marriages can serve as a barometer for social cohesion and integration, reflecting the degree to which diverse ethnic groups coexist harmoniously. In the Indonesian context, a country known for its vast ethnic diversity yet challenged by historical and ongoing ethnic tensions, understanding the dynamics of interethnic marriages becomes even more crucial. Furthermore, this chapter documents the establishment of new HEI and stresses the necessity for future research on their impact on economic development and regional growth. By examining how HEI act as economic engines for job creation, entrepreneurship, and technological innovation, future research could illuminate their crucial role in bolstering local and regional economies.

In summary, this dissertation offers a comprehensive analysis of the interplay between natural disasters, education policies, and socioeconomic outcomes in Indonesia, providing profound insights into the dynamics of human capital formation, educational attainment, and their broader implications for societal cohesion and economic development. Furthermore, it underscores the importance of robust educational frameworks capable of withstanding shocks, reducing inequalities, and fostering a more inclusive and harmonious society, thereby offering valuable information and practical guidance for policymakers, educators, and researchers.

References

A. Sathar, Z., Wazir, A. and Sadiq, M. (2013) 'Struggling against the Odds of Poverty, Access, and Gender: Secondary Schooling for Girls in Pakistan', *the Lahore Journal of Economics*, 18(Special Edition), pp. 67–92.

Acemoglu, D. and Angrist, J.D. (1999) *How Large are Human-Capital Externalities Evidence NBER, NBER Working Paper Series*. Working Paper 7444.

Adukia, A., Asher, S. and Novosad, P. (2020) 'Educational investment responses to economic opportunity: Evidence from Indian road construction', *American Economic Journal: Applied Economics*, 12(1), pp. 348–376.

Afkar, R., Yarrow, N., Surbakti, S., Cooper, R., Afkar, R. and Yarrow, N. (2020) *Inclusion in Indonesia's Education Sector: A Subnational Review of Gender Gaps and Children with Disabilities*, *Policy Research Working Paper*. 9282. Washington, DC: The World Bank Group.

Al-samarrai, S. and Lewis, B. (2021) *The Role of Intergovernmental Fiscal Transfers in Improving Education Outcomes*, *International Development in Focus*. Washington, D.C.: World Bank Group.

Alesina, A., Giuliano, P. and Reich, B. (2021) 'Nation-building and education', *The Economic Journal*, 131(638), pp. 2273–2303.

Allendorf, K. and Thornton, A. (2015) 'Caste and choice: The influence of developmental idealism on marriage behavior', *American Journal of Sociology*, 121(1), pp. 243–287.

Almond, D. and Doyle, J.J. (2011) 'After midnight: A regression discontinuity design in length of postpartum hospital stays', *American Economic Journal: Economic Policy*, 3(3), pp. 1–34.

Amin, V., Lundborg, P. and Rooth, D.O. (2015) 'The intergenerational transmission of schooling: Are mothers really less important than fathers?', *Economics of Education Review*, 47, pp. 100–117.

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*. Available at: https://inarisk.bnpb.go.id/pdf/Buku RBI_Final_low.pdf.

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. 1–67.

Angrist, J.D. and Krueger, A.B. (1991) 'Does Compulsory School Attendance Affect Schooling and Earnings?', *Quarterly Journal of Economics*, 106(4), pp. 979–1014.

Angrist, J.D. and Pischke, J.-S. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. 1st edn. Princeton University Press.

Attanasio, O.P., Meghir, C. and Santiago, A. (2012) 'Education choices in Mexico: Using a structural model and a randomized experiment to evaluate PROGRESA', *Review of Economic Studies*, 79(1), pp. 37–66.

Aurino, E. et al. (2023) 'Food for thought? Experimental Evidence on the Learning Impacts of a Large-Scale School Feeding Program', *Journal of Human Resources*, 58(4), pp. 74–111.

Awan, M., Malik, N., Sarwar, H. and Waqas, M. (2011) 'Impact of education on poverty reduction', *International Journal of Academic Research*, 3(1), pp. 659–664.

Aydemir, A. and Kirdar, M.G. (2017) 'Low Wage Returns to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey', *Oxford Bulletin of Economics and Statistics*, 79(6), pp. 1046–1086.

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. Available at: https://www.iza.org/publications/dp/5164/do-natural-disasters-affect-human-capital-an-assessment-based-on-existing-empirical-evidence.

