uc3m | Universidad Carlos III de Madrid

Master in Economic Analysis

2022-2023

Master Thesis

The Effect of Losing a Parent on Children's Educational Development

Sergio Blanco Piñeiro

Warn N. Lekfuangfu

Madrid 2023

SUMMARY

This research examines how the loss of a parent affects children's academic performance in developing countries such as Ethiopia, Peru, Vietnam, and India. Due to limited data, the study uses various methods to reveal causal relationships and their differences among gender and age groups. The study focuses on standardized test scores and uses a staggered difference-in-difference methodology to determine causality. However, the primary objective is to uncover how the effects of parental loss vary by gender and age. The study shows that there are significant gender differences and emphasizes the importance of the timing of parental loss. Future research directions include investigating the impact of parental education levels and exploring the influence of parental loss on educational attainment. This research is a valuable contribution to the literature on parental absence and its consequences, providing insights into education and economic development policies.

Keywords: Education, achievement, parental loss, gender, development, staggered difference-in-difference.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This master's thesis will investigate the impact of losing a parent on children's educational outcomes. The study will specifically analyze the effects of this loss in Ethiopia, Peru, Vietnam, and India, which are the developing countries that the study will focus on. Therefore, the causal channel studied in this paper can also be understood as a part of a much broader literature in education and development economics. By exploring the consequences of parental loss on children's educational outcomes in the countries previously cited, this research contributes to the existing literature and exposes the implications for education and economic development.

The objective of this research is to focus on one of the two crucial concepts that can effectively define a child's educational development during the initial stages of their educational life. These two topics are educational achievement and educational attainment. The distinction between these two concepts lies, following Blanden et al. (2022), in the fact that educational achievement refers to the child's academic performance, which is often measured through standardized tests. On the other hand, educational attainment involves focusing on the highest level of education attained by the child, reflecting the overall educational progress. By concentrating on the causal effect of losing a parent on educational achievement, this study aims to gain valuable insights into how this loss will be reflected in a child's cognitive skills.

This study holds significant importance as it aims to provide a comprehensive response to the previously mentioned research question. Therefore, it has profound policy implications. It can be hypothesized that there exists a negative impact of losing a parent on children's educational achievement. The crucial aspect lies not only in causally addressing this proposition but also in assessing how the extent of this loss varies based on an individual's life stage and gender.

To achieve this, the study utilizes longitudinal data, accounting for the dynamic nature of this effect over time. By adopting such an approach, the research shed light on the mechanisms through which policymakers can intervene to prevent declines in potential educational outcomes. This understanding of the temporal and individual dimensions enables the policymaker to implement relevant policy decisions.

This study aligns with a broader body of literature concerning parental absence and its impact on child outcomes. Parental absence can be manifested through various circumstances, such as parental death, which is the specific focus of this research, or it can also be explored in the context of parental migration. In the study of parental absence, the outcomes can vary across different dimensions of a child's life, as Lang and Zagorsky (2001) does for several outcomes such as achievement, attainment, marital status, annual wages, etc.

An essential benchmark of this study involves analyzing whether gender-based redistributive effects result from parental absence and whether these effects differ across different forms of parental absence. In the context of parental migration, several studies, such as Antman (2011, 2012), offer evidence of positive gender redistributive effects, particularly benefiting females.

Similarly, research by Sandefur, Wells, et al. (1997) employs a comparable methodology to the aforementioned studies. To establish causal effects within a longitudinal framework, Sandefur, Wells, et al. (1997) employ sibling fixed effects. However, the distinction lies in that this research focuses on the impact of family structure on educational attainment, revealing statistically significant although modest effects. This comparison provides context for the current study's objectives and sheds light on the broader academic environment investigating the effects of gender redistribution related to parental absence.

Analyzing research in the parental migration field, a seminal study by Booth (1995) stands out. This study is not only significant for being one of the pioneering works investigating the connection between parental migration and child achievement, but it also plays a crucial role in the analysis of formal schooling in developing countries, focusing on Swaziland as the study's context. Consequently, this research, along with the work of Lives (2020), illuminates the discussion in this paper about the comparability of test scores across diverse developing nations.

It is important to acknowledge specific research studies from a methodological perspective. One such study, conducted by Gertler et al. (2004), is similar to the current study in that it examines the impact of parental death on school enrollment with a focus on gender, differentiating between males and females. The researchers employ both parametric and semi-parametric matching techniques to analyze the causal effect of parental loss on school enrollment.

In a similar manner, papers like Lekfuangfu et al. (2018) and Todd and Wolpin (2003, 2007) explore a comprehensive characterization of the assumptions necessary for estimating the production function related to cognitive achievement. These papers are part of a much broader literature on early childhood development (ECD) and the education production function (EPF). These studies not only define the research questions within the context of a longitudinal setting but also employ panel data models to address them effectively. Likewise, Koedel et al. (2015) undertake a thorough review of the existing literature concerning value-added models, shedding light on both the merits and drawbacks of this methodology.

Additionally, related to the methodology that will be applied in this study, some papers deserve to be mentioned. In the analytical dataset subsection of this paper, a staggered diff-in-diff strategy will be employed for the identification of causal effects. In this case, the main references are Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020) for the application of the methodology and the utilization of a Doubly Robust estimator.

The gender-age analysis shows that there exists a negative association between the loss of a parent and standardized test scores. From this part of the analysis, a long-lasting effect is observed for males suffering from a father's death. For females instead, this dynamic effect is not observed but results are in line with what is hypothesized in Kailaheimo-Lönnqvist and Erola (2020). It is difficult to draw conclusions from the staggered difference-in-differences analysis due to data limitations. However, a negative effect exists even in the small analytical dataset.

The paper is structured as follows: Section 2 delves into the data sources used for the study. Section 3 describes the identification strategy. Section 4 presents the main findings of the research. Finally, Section 5 provides a summary of the key findings that have emerged from this investigation.

2. DATA

In this study, the primary source of data is The Young Lives datasets, which will provide comprehensive information on children, that constitute the individual identification of this study, across four developing countries mentioned earlier. These datasets cover data from five different waves, offering a longitudinal set-up. The information within the datasets comprises various covariates that can be further categorized into individual and family characteristics. To characterize this data set it must be said that the number of observations is almost 55,600 and the number of variables is around 360, after some modifications.

This dataset possesses a crucial characteristic that will play a fundamental role in this study. By stratifying the set of children into those belonging to the young cohort and those from the old cohort, a wider range of ages can be effectively categorized. As can be seen in Figures A1 and A2, the age range within the young cohort spans from 0 to nearly 17 years. In turn, the age range for the old cohort extends from approximately 7 years to nearly 26 years.

Following the analysis of age distribution in both cohorts, it is crucial to examine the average age of individuals within each cohort for each wave. This step is essential because it ensures a valid comparison of the educational outcomes in the analysis. Hence, Table A1 is fundamental in what follows.

Among the covariates mentioned earlier, a subset of variables will serve as outcomes, particularly related to educational achievement. Additionally, we will utilize two existing variables that indicate the year of a mother's or father's passing away to create a dummy variable representing family loss, that will serve as the treatment.

As shown in Table A1, the average age difference between individuals in the young and old cohorts is 7 years. This will give a sense of dynamics to this study. Distinguishing between the three educational achievement outcomes is fundamental. These outcomes include a math test score, a Peabody Picture Vocabulary Test (PPVT) score, and a reading score. Furthermore, it is crucial to consider the insights presented in Table A2, which provides valuable information on the timing and access to the different outcomes. For a further examination of the results check Revollo and Scott (2022).

