Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(With Specialization in Economics)

By

AQUIB PARVEZ
Reg. No.17SEPH16

SCHOOL OF ECONOMICS
UNIVERSITY OF HYDERABAD
HYDERABAD-500046, INDIA
March 2022

Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(With Specialization in Economics)

By

AQUIB PARVEZ
Reg. No.17SEPH16

Faculty Supervisor:
Prof. K. Laxminarayana

SCHOOL OF ECONOMICS
UNIVERSITY OF HYDERABAD
HYDERABAD-500046, INDIA
March 2022

DECLARATION

I, *Aquib Parvez*, hereby declare the research conducted in the present thesis entitled *"Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks"* is an original work of research carried out by me under the supervision of Prof. K. Laxminarayana, School of Economics, for the award of the Doctor of Philosophy degree from the University of Hyderabad. Furthermore, I declare that, to the best of my knowledge, that no part of this thesis is earlier submitted for the award of any research or diploma in full or partial fulfilment in any other university or institution.

Aquib 10/03/22

Place: Hyderabad
Date: 10/03/2022

Aquib Parvez
Registration No: 17SEPH16

CERTIFICATE

This is to certify that the thesis entitled *"Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks"* submitted by Mr. Aquib Parvez bearing registration number 17SEPH16, in partial fulfilment of the requirements for award of Doctor of Philosophy degree. This is a bonafide work carried out by him under my supervision. This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other university or institution for award of any degree or diploma.

A. Paper published from the thesis:

- Parvez, A., Laxminarayana, K. (2021). Mathematics learning inequality among children of private and public schools. *Asia Pacific Education Review* 1-13 *(Springer)*

B. Presented in the following conferences:

- Educational Achievement and Mothers' education: Evidence from India (Andhra Pradesh)' at The Indian Econometric Society (TIES) conference held at National Institute of Securities and Management, Mumbai, (January'2019)
- Mathematics Achievement of Public and Private School Children: Why is One Better?' at The Indian Econometric Society (TIES) conference held at Madurai Kamraj University, Madurai, (January'2020)

Furthermore, the student has passed the following courses towards fulfilment of the coursework requirements for PhD. He was exempted from doing coursework (recommended by the doctoral committee and the Dean) on the basis of the following courses passed during his M.Phil. program:

Sl. No.	Course	Name of the Course	Credits	Results
1	HS 605	Research Methods in Social Sciences	6.0	Pass
2	HS 615	Computer Aided Applied Statistics	6.0	Pass
3	HS 610	Cost Benefit Analysis	6.0	Pass
4	HS 633	Econometrics of Programme Evaluation	6.0	Pass
5	HS 821	Applied Econometrics	6.0	Pass
6	HS 606	Environmental Planning and Development	6.0	Pass
7	US 604	Management Techniques for Urban Development	6.0	Pass
8	HS 601	Development Planning and Policies: Issues and Alternatives	6.0	Pass
9	HS 603	Socio-Psychological Perspectives in Development and Change	6.0	Pass
10	HS 412	Social Movement and Social Change: Contemporary Reflections	6.0	Pass
11	HS 468	Philosophy of Religion	6.0	Pass
12	HS 602	Science and Technology in India	6.0	Pass
13	HS 694	Seminar	4.0	Pass
14	HS 693	R & D Project	12.0	Pass
15	HS 797	I Stage Project	24.0	Pass
16	HS 798	II Stage Project *(Dissertation)*	48.0	Pass

Dean
School of Economics

DEAN
SCHOOL OF ECONOMICS
UNIVERSITY OF HYDERABAD
Hyderabad-500 046. (INDIA)

Supervisor
Prof. K. Laxminarayana

SUPERVISOR
SCHOOL OF ECONOMICS
UNIVERSITY OF HYDERABAD
Hyderabad-500 046 (INDIA)

Acknowledgements

The thesis writing process was tiring at times but simultaneously enriching in a lot of senses. This thesis, in the form it is now, would not have been possible if the people I am surrounded by were not there.

I would like to convey my sincerest and deepest gratitude to my supervisor, Prof. K. Laxminarayana, who helped me through the entire process. He has been an extremely supportive person as a supervisor and as a person. He gave me sufficient amount of freedom to navigate through the writing process as well as providing me with directions in several ways. I am deeply thankful to him for all he has done without which this thesis would not be possible.

I would also like to mention my doctoral committee members, Mr. Limakumba Walling and Prof. G. Omkarnath, for reading my draft and giving me those invaluable comments. Lima Sir for all the discussions we have had over several cup of tea in exploration of understanding the numbers obtained in the results. I thank both of them for all the support and comments.

My parents for allowing me to pursue research as this was what I really wanted to do. I sincerely thank them for their patience that they have always shown towards me, succumbing to my stubbornness. Thank you very much.

One of my oldest friend, Tariq Azim for tolerating me for almost eight years now. Supporting me in ways which only you could have. For all the support that you have provided me is something I will always be grateful for. Thanks a lot friend.

My IIT (and now HCU) pal Irfan for his prayers plus the readings of first few drafts of the initial chapters were too helpful. MADS for happening in IIT, for making it possible in my betterment.

Mimansa, Ayushya, Titas, Anjana, Zeba, Anindita, Firdaus, Arif Da, Saarthak, Saurabh, Manaswini, Aswani and Anan for their friendship. I cannot see a life without you people around on campus. Thanks for all the fun that we have had, the discussions we have had, the time we have all spent. Thanks a lot for being understanding in ways I never knew was possible. Thanks for the love which has helped me in ways that has made this thesis possible.

Sadaf for just existing. You inspire me. And probably my other siblings too, especially the oldest one.

The data used in this thesis comes from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam (www.younglives.org.uk). Young Lives is core-funded by UK aid from the Department for International Development (DFID). The views expressed here are those of the author(s). They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders.'

And Lastly, I would like to extend my gratitude to ICSSR, for providing me with the fellowship for two years which has made it possible for me to keep working and finishing this work.

Aquib Parvez

Date:
SCHOOL OF ECNOMICS
UNIVERSITY OF HYDERABAD
C. R. RAO ROAD, GACHIBOWLI,
HYDERABAD, TELANGANA, INDIA
500046

Abstract

The thesis attempts to understand the mathematics learning inequality among the Indian children from the states of Andhra Pradesh and Telangana. We have used the longitudinal dataset, Young Lives Survey, for the older cohort for round three and four in the analysis. Based on the literature we have identified three benchmarks on which we explore the learning difference between these children. These benchmarks include the parental schooling status of the young lives children, the type of school they have last attended/or attending, and the division of their mathematics score in respective survey rounds.

In the first chapter we look into the mathematics learning inequality among the Indian children using the fourth round of Young Lives Survey data based on the schooling status of both of their parents. We look into the inequality alongside a set of background characteristics around which the children are growing. We first find that the children of schooled parents have better characteristics as compared to the children of non-schooled parents. We then study the mathematics learning gap with respect to background characteristics using the threefold Blinder-Oaxaca decomposition. We find that these differences play a crucial role in determining the extent of the gap in learning. The crucial findings in this chapter suggests that the differences in the background characteristics explain a part of the learning difference between the groups of children. Moreover, the differential returns to those background features explain the inequality as well. We also learn that the schooling cost (spent in the last academic year) is the consistent contributor to the learning difference between the children of schooled and non-schooled parents. The gap in average years of schooling between the children of schooled and non-schooled parents also contributes to the learning difference. English language score difference between them is another important contributor to the learning difference on mathematics. Gender of the child as well as the household quality explain a marginal part of the learning difference between them as well. In continuation, the next chapter discusses mathematics learning difference between them at two points in their lives, four years apart. We learn that the primary cause of the learning gap is the differences in average background features between them. We also learn that the schooling cost consistently contributes the most to the learning gap. The most important finding is the sudden significance of years of schooling on learning gap, the gap (in years of schooling) on which has magnified between these children in just four years.

Departing from the previous two chapters, in the next chapter we explore the mathematics learning difference between the children of private schools and public schools. We categorize the children into those who have attended/attending private school and public school attended/attending. The first thing that we notice is that there is a clear significant inequality in mathematics learning between the children of private schools and public school. Going into their background characteristics we learn that the children of private schools have better background characteristics compared to the children of public schools. After further exploration of the learning difference with respect to these background features difference, using again, the threefold Blinder-Oaxaca decomposition, we find that the entire difference between these children is because of the background features difference (endowments effect). Furthermore, we find how if the children of public school had had similar average features as the children of private schools, not only would they have performed better, in fact they would have performed better than the children of private schools. Furthermore, in the next chapter we research the same gap in average mathematics learning between the children of private and public schools at two points four years apart. We find that it is the difference in the average endowments between the two which consistently explains the gap in average performance between them. We also find the role of differential impact of the background characteristics on the average learning outcome of children on the first point. The most important and consistent contributor to the endowment effect is the schooling cost and the time allocation on studies. One striking result is the now significant contribution of the gap in average years of schooling which is worrying because these children are from the same age group. We find that with the average features and returns of the private school children, not only the gap between them would have been removed but, in fact, they would have performed better than the private school children.

In the next two chapter, unlike the previous chapters, we approach the children directly from the direction of their mathematics learning outcome. We divide the children into two groups, better-performing children and rest of the children. The children who have scored sixty percent and above marks have been categorized as better performing and the remaining as rest. The former have better background characteristics on an average as compared to the latter. Using, again, the Blinder-Oaxaca decomposition, we find that a large part of learning difference is explained by the differences in their characteristics, which is significantly better for the better performing children. Moreover, the impact of these differences in the form of average returns they reap for both the groups is significantly different where it is better for the former. And in continuation from this chapter, the next chapter studies mathematics learning gaps at two points in time where the gap between the points is four years. We find that when the children were younger the private schooling effect was the core contributor towards this learning gap. When these children got older, the effect vanished and the gap in average years of schooling, which has magnified during this time between these groups of children, contributes most to this learning gap.

JEL: I20, I21, I24, I25, I29

Keywords: Mathematics Learning; Blinder-Oaxaca Decomposition; Endowment Effect; Coefficient Effect; Years of schooling; Parental Schooling; Schooling Cost; Gender

List of Tables and Figures

List of Abbreviations

AP	Andhra Pradesh
AFQT	Armed Forces Qualifying Test
ASER	Annual Status of Education Report
CECED	Centre of Early Childhood Education and Development
CTT	Classical Test Theory
DHS	Demographic Health Survey
IHDS	India Human Development Survey
IRT	Item Response Theory
EA	Mathematics test scores
NAS	National Achievement Survey
NELS	National Educational Longitudinal Study
OCG	Occupational Changes in a Generation
PE	Parental Education
PISA	Programme for International Student Assessment
SAMCEQ-II	South Africa Consortium for Monitoring Educational Quality
SES	Socio-economic Status
TIMSS	Trends in International Mathematics and Science Study
UNICEF	United Nations Children’s Fund
YL	Young Lives
YLS	Young Lives Survey

Table of Contents

1. Introduction

The Trends in Mathematics and Science Study (TIMSS-2003) were applied on Indian children by the World Bank in 2006, where questions were asked to the secondary school going children from the states of Rajasthan and Orissa. The average learning outcome on mathematics and science of these children was found not only different but significantly poor than the international average on the same test (Kingdon, 2007). The international mean achievement in the mathematics test scores were 52 per cent for grade eighth students but the average scores of Rajasthan and Orissa students on the same test were 34 and 37 per cent respectively (Kingdon, 2007). Similarly, the international mean achievement was 57 per cent for grade 12 students but the corresponding scores for the Indian students were 44 and 38 per cent in Rajasthan and Orissa respectively. There are studies that have pointed out factors that have an impact on the academic performance of children with respect to various characteristics. The important factors that have been discussed includes socio-economic background of children, parental education, type of schooling, child related features etc.

The public expenditure on education in India as a percentage of GDP has been increasing over the years and marginally since 2010. The budgeted provision on the revenue account for education for the year 2018-19 by the education departments of the states/UTs and the Centre works out to be 539351.41 crores. This number stood at 433342.37 crores, during 2013-14 to 2015-16 according to the analysis of budgeted expenditure by the Ministry of Human Resource Development. As a percentage of GDP the government expenditure on GDP had been increasing marginally since 2010 to 2012, from 3.378% to 3.868% respectively.[1]

The Annual Status of Education Report (ASER) 2005 concludes that the increase in the financial resources towards education does not reflect in the levels of learning in primary education (ASER, 2005). ASER's later reports conclude that there is a decline in the ability to do basic math nationally, which is visible across all classes (ASER, 2010). However, there are states in India that perform better as compared to the national average. The National Achievement Survey (NAS, 2014) class III (cycle 3) finds no significant difference in the performance of students form Andhra Pradesh and the national average in language whereas former's performance is significantly better than the national average on mathematics.

The learning outcomes of children has been extensively researched in the previous literature. In those studies there are various internationally recognized tests that have been put to use to explore the learning outcomes. The Trends in International Mathematics and Science Study (TIMSS) is a cross-national comparative study of 4th and 8th grade students in mathematics in science. The test is coordinated by the International Association for Evaluation of Educational Achievements (IEA) where 46 countries participate. There are two dimensions to TIMSS, namely, the content dimension and the cognitive dimension. The first one concerns the subject matter to be assessed and the second one the thinking process. The mathematics part of TIMSS includes number, geometry (shapes and measures), data display, Algebra, data and chance, as well as reasoning. The data obtained from the tests has been widely used in several literature to explore the learning outcomes of children in varied contexts.

Similarly, the Programme for International Student Assessment (PISA) is another internationally recognized measure/data to collect information about students' achievement. PISA is a

[1] https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?end=2016&locations=IN&start=1997&view=chart

programme of OECD member countries to assess children at the age of 15 (because at this age majority of them are finishing schooling in the member countries) by combining the assessment of domain specific cognitive areas such as science, mathematics and reading. The test PISA are internationally standardized and are used in several literature on educational achievement literature since its inception in 2000. The first cycle of PISA was implemented in 43 countries (32 in 2000 and 11 in 2002). In the second cycle in 2003 41 countries participated and 56 countries participated in the third cycle in 2006. The idea of this programme is to explore the important knowledge and skills of the children form the member countries.

In India, Annual Status of Education Report (ASER) looks into the learning of the children from across all the states. They do a nationwide survey of children's ability to read simple text and do basic arithmetic to get an idea of how much the children are learning over the years. The first survey of ASER was done in late 2005 and they have done the same survey every year since. This is a largest household survey of children in India by citizens' groups where 25,000 volunteers approach 750,000 children in 15,000 villages each year. This is the only annual source of information regarding the learning levels of children available in India today. The reports provided by ASER are used widely to study and frame policy aspects on education. Similarly, the National Achievement Survey (NAS) is also available for the Indian children. NAS is a large-scale, nationally representative survey to capture the learning of students which is conducted by the Ministry of Education.

Borrowing the ideas and methods from TIMSS and PISA, Young Lives Survey (YLS-the dataset that has been used in this thesis) designed the survey questions on mathematics, reading (English), and Telegu. TIMSS and PISA questionnaires were a great influence on the framing of questionnaires that were used in the four rounds of YLS. There are other student assessment tests available which are also used in the previous literature such as National Achievement Survey (NAS), Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA). The educational outcome in general and learning outcome of children in particular has been studied in relation to several background characteristics of children in the previous literature. The upcoming sections talks in detail about the previous literature on the learning outcomes of children in relation to their varied background characteristics.

1.1 Learning outcomes and Background features

In India there are a number of studies that have explored the connection between various background characteristics and academic achievement of children. Chowdhury and Ghosh (2011) in their primary study of children from Purulia district of West Bengal, analyze the connection between cognitive development and Socio-economic Status (SES-which captures education, occupation, and income). They find that vulnerability in the SES of children is a core cause of poor cognitive development. On a different line, Das *et al.* (2013) in their experimental primary study of children from India and Zambia, research the connection between the anticipated and unanticipated school grants, household private educational spending and educational achievement. The results suggest that there is a connection between the grants and the educational outcomes of children through the household spending in both of the studied countries. Muralidharan and Sundararaman (2015) have done an experimental study of the Indian children from the state of Andhra Pradesh (AP) based on the AP school choice project (which included five district across AP-180 villages). They find that no difference in some situations on some aspects (subjects) of learning outcomes of AP. Using the YLS, India, Singh and Mukherjee (2019) have explored the connection between the type of pre-schools attended and cognitive skills of children. They find that there is a strong association between the type of pre-school attended and cognitive outcomes of children. Those who have attended private pre-school are more likely to achieve higher mathematics score.

The thesis has used the YL dataset, there are studies that have used this dataset to explore the academic achievement of children from Ethiopia, Peru, Vietnam, and India. For instance, Dercon and

Krishnan (2009) using the whole YL data for the four surveyed countries, explored the connection between poverty and psycho-social competencies (self-efficacy, sense of inclusion, self-esteem and educational aspirations) which also includes their educational aspirations. Similarly, Crookston *et al.* (2014) using the whole dataset studies the relationship between SES and child growth and changes in cognitive achievement scores in adolescents in resource poor settings. They find a consistent and strong association between parental schooling, wealth and child growth with their cognitive achievement. There are studies which have used just one YL country in their research. Glewwe *et al.* (2015) in their YL, Vietnam study researched the determinants of learning outcome gaps among ethnic minority students and Kinh using the Blinder Oaxaca decomposition. They find that the parental education explain a large gap between these children. Higher income among Kinh household, more time in school, and more years of schooling explains the gap in learning between them. Similarly, Using the Ethiopia YL, Sanfo and Ogawa (2021) using the Blinder-Oaxaca decomposition study the rural-urban learning gaps in Mathematics and English. They find that the majority of this gap is explained by the student background characteristics differences.