Baez, J. and Santos, I. (2008) 'On shaky ground: The effects of earthquakes on household income and poverty', *RPP-LAC-MDGs and Poverty* [Preprint]. Available at:

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a46fa2eb30205f04a062c9b233f2116f426992da.

Bandiera, O., Mohnen, M., Rasul, I. and Viarengo, M. (2019) 'Nation-building through compulsory schooling during the age of mass migration', *Economic Journal*, 129(617), pp. 62–109.

Bandyopadhyay, S. and Green, E. (2021) 'Explaining inter-ethnic marriage in Sub-Saharan Africa', *Journal of International Development*, 33(4), pp. 627–643.

Banerjee, A. V., Cole, S., Duflo, E. and Linden, L. (2007) 'Remedying Education: Evidence From Two Randomized Experiments in India* Abhijit V. Banerjee Shawn Cole', *The Quarterly Journal of Economics*, 122(3), pp. 1235–1264.

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.

Barakat, B. and Bengtsson, S. (2018) 'What do we mean by school entry age? Conceptual ambiguity and its implications: the example of Indonesia', *Comparative Education*, 54(2), pp. 203–224.

Barrera-Osorio, F., Bertrand, M., Linden, L.L. and Perez-Calle, F. (2011) 'Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia', *American Economic Journal: Applied Economics*, 3(2), pp. 167–195.

Barrera-Osorio, F., Linden, L.L. and Saavedra, J.E. (2019) 'Medium- and long-term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia', *American Economic Journal: Applied Economics*, 11(3), pp. 54–91.

Barro, R. and Sala-i-Martin, X. (1995) *Economic Growth*. New York: Cambridge University Press.

Barro, R.J. (1991) 'Economic Growth in a Cross Section of Countries', *The Quarterly Journal of Economics* [Preprint].

Barro, R.J. (2001) ‘Human capital and growth’, *American Economic Review*, 91(2), pp. 12–17.

Bazzi, S., Gaduh, A., Rothenberg, A.D. and Wong, M. (2019) ‘Unity in diversity? How intergroup contact can foster nation building’, *American Economic Review*, 109(11), pp. 3978–4025.

Becker, G.S. (1975) ‘Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition’, in *Human capital: A theorical and empirical Analysis*. 2nd Ed. The University of Chicago Press, pp. 45–70. Available at: http://www.nber.org/books/beck75-1.

Behrman, J.R. and Rosenzweig, M.R. (2005) ‘Does increasing women’s schooling raise the schooling of the next generation? Reply’, *American Economic Review*, 95(5), pp. 1745–1751.

Belloc, M., Drago, F. and Galbiati, R. (2016) ‘Earthquake, religion, and transition to self-government in Italian cities’, *The Quarterly Journal of Economics*, 131(4), pp. 1875–1926.

Belskaya, V., Peter, K.S. and Posso, C.M. (2020) ‘Heterogeneity in the effect of college expansion policy on wages: Evidence from the Russian labor market’, *Journal of Human Capital*, 14(1), pp. 84–121.

Benhabib, J. and Spiegel, M.M. (1994) ‘The role of human capital in economic development evidence from aggregate cross-country data’, *Journal of Monetary Economics*, 34, pp. 143–173.

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

Black, S.E., Devereux, P.J. and Salvanes, K.G. (2005) ‘Why the apple doesn’t fall far: Understanding intergenerational transmission of human capital’, *American Economic Review*, 95(1), pp. 437–449.

Blundell, R., Dearden, L. and Sianesi, B. (2004) *Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS, Working Paper No. 03/20*. IFS.

BMKG (2018) *Katalog Gempa Signifikan dan Merusak 1874-2017*. Jakarta, Indonesia. Available at: https://cdn.bmkg.go.id/Web/Katalog-Gempabumi-Signifikan-dan-Merusak-1821-2018.pdf.

Bratti, M., Cottini, E. and Ghinetti, P. (2022) *Education, Health and Health-Related Behaviors: Evidence from Higher Education Expansion, IZA Discussion Paper*.

de Brauw, A., Gilligan, D.O., Hoddinott, J. and Roy, S. (2015) 'The Impact of Bolsa Família on Schooling', *World Development*, 70, pp. 303–316.

Buchori, M. and Malik, A. (2004) 'The Evolution of Higher Education in Indonesia', in Altbach, P.G. (ed.) *Asian universities: Historical perspective and contemporary challenges*. Baltimore: The Johns Hopkins University Press. Available at: https://www.google.es/books/edition/Asian_Universities/AmORZ00knyEC?hl=en&gbpv=1&printsec=frontcover.