Something that must be pointed out is that the age interval during which the scores of each cohort are observed is (10, 15.7), (13,5, 17,3), (17.4, 22.7) years old for the old cohort, while the average age of each round is 12, 15 and 19, respectively. Whereas is (6.5, 11.5), (11.1, 13.7), (14, 16) for the young cohort, when the average age of each round is 8, 12, and 15, respectively. It should be noted that the age intervals selected are the same for all cohorts in this study and will be further explained. In Table A3, these age intervals are presented with the % of treated individuals from each cohort at each age.

To ensure that the results from different tests conducted across countries and rounds are comparable, an equating process was employed. This involved using a 3-parameter item response theory (IRT) model to generate standardized scores for the Reading and PPVT scores from the Young Lives datasets. For Maths scores, a 2-parameter model was used. The standardized scores align the outcomes within the range of values indicated in Table A4, providing a standardized measure that contributes to the comparability of the coefficients. It is worth noting that this process was emphasized by Revollo and Scott (2022).

In subsequent regression analyses, a set of covariates will be used. One significant challenge of this dataset is the notable presence of missing values. To deal with this issue, mean imputation will be adopted as a strategy. This technique entails substituting missing values with the mean value of the respective covariate.

3. EMPIRICAL METHODOLOGY

This section primarily centers on the empirical methods employed to address the question of whether losing a parent impacts a child's educational outcomes. Initially, the key emphasis will be on examining how the experience of parental loss influences child educational achievement across varying age groups. Additionally, the analysis will focus on investigating potential divergent effects by gender, as Antman (2011, 2012) claims. Finally, two subsamples of the original data will be used to apply Callaway and Sant'Anna (2021) approach. These two subsamples will be based on individuals belonging to either the young or old cohorts.

3.1. Gender-Age analysis

In this subsection, the goal is to carry out an analysis where it can be unrevealed the difference, if there is any, that individuals will face depending on their sex as well as the gender of the deceased parent. Furthermore, it is also essential in this study to deal with the dynamic effect that losing a parent might have on individuals. In this subsection Table A1 and Table A2, previously mentioned become fundamental as dynamics are handled in a very specific way in this research.

Initially, this analysis will concurrently employ both approaches. Furthermore, the study will partition the sample into two distinct subgroups, the young cohort and the old cohort. Since this research employs both approaches simultaneously, assimilating the estimations of these regressions will be clarified through the use of a more detailed explanation.

In this investigation, the outcomes will be initially categorized into three distinct groups. Consequently, the scores of each individual in each of these three outcomes during the last assessment of their achievements will be utilized as dependent variables. Concerning Table A2, it can be confirmed that the average ages for the young and old cohorts' outcomes will be 15 and 19 years old, respectively.

By consulting Table A1, it can be determined that all three tests will be applied to the young cohort, whereas the old cohort will be subject only to Math and Reading assessments. Finally, this methodology will incorporate the following age intervals for the timing of the treatment [0,2], (2,6], (6,11], (11,14] for the young cohort and (6,11], (11,14], (14, 17], (17,23) for the old cohort. This study examines the impact of losing a parent during common age intervals (6,11), (11,14) on individuals who are currently aged 15 and 19 on average. The dynamics of this effect vary depending on the cohort.

Therefore, it can be stated that the estimation strategy is based on the following equation:

$$y^{j}_{igc} = \theta_0 + \sum_{a^c=1}^{a^c=4} \theta_{a^c} 1\{A\}_{igc} + \mu \mathbf{X}_{ic} + \varepsilon_{igc} \tag{3.1}$$

where, for each individual i that belongs to cohort c and whose gender is $g = \{$ Male, Female $\}$, y^j is the educational outcome, where the index $j = \{$ Math, PPVT, Reading $\}$. Moreover, as explained before, the gender analysis is not only focused on the outcome but also on the treatment. Hence, it can define $1\{A\}_{igc} = 1\{\text{death g parent at } a^c\}_{igc}$ which equals one if the male/female parent died at age interval a^c for the individual i whose gender is g and that belongs to cohort c. In the case where both parents have passed away within the five waves of this analysis, only the initial parental death will be considered.

In this equation, there exist four age intervals ranging from $a^c = 1$ to $a^c = 4$, which, as can be observed, depend on $c = \{$ Young Cohort, Old Cohort $\}$. Consequently, these age intervals, specifically $a^c = 1$ to $a^c = 4$, will be contingent upon whether the individual belongs to the young cohort or the old cohort. The age intervals are the ones previously mentioned.

It is equally vital to unpack the content of the covariate vector $\mathbf{X}_{ic}$. In this context, following Lekfuangfu and Lordan (2023) approach, the covariate vector encompasses a range of control variables and a collection of indicator variables that delineate which of the previous controls have undergone imputation. This methodology is referred to as the Missing Indicator method. To apply this method the covariate vector must consist of continuous variables. In the context of this particular study, variables like BMI, travel time to school in minutes, time spent playing, time spent working at home, and similar attributes are part of the covariate vector.

Based on this estimation strategy, it can be readily ensured that a total of 8 regression equations, with covariate control will be established for both the Math and Reading tests. However, exclusively for the young cohort, 4 regressions will be feasible for the PPVT test. To summarize, these equations distinguish between father's death among both male and female individuals within both the young and old cohorts, as well as mother's death in both male and female individuals within both cohorts.

3.2. An extension analysis: a Staggerred Diff and Diff approach

In this subsection, the goal is to employ the methodology proposed by Callaway and Sant'Anna (2021). Since the aim is to determine the causal impact of parental loss on child educational outcomes, certain challenges arise from the dataset that need to be addressed and solved. Similar to the approach followed in the whole study, the primary partitioning of the sample will involve creating two distinct subsamples: one for the young cohort and another for the old cohort.

Consequently, after the initial formation of the earlier stated subsamples, the first challenge that requires attention is the necessity for further limiting the subsamples. This is because test scores are only available for the young cohort when individuals average 8, 12, and 15 years old, and for the old cohort when individuals average 12, 15, and 19 years old. Following Tables A1 and A2, this analytical dataset needs to be restricted to rounds 3, 4, and 5 for the young cohort, and rounds 2, 3, and 4 for the old cohort.

Building upon the preceding paragraph, it can be affirmed that the central concept of this analytical dataset is to investigate variations across the three previously mentioned rounds. In a way, this variation is constrained to a narrower subset due to the absence of scores in the initial waves. In the context of identifying the causal effect of the treatment, always treated individuals must be removed from these analytical subsets.

Table B5 and Table B6 provide insights into the observation that the percentage of treated individuals does not exceed 2% in any given round. This outcome arises from the necessity to apply Callaway and Sant'Anna (2021) to this specific dataset, requiring certain modifications, as previously discussed. Consequently, the underlying concept of this empirical approach is to shed light on the causal effect of this treatment, taking into account the data challenges.

In essence, this methodology aims to improve the identification of causal effects compared to the basic gender-age analysis, thus contributing to a more refined understanding of the research question.

The gender-age approach in math and reading scores is done for both cohorts, but the analysis of PPVT scores is only done for the young cohort due to data requirements. Similarly, the staggered difference-in-differences methodology is used for math scores across both cohorts. This methodology can also be extended to the PPVT scores for the young cohort.

An essential aspect of any difference-in-differences approach is to define the data's individual and time dimensions, if applicable. In this context, the individual dimension is straightforward, as it is closely associated with child observations. However, complications arise when delineating the time dimension since time is not construed conventionally in this analysis. In this study, the time dimension is determined by the variable indicating the specific round to which each observation belongs. Consequently, there will be three distinct periods, as can be observed in Tables B5 and B6, during which the initial period no individual is subjected to the treatment.