There are studies that are done in lower income countries to explore the connection between the academic achievement and background characteristics. Zhang (2006) in their research of Sub-Saharan Africa, researched the learning disadvantage of rural primary school students. Their study find that the rural students underperformed as compared to their urban counterparts by a large extent. And Li and Qiu (2018) in their study of Chinese children talk about the connection between the family background characteristics child's academic achievements at an early stage. Their results suggest that the urban students' academic performance are more severely influenced by their SES compared to the rural students.

A number of studies have established factors that have a key role to play in influencing the educational achievement of children in various socio-economic contexts. The socio-economic characteristics around which a child grows has an influence on the educational outcomes of children.

Studies (such as Witte, 1992; Kingdon, 1996; Kingdon, 2007; Desai *et al.* 2008; Goyal, 2009; Goyal and Pandey, 2009; Cobb *et al.* 2014; Arteaga and Glewwe, 2014; Glewwe *et al.* 2017) have talked about the socio-economic status of a family in relation to the academic achievement of children. Factors such as household expenditure/income, parental education level, household size etc. have been explored in these studies. There are studies (for instance Behram, 1996; Glewwe *et. al.*, 2001; Paxson and Schady, 2007; Frisvold, 2015; Glewwe *et. al.*, 2017; Belot and James, 2011; Mukherjee and Pal, 2016) that have discussed the child specific features with respect to the child's achievement. Broadly, these studies have talked about the health, gender, and nutrition of child and its influence on their achievement. There is a stream of literature (for example Kingdon, 2007; Wadhwa, 2009; Desai *et al.* 2008; Goyal and Pandey, 2009; Chudgar and Quin, 2012; Wamlawa and Burns, 2012; Singh, 2015) that talk about the association between attending private schools and the academic achievement of children. The majority of these studies have found a significant association between attending private schools and having a better performance score. A review of these studies establishes that there is a strong association between the educational achievement of child and the background features around which they are growing. This has been found in several studies along different contexts.

1.2 Educational outcome and background features

The educational outcomes of children in general are influenced by various factors.[2] The factors that the previous literature has discussed includes the parental education and involvement as an important influencer to the educational outcomes. The parental involvement for example has an influence on the decision to continue with successive grades on the ladder of education (for example, Mare, 1979, 1980,

[2]There are various educational indicators that have been taken as an outcome variable which includes the mathematics/language etc. test scores (measure of academic performance), choice of schooling, decision to continue on the ladder of education.

and 1981). The study by Mare (1979) analyzes the influence of socio-economic background on the grade progression rates. This study uses the data for white males born between 1907 and 1951 obtained from Occupational Changes in a Generation (OCG) survey, 1973. Similarly, Duncan (1967, 1972) includes father's occupation SES (Socio-economic Status) score among the socio-economic characteristics of the family. Apart from father's education, this study also talks about the influence of mother's education, the annual family income and the number of siblings, on the educational continuation decision. This study finds that the post-secondary progression rates are less responsive to changes in the family background characteristics compared to earlier schooling progresses.[3] Mare (1980), unlike the previous studies, uses the measure of child's ability by merging the OCG dataset used in Mare (1979) with the 1964 survey of 3000 veterans of United States' military.[4] The results in this study are similar to the ones reported in Mare (1979) i.e. the influence of these features on the educational continuation decision declines at successive grade transitions. The Armed Forces Qualifying Test (AFQT) scores in this study have been included as a measure of child's ability. The inclusion of the AFQT scores enhances the predictive power of the model as compared to Mare (1979), at each level of schooling up to college. (Mare, 1980, 1981) using the OCG data studied the econometric models used in the education literature and compared the results obtained.[5] The comparison of the results from different models suggest that the influence of the socio-economic background features declines at successive grade transitions. This means as the child progresses he/she

[3] The selected grades in Mare (1979) are whether the individual (1) completes 8th grade; (2) attends 9th grade given 8th grade completion; (3) completes 12th grade given 9th grade attendance; (4) attends 13th grade given 12th grade completion; (5) completes 16th grade given attends 13th grade; and (6) attends 17th grade given 16th grade completion. There were approximately 33500 males surveyed by OCG. The progression rate here is the proportion of the children surveyed in OCG continuing education at successive selected grades.

[4] This survey gives the score on the Armed Forces Qualifying Test (AFQT) which here is used as a measure of child's ability.

[5] The logistic response model of schooling continuation, linear probability model of schooling continuation decision, and linear model of highest grade completed.

becomes more independent from the family and the influence of the family on the probability that the child will continue further on the education ladder, declines.

The parental education and involvement has an influence on the educational achievement of children as well. Duncan (1967) analyses the influence of various socio-economic features of a family on the educational achievement of the children. The study uses 1962 survey by United States Bureau of the census obtained for five year birth cohorts of adult males. The factors that this study considers include the number of siblings, family type[6], family head's education, and family head's occupational SES score.[7] These influences have also been checked on the basis of the ethnic status of Americans surveyed. Overall, this study finds a stable relationship of family's socio-economic features on the educational achievements of children over the period of time. The net effect of growing up in an intact family and the family head's occupational SES score, both have a positive influence on the educational achievements of children. However, the effect of SES scores are greater for whites as compared to non-white Americans. This study points out the importance of growing up in an intact family, the family head's education and family head's economic status on the children's choice of continuing education. Using the similar factors and the same dataset as Duncan (1967), Duncan (1972) uses the Blau-Duncan model of status attainment for the American children. The Blau-Duncan model is a three-stage casual sequence of social mobility.[8] The first stage in this model is the level of education obtained by the father of the child. The second stage is the father's job status after getting education in the first stage. The third stage speaks of the educational attainment of the child given the educational status and occupational status of child's father. This study attempts to analyze the relationship between father's education (family head) with father's occupational

[6] Family members living together or they are living separately.
[7]Based on the occupational socio-economic status (SES) score which appears in J. Reiss, Jr., with the collaboration of Otis Dudley Duncan, Paul K. Hatt, and Cecil C. North, *Occupations and Social Status* (New York: Free Press of Glencoe, 1961).
[8] Peter M. Blau and Otis Dudley Duncan. *The American Occupational Structure*. New York: 1967

status and then with child's educational achievement. The findings of this study suggests that there is a direct link between the father's education and child's education, and father's education and child's educational attainment. There is a substantial difference between the economic achievements of whites and non-whites. When this study analyzed the results based on the ethnic background of the child, as presented in Duncan (1967), the influence of these factors is found to be greater for white Americans as compared to non-white Americans.

1.3 Socio-economic Status (SES) and educational achievement

There are socio-economic factors that are discussed in the previous literature that have a significant influence on the educational achievement of children. Fan and Chen (2001) in their meta-analysis conducted to synthesize the quantitative literature about the relationship between parental involvement and students' academic achievement. The findings reveal a small to moderate, and practically meaningful, relationship between parental involvement and academic achievement. Through moderator analysis, it was revealed that parental aspiration/expectation for children's education achievement has the strongest relationship, whereas parental home supervision has the weakest relationship, with students' academic achievement. Jeynes (2005) has done a meta-analysis of 41 studies examining the relationship between parental involvement and the academic achievement of urban elementary school children. Results indicate a significant relationship between parental involvement overall and academic achievement. Davis (2005) analyzes the academic achievement of children and how it is influenced by various socioeconomic characteristics of the family through parental beliefs. The data that has been used in the study is obtained from a national cross-sectional study of children of America. Jeynes (2007) had done a meta-analysis of 52 studies to determine the influence of parental involvement and the educational outcomes of urban secondary school children. This study concludes that there is a significant influence of parental involvement on the educational outcomes of urban secondary school children. Leibowitz (1977) in a study of children from California find that there is a significant positive relationship between mothers' schooling and children's test scores. Similarly, mothers' cognitive test scores have significant effects on the mathematics test scores of children (in Crane, 1996; Todd and Wolpin, 2007) using American National Longitudinal Survey.[9] Frisvold (2015) has pointed a connection between health/nutrition and learning outcome of American children. Artega and Glewwe (2014) studied the test score difference between the indigenous and non-indigenous children of Peru and tried to understand what is causing this difference between the performances of these two groups. The study, using the Blinder-Oaxaca decomposition on the Young Lives Survey of Peruvian children, finds that it is only parental education and child's health that has a significant role to play in determining their performances on mathematics and vocabulary. For instance, Carriea and Yelowitz (2000) finds a positive relationship between the housing quality and the educational achievement of children in poor families in the United States. Davis (2005) finds an indirect influence of socio-economic features of a household on a child's educational achievement.[10]

The previous literature (such as Goldhaber, 1996; Kingdon, 2007; Desai et. al. 2008; Goyal, 2009; Goyal and Pandey, 2009; Chudgar and Quin, 2012; Wamalwa and Burns, 2018) talk, in one way or the other, about the better performance of private school going children as compared to the public school going children. These studies make important points pertaining to differences in the educational performances of public and private school going children. The discussion over the better performance of latter attributed largely to the choice of schooling is dealt with. These studies mention that there are certain factors behind a particular choice of school, say the economic privilege or better cognitive skills etc., that have a role to play too apart from just the choice of schooling that would have an influence on the educational performance of children. For instance, Goldhaber (1996) says that the public-private choice

[9] Crane (1996) uses random sample of people from the entire country born between 1957 and 1964. Todd and Wolpin (2007) uses the National Longitudinal Surveys of Labor market Experience-Child sample.
[10] Mainly parental education and income.

of school as a policy initiative has the potential to improve the overall achievement of students. They have used the National Educational Longitudinal Study, 1988 (NELS88) of American children. In the context of India, Kingdon (1996) has also studied the Indian children for the state of Uttar Pradesh based on a prepared designed stratified random sample from the survey of schools in urban Lucknow. Desai *et al.* (2008), using the India Human Development Survey (IHDS 2005), provides a detailed description of public and private schools. Moreover, the considerations which eventually guides the parents to select private schools for their children. Goyal (2009) too studies the Indian children's test score data, on Grade 4 students, who are either attending public or private schools from Orissa, India. After correcting for selection, the results suggest an existence of private schooling effect. Goyal and Pandey (2009) have studied the systematic difference in then test scores of Indian children using the survey data of Indian children from the states of Uttar Pradesh and Madhya Pradesh. They have found that the private school children indeed have better scores than the public school children but the quality is low in both the types of schooling as they do not systematically differ in terms of the infrastructure. Similarly, Chudgar and Quin (2012) using the IHDS 2005 data, a nationally representative of Indian children explore the better performance of private school children on tests. They have found that the difference is insignificant when done using multivariate analysis on balanced data using the propensity score matching technique. There are other studies on India, for instance, Singh (2015) using the YLS of the Indian children. The author here has devised a value added model learning productivity (which, unlike previous studies, captures the whole history of a child's environment to capture its effect on the productivity of learning) in government and private schools. The results have found that there are substantial gain from private school attendance on both language and mathematics. The student achievement is measured using adoptions of standardized tests of numeracy. Mukherjee and Pal (2016) explore the relationship of the parental expectations and child labor and schooling of the Indian children using the YLS.

1.4 Type of school, child characteristics and educational outcomes

The enrolment in the private schooling in the rural India has been increasing over the years. One reason for this is the common perception that the private schools provide a better quality education (Wadhwa, 2009). There are studies that have found that the private school going children perform better than the public school going children. The private enrolment is associated with better child outcomes after controlling for family background characteristics (Desai, 2008). Similarly, (Chudgar and Quin, 2012; Wamlawa and Burns, 2012; Singh, 2015) have found an association between attending private schools and having better achievement outcomes. Wadhwa (2009) points out that the supplemental helps provided by parents at home like tuitions etc. also explains a part of this difference in the performance of these two types of schools. Uncontrolled differences are always greater than the actual attributable (to the choice of education) difference. There are studies that have found results unlike the ones discussed previously (for instance, Goldhaber, 1996) where the overall private schools have no significant advantages in education on mathematics and reading. Goyal and Pandey (2009) finds that the private performance is better than the public schools however the quality of both is low in both the cases. There are studies that talk about the characteristics that determine the differences in the performance of private and public school children, such as, the household and child characteristics (in Witte, 1992; Kingdon, 2007), cost effectiveness of private schools (in Kingdon, 1996; Goyal, 2009). Witte (1992) has also done an analysis of the test score difference between the public and private school children using the NELS88 data on the American children. Wamlawa and Burns (2018) have studied the connection between the private school attendance and literacy and numeracy skill acquisitions using the Kenyan third round Uwezo Survey (2012). They too have found, just like Singh (2015) that private school attendance entails substantial gains to language and numeracy.

There are studies on the educational achievement of the children as influenced by the child related features. Behrman (1996) finds that the health of a child and the nutrition have a strong positive association on their educational achievement. This study has used the survey of status on health and status in establishing the connection and causality. Glewwe *et. al*. (2001) on a similar line analyzes nutrition-learning connection, and they find that nourishment of a child is positively associated with the educational performance of a child in school using the longitudinal Filipino data (Cebu Longitudinal Health and Nutrition Survey). Glewwe et. al. (2014) looks also at the gender based differences in the educational achievement along with the nutritional difference and they find a similar influence like the previous studies. Mukherjee and Pal (2016) finds a positive influence of the health of child on the probability of school going children in India. The connection between the cognitive development and SES has also been explored by Paxson and Schady (2007) using a sample of 3000 children who were poor preschool aged. These Ecuadorian children have been found to have an association between the household wealth and parental education with higher scores.

1.5 Parental education and involvement, and educational outcomes

The previous literature suggests that the educational level of parents and the involvement from their side has an influence on the educational outcomes of children. Duncan (1967) in their study finds that there is a positive influence of family head's education on the educational achievement of five year birth cohorts of adult males in the United States of America. Similarly, Duncan (1972) finds that there is a direct relationship between the father's education and child's education, and father's education and child's educational achievement.[11] Duncan (1967, 1972) also finds a positive influence of mother's education on the probability of continuing education to further grades, given that the children are already

[11] Using 1962 survey by the United States Bureau as used in Duncan (1967).

in school. On a similar line, Mare (1980, 1981) finds a positive influence of the parental education on the educational continuation probability of school going children in the United States. Unlike the studies mentioned before, Fan and Chen (2001) have done a meta-analysis that reveals there is a significant and meaningful relationship between the involvement of parents and the academic achievement of elementary school going children.[12] The parental educational aspirations for children is found to have the strongest relationship. Just like Fan and Chen (2001), Jeynes (2005) in their meta-analysis of forty one studies conclude that there is a significant relationship between overall parental involvement and child's academic achievement of urban elementary school going children.[13] Davis (2005) again looks into the parental education like the previous studies find that there is an indirect positive influence of parental education on the educational achievement of a child. The study finds that the educational achievement relates indirectly to the parental beliefs and behavior. And Jeynes (2007) in their meta-analysis of fifty two studies find that the parental involvement influences all the academic variables and the influence of parental involvement overall is significant for secondary school. Artega and Glewwe (2014) concludes like the previous studies that the parental education is an important determinant of mathematics test score (a measure of academic achievement in their study) of the children in Peru.

1.6 Household factors and educational outcomes

Duncan (1967, 1972) finds that the influence of the household size on the educational continuation probability of school going children in the United States is negative. They find that the influence of the household expenditure has a negative influence on the probability of continuing education too. Carriea and Yelowitz (2000) finds a positive relationship between the housing quality and the educational

[12] The involvement is broadly defined as: parental aspiration/expectation for children's education achievement and parental home supervision.
[13] The study talks about the overall as well as subcategories of involvement from the side of the parents.

achievement of children in poor families in the United States. Davis (2005) just like the previous studies finds an indirect influence of socio-economic features of a household on a child's educational achievement.[14] Jeynes (2007) examines the socio-economic status of households and find its significant influence on the educational achievement of children. Just like Duncan (1967, 1972), Cobb *et. al.* (2014) too finds a negative influence of the household expenditure on the educational outcome of school going children. Glewwe *et. al.* (2014) analyzed the learning outcomes of a school from the perspective of advantaged and disadvantaged children in the developing countries. They find no significant discrimination but they do find a significant lack in learning outcomes of children coming from a lower cognitive skills. There are studies that suggest that the academic performance of children has a positive influence on the probability of continuing education. For instance (Mare, 1979; Mare, 1980; Breen and Jonsson, 2000; Mukherjee and Pal, 2016) have discussed the role academic performance of a child plays on the educational continuation decision of the children. The study by Mare (1979) analyzes the influence of socio-economic background on the grade progression rates. This study uses the data for white males born between 1907 and 1951 obtained from Occupational Changes in a Generation (OCG) survey, 1973. Similar to Duncan (1967, 1972), this study includes father's occupation Socio Economic Status (SES) score among the socio-economic characteristics of the family. Apart from father's education, this study also talks about the influence of mother's education, the annual family income and the number of siblings the child has, on the educational continuation decision. This study finds that the post-secondary progression rates are less responsive to changes in the family background characteristics compared to earlier schooling progresses. Mare (1980), unlike the previous studies, uses the measure of child's ability by merging the OCG dataset used in Mare (1979) with the 1964 survey of 3000 veterans of United States' military. The results in this study are similar to the ones reported in Mare (1979) i.e. the influence of these

[14] Mainly parental education and income.

features on the educational continuation decision declines at successive grade transitions. Behrman (1996) finds that the health of a child and the nutrition have a strong positive association on their educational achievement. Glewwe *et. al.* (2001) on a similar line analyzes nutrition-learning connection, and they find that nourishment of a child is positively associated with the educational performance of a child in school. Glewwe *et*. al. (2014) looks also at the gender based differences in the educational achievement along with the nutritional difference and they find a similar influence like the previous studies. Mukherjee and Pal (2016) finds a positive influence of the health of child on the probability of school going children in India.