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*. Available at: https://www.iza.org/publications/dp/6354/persistent-impact-of-natural-disasters-on-child-nutrition-and-schooling-evidence-from-the-1999-colombian-earthquake.

Buttenheim, A.M. and Nobles, J. (2009) 'Ethnic diversity, traditional norms, and marriage behaviour in Indonesia', *Population Studies*, 63(3), pp. 277–294.

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.

Calonico, S., Cattaneo, M. and Titiunik, R. (2014) 'Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs', *Econometrica*, 82(6), pp. 2295–2326.

Card, D. (1999) 'Chapter 30 The causal effect of education on earnings', in *Handbook of Labor Economics*. New York, pp. 1801–1863.

Carneiro, P.M., Liu, K. and Salvanes, K.G. (2023) 'The Supply of Skill and

Endogenous Technical Change: Evidence from a College Expansion Reform', *Journal ofthe European Economic Association*, 21(1), pp. 48–92.

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.

Cascio, E.U. and Lewis, E.G. (2006) 'Schooling and the Armed Forces Qualifying Test : Evidence from School-Entry Laws', *The Journal of Human Resources*, 41(2), pp. 294–318.

Cattaneo, M.D., Idrobo, N. and Titiunik, R. (2020) *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge: Cambridge University Press.

Cattaneo, M.D., Jansson, M. and Ma, X. (2018) 'Manipulation testing based on density discontinuity', *Stata Journal*, 18(1), pp. 234–261.

Cavallo, E., Bank, I.D., Galiani, S., Pantano, J., 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.

Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014) 'Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood', *American Economic Review*, 104(9), pp. 2633–2679.

Chevalier, A. (2004) *Parental Education and Child's Education: A Natural Experiment, IZA Discussion Paper No. 1153*.

Choi, Á. (2009) 'Sequential educational expansion, equality, and growth in the republic of Korea', *KEDI Journal of Educational Policy*, 6(2), pp. 21–47.

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

Clay, K., Lingwall, J. and Jr, M.S. (2021) ‘Laws, educational outcomes, and returns to schooling evidence from the first wave of U.S. state compulsory attendance laws’, *Labour Economics*, 68.

Corvalan, A. and Vargas, M. (2015) ‘Segregation and conflict: An empirical analysis’, *Journal of Development Economics*, 116, pp. 212–222.

Cottini, E., Ghinetti, P. and Moriconi, S. (2019) *Higher Education Supply, Neighbourhood Effects and Economic Welfare*, *CESifo Working Paper*. 7483.

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.

Cremin, P. and Nakabugo, M.G. (2012) ‘Education, development and poverty reduction: A literature critique’, *International Journal of Educational Development*, 32(4), pp. 499–506.

Crespin-Boucaud, J. (2020) ‘Interethnic and interfaith marriages in sub-Saharan Africa’, *World Development*, 125, p. 104668.

Croke, K., Grossman, G., Larreguy, H.A. and Marshall, J. (2016) ‘Deliberate disengagement: How education can decrease political participation in electoral authoritarian regimes’, *American Political Science Review*, 110(3), pp. 579–600.

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

Currie, J. and Moretti, E. (2003) ‘Mother’s education and the intergenerational transmission of human capital: Evidence from college openings’, *Quarterly Journal of Economics*, 118(4), pp. 1495–1532.

Dayioglu, M. and Kirdar, M.G. (2022) ‘Keeping Kids in School and Out of Work: Compulsory Schooling and Child Labor in Turkey’, *Journal of Human Capital*, 16(4).

Dee, T.S. (2004) ‘Are there civic returns to education?’, *Journal of Public Economics*, 88(9–10), pp. 1697–1720.

DGHE (2003) *Higher Education Long Term Strategy 2003-2010, Director General of Higher Education Ministry of National Education Republic of Indonesia*. Jakarta. Available at: https://luk.staff.ugm.ac.id/phk/helts/HELTS2003-2010E.pdf.

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.

Duflo, E. (2012) ‘Women empowerment and economic development’, *Journal of Economic Literature,* 50(4), pp. 1051–1079.

Duflo, E., Dupas, P. and Kremer, M. (2011) ‘Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya’, *American Economic Review,* 101(5), pp. 1739–1774.

Duflo, E., Dupas, P. and Kremer, M. (2015) ‘Education, HIV, and early fertility: Experimental evidence from Kenya’, *American Economic Review,* 105(9), pp. 2757–2797.