Following the notation of Callaway and Sant'Anna (2021), a staggered difference-in-differences approach requires the definition of a group variable. In this scenario, the group variable's nature will fluctuate based on whether an individual falls within the young or old cohort. In both instances, this group variable will assume a binary form, indicating the period during which the individual is initially treated, if indeed they are treated at any point in time.

4. EMPIRICAL RESULTS

This section will primarily deliver the empirical findings, providing essential insights to determine whether parental death has a detrimental impact on children's educational achievement. In the first subsection, descriptive statistics are presented. Subsequently, the empirical outcomes of the gender-age analysis will be showcased. Lastly, the results of the staggered difference-in-differences analysis for the analytical subsample will be uncovered within the concluding subsection.

4.1. Descriptive Statistics

Within this subsection, the primary objective is to examine certain descriptive statistics about the confounding vector in terms of treatment. The aim is to assess whether the covariate vector is equally distributed between treated and untreated individuals. To achieve this, the study will not only present summary statistics based on treatment but will also provide balancing tests for the controls. This approach ensures a comprehensive evaluation of the covariate distribution and helps determine the success of the treatment allocation process.

To start, Tables B1 and B2 display the covariate vector employed in this study. As previously noted, this confounding vector encompasses information concerning individual and family characteristics. Notably, within this covariate set, certain variables hold particular significance within the educational literature, such as study hours and other variables used to control for potential anticipation of the treatment effect as the shock-illness set of characteristics.

Upon reviewing these tables, a trend emerges where the average values of these variables appear to be not comparable. It's important to underline that this observation is not a statistically validated result. Thus, to ensure no comparability, further balancing tests are necessary.

As evident from Table B3, the null hypothesis of equal means between treated and non-treated groups can be refuted for each continuous covariate, except the Child's age at the start of grade 1 and Hours/day spent in leisure activities. This highlights the significance of incorporating controls within the estimation strategy to account for these differences in the gender-age analysis. The inclusion of these explanatory variables is pivotal for accurately controlling these variations and ensuring the validity of the analysis.

In the preceding balancing test, the null hypothesis being examined is whether the average values of the treated and untreated groups of individuals are equivalent, rather than examining whether the distributions of these two groups are identical or divergent. In order to improve the reliability of the statistical analysis, additional tests should be conducted. This will involve the utilization of non-parametric tests such as the Kolmogorov-Smirnov test for equality of distributions, along with the method introduced by Goldman and Kaplan (2018) that overcomes Kolmogorov-Smirnov limitations. This broader array of tests will thoroughly assess the distributional disparities or similarities between the treated and untreated groups. As can be seen in Table B4 K-S null hypothesis is rejected for every covariate. The same occurs with Goldman and Kaplan (2018) global test of equality of two CDFs that is rejected for $\alpha = 10\%, \alpha = 5\%, \alpha = 1\%$.

4.2. Gender-Age analysis

In this section, the outcomes of the gender-age analysis will be presented and examined. Initially, this analysis will be outlined for Math and Reading scores, as these two tests encapsulate the entire age range within the sample employed in this study. Subsequently, the outcomes for the PPVT score will be exposed, particularly focusing on the young cohort since it is the only cohort that provides scores for this test in the latest comparable wave. This approach ensures a comprehensive understanding of the gender-age analysis across various dimensions.

4.2.1. Standardized Math scores

As previously mentioned, the final goal of this section is to discuss gender-age regression estimates. To achieve this, a specific sequence of analyses for each outcome will be followed. First, a simple treatment analysis for family loss will be conducted. Next, a gender analysis will be performed. Then, this process will be repeated for age, considering different age intervals as treatment groups. Finally, once these preliminary analyses are completed, a comprehensive gender-age analysis will be conducted. Tables 4.1 to 4.4 present the results in the previously mentioned order for the standardized math score.

	(1)	(2)	(3)	(4)
Panel A				
Family loss	-0.31***	-0.24***	-0.43***	-0.38***
	(0.08)	(0.07)	(0.09)	(0.09)
Constant	0.02	0.49*	0.01	1.61***
	(0.02)	(0.25)	(0.02)	(0.30)
N	3692	3688	3412	3410
Covariates	No	Yes	No	Yes
Panel B				
Family loss	-0.41***	-0.33***	-0.49***	-0.39***
	(0.11)	(0.10)	(0.12)	(0.11)
Constant	0.37***	0.87*	0.21***	2.17***
	(0.03)	(0.35)	(0.03)	(0.34)
N	1580	1579	1552	1551
Covariates	No	Yes	No	Yes

*Notes: The table displays data on the impact of losing a family member on standardized math scores using a basic OLS specification. Panel A shows results for the younger cohort, while Panel B displays results for the older cohort. Columns (1) and (2) present data for males (with and without controls), and columns (3) and (4) present data for females (with and without controls). Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, * for p<0.01, and ** for p<0.001*

Table 4.1

Standardized math scores over simple treatment

Based on the data presented in Table 4.1 and after accounting for other factors, it is clear that losing a family member is associated with a negative impact on standardized math scores. Even in this simplified model, the effect remains statistically significant. Upon closer examination of columns (2) and (4), it can be seen that the negative impact is greater for females than for males. Additionally, it is worth noting that for males, the negative impact increases from 0.24 standard deviations to 0.33 standard deviations if individuals are on average 4 years older when taking the test. However, this age-related effect is not as evident for females.

The analysis in Table 4.2 focuses on gender-specific results, which provide important insights. It is worth noting that losing a father has a nearly double impact on females compared to males when individuals are around 15 years old (as seen in columns (1) and (2)). However, the gender gap narrows as individuals age increases. As in Panel B, the gender gap disappears. When examining columns (3) and (4), it becomes apparent that females' math scores are, on average, 0.31 standard deviations lower when they experience their mother's loss at the age of 15. There is no discernible effect on males. As individuals are evaluated at older ages, females show no significant effect, while experiencing a mother's loss is associated with a reduction of 0.46 standard deviations in males' math scores.

	(1)	(2)	(3)	(4)
Panel A				
Family loss*Parental gender	-0.26**	-0.45***	-0.14	-0.31*
	(0.08)	(0.10)	(0.10)	(0.15)
Constant	0.49*	1.61***	0.47	1.60***
	(0.25)	(0.30)	(0.25)	(0.31)
N	3688	3410	3688	3410
Panel B				
Family loss*Parental gender	-0.36**	-0.38**	-0.46**	-0.32
	(0.11)	(0.13)	(0.17)	(0.18)
Constant	0.88*	2.17***	0.84*	2.18***
	(0.35)	(0.35)	(0.35)	(0.34)
N	1579	1551	1579	1551

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss on standardized math scores, distinguishing between the loss of father and mother. Panel A presents results for the young cohort, while Panel B does it for the old cohort. Columns (1) and (2) present results for males and females, respectively, of father loss, while columns (3) and (4) present results for males and females, respectively, of mother loss. All specifications include covariates. Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.2

Standardized math scores over gender treatment

In Table 4.3, which presents age-specific results, several noteworthy observations can be made, particularly by examining columns (2) and (4). For males, it becomes evident that the effect of losing a parent is significant only within the age intervals of (2,6] and (6, 11] when individuals have an average age of 15. However, when individuals' average age increases to 19, the significance is retained only for the age interval (6, 11], and the associated negative effect becomes more pronounced. Specifically, the reduction in math scores shifts from 0.25 standard deviations for younger individuals to a substantial 0.66 standard deviations for older ones. This indicates a negative dynamic effect linked to parental loss. Nevertheless, when considering females and analyzing column (4), it is apparent that the earlier-mentioned effect does not hold. In Panel A, column (4), it is important to note that a decrease of 0.65 standard deviations in the standardized math score is linked to the loss of a parent during the age range of [0,2] years old. This supports the childhood hypothesis [1].