1.7 Research Objectives/Directions

In this thesis we study the mathematics learning outcome of Indian children (from united Andhra Pradesh) using the Young Lives Survey (YL henceforth) which is a longitudinal survey collecting information on children since 2002, every four years. The survey has so far released four rounds of data. In each round YL collected information about children's learning achievement with the help of language and mathematics tests (Galab *et al.* 2014a). The mathematics test for the older cohort in round 3 of the Indian children comprised of two sections. The first section included 20 items on addition, subtraction, multiplication, division, and square roots which allowed for eight minutes of solving time. The second section included 10 items on mathematics problem solving, which is publicly available on TIMSS and PISA (Cueto and Leon, 2012). This section allowed for ten minutes to attempt to the questions. Similarly, the round 4 test included problems on basic mathematics (addition, subtraction, division, and multiplication) which were more than double digit problems. This also included decimal problems, percentage, and mathematics problem solving.

The YL's preliminary findings for the latter round (fourth) points out that although there is an increasing choice of private schools, the gap is magnifying between the proportion of boys attending private schools compared to girls attending private schools over the years (Galab *et. al.* 2014a). We study the mathematics learning gaps among the children of united Andhra Pradesh based on three benchmark characteristics. The first one is the educational status of YL children's parents where we have divided the children into two groups, each on the basis of whether their mother/father is schooled or non-schooled. Second, we have divided them on the basis of the type of school they have last attended or attending, public or private school. And lastly, we have divided them directly on the basis of their mathematics test performance: better performing children (who have scored sixty percent and above marks on mathematics test at the time of third and fourth round survey) and rest of the children.

This research intends to look into the differences in the mathematics performance of the children of the Young Lives Survey from the united Andhra Pradesh. There are various aspects based on which there exists a gap in mean performance of children, for e.g. based on the mothers/fathers education, based on the type of schooling they receive, also based on the gender of the child, all of which have been thoroughly researched in the previous literature. The idea behind this research is to explore the mathematics learning inequality among the Indian children on the basis of three benchmarks. Furthermore, we have attempted to place those learning inequalities in the larger context of their background features differences. Starting with the usage of the fourth round YL data for the older cohort, India, we intend to use the third round of the data as well.[15] As the data that this research proposes to use is a longitudinal survey, the idea is to look into these factors first, at a particular point in time, starting from round 4, then extending it to two point by including the third round dataset as well. This would give an idea of the mean educational performance differences at a particular point in time and then putting the data of two rounds,

[15] The study proposes to use the Young Lives Survey (YLS henceforth) which is a longitudinal survey that has five rounds of data collected. The details of this is explained in the upcoming section.

would portray the pattern of educational performance differences overtime. The benchmarks based on which this research looks into the differences in the average mathematics performance includes the fathers (and mothers) schooling status, type of schooling attended/attending, and division of scores received.[16] Looking at the differences in the average mathematics performance based on various factors at a point in time and then putting the third round dataset too together and looking at the same factors overtime would portray a picture of the movement of the gap in average performance differences with respect to various factors, first at a particular point in time and then at two points of time. Moving ahead then this research would give an idea of point difference and shall also throw a picture of the pattern in the movement in those differences over the given period of this survey up until the fourth round.

We aim to decompose the gaps in these averages on these benchmarks. The methodology that we use to do this is borrowed from the wage literature. The decomposition methodology deployed here is the Blinder-Oaxaca decomposition. We have done the exercise using the threefold Blinder-Oaxaca decomposition. There have been previous studies which have used the Blinder-Oaxaca decomposition to study the achievement gaps among children. For instance, Ammermuller (2007b), using the PISA-2000, has studied the achievement gap between the German and Finnish children. Zhang and Lee (2011) have also used this decomposition technique, using PISA-2006, to study the achievement gap between the OECD countries. Burger (2011) too studies the achievement gaps between the urban and rural Zambian children using the SAMCEQ-II (South Africa Consortium for Monitoring Educational Quality) dataset. Baird (2012) studies the achievement gaps (using TIMSS for 2003) between the children with high socio-economic status (SES) and lower SES among 19 high income countries. Another interesting study using the Blinder-Oaxaca technique is by Barrera-Osorio *et al.* (2011) which studies the achievement improvement of the Indonesian children between two points using PISA 2003 and 2006. The two point

[16] The third benchmark has divided the data based directly on the performance of the children which is, who have scored sixty percent and above as better performing children and rest of the children as the other category.

studies that we have done in the context of India are none. This is where this study fills the gap in literature and provides valuable insights into the mathematics performance gaps among the Indian children. We decompose the gaps in the average mathematics performance among these Indian children based on the above discussed benchmarks using the threefold Blinder-Oaxaca decomposition.

1.8 Research chapters

In chapter 2.0 we describe the dataset that we have used for the purpose of our research. The source of the data along with the sampling methodology was described. We also provide details about the limitations of the dataset and also the limits of inference that is to be kept in mind while reading the results from the chapters. This chapter also details the variables that we have used in the thesis based on the literature and the context of India. We give basic descriptive statistics of the dataset as well as the preliminary observations from the statistics. Moving on, we give a detailed information on the decomposition methodology that we have used in all our main chapters where we derived the methodology and explained how the results obtained from this are to be read.

The mathematics learning gap between the children of schooled and non-schooled parents, which is this thesis' first benchmark, has been explored in chapters 3.0 and 3.1. In chapter 3.0 we use the fourth round of YLS for the older cohort. We see that the children of schooled parents (separately for mothers' and fathers' schooling status) have better scores on mathematics compared to the children of non-schooled parents which is significant. Looking into their different background features we see that the children of schooled parents have better background characteristics compared to the children of non-schooled parents. Using then, the threefold Blinder-Oaxaca decomposition, we place the significant learning difference between them in the context of their differences in the background. Similarly, chapter

3.1 extends this idea where we put to use the third round data, when these children were four years younger. Then we do the same exercise at that point as well and report our results.

The second benchmark on which we explore the mathematics learning difference between the YL children has been discussed in chapters 4.0 and 4.1. We have categorized the children who have attended private school last or are attending private schools currently have been categorized as one. And similarly the other are those who have attended public school last or are attending public schools currently. We use just the round 4 dataset for chapter 4.0. We learn from here that the children of private schools have better mathematics learning outcomes compared to the children of public schools. We also learn from this chapter that the children of private schools have, on an average, better background characteristics compared to the children of public schools. Using again the threefold Blinder-Oaxaca decomposition we breakdown the differences in average learning between these groups of children in the context of the differences in their average background. Extending that idea, in chapter 4.1, we repeat the same exercise by incorporating the round three data as well when these children were four years younger. And we read the results of both the rounds together and report our findings in this chapter.

The last benchmark on which we study the mathematics learning inequality in chapter 5.0 and 5.1, employs and unconventional approach in categorization of these children. Unlike the previous two benchmarks where we divided the children on the basis of their background (parental education and type of school attended), here we divide the directly on the basis of their mathematics score. The children who have scored sixty percent and above marks are categorized as the better performing children and the remaining as rest. And we learn from chapter 5.0, where we use only round four data, that the better performing children have better background features compared to the rest of the children. Using the Blinder-Oaxaca decomposition we breakdown the difference in average mathematics learning in the context of their background features. Similarly, in chapter 5.1 we repeat the same idea when the children were four years younger in round 3. We read the results of both the rounds together and report our findings and conclusion here.

And in chapter 6 we summarize the results obtained from the thesis and chart out the core contributor towards the mathematics learning. We also write about the policies that are directed by the results that we obtain form this thesis. The thesis also talks about the limitations and discussions in the very last section of this chapter.

2. Data, Variables, and Methodology

2.1 About Young Lives Survey, India

India is home to the second largest population more than a quarter of which are children. Despite huge population advantage India has, it still performs badly in terms of various socio-economic indicators. The stark inequalities between socio-economic groups in terms of wealth, education, welfare accessibility, and mobility in poverty are deeply entrenched in its fabric. A humungous population still lives on less than $2 a day couple that with a quarter of all the child death globally, occur in India. The Young Lives study had been doing a longitudinal survey in India in two southern states of Andhra Pradesh and Telangana. Together they have a population of 85 million. Young Lives Survey (YLS henceforth) is an international study (in four developing countries) with the aim to trace the lives of children from poor economic background, alongside tracking the changing nature of poverty. The lives of 12,000 children traced by YLS are from Ethiopia, India (Andhra Pradesh and Telangana), Peru, and Vietnam, in a collaborative research. The project is coordinated by a team based at the University of Oxford partnering with teams based in the surveyed countries (government, academics, and NGOs). The countries of focus were chosen to reflect the cultural, socio-economic, and political diversity in this project. With the help of large-scale household surveys of children, primary caregiver, as well as in depth interviews of groups of children (with a more detailed interactions with the teachers and community representatives) this survey has collected a plethora of information. The aim here is to record a holistic set of environment around which these children are growing up over the years. This includes information, not restricted to just their material and social circumstances, but to their own understanding of aspirations and current status in their lives. This survey has been tracing these children for over 15 years now. The changes occurring in their lives since they were children to now when they are adults has been recorded. The five rounds of data along with an in depth nested qualitative interviews is a rich set of information to explore and study. Our thesis uses two rounds of Young Lives Survey data for India (United Andhra Pradesh).[17] The survey is based on a pro-poor sample which has nearly an equal number of boys and girls, based both in rural and urban communities. The 3000 children in the Indian survey were selected from 20 sentinel sites spread now across two aforementioned southern Indian states. The sites have been defined specifically for each country. The sentinel site sampling methodology used in the YL is a form of purposive sampling. The observations in the sample represent a certain type of population.[18] As the YL is a longitudinal study, a prolonged contact with the observations had to be maintained. In order to ensure that YL is able to maintain this contact over a long period, this methodology was most suited. We must note here that the YL sample is not strictly statistically representative of the population. Keane *et al.* (2018) mentions that the YLS has a pro-poor sample bias. In our study, the eventual number of observation has reduced, primarily because of the missing information on certain variables. So, the analysis and inferences made in our chapters have to be seen in this light. This is bound to affect the generalizability of our study.

The sampling method deployed in the data collection is a form of purposive sampling (Galab *et. al.* 2014).[19] This survey is repeated on the same set of children every four years. We have made use of the third and fourth round data for the older cohort on India in our thesis released in 2010 and 2014 respectively. The survey in India was conducted on two age cohorts, young and old, where the younger cohort was approximately 8 months of age when the survey began in the year 2002 and the older cohort

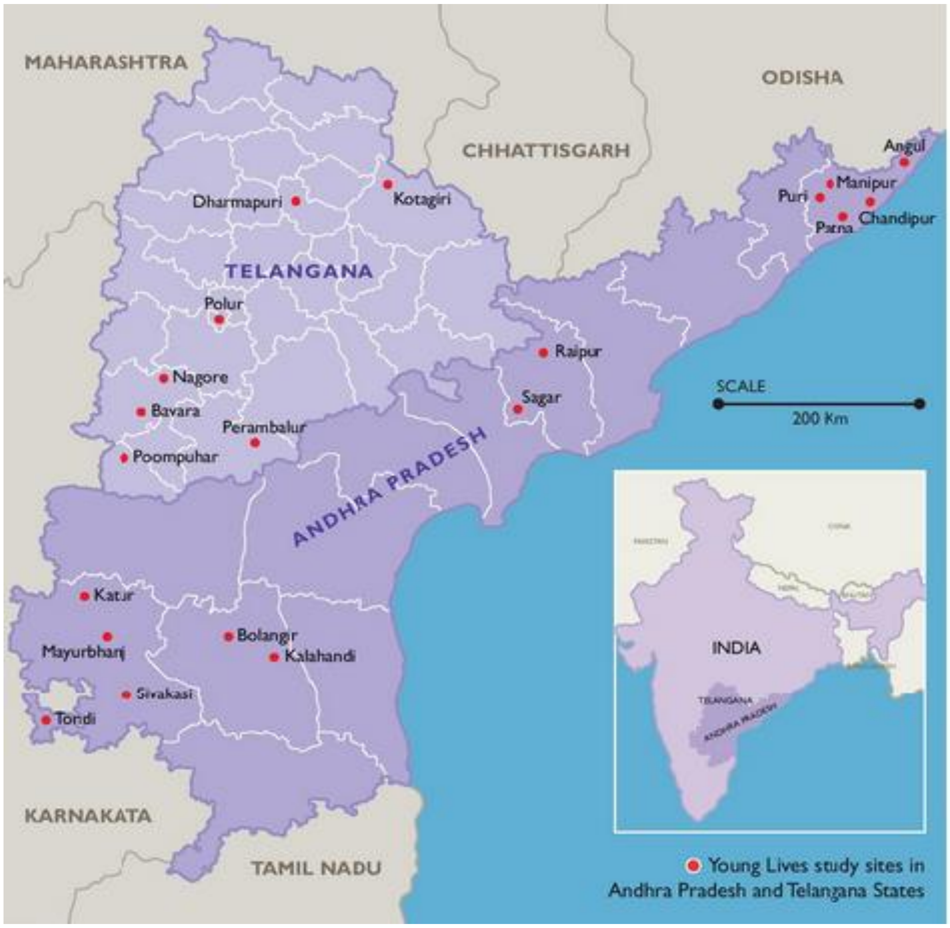
[17] The Indian state selected for this survey was Andhra Pradesh when the survey began in 2002. The state was bifurcated into two separate states in 2014, now the surveyed children belong to either one of the two states.
[18] Young Lives study has made use of 'sentinel site' methodology, which they have borrowed from health surveillance studies. This is a form of purposive sampling where these sites represent a certain type of population. The method produces broadly (but not strictly statistically) representative picture. In the Indian Young Lives, sentinel site was defined as a mandal. The state (then Andhra Pradesh) was divided into 23 administrative districts, each of which were again sub-divided into mandals depending on the size of the district. The state of Andhra Pradesh had 1125 mandals and about 27000 villages. More details are available at Young Lives Survey website here.
[19] Or 'cluster' in the language of sampling.

was around 8 years of age at that time. The YL employed a kind of purposive sampling they decided to cover a range of population had to be surveyed, although poorer families were over surveyed. The children were surveyed in a geographic clusters where the selection was done on the basis of semi-purposive approach. In each of those clusters, the selection of children was random. The two cohorts of children that were surveyed were intended for comparison. In each of these countries, 2000 children aged between 6-8 months were tracked and followed as they grew up to 15 years of age. A similar sample of 1000 children aged between7.5-8.5 years were selected as older cohort for comparison.

The sampling sites for the YL survey were chosen in 2001. First, the districts were chosen and then the 20 sentinel sites form within those districts were decided upon based on a set of criteria. In each of those sentinel sites, 100 households where with a child born in 2001-02 and 50 households with a child born during 1994-95 were taken into consideration where both were randomly selected. If there was a case of where a household had children who fell in both of the categories, the younger ones were chosen since they comprised a greater number of surveyed children. The sentinel sites in India were defined as Mandal. Where the old state of Andhra Pradesh was divided into 23 administrative districts, and each of which was divided into Mandals based on the population size of those. In sum, 1125 Mandals and around 27,000 villages were considered, where generally 20-40 villages were there in each Mandal. However, in the tribal areas there were as many as 200 villages.[20] The 20 sentinel sites are shown in the map below.

[20] For more details on caste and other social background details of the surveyed children see (click here).

Fig: 1 Surveyed fields on the States' Map



Source: Young Lives Survey, India

The two states were divided into three regions in order to choose 20 sentinel sites. These regions were the Telangana state (Mahboobnagar and Karimnagar and this region also included the state capital Hyderabad), Rayalaseema (YSR Kadapa and Anantapur), and Coastal Andhra (West Godavari and Srikakulam). The sentinel sites (7 locations) that were eventually chosen from districts in Telangana include Mahboobnagar (Nagore, Bavara, Perambalur, and Poompuhar), Hyderabad (Polur), and Karimnagar (Dharmapuri and Kotagiri). The sites chosen from Andhra state (13 locations) include Anantapur (Katur, Mayurbhanj, Sivakasi, and Tondi), YSR Kadapa (Bolangir and Kalahandi), West Godavari (Raipur and Sagar), and Srikakulam (Angul, Manipur, Puri, Chandipur, and Patna). The site names inside these districts are pseudonyms used by YLS to protect the anonymity of the surveyed children.

2.2 Topics covered by YLS by Round 4 survey

There were several questionnaires for different types of information that the YL intended to collect. The core questionnaire was for the YL child and a questionnaire also for the primary caregiver. The latter focused on the household circumstances around which the YL child was growing up. A questionnaire was also used for the local community representative to collect the information about the local economy, and various other issues that were affecting the situation of child within the community. The household questionnaire for both the cohorts asked questions pertaining to the parental background, household and child education level, livelihoods etc. This also included information on household food and non-food expenditure, the social capital, recent life history, child health, caregiver's attitude and perceptions. The older cohort child questionnaire included sections on parents and caregivers update, mobility, subjective well-being, employment, health and nutrition, time-use cognitive tests (Telegu reading comprehension, English, Mathematics, and Self-administered questions). The younger cohort child questionnaire included sections on schooling, time-use, health, social networks, feelings and attitudes, cognitive tests (Peabody Picture Vocabulary test, Telugu reading comprehension, English, Mathematics, and Self-administered questions). And the community questionnaire included information on the characteristics of the locality which included the social environment, access to services, economy, educational and health facilities.