Dursun, B., Cesur, R. and Kelly, I.R. (2022) ‘Mandatory Schooling of Girls Improved Their Children’s Health: Evidence from Turkey’s 1997 Education Reform’, *Journal of Policy Analysis and Management,* 41(3), pp. 824–858.

Dustan, A. (2020) ‘Can large, untargeted conditional cash transfers increase urban high school graduation rates? Evidence from Mexico City’s Prepa Sí’, *Journal of Development Economics,* 143(September 2019), p. 102392.

Eble, A. and Hu, F. (2019) ‘Does primary school duration matter? Evaluating the consequences of a large Chinese policy experiment’, *Economics of Education Review,* 70(March), pp. 61–74.

Elsayed, M.A.A. (2019) ‘Keeping kids in school: The long-term effects of extending compulsory education’, *Education Finance and Policy,* 14(2), pp. 242–271.

Erten, B. and Keskin, P. (2019) ‘Compulsory schooling for whom? The role of gender, poverty, and religiosity’, *Economics of Education Review,* 72(May), pp. 187–203.

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.

Esteban, J., Mayoral, L. and Ray, D. (2012) ‘Ethnicity and conflict: An empirical study’, *American Economic Review*, 102(4), pp. 1310–1342.

Fang, H., Eggleston, K., Rizzo, J., Rozelle, S. and Zeckhauser, R. (2012) *The Returns to Education in China: Evidence from the 1986 Compulsory Education Law, NBER Working Paper Series*. 18189.

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.

Fryer, R.G. (2011) ‘Financial incentives and student achievement: Evidence from randomized trials’, *Quarterly Journal of Economics*, 126(4), pp. 1755–1798.

Fukushima, K., Mizuoka, S., Yamamoto, S. and Iizuka, T. (2016) ‘Patient cost sharing and medical expenditures for the Elderly’, *Journal of Health Economics*, 45, pp. 115–130.

Furtado, D. and Song, T. (2022) ‘Interethnic Marriages’, in Zimmermann, K.F. (ed.) *Handbook of Labor, Human Resources and Population Economics*. Cham: Springer International Publishing, pp. 1–19.

García, J. and 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.

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.

Gazeaud, J. and Ricard, C. (2024) ‘Learning effects of conditional cash transfers: The role of class size and composition’, *Journal of Development Economics*, 166(September 2023), p. 103194.

Ghosh, T. and Ramanayake, S.S. (2020) ‘The macroeconomics of gender equality’, *International Journal of Finance & Economics*, 26(2), pp. 1955–1977.

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.

Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E. (2004) ‘Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya’, *Journal of Development Economics*, 74(1), pp. 251–268.

Glewwe, P., Siameh, C., Sun, B. and Wisniewski, S. (2011) *School resources and educational outcomes in developing countries, NBER Working Paper No. 17554*. 17554.

Goldin, C. and Katz, L.F. (2011) ‘Mass Secondary Schooling and the State: The Role of State Compulsion in the High School Movement’, in Costa, D.L. and Lamoreaux, N.R. (eds) *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, pp. 275–310.

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.

Goode, R.B. (1959) ‘Adding to the Stock of Physical and Human Capital’, *The American Economic Review*, 49(2), pp. 147–155.

Grenet, J. (2013) ‘Is Extending Compulsory Schooling Alone Enough to Raise Earnings? Evidence from French and British Compulsory Schooling Laws’, *Scandinavian Journal of Economics*, 115(1), pp. 176–210.

Griliches, Z. (1996) *Education, Human Capital, and Growth: A Personal Perspective, Working Paper No. 5426*. NBER. Cambridge, MA.

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.

Hampf, F. (2019) *The Effect of Compulsory Schooling on Skills: Evidence from a Reform in Germany, Working Paper No. 313*. Munich.

Hannum, E. and Buchmann, C. (2005) 'Global educational expansion and socio-economic development: An assessment of findings from the social sciences', *World Development*, 33(3), pp. 333–354.

Hanushek, E.A. (2020) 'Education production functions', in Bradley, S. and Green, C. (eds) *The Economics of Education A Comprehensive Overview*. 2nd edn. Academic Press, pp. 161–170.

Hanushek, E.A., Machin, S. and Woessmann, L. (eds) (2006a) *Handbook of the Economics of Education*. Amsterdam, The Netherlands: North Holland.

Hanushek, E.A., Machin, S. and Woessmann, L. (eds) (2006b) *Handbook of the Economics of Education*. Amsterdam, The Netherlands: North Holland.