[1] Following Kailaheimo-Lönnqvist and Erola (2020) the childhood hypothesis states that parental death is more harmful the younger the child is.

	(1)	(2)	(3)	(4)
Panel A				
[0,2]	-0.43	-0.17	-0.56**	-0.65***
	(0.23)	(0.19)	(0.18)	(0.18)
(2,6]	-0.33*	-0.35***	-0.45**	-0.23
	(0.13)	(0.10)	(0.15)	(0.13)
(6, 11]	-0.38***	-0.25**	-0.34*	-0.40**
	(0.10)	(0.09)	(0.15)	(0.15)
(11, 14]	0.21	0.11	-0.77***	-0.57**
	(0.33)	(0.29)	(0.18)	(0.20)
Constant	0.02	0.49*	0.01	1.62***
	(0.02)	(0.25)	(0.02)	(0.30)
N	3692	3688	3412	3410
Covariates	No	Yes	No	Yes
Panel B				
(6, 11]	-0.77***	-0.66***	-0.46**	-0.36*
	(0.13)	(0.14)	(0.16)	(0.18)
(11, 14]	-0.23	-0.26	-0.49*	-0.28
	(0.20)	(0.16)	(0.20)	(0.19)
(14, 17]	-0.17	-0.07	-0.67**	-0.60*
	(0.22)	(0.20)	(0.23)	(0.23)
(17, 23)	-0.44	-0.34	-0.28	-0.31
	(0.25)	(0.21)	(0.23)	(0.21)
Constant	0.38***	0.83*	0.22***	2.16***
	(0.03)	(0.36)	(0.03)	(0.34)
N	1580	1579	1552	1551
Covariates	No	Yes	No	Yes

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss across different age intervals on standardized math scores. Panel A presents results for the young cohort, while Panel B does it for the old cohort. Columns (1) and (2) present data for males (with and without controls), and columns (3) and (4) present data for females (with and without controls). Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.3

Standardized math scores over age treatment

	(1)	(2)	(3)	(4)
Panel A				
[0,2]	-0.34	-0.48*	0.09	-0.98***
	(0.23)	(0.22)	(0.29)	(0.22)
(2,6]	-0.32**	-0.34*	-0.30	0.06
	(0.12)	(0.14)	(0.20)	(0.22)
(6, 11]	-0.31*	-0.52**	-0.10	-0.45
	(0.13)	(0.17)	(0.10)	(0.25)
(11, 14]	0.10	-0.65**	-0.59***	-0.43
	(0.26)	(0.24)	(0.04)	(0.32)
Constant	0.48	1.61***	0.48	1.61***
	(0.25)	(0.30)	(0.25)	(0.31)
N	3688	3410	3688	3410
Panel B				
(6, 11]	-0.69***	-0.24	-0.62*	-0.48
	(0.15)	(0.18)	(0.25)	(0.34)
(11, 14]	-0.33	-0.15	-0.40*	-0.59*
	(0.20)	(0.24)	(0.20)	(0.24)
(14, 17]	-0.22	-0.84***	-0.08	0.41
	(0.23)	(0.25)	(0.57)	(0.39)
(17, 23)	-0.27	-0.30	-0.99***	-0.28
	(0.22)	(0.28)	(0.29)	(0.21)
Constant	0.86*	2.16***	0.89*	2.17***
	(0.35)	(0.34)	(0.35)	(0.35)
N	1579	1551	1579	1551

*Notes: The effect of family loss across different age intervals on standardized math scores, distinguishing between the loss of father and mother. Panel A presents results for the young cohort, while Panel B does it for the old cohort. Columns (1) and (3) show data for males who experienced father loss and mother loss, respectively. Columns (2) and (4) show data for females who experienced father loss and mother loss, respectively. All specifications include covariates. Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.4

Standardized math scores over gender-age treatment

Finally, in Table 4.4, the gender-age analysis is conducted. Initially, it must be examined the effect of losing a father for males and females in columns (1) and (2), respectively. After analyzing the data in column (1), it is evident that males who experienced the loss of their fathers at the age interval (6, 11] suffer from a long-lasting negative impact. Furthermore, this negative effect is not only prevalent but also gender-biased, with females in the same age group exhibiting almost 50% less of this negative impact. It should be noted that, after testing, losing a parent does not have a significantly different effect on males within the age intervals (2, 6] and (6, 11]. Turning to column (2), it becomes clear that the previously mentioned dynamic effect observed for males does not apply to females.

Comparing columns (2) and (4) of Panel A it can be ensured that the gender of deceased parent and child's age hypothesis [2] is present for our data on females. Losing a mother at the age interval [0,2] is associated with a reduction of almost one standard deviation in the standardized math scores. On the other hand, the negative effect of losing a father is more pronounced for the age intervals (6, 11] and (11, 14].

4.2.2. Standardized Reading scores

In this analysis, we will examine the effects of parental loss on standardized reading scores. This assessment follows a similar approach to the one employed in the previous section. Thus, we present Table 4.5 to illustrate the analysis of the simple treatment impact. When scrutinizing columns (2) and (4), it becomes apparent that losing a parent is linked to a decline of 0.19 and 0.28 standard deviations in standardized reading scores for males and females, respectively. However, it is important to note that this effect is only statistically significant when individuals have an average age of 15 years.

The results from the regressions in Table 4.6, the gender-specific analysis, indicate that only columns (1) and (2) for Panel A show statistically significant effects. These results suggest that the negative impact of losing a father is more substantial for females than for males. To be precise, losing a father is linked to a decrease of 0.34 standard deviations for females, whereas for males, it is only a reduction of 0.22 standard deviations.

Once the gender-specific analysis has been conducted, the subsequent step is to analyze the effects of age treatment, which can be found in Table 4.7. It is essential to concentrate on columns (2) and (4) in this table since the covariates have a significant impact on this research. In Panel A, column (2) reveals that, upon testing, there is no significant difference between the effects of the treatment on the age intervals (2,6] and (6, 11]. It's worth noting that a similar testing procedure was applied to assess whether the differences between the age intervals (11,14] and (17, 23) in column (4) of Panel B are significant, and in this case, the null hypothesis of equal effects can be rejected.

[2]Following Kailaheimo-Lönnqvist and Erola (2020) the gender of deceased parent and child's age hypothesis states that the death of a mother during childhood is more adverse than the death of a father, while the death of a father is more adverse during adolescence than maternal death.

	(1)	(2)	(3)	(4)
Panel A				
Family loss	-0.28**	-0.19*	-0.26**	-0.28**
	(0.09)	(0.09)	(0.10)	(0.10)
Constant	-0.06***	0.27	0.08***	0.25
	(0.02)	(0.28)	(0.02)	(0.36)
N	3260	3257	3067	3065
Covariates	No	Yes	No	Yes
Panel B				
Family loss	-0.05	0.05	-0.21	-0.11
	(0.12)	(0.10)	(0.12)	(0.12)
Constant	0.42***	0.58	0.42***	1.24***
	(0.03)	(0.42)	(0.03)	(0.33)
N	1461	1460	1464	1463
Covariates	No	Yes	No	Yes

*Notes: The table displays data on the impact of losing a family member on standardized reading scores using a basic OLS specification. Panel A shows results for the younger cohort, while Panel B displays results for the older cohort. Columns (1) and (2) present data for males (with and without controls), and columns (3) and (4) present data for females (with and without controls). Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.5