We must point it out here that the Young Lives Survey is not a nationally representative sample like others (for instance Demographic Health Survey). Instead, this survey intends to trace the changes occurring in the lives of a set of children over a period of time (longitudinal survey). Moreover, the changes in different outcomes of interest as a result of the changing circumstances were also intended to be recorded. The YLS sample, compared to DHS 1998/99 (the year closest to Round 1 YLS) includes household with better access to service and more ownership of assets which indicates some bias. This is true for the wealth index of the YL children which is more than the DHS (Andhra Pradesh). Despite these biases YL is shown to have covered the diversity of children in poor households in Andhra Pradesh. One of the important statistical feature of the dataset is that the survey has tried to keep the attrition rate low by keeping track of the chosen children. This is 2.6% for the younger cohort and 4.3% for the older cohort. There were several reasons for the attrition in the YLS. To point out a few, migration, marriage, and pointlessness (as it did not bring any benefit according to the YL parents) for the surveyed children.

The older cohort which makes the Indian sample one-third of the total surveyed children, were born between Jan'94 and June'95. The remaining children, younger cohort, makes two-third of the Indian sample and they were born between Jan'01 and June'02. This thesis uses the data on the older cohort for the third and fourth round when the children were around 15-16 and 19-20 years of age respectively. The third round covered 977 and fourth round covers 952 children from the older cohort. The attrition rate in the Indian survey has been kept really low (Singh, 2015). Our study had to drop out observations as information on various variables for many observations were missing. The thesis considered only those children who are studying/studied in purely public or private schools and we had to drop children who have studied in different schools. We are only studying those children who have participated and on whom the information on the required variables are available at both the rounds (third and fourth rounds) and we ended up with 522 observations for the analysis. The Young Lives Survey provides information on various background characteristics of children and detailed record of necessary information needed for this research. The variable of interest here is mathematics test scores of children. In each round YL collected information about children's learning achievement with the help of language and mathematics tests (Galab *et al.* 2014a). The tests that were administered to YL children were different for each round, though there were few similarities. The detailed information on these tests are available in YLS questionnaires. In round

3, the test for the older cohort had two sections. The first section (which included 20 items) asked questions pertaining to addition, subtraction, multiplication, division, square roots (which included both whole numbers as well as fractions). The second section (10 items) included questions on mathematics problem solving which were developed using PISA and TIMSS. This section's questions were on data interpretation, number problem solving, measurement, and geometry. Similarly, the round 4 test had five sections which asked problems on similar aspects as round 3. The difficulty level was adjusted, as there was a gap of four years between the two rounds (Dawes, 2020). In total there were 30 questions in each round and the scores were obtained by adding the correct responses out of 30. We must note here that although there were similarities in the questions between the rounds (Dawes, 2020), they cannot be compared directly (Dawes, 2020; Rolleston, 2014). In order to make these tests comparable, especially mathematics, between the rounds, the reliability and validity has to be checked. There are several studies (Cueto and Leon, 2012; Azubuike *et al.* 2017) which have done that using the Classical Test Theory (CTT) and Item Response Theory (IRT). The application of these tests brought them to a uniform comparable scale, which allowed for a comparable study between the rounds. Our study does not directly do a comparison between the rounds hence we have not used this in our study.

The Table 1 gives description of the variables selected in our study based on the literature reviewed and the available data with the YLS. The Table 2 (2.0, 2.1, and 2.2) presents the overall descriptive statistics of all the variables described in Table 1 for each round of the survey, and by the three sub-group of children that we have created (which we call the benchmark characteristics). What is immediately noticeable from these tables is that the average score of these children has improved from round three to round four (although we cannot compare the average of the test directly between rounds) on the raw test score average. The sub-sample of our children has been created based on three benchmark characteristics. The first subgroup is created based on the schooling status of child's parents (individual parents). The children of schooled *(+)* mothers (and fathers) as against the children of non-schooled *(0)* mothers (fathers). Similarly, the children of private schools *(Pr)* as against the children of public schools *(Pub)*. And lastly, the children who have scored 60 percent and above marks as better performing/first division children *(F)* as against the remaining children as rest *(R)* who have scored less than 60 percent marks We see very clearly from these tables that the mathematics learning is different within each sub groups at each survey points. For instance, there is a clear learning gap between the children of schooled and non-schooled parents. There is also a learning difference between the children of private schools and public schools. Similarly, there is a visible difference in learning between the better performing children and rest of the children. For the ease of understanding at a point, the scores have been normalized (Table 3.0, 3.1, and 3.2) by dividing with the standard deviation of the overall sample of respective rounds. We must note that comparison of average raw/standardized mathematics score at two points is not possible because of the reasons mentioned in the previous discussion. This has been pointed out earlier that though there are similarities between the tests, they cannot be directly compared between rounds unless they are brought together on the same comparable scale. Moreover, the children were four years older in round 4 survey, we cannot, with certainty, say that the improvement in the average score (although incomparable) as an improvement. Glewwe *et al.* (2015), for instance, points out that there is a connection between age of the child and their learning outcome. We have to be careful while interpreting the averages at these two points since we do expect a change in score simply because the children are older. We considered the age of child at the time of both of these surveys in our analysis since there is not significant variation in their age, we do not find a significant role of this variable at these points.

Table 1: Description of the Variables

Variables		*Description*
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of surveys out of 30 marks. This has also been standardized using the standard deviation of whole sample in each round for ease of comparison.
Parental Education (PE)[21]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets (Round 4)	Value of the five most valuable assets owned, rented or borrowed in the YL child's household
	Assets (Round 3)	Value of assets owned, rented or borrowed in the YL child's household
	School Cost (Round 4)	Total expenditure incurred on school in the last academic year.[22]
	School Cost (Round 3)	How much has the YL household spent on school fees and extra tuition for the child per year
	Expenditure (Round 4)	The log of per capita monthly expenditure of YL child's household
	Expenditure (Round 3)	The log of real per capita monthly expenditure of YL child's household; base 2006 prices
	Years	Years of schooling received by YL child at the time of surveys
Child Specific Features (C)	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey out of 30 marks[23]
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of surveys
	Body Mass Index	Body mass index of the YL child at the time of surveys
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[24]
	Type of School	=1 if attended/attending public school last, 0 if private.

[21] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[22] In the fourth round this cost is the sum total of the tuition fees, education charges, private tuition, accommodation, transportation, uniforms, stationary etc. in the last academic year. The school level heterogeneity is captured by the variance in the sum total of these costs.

[23] This information is available only in the fourth round survey.

[24] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

Table 2.0: Descriptive statistics of the Variables

	Round 4						Round 3					
	Overall		Schooled Mothers		Non-Schooled Mothers		Overall		Schooled Mothers		Non-Schooled Mothers	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Math Score	14.478	7.074	16.320	6.644	12.522	7.005	11.312	6.156	12.996	5.950	9.521	5.869
Mothers Education	3.632	4.314	7.048	3.466	0	0	3.632	4.314	7.0483	3.466	0	0
Fathers Education	5.644	5.045	8.160	4.581	2.968	4.047	5.643	5.045	8.159	4.580	2.968	4.047
Household Size	4.753	1.908	4.691	1.856	4.818	1.964	5.086	1.974	5.115	2.245	5.055	1.641
Assets	30018.64	44478.14	34932.97	43939.72	24793.52	44537.56	9291.452	52007.87	11312.16	54614.45	7142.957	49101.88
School Cost	19226.17	26380.52	25337.38	30585.59	12728.48	19018	2723.567	5344.192	3835	6151.854	1541.846	4009.821
Expenditure	7.170	.588	7.240	.575	7.095	.593	6.776	.561	6.866	.527	6.680	.580
Years	13.448	1.745	13.870	1.569	13	1.813	8.641	1.308	8.728	1.399	8.549	1.199
Time Allocation	2.013	1.442	2.316	1.437	1.691	1.377	2.626	1.293	2.921	1.342	2.312	1.162
English Score	16.513	3.306	17.405	3.104	15.565	3.255	-	-	-	-	-	-
Gender	.549	.498	.509	.501	.593	.492	.547	.498	.509	.500	.588	.4930
Age in Months	227.877	4.134	227.580	4.195	228.193	4.053	179.197	4.105	178.825	4.180	179.592	3.994
Body Mass Index	19.934	8.743	19.966	4.296	19.899	11.765	17.608	2.767	17.969	2.875	17.224	2.598
Drinking Water	.981	.137	.993	.0861	.968	.175	.963	.187	.985	.121	.940	.236
Household Quality	.725	.217	.755	.200	.694	.230	.604	.298	.622	.252	.585	.340
Type of School	.425	.495	.308	.463	.549	.499	.622	.485	.505	.500	.747	.435
Observations	**522**		**269**		**253**		**522**		**269**		**253**	

Table 3.0: The Mean Math Test score by the groups

Round 4

	Overall			Schooled Mothers			Non-Schooled Mothers			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(+)-(0)	t-test
Math Score	522	2.04	1.00	269	2.3017	.937	253	1.766	.988	.536	6.36

Round 3

	Overall			Schooled Mothers			Non-Schooled Mothers			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(+)-(0)	t-test
Math Score	522	1.839	1.00	269	2.113	0.967	253	1.548	0.954	0.564	6.711

*the test scores have been normalized by dividing it with the standard deviation of the overall sample for ease

Table 2.1: Descriptive statistics of the Variables

	Round 4						Round 3					
	Overall		Private		Public		Overall		Private		Public	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Math Score	14.479	7.07	16.38667	6.744	11.90	6.70	11.312	6.156	14.279	6.009	9.513	5.523
Mothers Education	3.632	4.314	4.66	4.514	2.24	3.60	3.632	4.314	5.532	4.676	2.48	3.629
Fathers Education	5.643	5.045	6.75	5.070	4.15	4.62	5.643	5.045	8.182	4.820	4.104	4.537
Household Size	4.753	1.908	4.73	2.036	4.78	1.73	5.086	1.974	5.162	2.117	5.04	1.884
Assets	30018.64	44478.14	35547.63	44685.45	22547.03	43181.86	9291.452	52007.87	12596.24	60446.55	7288.24	46134.97
School Cost	19226.17	26380.52	28759.59	30584.56	6343.17	9218.29	2723.567	5344.192	6752.056	6904.182	281.683	1114.053
Expenditure	7.170	0.587	7.23	0.580	7.09	0.59	6.776	.561	6.931	.487	6.682	.582
Years	13.448	1.745	13.99	1.376	12.71	1.92	8.641	1.308	8.761	1.494	8.569	1.178
Time Allocation	2.013	1.442	2.3	1.379	1.63	1.44	2.626	1.293	2.766	1.342	2.541	1.257
English Score	16.513	3.306	17.27	2.924	15.50	3.52	-	-	-	-	-	-
Gender	.550	.498	0.58	0.494	0.51	0.50	.547	.498	.583	.494	.526	.500
Age in Months	227.88	4.134	227.76	4.143	228.03	4.13	179.197	4.105	179.376	4.130	179.092	4.092
Body Mass Index	19.933	8.744	19.74	3.489	20.20	12.79	17.608	2.767	18.063	3.111	17.333	2.501
Drinking Water	0.981	0.137	0.99	0.115	0.97	0.16	.963	.187	.974	.157	.956	.203
Household Quality	0.725	.217	0.75	0.204	0.69	0.23	.604	.298	.691	.202	.551	.333
Type of School	.425	.495	-	-	-	-	.622	.485	-	-	-	-
Observations	**522**		**300**		**222**		**522**		**197**		**325**	

Table 3.1: The Mean Math Test score by the groups

Round 4	Overall			Private			Public			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(Pr)-(Pub)	t-test
Math Score	522	2.047	.953	300	2.317	1.00	222	1.683	.947	.634	7.53

Round 3	Overall			Private			Public			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(Pr)-(Pub)	t-test
Math Score	522	1.839	1.00	197	2.321	.977	325	1.546	.898	.774	9.240

*the test scores have been normalized by dividing it with the standard deviation of the overall sample for ease

Table 2.2: Descriptive Statistics of the Variables

	Round 4						Round 3					
	Overall		First Division		Rest		Overall		First Division		Rest	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Math Score	14.479	7.074	22.289	2.482	10.368	4.908	11.312	6.156	21.044	2.546	9.258	4.504
Mothers Education	3.632	4.314	5.161	4.479	2.827	4.003	3.632	4.314	6.043	4.730	3.122	4.047
Fathers Education	5.644	5.046	6.872	5.302	4.997	4.788	5.643	5.045	7.736	5.305	5.201	4.882
Household Size	4.753	1.908	4.600	2.116	4.833	1.787	5.086	1.974	5.065	2.421	5.090	1.869
Assets	30018.640	44478.140	35562.420	48508.220	27100.860	41983.730	9291.452	52007.87	8172.802	33687.21	9527.64	55130.67
School Cost	19226.170	26380.520	31147.950	33431.980	12951.550	19028.050	2723.567	5344.192	5979.121	7369.401	2036.2	4530.036
Expenditure	7.170	0.588	7.193	0.570	7.158	0.598	6.776	.561	6.977	.455	6.734	.572
Years	13.448	1.745	14.411	1.056	12.942	1.823	8.641	1.308	8.978	1.282	8.571	1.304
Time Allocation	2.013	1.442	2.461	1.283	1.778	1.466	2.626	1.293	3.066	1.459	2.534	1.238
English Score	16.513	3.306	18.700	2.347	15.363	3.155	-	-	-	-	-	-
Gender	0.550	0.498	0.650	0.478	0.497	0.501	.550	.498	.703	.459	.515	.500
Age in Months	227.877	4.134	228.289	4.170	227.661	4.104	179.197	4.105	179.376	3.784	179.160	4.172
Body Mass Index	19.934	8.744	19.583	3.272	20.118	10.540	17.608	2.767	17.709	2.905	17.588	2.741
Drinking Water	0.981	0.137	0.994	0.075	0.974	0.160	.963	.187	.989	.105	.958	.200
Household Quality	0.725	0.217	0.782	0.182	0.695	0.228	.604	.298	.712	.213	.582	.309
Type of School	0.425	.494	0.255	0.437	0.515	0.500	.622	.485	.297	.459	.691	.462
Observations	**522**		**180**		**342**		**522**		**91**		**431**	

Table 3.2: The Mean Math Test score by the groups

Round 4

	Overall			First Division			Rest			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(F)-®	t-test
Math Score	522	2.047	1.00	180	3.152	.351	342	1.466	.694	1.686	30.58

Round 3

	Overall			First Division			Rest			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(F)-®	t-test
Math Score	522	1.839	1.00	91	3.422	.414	431	1.505	.732	1.916	24.148

*the test scores have been normalized by dividing it with the standard deviation of the overall sample for ease

2.3 Methodological Framework and Empirical Approach

The learning outcome of children depends on various background characteristics. This relationship could be depicted in a mathematical relationship. A simple production function can be used to express this relationship between mathematics learning and a set of background characteristics. The mathematics learning (EA) is a function of input variables like parental education (PE), household factors (H), Child specific features (C), and other features (O).

$$EA = f(PE, H, C, O) \qquad (1)$$

Yet even when an educational production function exists, there is no guarantee that one can estimate it (Artega & Glewwe, 2014). The linear regression equation for the whole dataset is shown below:

$$EA = \theta_0 + \theta_1(PE) + \theta_2(H) + \theta_3(C) + \theta_4(O) + \epsilon_i \qquad (2)$$

The estimation of this equation would tell that how different variables influence the learning outcome, holding other variables constant. As discussed previously we have three benchmark characteristics or three subgroups (two in each) based on those benchmarks. To understand the development of the threefold Blinder-Oaxaca decomposition methodology that we use in this thesis, let us derive a generic model. The children are divided into two groups (based on each benchmark characteristic). Let us call them group *A* and *B*. The outcome variable here in equation 2 is the raw (and standardized) mathematics test scores (EA) of the children collected during the time of fourth and third round survey. The objective of this equation is to estimate the determinants that determine/influence the difference in mean learning outcome between group A and group B children. The difference of the mean test scores between the two groups (*A* and *B*) can be expressed as:

$$D = E(EA_A) - E(EA_B) \qquad (3)$$

Where $E(EA)$ denotes the expected (mean) value of the mathematics test score. D denotes the difference in the test score between the two groups. Based on the above discussion we can write the generalized linear equation as:

$$EA_i = X_i'\gamma_i + \varphi_i, \tag{4}$$

Where, $E(\varphi_i) = 0, i \in (A, B)$ and X is a vector of all the predictors that includes different explanatory variable categories from Table 1.

Equation (2) can be re-written for the two groups separately based on the discussion. We get the following two regression equation for the two sub-groups:

$$EA_A = \alpha_0 + \alpha_1 PE + \alpha_2 H + \alpha_3 C + \alpha_4 O + \omega_i$$

and,

$$EA_B = \beta_0 + \beta_1 PE + \beta_2 H + \beta_3 C + \beta_4 O + \vartheta_i$$

The vector X contains set of predictors (from Table 1) and a constant, Jann (2008), α_is and β_is are the slope parameters and the intercept, and ω_i and $\vartheta_i$are the error terms. The difference in the mean of learning outcomes can be expressed as the difference in the linear prediction at the group specific mean of the regressors.

$$D = E(EA_A) - E(EA_B)$$

$$= E(X_A)'\alpha - E(X_B)'\beta$$

Since,

$$E(EA_i) = E(X_i\gamma_i + \varphi_i)$$

$$= E(X_i'\gamma_i) + E(\varphi_i)$$

$$= E(X_i)'\gamma_i$$

Because, $E(\varphi_i) = 0$, and $E(\gamma_i) = \gamma_i$, by assumption.