Hanushek, E.A., Machin, S. and Woessmann, L. (eds) (2011a) *Handbook of the Economics of Education*. Amsterdam, The Netherlands: North Holland.

Hanushek, E.A., Machin, S. and Woessmann, L. (eds) (2011b) *Handbook of the Economics of Education*. Amsterdam, The Netherlands: North Holland.

Hanushek, E.A., Machin, S. and Woessmann, L. (eds) (2016) *Handbook of the Economics of Education*. Amsterdam, The Netherlands: North Holland.

Harmon, C. and Walker, I. (1995) 'Estimates of the Economic Return to Schooling for the United Kingdom', *The American Economic Review*, 85(5), pp. 1278–1286.

Harmon, C. and Walker, I. (1999) 'The marginal and average returns to schooling in the UK', *European Economic Review*, 43(4–6), pp. 879–887.

Harrendorf, S., Heiskanen, M. and Malby, S. (2010) *International Statistics on Crime and Justice*, *Heuni Publication Series*. Helsinki.

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.

Holmlund, H., Lindahl, M. and Plug, E. (2011) 'The causal effect of parents' schooling on children's schooling: A comparison of estimation methods', *Journal of Economic Literature*, 49(3), pp. 615–651.

Hoskins, B., D'Hombres, B. and Campbell, J. (2008) 'Does formal education have an impact on Active Citizenship behaviour?', *European Educational Research Journal*, 7(3), pp. 386–402.

Hoxby, C.M. and Bulman, G.B. (2016) 'The effects of the tax deduction for postsecondary tuition: Implications for structuring tax-based aid', *Economics of Education Review*, 51, pp. 23–60.

Ida Bagus, M. (2008) 'Mixed marriages in jembrana, bali: Mediation and fragmentation of citizenship and identity in the post-bomb(s) bali world', *Asia Pacific Journal of Anthropology*, 9(4), pp. 346–362.

Imbens, G.W. and Lemieux, T. (2008) 'Regression discontinuity designs: A guide to practice', *Journal of Econometrics*, 142(2), pp. 615–635.

Jagnani, M. and Khanna, G. (2020) 'The effects of elite public colleges on primary and secondary schooling markets in India', *Journal of Development Economics*, 146, p. 102512.

Jensen, R. (2000) 'Agricultural volatility and investments in children', *American Economic Review*, 90(2), pp. 399–404.

Kamhöfer, D.A., Schmitz, H. and Westphal, M. (2019) 'Heterogeneity in marginal non-monetary returns to higher education', *Journal of the European Economic Association*, 17(1), pp. 205–244.

Kazianga, H., Levy, D., Linden, L.L. and Sloan, M. (2013) 'The effects of "girl-friendly" schools: Evidence from the bright school construction program in burkina faso', *American Economic Journal: Applied Economics*, 5(3), pp. 41–62.

Khalid, A. (2023) 'Mothers and their daughters' education: a comparison of global and local aspirations', *Comparative Education*, 59(2), pp. 259–281.

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

Kirdar, M.G., Dayioglu, M. and Koç, I. (2016) 'Does longer compulsory education equalize schooling by gender and rural/urban residence?', *World Bank Economic Review*, 30(3), pp. 549–579.

Kirdar, M.G., Dayıoğlu, M. and Koç, İ. (2018) 'The effects of compulsory-schooling laws on teenage marriage and births in Turkey', *Journal of Human Capital*, 12(4), pp. 640–668.

Klasen, S. and Lamanna, F. (2009) 'The impact of gender inequality in education and employment on economic growth: New evidence for a panel of countries', *Feminist Economics*, 15(3), pp. 91–132.

Korwatanasakul, U. (2019) *Revisiting the returns to education during rapid structural and rural transformation in Thailand A regression discontinuity approach*. 105.

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

Kremer, M., Brannen, C. and Glennerster, R. (2013) 'The challenge of education and learning in the developing world', *Science*, 340(6130), pp. 297–300.

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.

Kukić, L. (2023) 'The last Yugoslavs: Ethnic diversity and national identity', *Explorations in Economic History*, 88(November 2021), p. 101504.

Kyui, N. (2016) 'Expansion of higher education, employment and wages: Evidence from the Russian Transition', *Labour Economics*, 39, pp. 68–87.

de la Cuesta, B. and Imai, K. (2016) 'Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections∗', *Annual Review of Political Science*, 19, pp. 375–396.

Larreguy, H. and Liu, S.X. (2023) ‘When does education increase political participation? Evidence from Senegal’, *Political Science Research and Methods*, pp. 1–18.