Standardized reading scores over simple treatment

	(1)	(2)	(3)	(4)
Panel A				
Family loss*Parental gender	-0.22*	-0.34**	-0.21	-0.17
	(0.10)	(0.11)	(0.19)	(0.18)
Constant	0.27	0.25	0.26	0.24
	(0.28)	(0.36)	(0.28)	(0.36)
N	3257	3065	3257	3065
Panel B				
Family loss*Parental gender	0.08	-0.25	-0.13	0.19
	(0.11)	(0.13)	(0.22)	(0.19)
Constant	0.58	1.24***	0.58	1.25***
	(0.42)	(0.33)	(0.42)	(0.33)
N	1460	1463	1460	1463

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss on standardized reading scores, distinguishing between the loss of father and mother. Panel A presents results for the young cohort, while Panel B does it for the old cohort. Columns (1) and (2) present results for males and females, respectively, of father loss, while columns (3) and (4) present results for males and females, respectively, of mother loss. All specifications include covariates. Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.6

Standardized reading scores over gender treatment

	(1)	(2)	(3)	(4)
Panel A				
[0,2]	0.14	0.27	-0.23	-0.26
	(0.20)	(0.19)	(0.27)	(0.25)
(2,6]	-0.27*	-0.28*	-0.13	-0.10
	(0.13)	(0.14)	(0.15)	(0.15)
(6, 11]	-0.53**	-0.34*	-0.33**	-0.40**
	(0.16)	(0.16)	(0.13)	(0.13)
(11, 14]	0.23	0.18	-0.87	-0.70
	(0.20)	(0.21)	(0.72)	(0.76)
Constant	-0.06***	0.27	0.08***	0.26
	(0.02)	(0.28)	(0.02)	(0.36)
N	3260	3257	3067	3065
Covariates	No	Yes	No	Yes
Panel B				
(6, 11]	-0.23	-0.18	-0.25	-0.08
	(0.15)	(0.14)	(0.25)	(0.27)
(11, 14]	0.26	0.17	0.24	0.40***
	(0.22)	(0.16)	(0.15)	(0.11)
(14, 17]	-0.14	0.07	-0.36	-0.26
	(0.25)	(0.23)	(0.23)	(0.21)
(17, 23)	-0.25	-0.02	-0.42*	-0.43*
	(0.22)	(0.20)	(0.20)	(0.17)
Constant	0.42***	0.55	0.42***	1.27***
	(0.03)	(0.42)	(0.03)	(0.33)
N	1461	1460	1464	1463
Covariates	No	Yes	No	Yes

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss across different age intervals on standardized reading scores. Panel A presents results for the young cohort, while Panel B does it for the old cohort. Columns (1) and (2) present data for males (with and without controls), and columns (3) and (4) present data for females (with and without controls). Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.7

Standardized reading scores over age treatment

	(1)	(2)	(3)	(4)
Panel A				
[0,2]	0.19	-0.35	0.38	-0.10
	(0.23)	(0.35)	(0.32)	(0.31)
(2,6]	-0.31*	-0.19	-0.25	0.27
	(0.14)	(0.17)	(0.32)	(0.25)
(6, 11]	-0.37*	-0.46**	-0.35	-0.49
	(0.17)	(0.15)	(0.29)	(0.26)
(11, 14]	0.07	-0.39	-0.14**	-0.90
	(0.22)	(0.94)	(0.05)	(1.09)
Constant	0.27	0.25	0.25	0.25
	(0.28)	(0.36)	(0.28)	(0.36)
N	3257	3065	3257	3065
Covariates	Yes	Yes	Yes	Yes
Panel B				
(6, 11]	-0.15	-0.36	-0.53	0.25
	(0.13)	(0.36)	(0.42)	(0.37)
(11, 14]	0.18	0.49***	0.14	0.08
	(0.18)	(0.12)	(0.23)	(0.19)
(14, 17]	0.09	-0.52*	-0.01	0.56
	(0.27)	(0.21)	(0.61)	(0.44)
(17, 23)	0.04	-0.49*	-0.54	-0.17
	(0.20)	(0.21)	(0.35)	(0.29)
Constant	0.56	1.26***	0.59	1.25***
	(0.42)	(0.33)	(0.41)	(0.33)
N	1460	1463	1460	1463
Covariates	Yes	Yes	Yes	Yes

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss across different age intervals on standardized reading scores, distinguishing between the loss of father and mother. Panel A presents results for the young cohort, while Panel B does it for the old cohort. Columns (1) and (3) show data for males who experienced father loss and mother loss, respectively. Columns (2) and (4) show data for females who experienced father loss and mother loss, respectively. All specifications include covariates. Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.8

Standardized reading scores over gender-age treatment

In the gender-age analysis, it is noteworthy that only a few parameters are significant, and these parameters are related to father loss. In Panel A, column (1), after testing, it's worth highlighting that there is no significant difference between the negative effect of losing a father in the age interval (2, 6] and losing a father during (6, 11]. On the other hand, in Panel B, column (2), after testing, it should be noted that there is no significant difference between the effects of losing a parent in the following age intervals: (11, 14], (14, 17], and (17, 23].

4.2.3. Standardized PPVT scores

The gender-age section concludes by examining the standardized PPVT scores, following a similar approach to the previous two subsections. However, it is important to note that for the PPVT score, only Panel A is available, so the analysis will be focused solely on this panel, and the dynamic interpretation will not be possible.

Upon analyzing columns (2) and (4) of Table 4.9, it becomes evident that a family loss is associated with a more pronounced reduction in standardized PPVT scores for males compared to females. Specifically, the reduction for males amounts to 0.25 standard deviations, whereas for females, it is only 0.18. Table 4.10 reveals that only the treatment variable is significant in column (3). This implies that experiencing the loss of a mother is significantly associated with a negative effect, leading to a reduction of 0.34 standard deviations in the standardized PPVT score.

	(1)	(2)	(3)	(4)
Panel A				
Family loss	-0.28**	-0.25**	-0.18	-0.18*
	(0.09)	(0.08)	(0.09)	(0.09)
Constant	0.03	-0.51	-0.01	-0.89*
	(0.02)	(0.29)	(0.02)	(0.37)
N	3586	3583	3284	3283
Covariates	No	Yes	No	Yes

*Notes: The table displays data on the impact of losing a family member on standardized PPVT scores using a basic OLS specification. Panel A shows results for the younger cohort. Columns (1) and (2) present data for males (with and without controls), and columns (3) and (4) present data for females (with and without controls). Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.9

Standardized PPVT scores over simple treatment

	(1)	(2)	(3)	(4)
Panel A				
Family loss*Parental gender	-0.15	-0.18	-0.34*	-0.22
	(0.10)	(0.11)	(0.14)	(0.14)
Constant	-0.53	-0.89*	-0.53	-0.89*
	(0.29)	(0.37)	(0.29)	(0.3)
N	3583	3283	3583	3283

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss on standardized PPVT scores, distinguishing between the loss of father and mother. Panel A presents results for the young cohort. Columns (1) and (2) present results for males and females, respectively, of father loss, while columns (3) and (4) present results for males and females, respectively, of mother loss. All specifications include covariates. Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.10

Standardized PPVT scores over gender treatment

	(1)	(2)	(3)	(4)
Panel A				
[0,2]	-0.16	-0.11	0.18	0.10
	(0.27)	(0.29)	(0.24)	(0.22)
(2,6]	-0.31	-0.33*	-0.12	-0.07
	(0.16)	(0.15)	(0.16)	(0.15)
(6, 11]	-0.42***	-0.30**	-0.27*	-0.30*
	(0.12)	(0.11)	(0.13)	(0.13)
(11, 14]	0.33	0.17	-0.90*	-0.80*
	(0.21)	(0.15)	(0.36)	(0.34)
Constant	0.03	-0.51	-0.01	-0.88*
	(0.02)	(0.29)	(0.02)	(0.37)
N	3586	3583	3284	3283
Covariates	No	Yes	No	Yes

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss across different age intervals on standardized PPVT scores. Panel A presents results for the young cohort. Columns (1) and (2) present data for males (with and without controls), and columns (3) and (4) present data for females (with and without controls). Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.11

Standardized PPVT scores over age treatment

From Table 4.11, where the age-specific analysis is conducted, columns (2) and (4) deserve particular attention. In column (2), after testing, it is evident that there is no significant difference between the treatment's effect within the age intervals (2, 6] and (6, 11]. However, when turning to columns (3) and (4), after testing, it should be noted that without covariate conditioning, there is a significant difference between the coefficients for the effect within the age intervals (6, 11] and (11, 14]. This significant difference vanishes after covariate conditioning.