Following (Blinder, 1973; Oaxaca, 1973; Jann, 2008) the contribution of the group difference in the predictors to the overall outcome difference can be rearranged as follows:

$$D = \{E(X_A) - E(X_B)\}'\beta + E(X_B)'(\alpha - \beta) + \{E(X_A) - E(X_B)\}'(\alpha - \beta)$$

This is referred to as threefold decomposition. To explain the results obtained from the decomposition we need a reference category. Let us say we take group *A* as that reference group. The point of analysis will be group *B*, as we will see in the explanation of the three components of the previous equation.

$$D = E + C + I$$

The first component,

$$E = \{E(X_A) - E(X_B)\}'\beta$$

Amounts to the part of the difference that is due to the group difference in the predictors which is called the "endowment effects". That part of the difference in the mean learning outcome of *A* and *B* which is explained because of the differences in the mean values of the background characteristics. In other words, if group *B* had had same average characteristics as group *A*, given their returns (β) stays the same, this is how much their scores would have improved. Differences in the average features applied to the impacts of group *B*.

The second component,

$$C = E(X_B)'(\alpha - \beta)$$

Measures the contribution of the differences in the coefficients which includes also the intercept. It captures the different returns to individual features for each group. That part of the difference in the learning outcome of these two groups of children explained by the difference between the groups' impact of the background characteristics If group B had had same returns as A, assuming their average features (X_B) stays at current level, this is how much their scores would have improved. Difference in the returns applied to the average features of group B children.

And the third component,

$$I = \{E(X_A) - E(X_B)\}'(\alpha - \beta)$$

This is an interaction term which accounts for the fact that differences in endowments and coefficients exist simultaneously between the two groups and is causing the difference in the mean mathematics learning simultaneously. The study estimates the threefold decomposition and discusses about the three effects in detail at both the survey points. Based on this discussion, we will get three different equations for each benchmark characteristics.

$$D_0 = \{E(X_+) - E(X_0)\}'\beta + E(X_0)'(\alpha - \beta) + \{E(X_+) - E(X_0)\}'(\alpha - \beta)$$

$$D_1 = \{E(X_{Pr}) - E(X_{Pub})\}'\beta + E(X_{Pub})'(\alpha - \beta) + \{E(X_{Pr}) - E(X_{Pub})\}'(\alpha - \beta)$$

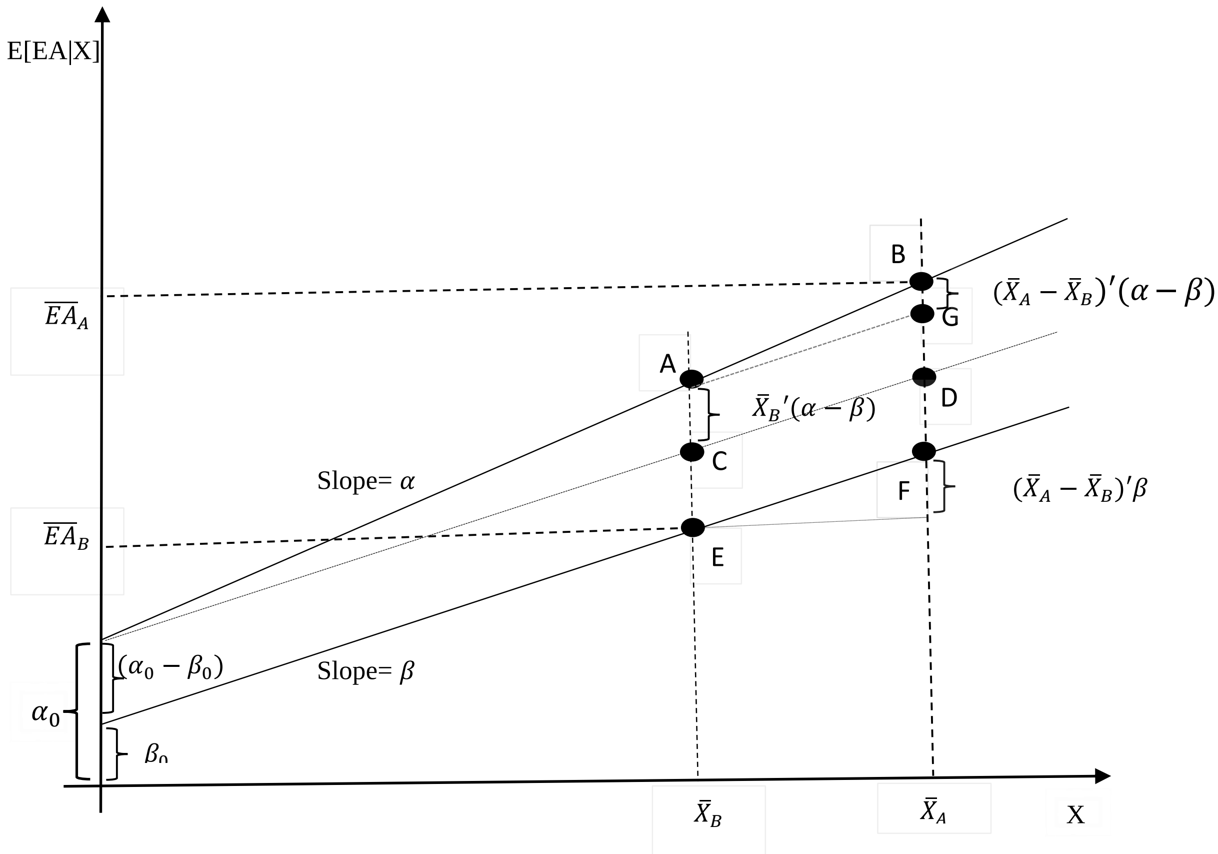
$$D_2 = \{E(X_F) - E(X_R)\}'\beta + E(X_R)'(\alpha - \beta) + \{E(X_F) - E(X_R)\}'(\alpha - \beta)$$

The βs are the coefficients for the second group(s) of children in each of the equations mentioned. Those groups are (the group on which the point of analysis is based) children of non-schooled parents (*0*), public school children (*Pub)*, and rest of the children (*R*). The first of the three equations above, decomposes the difference in mean mathematics learning between the children of schooled parents and non-schooled parents. The second equation decomposes the difference in the mean mathematics learning between the private and public school children. The last one breaks down the difference in mean learning between the first division scoring and rest of the children.

The previous studies have largely focused on the factors influencing the learning outcomes not the extent to which they cause or explain the differences in the learning outcomes between groups of children. Our thesis is an attempt to add to the growing body of literature on the differences in average learning outcome on three different basis (the benchmark characteristics) to point out the contributors to the differences in learning. The need to study these specific children arise from the growing reports (National Achievement Survey and ASER) suggesting that the mathematics learning of the children of united Andhra Pradesh is significantly better than the national average which calls for particular research attention. Our study is for a specific population at two specific points in time i.e. when they were 15-16 and 19-20 years old (during the third and fourth round surveys respectively). The attempt of this thesis is to look into the learning inequality in mathematics between the groups of children with respect to a set of background characteristics. The thesis aims to place those learning inequalities in the larger context of their background differences. The idea is also to see whether their respective differences (in the average background features, if any) explain the differences in the average learning outcomes and if so, to what extent they explain their differences in the performance. This we have done using the threefold Blinder-Oaxaca decomposition. There have been studies which have used the Blinder-Oaxaca decomposition to study the achievement gaps among children. For instance, Ammermuller (2007b), using the PISA-2000, has studied the achievement gap between the German and Finnish children. Zhang and Lee (2011) have also used this decomposition technique, using PISA-2006, to study the achievement gap between the OECD countries. Burger (2011) too studies the achievement gaps between the urban and rural Zambian children using the SAMCEQ-II (South Africa Consortium for Monitoring Educational Quality) dataset. Baird (2012) studies the achievement gaps (using TIMSS for 2003) between the children with high socio-economic status (SES) and lower SES among 19 high income countries. Another interesting study using the Blinder-Oaxaca technique is by Barrera-Osorio *et. al.* (2011) which studies the achievement improvement of the Indonesian children between two points using PISA 2003 and 2006. The two point studies that we have done in the context of India are rare and this is where this study fills the gap in literature. In this study, we further explore the Indian children using YLS for older (Round 3 and Round 4) cohort. The studies of similar sort are minimal that try to explain the learning outcome gaps within Indian children. Using the threefold Blinder-Oaxaca decomposition we break down the mean mathematics learning gap between them at two points based on the aforementioned benchmark characteristics. The Blinder-Oaxaca decomposition has been previously used (in Ammermueller, 2007b; Zhang and Lee, 2011; Barrera-Osorio *et al.*, 2011; Burger, 2011; Baird, 2012; Arteaga and Glewwe, 2014; Glewwe *et al.*, 2015; Sanfo and Ogawa 2021) to study the learning outcome differences between groups of children.

The figure 1 portrays these three effects of Blinder-Oaxaca decomposition borrowing from Arteaga and Glewwe (2014) which has been modified slightly with respect to our study. The decomposition of the test score gap between the groups at a point is shown in the figure 1. The first composition of the equation, the endowment effect, is shown by the move from E to F. This shows the increase in the score of an average *B* group child if they had similar characteristics as an average *A* group child. The second term of the equation, the coefficient effect is shown by the vertical distance between A and C which is the increase in the test scores of an average *B* group child if they had similar returns as an average *A* group child. The last part of the equation, the interaction of the previous two effects, is shown by the distance between B and G.

Figure 1: Decomposition of Test score gaps between children



The following table 4 describes the whole dataset where it has been pointed out about the availability of information on the variables. As we see that there are several variables which have missing information which has eventually reduced the study sample to just 522. We have used only those children on whom the information is available on all the variables at both of the survey points.

Table 4: Descriptive statistics (full dataset) of variables for each round

	Round 3			*Round 4*		
Variables	**Mean**	**Std. Dev.**	**N**	**Mean**	**Std. Dev.**	**N**
Math Score	9.47	6.35	920	14.01	7.256	899
Mothers Education	3.47	4.28	642	3.47	4.28	642
Fathers Education	5.48	4.98	642	5.48	4.98	642
Household Size	5.05	1.90	976	4.72	1.91	952
Assets*	7886.948	44862.07	808	26701.03	38720.52	933
School Cost*	2534.85	5542.55	800	20287.16	31761.27	581
Expenditure*	1013.14	705.87	976	7.17	0.58	951
Years	8.80	7.42	971	13.51	1.72	581
Time Allocation	2.01	1.54	975	2.03	1.46	642
English Score	-	-	-	15.08	4.47	886
Gender	0.49	0.50	976	0.49	0.50	952
Age in Months	179.25	4.07	975	227.91	4.14	944
Body Mass Index	17.62	2.77	974	23.81	95.93	943
Drinking Water	0.97	0.17	975	0.98	0.13	952
Household Quality	0.58	0.26	973	0.71	0.23	952
Type of School#	-	-	-	-	-	-

Types of Schools	*Round 3*		*Round 4*	
	N	*%*	*N*	*%*
Public (Government)	667	68.900	276	40.07
Private (Unaided)	294	30.370	337	49.05
NGO/Charity/Not for Profit/ Religious	4	0.004	2	0.29
Informal or Non-formal	-	-	3	0.44
Charitable Trust	2	0.002	1	0.15
Mix of Public Private	-	-	65	9.46
India Bridge School	1	0.001	-	-
Others	-	-	3	0.44
Total	**968**		**687**	

*the definition of these variables is slightly different for each round so we have to be careful in interpreting these variables.

the details of the descriptive statistics only on public and private schools is shown in Table 2.

3.0 Parental Education and Child's Mathematics Achievement: Evidence from United Andhra Pradesh (India)

Abstract

This chapter looks into the mathematics learning inequality among the Indian children using the fourth round of Young Lives Survey data. The mathematics learning inequality between the children of schooled and non-schooled parents has been explored and placing them in the larger context of their background features has been done here. The children of schooled parents have better characteristics as compared to the children of non-schooled parents at the time of fourth round survey. The chapter finds, using threefold Blinder-Oaxaca decomposition, that these differences in their backgrounds play a crucial role in determining the extent of the gap in mathematics learning. The crucial findings in this chapter suggests that the differences in the background characteristics explain a part of the learning difference between the groups of children. Moreover, the differential returns/impacts to those background features explain the inequality in learning as well. The schooling cost (spent in the last academic year) is the consistent contributor to the learning difference between them. Moreover, the gap in average years of schooling between the children of schooled and non-schooled parents also contributes to the learning difference. English language score difference between them is another important contributor to the learning difference on mathematics. Gender of the child as well as the household quality explain a marginal part of the learning difference between them as well.

Keywords: Mathematics Achievement; India; Mothers' Education; Blinder-Oaxaca Decomposition; Endowment Effect; Coefficient Effect

3.0.1 Introduction

The Trends in Mathematics and Science Study (TIMSS) were used on the secondary school going Indian children from the states of Rajasthan and Orissa. The findings suggest that the Indian children performed significantly poorly as compared to the international average (Kingdon, 2007). There are various studies that talk about what influences academic achievement of children. The Annual Statistics of Education Report (ASER), 2005 concludes that the increase in the financial resources towards education do not reflect in terms of children's learning. ASER's latter report also finds a decline in the ability to do basic math nationally, across classes (ASER, 2010). The previous literature has captured the quality of education in India using children's test scores on various subjects (Kingdon, 1996; Goyal, 2009; Goyal and Pandey, 2009; Singh, 2015; Singh and Mukherjee, 2019). One of the core factor that influences the learning outcome of a child is the parental education which has been frequently discussed in the previous literature across varied contexts (Crane, 1996; Davis, 2005; Paxson and Schady, 2007; Brown and Iyengar, 2008; Desai *et al.,* 2008; Goyal, 2009; Barrera-Osorio *et al.,* 2011; Burger, 2011; Holmlund *et al.,* 2011; Arteaga and Glewwe, 2014; Crookston *et al.,* 2014; Sharma, 2014; Glewwe *et al.,* 2015; Li and Qiu, 2018; Sanfo and Ogawa 2021). In this chapter, we explore the mathematics learning outcome of Indian children with respect to the parental schooling status, using the fourth round dataset of older cohort of Young Lives Survey (YLS/YL henceforth). There have been studies that have talked about the connection between the parental education and learning outcome of children in the context of India (for instance Crookston *et al.,* 2014; Desai *et al.,* 2008; Goyal, 2009; Brown and Iyengar, 2008; Sharma, 2014).

As mentioned before, we have used the YL dataset, which has been previously used to study the learning outcomes of children. Dercon and Krishnan (2009) and Crookston *et al.,* (2014) have studied several learning aspects of YL children of all the four surveyed countries (Ethiopia, India, Peru, and Vietnam). Arteaga and Glewwe (2014), using the Peruvian YL data, explore the learning inequality between the indigenous and non-indigenous children. Similarly, Glewwe *et al.,* (2015) studies the learning difference among the Vietnamese children. Using the Vietnamese YL dataset, they explore the learning differential between the Kinh community and ethnic minority children. Singh and Mukherjee (2019) have used the Indian YL data and explored the connection between attending a particular type of school and learning. Ethiopian children's dataset is used in Sanfo and Ogawa (2021) to look into the learning difference between urban and rural children.

There have been several countries that have been researched in the previous literature on these aspects. Among the lower income countries, several socio-economic aspects have been talked about in relation to the learning outcome of children. Zhang (2006), for instance, looks into the learning disadvantages of children from Sub-Saharan Africa. The Zambian children's learning outcome (in Burger, 2011; Das *et al.,* 2013) has been studied as well. Sakellariou (2008) have studied the learning difference between the indigenous and non-indigenous Peruvian children, similar to Arteaga and Glewwe (2014). The connection between the parental education and the learning outcome of Ecuadorian children (in Paxson and Schady, 2007) has been studied as well. Chinese children (in Li and Qiu, 2018) and Indonesian children (in Barrera-Osorio *et al.,* 2011) are discussed in the previous literature as well.

Similarly, the learning outcomes of children from high income countries have been studied in depth as well. Among those studies, the children from the United States of America (in Leibowitz, 1977; Crane, 1996; Currie and Yelowitz, 2000; Davis, 2005; Todd and Wolpin, 2007) have been frequently studied. The learning outcome of Australian children (in Cobb et al., 2014) have been studied with respect to the background features of children. Similarly, the learning outcome of children from Germany and Finland has been studied in depth (Ammermueller, 2007b). On the same line, learning outcome of children from the OECD countries (Zhang and Lee, 2011) as well as Swedish children (Holmlund et al., 2011) have been researched too.

Although the previous literature has talked about several background characteristics in relation to the learning outcome of children, there are very few studies that decompose the learning difference between certain groups of children to explain what causes that difference in the first place. In this chapter, we add on that aspect, where we attempt to explain the learning difference between the children of schooled mothers (fathers) and non-schooled mothers (fathers). We specifically look into that learning difference with respect to the background characteristics and try to associate/decompose the difference into its explainers. The studies mentioned before have largely focused on the factors that have an influence on the learning outcome of children. The extent and direction of influence have been talked about in detail in these studies. In this chapter, we slightly depart from these literature and add to the growing body of literature on learning difference decomposition. Keeping the parental schooling status as the benchmark for categorization of children, we put the children's background features as against their learning outcomes to understand the association. This chapter is for a specific set of population at a specific point in time, when the children were 19-20 years old (during the fourth round YL survey). The children in the YL dataset are growing up in the context of poverty and the dataset traces the changes occurring in their lives over a period of time. The idea in our study is to see if there is any significant background difference in the background characteristics between the children of schooled mothers (fathers) and non-schooled mothers (fathers). To see whether these differences (if any) in the background characteristics explain the learning difference between the groups and to what extent. To explore these questions, we have made use of the threefold Blinder-Oaxaca decomposition technique. This technique has been previously used in the previous literature to study the learning difference between groups of children. For instance, Ammermuller (2007b), using the PISA-2000, has studied the achievement gap between the German and Finnish children. Zhang and Lee (2011) have also used this decomposition technique, using PISA-2006, to study the achievement gap between the OECD countries. Burger (2011) too studies the achievement differences between the urban and rural Zambian children using the SAMCEQ-II (South Africa Consortium for Monitoring Educational Quality) dataset. Baird (2012) studies the achievement gaps (using TIMSS for 2003) between the children with high socio-economic status (SES) and lower SES among 19 high income countries. Another interesting study using the Blinder-Oaxaca technique is by Barrera-Osorio et. al. (2011) which studies the achievement improvement of the Indonesian children between two points using PISA 2003 and 2006. In our chapter, we decompose the average learning outcome difference between the private and public school children and attribute it to the factors responsible for explaining the difference at this point.