Lee, D.S. and Card, D. (2008) ‘Regression discontinuity inference with specification error’, *Journal of Econometrics*, 142(2), pp. 655–674.

Lewis, B.D. and Nguyen, H.T.M. (2020) ‘Assessing the causal impact of compulsory schooling policy in Indonesia’, *International Journal of Educational Research*, 104(June), p. 101693.

Lleras-Muney, A. (2002) ‘Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939’, *Journal of Law and Economics*, 45(2 I), pp. 401–435.

Lleras-Muney, A. (2005) ‘The relationship between education and adult mortality in the United States’, *Review of Economic Studies*, 72(1), pp. 189–221.

Lleras-Muney, A. and Shertzer, A. (2015) ‘Did the Americanization movement succeed? An evaluation of the effect of english-only and compulsory schooling laws on immigrants’, *American Economic Journal: Economic Policy*, 7(3), pp. 258–290.

Lochner, L. (2011) *Non-Production Benefits of Education: Crime, Health, And Good Citizenship, Working Paper No. 16722*. NBER. Cambridge, MA.

Lucas, R.E. (1988) ‘On the Mechanics of Economic Development’, *Journal of Monetary Economics*, 22(February), pp. 3–42.

Marchetta, F. and Sahn, D.E. (2016) ‘The role of education and family background in marriage, childbearing, and labor market participation in Senegal’, *Economic Development and Cultural Change*, 64(2), pp. 369–403.

Masiero, G. and Santarossa, M. (2021) ‘Natural disasters and electoral outcomes’, *European Journal of Political Economy*, 67(March), pp. 1–19.

McCrary, J. (2008) ‘Manipulation of the running variable in the regression discontinuity design: A density test’, *Journal of Econometrics*, 142(2), pp. 698–714.

McDermott, T.K.J. (2012) *The Effects of Natural Disasters on Human Capital Accumulation, Institute for International Integration Studies 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.

Mincer, J. (1958) ‘Investment in Human Capital and Personal Income Distribution’, *Journal of Political Economy*, 66(4), pp. 281–302.

Momo, M.S.M., Cabus, S.J. and Groot, W. (2021) ‘Evidence on the marginal impact of a compulsory secondary education reform in Senegal on years of education and changes in high school decisions’, *International Journal of Educational Research Open*, 2(June), p. 100058.

Montenegro, C.E. and Patrinos, H.A. (2014) *Comparable estimates of returns to schooling around the world, World Bank Group: Education Global Practice Group*. Education Global Practice Group.

Mukminin, A., Habibi, A., Prasojo, L.D., Idi, A. and Hamidah, A. (2019) ‘Curriculum reform in indonesia: Moving from an exclusive to inclusive curriculum’, *Center for Educational Policy Studies Journal*, 9(2), pp. 53–72.

Natarajan, M. (2016) ‘Crime in developing countries: The contribution of crime science’, *Crime Science*, 5(8), pp. 2–5.

Nelson, R.R., Phelps, E.S., Plelps, E.S. and Phelps, E.S. (1966) ‘Investment in Humans, Technological Diffusion, and Economic Growth’, *The American Economic Review*, 56(1/2), pp. 69–75. Available at: https://www.jstor.org/stable/1821269.

Ngo, J. and Meek, L. (2019) ‘Higher Education Governance and Reforms in Indonesia: Are the Matrices of Autonomy Appropriate?’, *Journal of International and Comparative Education*, 8(1), pp. 17–26.

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.

Oreopoulos, P. (2006a) ‘Estimating average and local average treatment effects of education when compulsory schooling laws really matter’, *American Economic Review*, 96(1), pp. 152–175.

Oreopoulos, P. (2006b) ‘The compelling effects of compulsory schooling: Evidence from Canada’, *Canadian Journal of Economics*, 39(1), pp. 22–52.

Ortiz-Ospina, E. and Roser, M. (2020) *Global Health - Our World in Data*, *Our World In Data*. Available at: https://ourworldindata.org/health-meta#citation.

Pannen, P. (2018) ‘Higher Education Systems and Institutions, Indonesia’, in Amaral, A. et al. (eds) *Encyclopedia of International Higher Education Systems and Institutions*. Springer Dordrecht, pp. 1143–1151.

Parinduri, R.A. (2014) ‘Do children spend too much time in schools? Evidence from a longer school year in Indonesia’, *Economics of Education Review*, 41, pp. 89–104.