Finally, based on the gender-age analysis presented in Table 4.12, it is evident that only the effect of father death on female individuals yields significant coefficients. Consequently, limited conclusions can be drawn solely from column (2). It is noteworthy that losing a father in the age intervals (6, 11] and (11, 14] is correlated with a reduction in the standardized PPVT scores of 0.34 and 0.72 standard deviations, respectively. These findings appear to support the hypothesis concerning the gender of the deceased parent and the child's age, as explained earlier.

	(1)	(2)	(3)	(4)
Panel A				
[0,2]	-0.12	0.14	-0.08	0.06
	(0.37)	(0.31)	(0.45)	(0.26)
(2,6]	-0.25	-0.02	-0.52	-0.09
	(0.18)	(0.19)	(0.27)	(0.20)
(6, 11]	-0.24	-0.34*	-0.32	-0.35
	(0.14)	(0.14)	(0.18)	(0.22)
(11, 14]	0.28	-0.72*	-0.01	-0.96
	(0.15)	(0.36)	(0.05)	(0.72)
Constant	-0.53	-0.88*	-0.52	-0.89*
	(0.29)	(0.37)	(0.29)	(0.37)
N	3583	3283	3583	3283
Covariates	Yes	Yes	Yes	Yes

*Notes: This table presents the results of a simple OLS analysis on the effect of family loss across different age intervals on standardized PPVT scores, distinguishing between the loss of father and mother. Panel A presents results for the young cohort. Columns (1) and (3) show data for males who experienced father loss and mother loss, respectively. Columns (2) and (4) show data for females who experienced father loss and mother loss, respectively. All specifications include covariates. Robust standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.12

Standardized PPVT scores over gender-age treatment

To summarize, the regression analysis reveals that the standardized math score has more significant estimates compared to the other two test scores. In contrast, most coefficients of the other two tests are no different from zero. Hence, it must be highlighted that standardized test scores are negatively affected by the loss of a family member. Males suffer from a negative dynamic effect that is not clear for females. It should be noted that the loss of a father has a greater negative impact on females compared to males. However, males experience long-lasting effects.

4.3. An extension analysis: a Staggerred Diff and Diff approach

In this section, the primary goal is to establish causality between the treatment and standardized test scores. It is essential to note that due to data limitations, this section serves as supplementary evidence to support the findings of the gender-age analysis, rather than being the primary focus of this research. Here, the results concerning the impact of family loss on standardized math and PPVT scores will be presented. It is worth emphasizing that, in this section, the time dimension is distinct from the individual's age when they first experienced the treatment; instead, it corresponds to the various rounds within the dataset.

To examine this causal effect, it is crucial to refer to Tables B4 and B4, where the staggered structure of the treatment becomes apparent. Given the relatively low number of treated individuals in certain rounds, especially for Table B4, the most appropriate policy parameter that can be derived is a simple weighted average of ATT's (Average Treatment Effects on the Treated). This approach is chosen due to the impracticality of presenting other estimates that would be meaningful and interpretable given the data constraints.

Firstly, Table 4.13 and 4.14 presents the results for the conditional and unconditional staggered difference-in-differences. It can be immediately ensured that only the analysis of Panel A for Table 4.13 can contribute to the analysis as the rest of the policy parameters are not significantly different from zero. It is important to note that losing a parent can have a significant impact on standardized math scores, resulting in a decrease of 0.23 standard deviations. However, after accounting for other factors, the reduction is lessened to only 0.17 standard deviations.

Something, that must be mentioned are Figures B1, B2, and B3. In some of these figures, it can be stated that there exists an effect of treatment that is negatively associated with each of the standardized test scores analyzed. There is an issue related to the rejection of the null hypothesis that all pre-treatments are equal to 0. Specifically, this issue pertains to the standardized PPVT for the after-covariate condition and the standardized math score for the old cohort. The null hypothesis is not rejected only for the regression analysis that presents no significant coefficient. This can be explained by the fact that there exists only one pretreatment period due to data limitations and hence the test becomes more demanding.

	(1)	(2)
Panel A		
Simple weighted average	-0.23***	-0.17**
	(0.06)	(0.06)
N	21,191	21,191
Covariates	No	Yes
Panel B		
Simple weighted average	-0.10	-0.07
	(0.10)	(0.09)
N	9,252	9,252
Covariates	No	Yes

*Note: This table presents the standardized math score treatment effects under unconditional and conditional parallel trends. Column (1) presents the unconditional version, whereas column (2) is the conditional one. In this case, the only reported parameter is the simple weighted average that as stated in Callaway and Sant'Anna (2021) is the weighted average (by group size) of all available group-time average treatment effects. Panel A shows results for the younger cohort, while Panel B displays results for the older cohort. Both estimates use the doubly robust version from Sant'Anna and Zhao (2020). Wild bootstrap standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.13

Staggered Difference-in-Difference for standardized math scores

	(1)	(2)
Panel A		
Simple weighted average	-0.08	-0.05
	(0.09)	(0.09)
N	19,736	19,736
Covariates	No	Yes

*Note: This table presents the standardized PPVT score treatment effects under unconditional and conditional parallel trends. Column (1) presents the unconditional version, whereas column (2) is the conditional one. In this case, the only reported parameter is the simple weighted average that as stated in Callaway and Sant'Anna (2021) is the weighted average (by group size) of all available group-time average treatment effects. Panel A shows results for the younger cohort, while Panel B displays results for the older cohort. Both estimates use the doubly robust version from Sant'Anna and Zhao (2020). Wild bootstrap standard errors are presented in brackets and with respect to the significant levels: * for p<0.05, ** for p<0.01, and *** for p<0.001*

Table 4.14

Staggered Difference-in-Difference for standardized math scores

5. CONCLUSIONS

This paper examines how losing a parent affects academic achievement. Due to limited data, it is difficult to establish causality and accurately identify the policy parameter of interest. To address this challenge, this research takes a multifaceted approach by examining the research question from various angles. While the study utilizes a staggered difference-in-difference methodology to address causality, it is important to emphasize that the primary focus of this research extends far beyond this aspect. Instead, the study's broader scope centers on investigating the gender-age dimension of the effects associated with losing a parent.

The analysis of gender and age highlights the important role that gender plays in this investigation. It shows that there are significant differences in the impact of losing a parent based on the individual's gender and even the gender of the parent who passed away. Comparing people who are, on average, four years apart in age brings attention to the importance of considering dynamic effects when trying to understand this research question.

Several questions are left unanswered and open for future research. One promising area for future investigation is to examine the research question from different angles. For example, distinguishing the effects of losing a parent based on their education level could reveal potential variations in the impact of parental loss. It's possible that there may be a redistributive effect in education, where individuals from lower socioeconomic backgrounds are not as affected as those whose parents are a source of cognitive achievement in the early stages of their lives. Another aspect that warrants further exploration is the impact of parental loss on educational attainment. This aspect was not directly addressed in the current research. However, future analyses could delve into this area, given the dataset's potential to shed light on how parental loss may influence an individual's educational achievements.