The rest of the chapter is organized as follows. Section 2 gives the description of data, source and variables used in our study. Section 3 presents the theoretical framework, in section 4 we present the decomposition methodology and decomposition results followed by summary and conclusion of the chapter in Section 5.

3.0.2 Data and Variables

The study uses the longitudinal survey conducted by the University of Oxford. The survey includes four developing countries Ethiopia, India, Peru, and Vietnam. The survey has so far done five rounds across these countries and traces of the lives of 12000 children growing up in the context of poverty. The 3000 children in the Indian survey were selected from 20 sentinel sites spread now across two southern states of Andhra Pradesh and Telangana. The sampling method deployed in the data collection is a form of purposive sampling (Galab *et. al.* 2014). The chapter has used of the fourth round data for the older cohort on India (united Andhra Pradesh) which was released in 2014. This study uses the data on the older cohort for the fourth round when the children were around 20 years of age. The round covers 952 children from the older cohort. Our study had to drop out observations as many of the information on various variables were missing. Eventually we ended up with 522 observations for the analysis. More detailed information on the YLS is available in the previous chapter. The Table 1 gives the description of the variables used in the analysis and Table 2 gives the descriptive statistics of the variables.

Table 1: Description of the Variables

Variables		*Description*
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of the survey out of 30 marks. This score has also been standardized using the standard deviation of the whole sample for comparison and ease of understanding.
Parental Education (PE)[25]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets	Value of the five most valuable assets owned in the YL child's Household
	School Cost	Total expenditure incurred on school in the last academic year.
	Expenditure	The log of per capita monthly expenditure of YL child's household
Child Specific Features (C)	Years	Years of schooling received by YL child at the time of fourth round survey
	Type of School	=1 if attended/attending public school; 0 if private.
	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of fourth round survey
	Body Mass Index	Body mass index of the YL child at the time of fourth round survey
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[26]

In our thesis, the outcome of interest is the mathematics test scores of children at the fourth round survey point. The YL has collected the information on child's achievement with the help of language and mathematics tests (Galab *et al.* 2014a). The mathematics test that we use in this study was designed for each round separately and there were certain similarities between the rounds. More details on the test content and the questions that were asked to the surveyed children is available in the previous chapter in detail.

[25] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[26] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

Table 2: Descriptive Statistics (means) of overall and sub-samples (Round4)

Variable Means	Overall	Schooled and Non-Schooled Mothers (respectively)		Schooled and Non-Schooled fathers (respectively)	
Math Score	14.479	16.319	12.521	15.091	13.232
	(7.074)	(6.644)	(7.004)	(7.028)	(7.024)
Fathers Education	5.644	8.160	2.968	8.417	-
	(5.045)	(4.580)	(4.047)	(3.820)	-
Mothers Education	3.632	7.048	-	4.926	1
	(4.314)	(3.466)	-	(4.429)	(2.506)
Household Size	4.753	4.691	4.818	4.714	4.831
	(1.908)	(1.856)	(1.963)	(1.851)	(2.023)
Assets	30018.64	34932.97	24793.52	36274.49	17288.72
	(44478.14)	(43939.72)	(44537.56)	(50215.62)	(25257.2)
School Cost	19226.17	25337.38	12728.48	23184.22	11171.99
	(26380.52)	(30585.59)	(19018)	(29750.13)	(14719.68)
Expenditure	7.170	7.240	7.095	7.237	7.035
	(.588)	(.575)	(.593)	(.587)	(.568)
Years	13.448	13.869	13	13.58	13.180
	(1.745)	(1.569)	(1.813)	(1.711)	(1.789)
Time Allocation	2.013	2.316	1.691	2.1	1.837
	(1.441)	(1.437)	(1.377)	(1.467)	(1.375)
English Score	16.513	17.405	15.565	17.111	15.297
	(3.306)	(3.104)	(3.255)	(3.073)	(3.437)
Gender	.550	.509	.588	.497	.651
	(.498)	(.501)	(.492)	(.501)	(.478)
Age in Months	227.877	227.579	228.193	227.666	228.308
	(4.134)	(4.194)	(4.053)	(4.153)	(4.073)
Body Mass Index	19.933	19.966	19.899	20.466	18.851
	(8.743)	(4.295)	(11.765)	(10.501)	(2.474)
Drinking Water	.980	.992	.968	.986	.971
	(.137)	(.086)	(.175)	(.119)	(.168)
Household Quality	.725	.754	.693	.756	.663
	(.217)	(.200)	(.2297)	(.187)	(.258)
Type Of School	.425	.308	.549	.366	.547
	(.494)	(.462)	(.498)	(.482)	(.499)
Observations	**522**	**269**	**253**	**350**	**172**

Note: Standard Deviations are in parentheses

3.0.3 Theoretical Framework

Based on the objective of the chapter, we employ behavioral framework in order to carry out the empirical analysis. There are various factors that have an influence on the mathematics achievement of the school going children based on the review of literature. We have included several variables which are presented in Table 1. The description of which is given against each of these variable. The mathematics achievement (EA) can be depicted as a structural relationship between various variables. The input variables in this relationship includes the parental education (PE), household variables (H), child specific variables (C), and **o**thers (O).

$$EA = f(PE, H, C, O) \tag{1}$$

The regression would take the following form:

$$EA = \alpha_0 + \alpha_1(PE) + \alpha_2(H) + \alpha_3(C) + \alpha_{4(}O) + \epsilon_i \tag{2}$$

The α_is are the row vectors of the coefficients that captures the total influence (impact) of the column vectors of set of variables in the bracket.

3.0.4 Methodology for Decomposition and Decomposition Analysis

The sub-samples are the children whose mothers (and fathers) have positive ($_+$) years of education/schooling and the rest whose mothers (and fathers) have no ($_0$) education/schooling. The subsample descriptive statistics (Table 3) presents a difference in the mean mathematics outcome of both the groups where the YL children whose mothers (fathers) have positive years of education have greater mean mathematics scores compared to those whose mothers (fathers) do not have any education. Mother (R4) in Table 3 represents the difference in the mean value of each of the background variable between the children of schooled and non-schooled mothers from Table 2. Similarly, Father (R4) captures the difference between the children of schooled and non-schooled fathers. We see from table 3 that children whose parents are schooled have better mathematics score along with better background characteristics compared to the children whose parents are non-schooled. These differences in background characteristics are highly significant between them.

Table 3 Sub-sample difference in mean background features

Variable	Overall Mean	Difference in mean background characteristics between the children of Schooled mother (and father) and non-schooled mother (and father)	
	R 4	Mother (R4)	Father (R4)
Math Score	14.479 (7.074)	3.797***	1.859***
Fathers Education	5.644 (5.045)	5.191***	-
Mothers Education	3.632 (4.314)	-	3.936***
Household Size	4.753 (1.908)	-.126	-0.117
Assets	30018.64 (44478.14)	10139.46***	18985.76***
School Cost	19226.17 (26380.52)	12608.9***	12012***
Expenditure	7.170 (.588)	.145***	0.202***
Years	13.448 (1.745)	.870***	0.400***
Time Allocation	2.013 (1.441)	.624***	0.263**
English Score	16.513 (3.306)	1.840***	1.814***
Gender	.550 (.498)	-.0796**	-0.156***
Age in Months	227.877 (4.134)	-.614	-0.642**
Body Mass Index	19.933 (8.743)	.067	1.615**
Drinking Water	.980 (.137)	.0242**	0.015
Household Quality	.725 (.217)	.061***	0.092***
Type Of School	.425 (.494)	-.240***	-0.181***
Observations		**522**	

Note: ***p<0.01, ** p<0.05, * p<0.10: Standard Deviations are in parentheses
The mean values of each background variables for each group is shown in Table 2

The Table 4 summarizes the outcome variable, the mathematics test score, which in this table is standardized. The normalization of the scores is done by dividing the respective group scores by the standard deviation of overall sample for the ease of comparison. The test score is greater for the children whose mothers (fathers) have positive years of education as compared to those children whose mothers (fathers) do not have any education. We learn from here that the mathematics learning gap is significant between the children based on each parents' schooling status. We learn from here that the difference is more magnified between the children of schooled and non-schooled mothers compared to the difference between the children of schooled and non-schooled fathers.

Table 4: The Mean Math Test score by the groups (Round 4)

	Overall	Schooled Mothers (+)	Non-Schooled Mothers (0)	Difference (+)-(0)	Schooled Fathers (+)	Non-Schooled Fathers (0)	Difference (+)-(0)
Math Score (Mean)	2.04	2.30	1.771	0.537***	2.135	1.872	0.263***
S. Dev.	1.00	0.94	0.991	(0.084)	0.994	0.994	(0.093)
N	522	269	253		350	172	

Note: ***p<0.01; Figures in parenthesis are standard errors; The test scores have been normalized by dividing it with the standard deviation of the overall sample for ease.

We find evidence that the features of these two groups of children are significantly different. How much of this difference explains the difference in the mathematics score is what we deal with in the remaining section. The primary objective of the chapter is to study the difference in the mean learning on mathematics between the children of schooled and non-schooled parents by decomposing it with respect to a set of background characteristics. We have done this using the threefold Blinder-Oaxaca decomposition to place the difference in mean learning in a larger context.

As we have divided the data into two groups which are the mothers/fathers who have positive years of education ($_+$) and the rest whose mothers/fathers have no education ($_0$). The outcome variable here is the mathematics test scores (EA) of the children collected during the time of fourth round survey. The difference of the mean test scores between the two groups can be expressed as:

$$D = E(EA_+) - E(EA_0) \tag{3}$$

Using (2) for each group separately (and following Blinder, 1973; Oaxaca, 1973; Jann, 2008) the contribution of the group difference in the predictors to the overall outcome difference can be rearranged as follows:

$$D = \{E(X_+) - E(X_0)\}'\beta + E(X_0)'(\alpha - \beta) - \{E(X_+) - E(X_0)\}'(\alpha - \beta)$$

This is referred to as threefold decomposition. A more detailed derivation and discussion of the equation is discussed in the previous chapter. The reference category in our analysis is the children of private schools and the point of analysis in the empirical results is the children of public schools.

$$D = E + C + I$$

The first component,

$$E = \{E(X_{+}) - E(X_{0})\}'\beta$$

The part of the learning difference which is due to the group difference in the mean background features and it is called "endowments effect". This means that if the children of non-schooled mothers (fathers) had had similar average characteristics ($\bar{X}_{+}$) as the children of schooled mothers (fathers), keeping their returns (βs) at the current level, this is how much their scores would have improved. In other words, the difference in mean background characteristics of the children of schooled mothers (fathers) and the children of non-schooled mothers (fathers) applied to the impact of the children of non-schooled mothers (fathers).

$$C = E(X_{0})'(\alpha - \beta)$$

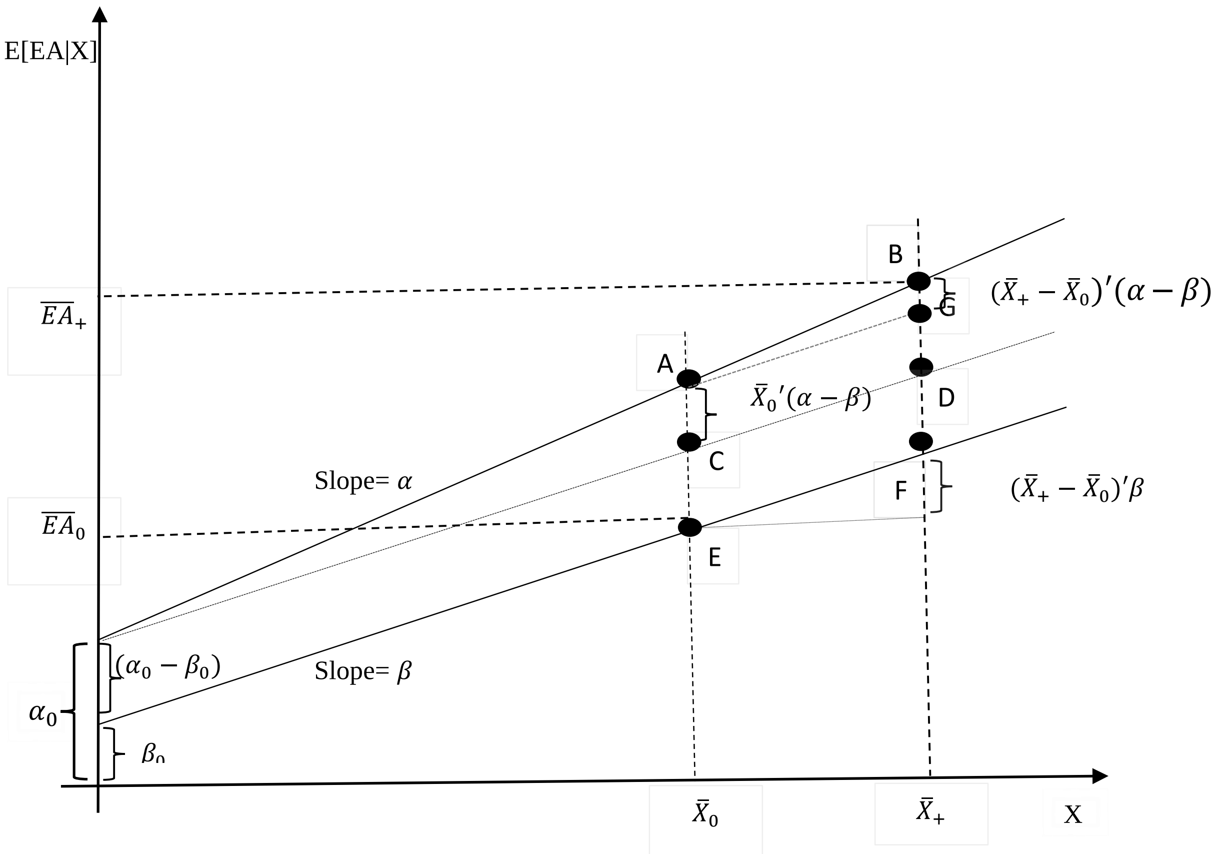
This part measures the contribution to the learning gap by the differences in the coefficients/returns $(\alpha - \beta)$ of the two groups which also includes the intercepts. The part of the gap in learning outcome of the two groups of children explained by the differences in groups' impact of the background features. If the children of non-schooled *(0)* mothers (fathers) had had similar average returns (α) as the children of schooled *(+)* mothers (fathers), keeping their features *($\bar{X}_{0}$)* at the current level, this is how much their scores would have improved. The gap in the average returns to the background features applied to the average features of the children of non-schooled mothers (fathers).

$$I = \{E(X_{+}) - E(X_{0})\}'(\alpha - \beta)$$

And the third component, is an interaction term which accounts for the fact that the differences in endowments and coefficients exist simultaneously between the two groups. In our study, we estimate these three effects (at two points) based both on the schooling status of mothers and fathers and discuss the results in detail.

These three effects (from equation 2) are portrayed in figure 1, which has been borrowed from Arteaga and Glewwe (2014) and has been modified slightly based on our needs. The decomposition of tests score gap between the groups of children (in total, four based on each parents schooling status) at each point is shown. The first component, endowment effect is shown by the move from point E to F. This shows the increase in the test score of an average child of non-schooled mother (father) if they had had similar average features as the child of schooled mother (father). The second term of the equation, the coefficient effect is shown by the vertical distance between A and C, which is the increase in the test score of an average child of a non-schooled mother (father) if they had similar returns as an child of a schooled mother (father). The interaction effect, which is the last part of the equation, an interaction of the previous two effects is displayed by the distance between B and G.

Figure 1: Decomposition of Test score gaps between children



3.0.4.1 Empirical Results

The Table 5 shows the actual Blinder-Oaxaca decomposition of the mean difference in the mathematics test scores for the two groups (based each on individual parents' schooling status). The idea as derived in the methodology was to break the difference into its various components which would tell how much of the difference is caused by the differences in the endowment of the two groups, how much of the difference is because of the differences in the coefficients (or returns to the background characteristics), and how much is because of the interaction of endowments and coefficients. The difference in the mean test score of the two groups (on the basis of mothers' schooling) is approximately 3.79 points (or 0.537 standard deviations). This difference on the basis of fathers schooling stands at 1.859 points (or 0.263 standard deviations).

The first thing that we notice from table 5,which we have pointed out earlier as well, that this difference is more magnified based on mothers' schooling status as compared to the difference based on the schooling status of fathers. On the standardized score, the learning gap stood at 0.537 standard deviation between the children of schooled and non-schooled mothers. Whereas, this difference between the children of schooled and non-schooled fathers stood at 0.263 standard deviation. In the same table we have also shown the results on the decomposition of these learning difference between these groups of children using then threefold Blinder Oaxaca decomposition. We see that the differences in the endowments (which we talked about earlier from Table 4) is a consistent influencer of the learning gap between these groups of children. What this effect (endowments) means is that if the children of non-schooled parents had had similar average features as the children of schooled parents ($\bar{X}_{+}$), keeping the returns β for them at the current level, a large part of the learning difference between them would have been removed. We also learn from here that this effect is stronger between the children of schooled and non-schooled fathers as compared the same between the children of schooled and non-schooled mothers. This removal of differences in endowments would have removed the gap in learning between the children of schooled and non-schooled mothers by 65 percent (0.35 standard deviation out of total 0.537 standard deviation). Whereas this removal would have actually made the children of non-schooled fathers to perform better than the children of schooled fathers. In sum the differences in endowments is causing the children of non-schooled parents to underperform in general. This is also leading to the children of non-schooled fathers to underperform and stopping them to outperform the children of schooled fathers. The differences in returns to background characteristics does not operate between the children of schooled and non-schooled fathers, however, it does have a role to play between the children of schooled and non-schooled mothers. 32 percent (0.167) of the total difference in learning (0.537 standard deviation) is explained by the differential returns to the background characteristics in favor of the children of schooled mothers.