Parinduri, R.A. (2017) ‘Does Education Improve Health? Evidence from Indonesia’, *Journal of Development Studies*, 53(9), pp. 1358–1375.

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.

Parker, L., Hoon, C.Y. and Raihani (2014) ‘Young people’s attitudes towards inter-ethnic and inter-religious socializing, courtship and marriage in Indonesia’, *South East Asia Research*, 22(4), pp. 467–486.

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

PDSP Kemdikbud (2013) *Statistik Perguruan Tinggi (PT) 2012/2013*. Available at: https://publikasi.data.kemdikbud.go.id/upload/file/isi_11800868-A25A-4153-BBBB-C416C54BCA5F_.pdf.

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(November), pp. 2–17.

Di Pietro, G. (2017) '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.

Pischke, J.S. and Von Wachter, T. (2008) 'Zero returns to compulsory schooling in Germany: Evidence and interpretation', *Review of Economics and Statistics*, 90(3), pp. 592–598.

Powdthavee, N. (2010) 'Does education reduce the risk of hypertension? Estimating the biomarker effect of compulsory schooling in England', *Journal of Human Capital*, 4(2), pp. 173–202.

Pronzato, C. (2012) 'An examination of paternal and maternal intergenerational transmission of schooling', *Journal of Population Economics*, 25(2), pp. 591–608.

Psacharopoulos, G. and Patrinos, H.A. (2004) 'Returns to investment in education: A further update', *Education Economics*, 12(2), pp. 111–134.

Psacharopoulos, G. and Patrinos, H.A. (2018) *Returns to Investment in Education A Decennial Review of the Global Literature*, *Working Paper No. 8402*. Education Global Practice.

Rademakers, R. and van Hoorn, A. (2021) 'Ethnic switching: Longitudinal evidence on prevalence, correlates, and implications for measuring ethnic segregation', *Journal of Development Economics*, 152(June).

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.

Ray, R. and Lancaster, G. (2005) *The impact of children’s work on schooling: Multi-country evidence*, *International Labour Review*.

Ray, T., Roy Chaudhuri, A. and Sahai, K. (2020) ‘Whose education matters? An analysis of inter caste marriages in India’, *Journal of Economic Behavior and Organization*, 176, pp. 619–633.

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.

Ridwan, N. haeriyah (2017) ‘Kesetaraan Gender Pendidikan Di Indonesia’, *Jurnal Manajemen Pendidikan Islam*, 1(1), pp. 35–52.

Rohner, D. and Saia, A. (2019) *Education and Conflict: Evidence from a Policy Experiment in Indonesia*, *CEPR Discussion Paper No. 13509*. Available at: https://cepr.org/publications/dp13509.

Roth, C. and Sumarto, S. (2015) *Does education increase interethnic and interreligious tolerance? Evidence from a natural experiment*, *MPRA Paper No. 64558*. Available at: https://mpra.ub.uni-muenchen.de/64558/.

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

Sabates, R. and Feinstein, L. (2007) ‘Effects of government initiatives on youth crime’, *Oxford Economic Papers*, 60(3), pp. 462–483.

Sanjaya, M.R., Chuah, S.H., Feeny, S. and Hoffmann, R. (2022) ‘The Impact of Cultural Heterogeneity on Violence in Indonesia: Fractionalisation versus polarization’, *Applied Economics*, 55(16), pp. 1790–1806.

Schultz, T.W. (1961) ‘Investment in Human Capital’, *The American Economic Review*, 51(1), pp. 1–17.

Shanan, Y. (2021) ‘The effect of compulsory schooling laws and child labor restrictions on fertility: evidence from the early twentieth century’, *Journal of Population Economics* [Preprint].

Silles, M.A. (2008) ‘The causal effect of education on health: Evidence from the United Kingdom’, *Economics of Education Review*, 28(1), pp. 122–128.

Spohr, C.A. (2003) ‘Formal schooling and workforce participation in a rapidly developing economy: Evidence from “compulsory” junior high school in Taiwan’, *Journal of Development Economics*, 70(2), pp. 291–327.

Statistics Indonesia (2010) *Sensus Penduduk 2010*. Available at: https://sensus.bps.go.id/main/index/sp2010.

Stella, L. (2013) ‘Intergenerational transmission of human capital in Europe: evidence from SHARE’, *IZA Journal of European Labor Studies*, 2(13), pp. 1–24.

Stephens, M. and Yang, D.Y. (2014) ‘Compulsory education and the benefits of schooling’, *The American Economic Review*, 104(6), pp. 1777–1792.