To sum up, this research is a significant addition to the existing literature that has mainly focused on developed countries. By exploring a well-known inquiry from a gender and age perspective in different developing nations, this study is a crucial move in improving external validity and broadening our comprehension of the intricate dynamics related to the loss of a parent and its influence on cognitive accomplishments.

BIBLIOGRAPHY

Antman, F. M. (2011). International migration and gender discrimination among children left behind. *American Economic Review*, *101*(3), 645–649.

Antman, F. M. (2012). Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics*, *25*, 1187–1214.

Blanden, J., Doepke, M., & Stuhler, J. (2022). *Educational inequality* (tech. rep.). National Bureau of Economic Research.

Booth, M. Z. (1995). Children of migrant fathers: The effects of father absence on swazi children's preparedness for school. *Comparative Education Review*, *39*(2), 195–210.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Gertler, P., Levine, D. I., & Ames, M. (2004). Schooling and parental death. *Review of Economics and Statistics*, *86*(1), 211–225.

Goldman, M., & Kaplan, D. M. (2018). Comparing distributions by multiple testing across quantiles or cdf values. *Journal of Econometrics*, *206*(1), 143–166.

Kailaheimo-Lönnqvist, S., & Erola, J. (2020). Child's age at parental death and university education. *European Societies*, *22*(4), 433–455. https://doi.org/10.1080/14616696.2020.1719179

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, *47*, 180–195. https://doi.org/https://doi.org/10.1016/j.econedurev.2015.01.006

Lang, K., & Zagorsky, J. L. (2001). Does growing up with a parent absent really hurt? *Journal of human Resources*, 253–273.

Lekfuangfu, W. N., & Lordan, G. (2023). Documenting occupational sorting by gender in the uk across three cohorts: Does a grand convergence rely on societal movements? *Empirical Economics*, *64*(5), 2215–2256.

Lekfuangfu, W. N., Powdthavee, N., Warrinnier, N., & Cornaglia, F. (2018). Locus of control and its intergenerational implications for early childhood skill formation. *The Economic Journal*, *128*(608), 298–329.

Lives, Y. (2020). Equating cognitive scores across rounds and cohorts for young lives in ethiopia, india, peru and vietnam.

Revollo, P. E., & Scott, D. (2022). Cognitive and achievement tests in the young lives study.

Sandefur, G. D., Wells, T., et al. (1997). *Using siblings to investigate the effects of family structure on educational attainment*. Institute for Research on Poverty, University of Wisconsin–Madison.

Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3–F33.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, *1*(1), 91–136.

APPENDIX A

Year	Round	Average age (years)	
		Young Cohort	Old Cohort
2002	1	1	8
2006	2	5	12
2009	3	8	15
2013	4	12	19
2016	5	15	22

Table A1
Average age of each cohort by round.

	Average age (years)	Observed outcomes
Young Cohort	8	Math/PPVT
	12	Math/PPVT/Reading
	15	Math/PPVT/Reading
Old Cohort	12	Math/PPVT
	15	Math/PPVT
	19	Math/Reading

Table A2
Outcomes by cohort and by average age.

Age intervals	[0,2]	(2,6]	(6,11]	(11,14]	(14,17]	(17,23)
Treated individuals Young Cohort (%)	0.033	0.093	1.475	0.066	NA	NA
Treated individuals Old Cohort (%)	NA	NA	1.376	0.066	0.047	0.048

Table A3
Porcentage of treated individuals by age internal and by cohort

	N	Mean	Sd	Min	Max
Math (standardize measure)	31447	0.00	1.00	-3.315805	3.739466
Read score (standardize measure)	16418	0.00	1.00	-3.664618	2.991497
PPVT score (standardize measure)	32959	0.00	1.00	-5.73088	3.607101

Table A4
Descriptive statistics of outcomes.

Figure A1

Span of ages in Young Cohort

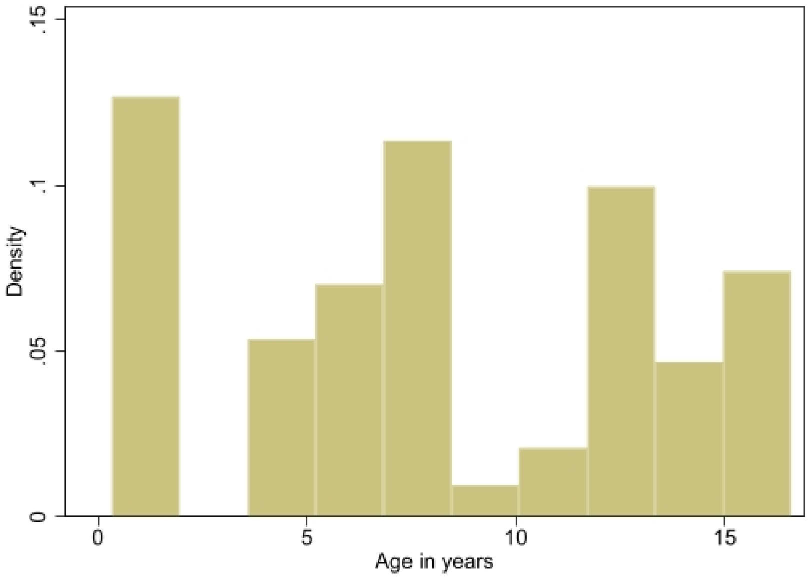
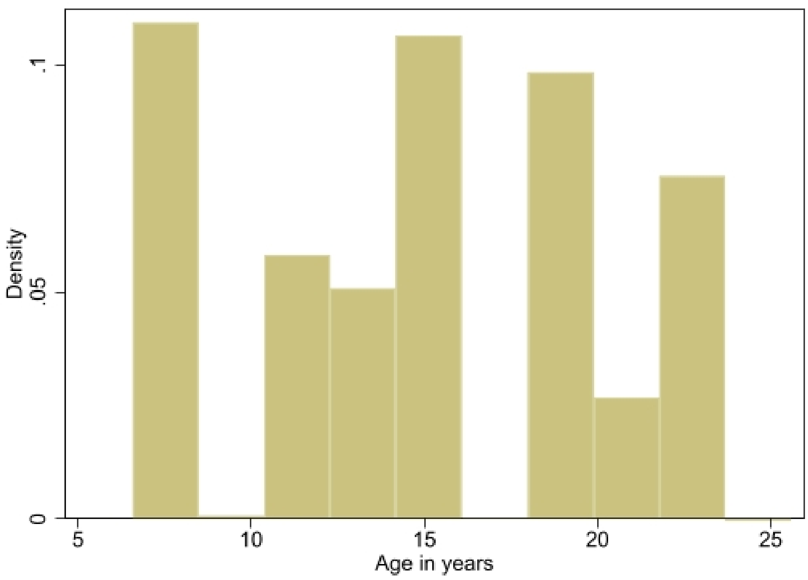


Figure A2

Span of ages in Old Cohort



APPENDIX B

	N	Mean	SD	Min	Max
Child's age at start of grade 1	54140	6.03	1.44	2	20
Calculated BMI=weight/squared(height)	54140	17.02	27.85	-3517.559	2917.728
Commuting time to school (out and return, in minutes)	54140	36.53	17.56	0	480
Hours/day spent in caring for hh members	54140	0.50	1.02	0	19
Household size	54140	5.24	1.97	0	30
Hours/day spent in hh chores	54140	1.20	1.09	0	16
Hours/day spent in leisure activities	54140	4.56	2.19	0	24
Hours/day spent at school	54140	5.20	2.47	0	17
Hours/day spent sleeping	54140	9.09	1.08	0	20
Hours/day spent studying outside school	54140	1.75	1.29	0	14
Hours/day spent in paid activity	54140	0.61	1.94	0	20
shock-death of other household member	41799	0.04	0.19	0	1
shock-illness of father	41799	0.08	0.27	0	1
shock-illness of mother	41799	0.10	0.30	0	1
shock-illness of other household member	39028	0.11	0.31	0	1