Table 5: The Decomposition Results

	Mothers		Fathers	
	Coefficient (on raw score)	**Coefficients (on Standardized Math Score)**	**Coefficient (on raw score)**	**Coefficients (on Standardized Math Score)**
	R4	**R4**	**R4**	**R4**
Mean Score-Schooled (+)	16.320	2.308	15.091	2.135
	(0.000)	(0.000)	(0.000)	(0.000)
Mean Score- Non-Schooled (0)	12.522	1.771	13.233	1.872
	(0.000)	(0.000)	(0.000)	(0.000)
Difference	3.798	.537	1.859	0.263
	(0.000)	(0.000)	(0.005)	(0.005)
Endowment Effect	2.472	.349	2.769	0.392
$\{E(X_+) - E(X_0)\}'\beta$	(0.000)	(0.000)	(0.002)	(0.002)
Coefficient Effect	1.177	.167	-0.472	-0.067
$E(X_0)'(\alpha - \beta)$	(0.047)	(0.047)	(0.404)	(0.404)
Interaction effect	.1489	.021	-0.438	-0.062
$\{E(X_+) - E(X_0)\}'(\alpha - \beta)$	(0.810)	(0.810)	(0.595)	(0.595)

Note*: p-values are in parenthesis.*

Table 6: Detailed Decomposition Results (Round 4)

Variable	Endowments Effect		Coefficient Effect		Interaction Effect	
	Mother	Father	Mother	Father	Mother	Father
Mothers Education	-	.091 (.090)	-	-.025 (.025)	-	-.096 (.098)
Fathers Education	-.077 (.063)	-	-.001 (0.048)	-	-.001 (0.084)	-
Household Size	.004 (0.006)	.004 (.007)	.201 (0.173)	.172 (.186)	-.005 (0.008)	-.004 (.008)
Assets	-.012 (0.012)	-.005 (.044)	.032 (0.040)	-.010 (.043)	.013 (0.017)	-.011 (.047)
School Cost	.064* (0.039)	.091* (.053)	-.020 (0.044)	-.048 (.051)	-.019 (0.043)	-.051 (.055)
Expenditure	.001 (0.014)	.006 (.023)	-.548 (0.902)	-.690 (.951)	-.011 (0.019)	-.020 (.028)
Years	.135*** (0.036)	.079** (.036)	.306 (0.593)	-.675 (.635)	.020 (0.040)	-.020 (.021)
Time Allocation	.0220 (0.025)	-.005 (.012)	.057 (0.085)	.192* (.100)	.021 (0.032)	.027 (.020)
English Score	.211*** (0.044)	.178*** (.045)	-.032 (0.368)	.460 (.367)	-.004 (0.044)	.055 (.044)
Gender	-.030* (0.018)	-.053** (.025)	-.098 (0.077)	-.062 (.097)	.014 (0.013)	.015 (.024)
Age in Months	-.018 (0.013)	-.017 (.013)	-5.105 (3.574)	-3.233 (3.867)	.014 (0.013)	.009 (.012)
Body Mass Index	0 (0.004)	-.023 (.038)	-.026 (0.215)	.140 (.436)	0 (0.001)	.011 (.038)
Drinking Water	-.001 (0.006)	-.001 (.005)	.112 (0.563)	.095 (.461)	.003 (0.014)	.001 (.007)
Household Quality	.029* (0.016)	.041* (.022)	-.006 (0.218)	.0747 .(206)	-.001 (0.019)	.010 (.029)
Type Of School	.023 (0.026)	.005 (.022)	-.042 (0.067)	.028 .071	-.022 (0.036)	.011 (.028)
Constant	-	-	5.350 (3.775)	3.513 (3.961)	-	-
Total	**.349*****	**.392*****	**.167****	**-.067**	**.021**	**-.062**

Note: ***p<0.01, ** p<0.05, *p<0.10: Standard errors are in parentheses

When we look at the detailed decomposition results of the Blinder-Oaxaca decomposition, they identify the individual factors that contribute to the three aforementioned effects. We find that the difference in the average background characteristics and differential returns to those background characteristics is the core cause of the learning difference between the children of schooled and non-schooled parents. We learn from our results that the endowments effect (*E*) explains the majority (65 percent of the total standardized difference) of the learning difference between the children of schooled and non-schooled mothers. Of the total *E* of 0.349 standard deviation between the children of schooled and non-schooled mothers, schooling cost incurred in the last academic year explains 17 percent (.064 standard deviation). This variable explains a part of *E* between the children of schooled and non-schooled fathers (.09 standard deviation or 23 percent of total *E*, 0.392 standard deviation) as well, making schooling cost a consistent explainer of the learning difference between the children of schooled and non-schooled parents. This means that better spending capacity of the children of schooled parents is one core reason behind the gap in mathematics learning between the children of schooled and non-schooled parents. We find one worrying variable which explains the learning difference between the children of schooled and non-schooled parents. The years of schooling which captures the highest grade completed at the time of fourth round survey, explaining a large chunk of *E* between the children of schooled and non-schooled parents. This explains 21 percent of *E* or .08 standard deviation between the children of schooled and non-schooled fathers, whereas it explains 40 percent of *E* or 0.135 standard deviation between the children of schooled and non-schooled mothers. We see that the variable is extremely important in explaining the learning difference between the children of schooled and non-schooled parents. What is worrying here is the existence of gap in average years of schooling between the children of schooled and non-schooled parents in the first place. These children fall in the same age group and the fact that some have greater years of schooling completed compared to others, means that there are several children lagging behind in continuing in education. One interesting result that comes out of the detailed decomposition is the extreme significance of the English language score on the mathematics learning gap between the children of schooled and non-schooled parents. This variable too is consistent where it explains 46 percent of *E* between the children of schooled and non-schooled fathers. Whereas this explains 60 percent of *E* between the children of schooled and non-schooled mothers. We must point this out here that the language in which the questionnaire was asked was in Telegu (local language) as well as English. Moreover, the children of schooled parents were able to perform better on mathematics, partly because they were better versed in English. There are other consistent contributor which explain percent a marginal part of the learning difference through *E*. This includes the gender of a child which explain 8 percent of *E* between the children of schooled and non-schooled mothers. This stands at 12 percent between the children of schooled and non-schooled fathers. The household quality index explains 10 percent of *E* between the children of schooled and non-schooled fathers whereas this explains 8 percent *E* between the children of schooled and non-schooled mothers. We also find that the differential returns to background characteristics have role to play between the children of schooled and non-schooled mothers only. However, our results do not identify the factors that contribute to this effect in our detailed results.

3.0.5 Summary and Conclusions

In this chapter we have focused on the mathematics learning difference between the children of schooled and non-schooled parents. We have made use of the Young Lives Survey for the older cohort of the Indian children. Using the fourth round dataset, we categorize the children based on the schooling status of their parents individually. We see that there is a clear gap in mathematics learning between the children of schooled and non-schooled parents. We then put them together with respect to a set of background characteristics and we first notice that the children of schooled parents have better background characteristics compared to the children of non-schooled parents. To see how much of the mathematics learning between the children of schooled and non-schooled parents is because of these background differences, we apply the Blinder-Oaxaca decomposition. We find that the differences in the background characteristics are the core cause of the learning difference between these groups of children.

We find that the endowments difference between the children of schooled and non-schooled parents is the most important reason that there exist a learning difference between them. Looking at the detailed results of decomposition we find that there are several consistent factors that contribute to the endowments effect (or the learning difference). The foremost contributor to the endowments effect between the children of schooled and non-schooled parents is the schooling cost incurred by the children in the last academic year. We also find that a major part of the endowments effect between these children is because of the existence of a gap in years of schooling between these children which is worrying because these children fall in the same age group. This means that there must be a drop out from education (more so among the children of non-schooled parents) which is bringing down the average years of schooling and magnifying the gap between them. We also find that the English language score difference between the children of schooled and non-schooled parents is also an important contributor to the mathematics learning difference between them. Those children who scored well in English have scored well in mathematics as well. In the test questionnaire the children were asked questions in the native language (Telegu) as well as in English. If the questions were asked only in English, the explanation to the previously mentioned association between mathematics and English could be explained. "Children whose native language is not English scored lower than other children with similar backgrounds" (Leibowitz, 1977). Those children who are better versed in English scored better than those who are not that well versed. The gender of the child as well as the household quality index have a role, though marginal compared to other variables, to play as well in determining the mathematics learning gap between the children of schooled and non-schooled parents. We find that the differential returns to the background characteristics between the children of schooled and non-schooled mothers however, our results do not identify the factors that identify the contributors to *C* here.

The results presented in this chapter give one important policy implication. This pertains to the framing of policies which would make it possible to obtain better schooling quality without having to spend as much. As we have seen in the results that one core reason that the children of schooled parents perform better is because they could spend more on education as against the children of non-schooled parents. There are two important exploratory implication of this chapter. Why is it that the children, though they fall in the same age group, a specific group has lesser average years of schooling. Is this because there is a dropout in this category or because staying in education is costly (or monetary opportunity cost attached to it). This must be further researched and explored to establish as years of schooling variable is a core factor behind the existence of mathematics learning gap. The score on English language and its relationship with the mathematics score needs to be established as we, from our results, cannot say this for certain that the causality is from the side of English language.

3.1 Mathematics Learning Inequality among Indian Children: An Insight into Child Learning at two points with respect to Parental Schooling Status

Abstract

This chapter discusses mathematics learning difference between the children of schooled and non-schooled parents in India, at two points. Using threefold Blinder-Oaxaca decomposition we break down the learning difference into the components that explain this gap. We learn that the primary cause of learning gap is the differences in average background features between them. In that, we also learn that the schooling cost contributes the most to the learning gap. The most important finding is the sudden significance of years of schooling on learning gap, the gap (in years of schooling) on which has magnified between these children in just four years.

Keywords: Mathematics Achievement; India; Parental Education; Blinder-Oaxaca Decomposition; Endowment Effect; Coefficient Effect

3.1.1 Introduction

In this chapter we continue from where we left in the previous chapter. We have only used the fourth round of the YL dataset in the previous chapter. However, in this chapter we add on to the exploration of mathematics learning inequality using the third round of dataset as well. We do this in order to see how the learning inequality looks at two points where the gap between the points is four years. We looked at the learning difference just at one point in time whereas here we repeat the same exercise when the YL children were four years younger. Doing this, we aim to draw a pattern of influence of the background characteristics which have a consistent significant influence on the learning inequality between the children of schooled and non-schooled parents. As we mentioned previously, the Trends in Mathematics and Science Study (TIMSS) were used on secondary school going Indian children from the states of Rajasthan and Orissa. The findings suggest that the Indian children performed differently and significantly poorly as compared to the international average (Kingdon, 2007). The Annual Statistics of Education Report (ASER), 2005 concludes that the increase in financial resources towards education do not reflect in children's learning. ASER's latter report also finds a decline in the ability to do basic math nationally, across classes (ASER, 2010). The previous literature has clearly informed us, across varied contexts (as discussed in detail in the first chapter as well as in the previous chapter) that there is a clear connection between the parental education and child achievement. (Crane, 1996; Davis, 2005; Paxson and Schady, 2007; Brown and Iyengar, 2008; Desai *et al.*, 2008; Goyal, 2009; Barrera-Osorio *et al.*, 2011; Burger, 2011; Holmlund *et al.*, 2011; Arteaga and Glewwe, 2014; Crookston *et al.*, 2014; Sharma, 2014; Glewwe *et al.*, 2015; Li and Qiu, 2018; Sanfo and Ogawa 2021). In this chapter, we explore the mathematics learning outcome of Indian children with respect to the schooling status of their parents. We do this using two rounds of Young Lives Survey dataset for the older cohort in this chapter. There are studies that have talked about the connection between the parental education and learning outcome of children in the context of India (for instance Crookston *et al.,* 2014; Desai *et al.,* 2008; Goyal, 2009; Brown and Iyengar, 2008; Sharma, 2014). In this chapter, we study the mathematics learning inequality between the children of schooled mothers (fathers) and non-schooled mothers (fathers). We explore the mathematics learning inequalities at two points in these children's lives, where the gap between the points is four years. In other words, we delve into the learning inequality of children based on individual parents' schooling status at two points. We explore this inequality in learning with respect to a set of background characteristics of children in the empirical analysis of our research.

The previous literature has talked about several background characteristics in relation to the learning outcome of children (see chapter 1 for more details). There are very few studies on Indian children that decompose the learning difference between certain groups of children to explain what causes that difference. In this chapter, we add on that aspect, where we specifically look into learning gap (on the basis of schooling status of parents) with respect to their background characteristics and try to associate/decompose the learning difference into parts/features that explain those. In this chapter, we slightly depart from these literature and add to the growing body of literature on the decomposition of differences in mean mathematics learning between children of schooled mothers (fathers) and children of non-schooled mothers (fathers). Our chapter is for a specific set of population at two specific points in time, when the children were 15-16 years old and 19-20 years old (during the third and fourth round YL surveys respectively). The children in the YL dataset are growing up in the context of poverty and the dataset traces the changes occurring in their lives over a period of time, since 2002. The idea in our chapter is to see if there is any significant background difference in the average characteristics between the children of schooled mothers (fathers) and non-schooled mothers (fathers). Furthermore, to see whether these differences (if any) in the background characteristics explain the learning difference between the groups and to what extent. To explore these questions, we have made use of threefold Blinder-Oaxaca decomposition technique. This technique has been previously used in the previous literature to study the learning difference between groups of children in many contexts (for instance Ammermueller, 2007b; Zhang and Lee, 2011; Burger, 2011; Baird, 2012; Barrera-Osorio *et al.*, 2011) but not in the context of India. The two point research that we have done in the context of India is rare and this is where our research fills the gap in Indian literature on learning inequality.

The rest of the chapter is organized as follows. Section 2 gives the description of data, source and variables used in our research along with descriptive statistics. Section 3 presents the methodological framework and empirical approach along with the results followed by summary and conclusion of the chapter in Section 4.

3.1.2 Data and Variables

In this chapter, we have used two rounds of Young Lives Survey data for the older cohort of Indian children. The longitudinal survey is a collaborative research project coordinated by the University of Oxford. This survey includes four developing countries, namely, Ethiopia, India, Peru, and Vietnam. To trace the changing nature of poverty, the survey has done five rounds so far, every four years since 2002. The survey sample in the YL is based on a pro-poor tendency which draws an equal number of boys and girls, based both in rural and urban communities. In the Indian survey, the 3000 children were selected from 20 sentinel sites which are now spread across two southern states of, Andhra Pradesh and Telangana. The sentinel site sampling methodology is a form of purposive sampling which represents a certain type of population. As the YL is a longitudinal study, a prolonged contact with the observations had to be maintained. In order to ensure that this has been maintained, the sentinel site methodology was found most suitable.

The third and fourth round dataset for the older cohort of the Indian children were released in 2010 and 2014 respectively. The survey in India was conducted on two age cohorts, young and old. The younger cohort of the surveyed children were 8 months of age when the survey began in 2002 and the older cohort was 8 years old at that time. The older cohort is one-third of the total surveyed children in India, and they were born between Jan'94-June'95. The younger cohort, which makes the remaining sample of children, were born between Jan'01-June'02. Our study uses the dataset for the older cohort for the third and fourth round. In the third round the number of observations available were 977 whereas in the fourth round this was 952 from the older cohort. The attrition rate in the dataset has been kept very low (Singh, 2015). In this study we had to drop out some observations as the information required on various variable was missing. Moreover, we are also considering only those children who have either studied in public or private school. There were various other types of school that were attended by these children apart from these two but the majority were either in public or private schools.[27] The information on various required variables were missing, which is why we ended up with a 522 workable observations from our dataset. If the information on any variable for a child was missing in either round, we dropped those observations as well. In this chapter we have done a two point study of the same set of children with a gap of four years between the points, we had to consider that the required information on all the variables were available for each of these round.

In our chapter, the outcome of interest is the mathematics test scores of children at both of these surveyed points. The YL has collected the information on child's achievement with the help of language and mathematics tests (Galab *et al.* 2014a). This mathematics test that we use in this study was designed for each round separately, however there were certain similarities between the rounds. In round 3, the mathematics test for the older cohort was divided into two sections. In the first one, there were 20 items pertaining to addition, subtraction, division, multiplication, and square roots (of both whole numbers and

[27] The other type of schools that were attended by these children apart from public or private schools included private aided, NGO/charity/Religious, informal or non-formal, charitable trust, bridge schools, and mix of public-private. In this study we have considered only those who have attended/attending purely public or private schools.

fractions). The 10 items of the second section involved mathematics problem solving, geometry, and measurement. Similarly, in the fourth round test, the mathematics test section was divided into five sections which asked questions pertaining to same aspects as the test in the round 3. Since there was a gap of four years between the rounds, difficulty level was adjusted (Dawes, 2020). There were 30 questions in total for each round and the scores were obtained by adding the correct response out of 30 marks. A more detailed information on the content of mathematics test is available in the questionnaire for each round. It must be noted here that there indeed were similarities in these tests between the rounds (Dawes, 2020), the marks obtained cannot be directly compared between rounds (Dawes, 2020; Rolleston, 2014). To make the tests comparable between the rounds, the reliability and validity of these has to be checked. There are studies (for instance Cueto and Leon, 2012; Azubuike *et al.*, 2017) that have done a comparison between rounds using the Classical Test Theory (CTT) and Item Response Theory (IRT). The usage of these tests brought the test scores to a uniform comparable scale making comparisons between the rounds possible. In our study, we have not done a direct comparison between the rounds hence we have not applied these tests here.