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.

Sukamdi and Mujahid, G. (2015) Internal Migration in Indonesia, UNFPA Indonesia Monograph Series No.3. Available at: http://indonesia.unfpa.org/sites/default/files/pub-pdf/FA_Isi_BUKU_Monograph_Internal_Migration_ENG.pdf.

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.

Suryadarma, D., Suryahadi, A., Sumarto, S. and Rogers, F.H. (2006) ‘Improving student performance in public primary schools in developing countries: Evidence from Indonesia', *Education Economics*, 14(4), pp. 401–429.

Tang, C., Zhao, L. and Zhao, Z. (2020) 'Does free education help combat child labor? The effect of a free compulsory education reform in rural China', *Journal of Population Economics*, 33(2), pp. 601–631.

Tian, X., Gong, J. and Zhai, Z. (2022) 'Natural disasters and human capital accumulation: Evidence from the 1976 Tangshan earthquake', *Economics of Education Review*, 90(c), pp. 1–14.

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

Tsai, W.J., Liu, J.T., Chou, S.Y. and Thornton, R. (2009) 'Does educational expansion encourage female workforce participation? A study of the 1968 reform in Taiwan', *Economics of Education Review*, 28(6), pp. 750–758.

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

UIS (2022) *Unesco Institute for Statistics Education and Literacy Data*. Available at: https://uis.unesco.org/.

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

Unterhalter, E., North, A., Arnot, M., Lloyd, C., Moletsane, L., Murphy-Graham, E., Parkes, J. and Saito, M. (2014) *Interventions to enhance girls' education and gender equality. Education rigorous literature review*. London, UK. Available at: http://r4d.dfid.gov.uk/andtheEPPI-Centrewebsite:http://eppi.ioe.ac.uk/.

Utomo, A. and McDonald, P. (2016) 'Who marries whom?: Ethnicity and marriage pairing patterns in Indonesia', *Asian Population Studies*, 12(1), pp. 28–49.

Utomo, A.J. (2019) 'Love in the melting pot : ethnic intermarriage in Jakarta', *Journal of Ethnic and Migration Studies*, 46(14), pp. 1–18.

Utomo, A.J. and McDonald, P. (2020) ‘Internal migration , group size , and ethnic endogamy in Indonesia’, *Geographical Research*, 59(1), pp. 56–77.

Valero, A. and Van Reenen, J. (2019) ‘The economic impact of universities: Evidence from across the globe’, *Economics of Education Review*, 68(January 2018), pp. 53–67.

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.

Wantchekon, L., Klašnja, M. and Novta, N. (2015) ‘EDUCATION AND HUMAN CAPITAL EXTERNALITIES : EVIDENCE FROM COLONIAL BENIN’, *The Quarterly Journal of Economics*, 130(2), pp. 703–757.

Welch, A.R. (2007) ‘Blurred vision?: Public and private higher education in Indonesia’, *Higher Education*, 54(5), pp. 665–687.

Westphal, M., Kamhöfer, D.A. and Schmitz, H. (2022) ‘Marginal college wage premiums under selection into employment’, *The Economic Journal*, 132(646), pp. 2231–2272.

Wilkinson, T.J., Ali, A.N., Bell, C.J., Carter, F.A., Frampton, C.M. and Mckenzie, J.M. (2013) ‘The impact of learning environment disruption on medical student performance’, *Medical Education*, 47(2), pp. 210–213.

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 (1993) *The East Asian miracle : Economic growth and public policy*. Oxford University Press.

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 2007*. 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.

World Bank (2019) *The Promise of Education in Indonesia: Consultation Eduction (Overview)*. Washington, D.C.

World Bank (2022a) *Lower secondary completion rate, total (% of relevant age group) - Indonesia | Data*. Available at: https://data.worldbank.org/indicator/SE.SEC.CMPT.LO.ZS?locations=ID.

World Bank (2022b) *Primary completion rate, total (% of relevant age group) | Data*. Available at: https://data.worldbank.org/indicator/SE.PRM.CMPT.ZS.

World Bank (2023) *Mean age at first marriage, Gender Data Portal*. Available at: https://genderdata.worldbank.org/indicators/sp-dyn-smam/.

Yeom, M., Acedo, C., Utomo, E. and Yeom, M. (2002) 'The reform of secondary education in indonesia during the 1990s: Basic education expansion and quality improvement through curriculum decentralization', *Asia Pacific Education Review*, 3(1), pp. 56–68.

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.