Table B1

Summary statistics for untreated individuals

	N	Mean	SD	Min	Max
Child's age at start of grade 1	1454	6.03	1.70	2	15
Calculated BMI=weight/squared(height)	1454	19.29	71.72	.1956198	2749.108
Commuting time to school (out and return, in minutes)	1454	37.45	22.94	0	240
Hours/day spent in caring for hh members	1454	0.59	1.46	0	12
Household size	1454	4.69	2.16	1	17
Hours/day spent in hh chores	1454	1.53	1.50	0	12
Hours/day spent in leisure activities	1454	4.57	2.60	0	15
Hours/day spent at school	1454	4.62	3.39	0	13
Hours/day spent sleeping	1454	8.90	1.19	1	14
Hours/day spent studying outside school	1454	1.43	1.48	0	12
Hours/day spent in paid activity	1454	1.35	3.19	0	14.5
shock-death of other household member	1445	0.06	0.24	0	1
shock-illness of father	1445	0.08	0.28	0	1
shock-illness of mother	1445	0.11	0.32	0	1
shock-illness of other household member	1317	0.11	0.32	0	1

Table B2

Summary statistics for treated individuals

	P value
Child's age at start of grade 1	0.95
Calculated BMI=weight/squared(height)	0.00
Commuting time to school (out and return, in minutes)	0.05
Hours/day spent in caring for hh members	0.00
Household size	0.00
Hours/day spent in hh chores	0.00
Hours/day spent in leisure activities	0.86
Hours/day spent at school	0.00
Hours/day spent sleeping	0.00
Hours/day spent studying outside school	0.00
Hours/day spent in paid activity	0.00

Table B3

P values of the t-test for difference in means

	P values
Child's age at start of grade 1	0.00
Calculated BMI=weight/squared(height)	0.00
Commuting time to school (out and return, in minutes)	0.05
Hours/day spent in caring for hh members	0.00
Household size	0.00
Hours/day spent in hh chores	0.00
Hours/day spent in leisure activities	0.00
Hours/day spent at school	0.00
Hours/day spent sleeping	0.00
Hours/day spent studying outside school	0.00
Hours/day spent in paid activity	0.00

Table B4

Approximate asymptotic p-value for the two-sample Kolmogorov–Smirnov test for equality of distribution functions, whereas

	Frequency	Porcentage
Never treated	22244	98.79
Treated at round 4	270	1.20
Treated at round 5	3	0.01
Total	22517	100.00

Table B5

Percentage of treated individuals of the Young Cohort in the analytical dataset

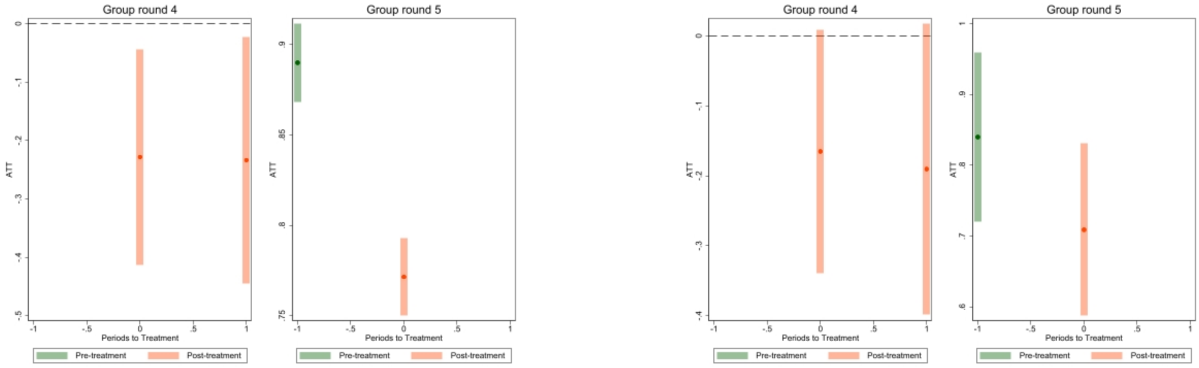
	Frequency	Porcentage
Never treated	9696	97.09
Treated at round 3	99	0.99
Treated at round 4	192	1.92
Total	9987	100.00

Table B6

Percentage of treated individuals of the Old Cohort in the analytical dataset

Figure B1

Event study of the standardized math score for the young cohort

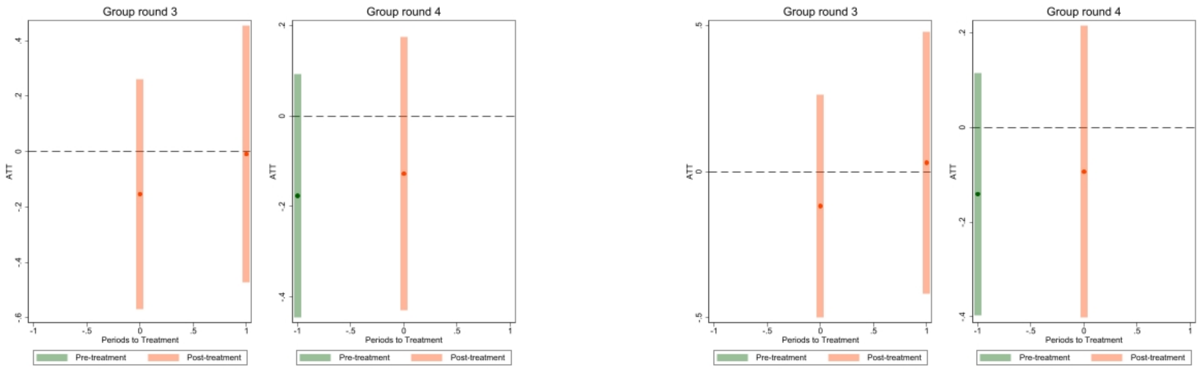


(a) *Without conditioning on covariates*

(b) *Conditioning on covariates*

Figure B2

Event study of the standardized math score for the old cohort

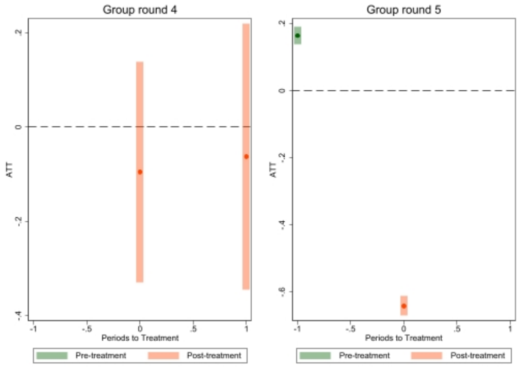


(a) *Without conditioning on covariates*

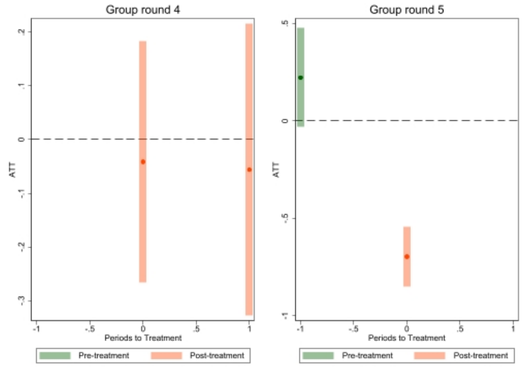
(b) *Conditioning on covariates*

Figure B3

Event study of the standardized PPVT score for the young cohort



(a) *Without conditioning on covariates*



(b) *Conditioning on covariates*