Table 1 gives a description of the variables used in our chapter based on the review of literature. Please note that that there are some variables which have slightly different description and composition for both rounds. The descriptive statistics of these variables is presented in Table 2 for each round. Table 2 also shows the descriptive statistics based on the sub-groups of children. The test score that we use in our study has either been standardized or used in its raw form. The standardized test scores (for comparison between groups in a particular round) are shown in Table 3 for the whole sample as well as for the sub-samples. Moreover, this is also shown for both the rounds separately. We can see from this table that there is a gap in average learning based on each parents' schooling status. Moreover, this gap is more magnified among children, based on mothers' schooling status as compared to the children based on fathers' schooling status at both the survey points. In Table 4, we have presented the mean values of each of the variables used in our study for the whole sample as well as the sub-sample. We see that the majority of the variables, on an average, are significantly better for the children of schooled parents at both the survey points. The aim now is to see which of these differences (in characteristics) contribute to the learning difference between the children of schooled mothers (and fathers) and non-schooled mothers (and fathers). The round four part of the tables has also been included in here for comparison, however, this part is discussed in isolation in the previous chapter.

Table 1: Description of the Variables

	Variables	Description
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of surveys out of 30 marks. This has also been standardized using the standard deviation of whole sample in each round for ease of comparison.
Parental Education (PE)[28]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets (Round 4)	Value of the five most valuable assets owned, rented or borrowed in the YL child's household (in rupees)
	Assets (Round 3)	Value of assets owned, rented or borrowed in the YL child's household (in rupees)
	School Cost (Round 4)	Total expenditure (in rupees) incurred on school in the last academic year.[29]
	School Cost (Round 3)	How much has the YL household spent on school fees and extra tuition for the child per year (in rupees). The composition of this cost is different from the round 4 cost.
	Expenditure (Round 4)	The log of per capita monthly expenditure of YL child's household (in rupees)
	Expenditure (Round 3)	The log of real per capita monthly expenditure (in rupees) of YL child's household; base 2006 prices
Child Specific Features (C)	Years	Years of schooling received by YL child at the time of surveys
	Type of School	=1 if attended/attending public school last, 0 if private.
	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey out of 30 marks[30]
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of surveys
	Body Mass Index	Body mass index of the YL child at the time of surveys
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[31]

[28] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[29] In the fourth round this cost is the sum total of the tuition fees, education charges, private tuition, accommodation, transportation, uniforms, stationary etc. in the last academic year. The school level heterogeneity is captured by the variance in the sum total of these costs.

[30] This information is available only in the fourth round survey.

[31] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

Table 2: Descriptive Statistics (mean) of overall and sub-samples

Variable Means	Overall		Schooled Mothers		Non-Schooled Mothers		Schooled Fathers		Non-Schooled Fathers	
	R 4	R3	R 4	R3	R 4	R3	R 4	R3	R4	R3
Math Score	14.479	11.312	16.319	12.996	12.521	9.521	15.091	11.871	13.232	10.174
	(7.074)	(6.156)	(6.644)	(5.950)	(7.004)	(5.869)	(7.028)	(6.280)	(7.024)	(5.749)
Fathers Education	5.644	5.643	8.160	8.160	2.968	2.968	8.417	8.417	-	-
	(5.045)	(5.045)	(4.580)	(4.580)	(4.047)	(4.047)	(3.820)	(3.820)	-	-
Mothers Education	3.632	3.632	7.048	7.048	-	-	4.926	4.926	1	1
	(4.314)	(4.314)	(3.466)	(3.466)	-	-	(4.429)	(4.429)	(2.506)	(2.506)
Household Size	4.753	5.086	4.691	5.115	4.818	5.055	4.714	5.074	4.831	5.110
	(1.908)	(1.974)	(1.856)	(2.245)	(1.963)	(1.641)	(1.851)	(2.079)	(2.023)	(1.749)
Assets	30018.64	9291.452	34932.97	11312.16	24793.52	7142.957	36274.49	13128.39	17288.72	1483.738
	(44478.14)	(52007.87)	(43939.72)	(54614.45)	(44537.56)	(49101.88)	(50215.62)	(63122.31)	(25257.2)	(4194.275)
School Cost	19226.17	2723.567	25337.38	3835	12728.48	1541.846	23184.22	3562.071	11171.99	1017.308
	(26380.52)	(5344.192)	(30585.59)	(6151.854)	(19018)	(4009.821)	(29750.13)	(6106.549)	(14719.68)	(2558.033)
Expenditure	7.170	6.776	7.240	6.866	7.095	6.680	7.237	6.887	7.035	6.552
	(.588)	(.561)	(.575)	(.527)	(.593)	(.580)	(.587)	(.524)	(.568)	(.568)
Years	13.448	8.641	13.869	8.728	13	8.549	13.58	8.666	13.180	8.593
	(1.745)	(1.308)	(1.569)	(1.399)	(1.813)	(1.199)	(1.711)	(1.331)	(1.789)	(1.265)
Time Allocation	2.013	2.626	2.316	2.921	1.691	2.312	2.1	2.703	1.837	2.471
	(1.441)	(1.293)	(1.437)	(1.342)	(1.377)	(1.162)	(1.467)	(1.310)	(1.375)	(1.249)
English Score	16.513	-	17.405	-	15.565	-	17.111	-	15.297	-
	(3.306)	-	(3.104)	-	(3.255)	-	(3.073)	-	(3.437)	-
Gender	.550	.550	.509	.509	.588	.588	.497	.497	.651	.651
	(.498)	(.498)	(.501)	(.500)	(.492)	(.493)	(.501)	(.501)	(.478)	(.478)
Age in Months	227.877	179.197	227.579	178.825	228.193	179.592	227.666	178.954	228.308	179.692
	(4.134)	(4.105)	(4.194)	(4.180)	(4.053)	(3.994)	(4.153)	(4.072)	(4.073)	(4.140)
Body Mass Index	19.933	17.608	19.966	17.969	19.899	17.224	20.466	17.872	18.851	17.07
	(8.743)	(2.767)	(4.295)	(2.875)	(11.765)	(2.598)	(10.501)	(2.899)	(2.474)	(2.399)
Drinking Water	.980	.963	.992	.985	.968	.940	.986	.974	.971	.9419
	(.137)	(.187)	(.086)	(.121)	(.175)	(.236)	(.119)	(.159)	(.168)	(.235)
Household Quality	.725	.604	.754	.622	.693	.585	.756	.631	.663	.550
	(.217)	(.298)	(.200)	(.252)	(.2297)	(.340)	(.187)	(.235)	(.258)	(.393)
Type Of School	.425	.60	.308	.505	.549	.747	.366	.522	.547	.826
	(.494)	(.485)	(.462)	(.500)	(.498)	(.435)	(.482)	(.500)	(.499)	(.381)
Observations	**522**		**269**		**253**		**350**		**172**	

Note: Standard Deviations are in parentheses

Table 3: The Mean Math (standardized) Test score by the groups

	Overall		Schooled Mothers (+)		Non-Schooled Mothers (0)		Difference (+)-(0)		Schooled Fathers (+)		Non-Schooled Fathers (0)		Difference (+)-(0)	
	R4	**R3**	**R4**	**R3**	**R4**	**R3**	**R4**	**R3**	**R4**	**R3**	**R4**	**R3**	**R4**	**R3**
Math Score (Mean)	2.04	1.839	2.30	2.113	1.771	1.548	0.537***	0.565***	2.135	1.930	1.872	1.654	0.263***	0.276***
S. Dev.	1.00	1.00	0.94	0.968	0.991	0.954	(0.084)	(0.084)	0.994	1.02	0.994	.935	(0.093)	(0.093)
N	522		269		253				350		172			

Note: ***p<0.01; Figures in parenthesis are standard errors; The test scores have been normalized by dividing it with the standard deviation of the overall sample for ease

Table 4: The Difference in the Means of explanatory variables between the groups

Variable	Overall Mean		Difference in mean background characteristics between the children of Schooled mothers (and fathers) and non-schooled mothers (and fathers)			
	R 4	R3	Mother (R4)	Mother (R3)	Father (R4)	Father (R3)
Math Score (Raw)	14.479 (7.074)	11.312 (6.156)	3.797***	3.474***	1.859***	1.70***
Fathers Education	5.644 (5.045)	5.643 (5.045)	5.191***	5.191***	-	-
Mothers Education	3.632 (4.314)	3.632 (4.314)	-	-	3.936***	3.936***
Household Size	4.753 (1.908)	5.086 (1.974)	-.126	.0599	-0.117	-0.036
Assets	30018.64 (44478.14)	9291.452 (52007.87)	10139.46***	4169.2	18985.76***	11644***
School Cost	19226.17 (26380.52)	2723.567 (5344.192)	12608.9***	2293.154***	12012***	2544***
Expenditure	7.170 (.588)	6.776 (.561)	.145***	.186***	0.202***	0.335***
Years	13.448 (1.745)	8.641 (1.308)	.870***	.179*	0.400***	0.073
Time Allocation	2.013 (1.441)	2.626 (1.293)	.624***	.609***	0.263**	0.232**
English Score	16.513 (3.306)	- -	1.840***	-	1.814***	-
Gender	.550 (.498)	.550 (.498)	-.0796**	-.0796**	-0.156***	-0.156***
Age in Months	227.877 (4.134)	179.197 (4.105)	-.614	-.767**	-0.642**	-0.738**
Body Mass Index	19.933 (8.743)	17.608 (2.767)	.067	.744***	1.615**	0.800***
Drinking Water	.980 (.137)	.963 (.187)	.0242**	.044***	0.015	0.032**
Household Quality	.725 (.217)	.604 (.298)	.061***	.0376*	0.092***	0.081***
Type Of School	.425 (.494)	.622 (.485)	-.240***	-.241***	-0.181***	-0.303***
Observations	522					

Note: ***p<0.01, ** p<0.05, * p<0.10: Standard Deviations are in parentheses
The mean values of each background variables for each group is shown in Table 2

3.1.3 Methodological Framework and Empirical Approach

As we have mentioned in the previous chapter that we have used the threefold Blinder-Oaxaca decomposition in the analysis, we do not detail the methodology here in the chapter as it has already been done in the previous chapter as well as in the second chapter. The reduced form equation for this decomposition (following Blinder, 1973; Oaxaca, 1973; Jann, 2008) could be depicted as follows:

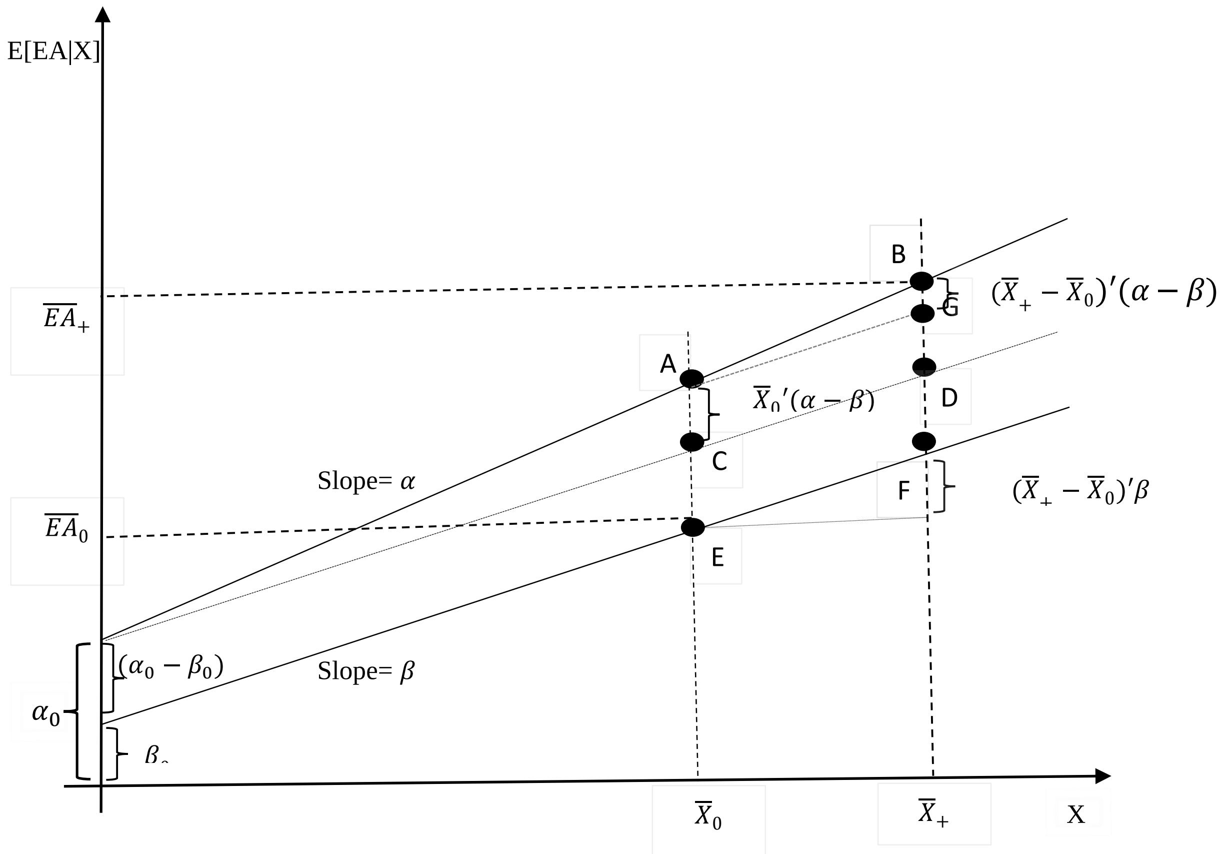
$$D = \{E(X_+) - E(X_0)\}'\beta + E(X_0)'(\alpha - \beta) + \{E(X_+) - E(X_0)\}'(\alpha - \beta) \qquad (2)$$

Here D stands for the difference in the mean mathematics score (raw or standardized) between the children of schooled mothers (and fathers) and non-schooled mothers (and fathers). *Xs* are the set of (vector) of background characteristics, β is the vector of slope coefficients for the children of non-schooled mothers (fathers) and α is for the children of schooled mothers (fathers). (+) denotes that the parents have been in school as against (0) which means they have no years of schooling. The coefficients include intercepts as well. We estimate Equation (2) separately, first, based on the schooling status of mothers followed by the schooling status of fathers. In the empirical analysis we need a reference category from the two groups (+ and 0) and we have taken (+) children to be that and the point of analysis is (0) children.

$$D = E + C + I$$

The three effects (at one point) have been described in detail in the previous chapter already. These three effects (from equation 2) are portrayed in figure 1, which has been borrowed from Arteaga and Glewwe (2014) and has been modified slightly based on our needs. The decomposition of tests score gap between the groups of children (in total, four based on each parents schooling status) at each point is shown. The first component, endowment effect is shown by the move from point E to F. This shows the increase in the test score of an average child of non-schooled mother (father) if they had had similar average features as the child of schooled mother (father). The second term of the equation, the coefficient effect is shown by the vertical distance between A and C, which is the increase in the test score of an average child of a non-schooled mother (father) if they had similar returns as an child of a schooled mother (father). The interaction effect, which is the last part of the equation, an interaction of the previous two effects is displayed by the distance between B and G.

Figure 1: Decomposition of Test score gaps between children



3.1.3.1. Empirical Results

The results for decomposition equation (2) are shown in table 5. On the standardized scores, we see that in round 4, the learning difference between the children of schooled and non-schooled mothers (0.537 standard deviations) is double the difference between the children of schooled and non-schooled fathers (0.263 standard deviations). Same is true for this difference in round 3, where they stood at 0.565 standard deviations and 0.276 standard deviations respectively. We see here that the learning differences are more magnified (at both points) based on mothers' schooling status as compared to fathers' schooling status. This difference does not go away even when the children are older by four years in round four. Our findings decompose these learning differences based on each parent's schooling status and for each round separately. In table 5 we see that the most consistent contributor to these learning differences is the endowments effect (*E*). The results for round 4 find that, out of total difference of 0.537 standard deviations between the children of schooled and non-schooled mothers, *E* explains 65 percent (0.35 standard deviation) of the difference here. Whereas *E* explained the entirety plus one-third more (0.39 standard deviation) of the total learning difference (0.263 standard deviations) between the children of schooled and non-schooled fathers. The difference in average background explain only a part of the learning difference between the children of schooled and non-schooled mothers but it over-explained the same between the children of schooled and non-schooled fathers. If the gaps in average background were to be removed the gap in learning between the children of schooled and non-schooled mothers would have been cut short by 65 percent. On the contrary, this gap between the children of schooled and non-schooled fathers would have been removed and moreover, the children of non-schooled fathers would have performed better than their counterparts. In round 3 results, *E* does not explain any part of learning difference between the children of schooled and non-schooled mothers. Similar to round four result, between the children of schooled and non-schooled fathers, *E* explains the entirety as well as a part more (over-explains). We notice here the sudden significance (from round 3 to 4 results) of *E* between the children of schooled and non-schooled mothers whereas the consistent (and sole) significance between the children of schooled and non-schooled fathers. The difference in average features between the children of schooled and non-schooled mothers is causing the children of non-schooled mothers to underperform. However, these differences between the children of schooled and non-schooled fathers is not only causing the children of non-schooled fathers to underperform but stopping them to perform better than the children of schooled fathers. We do not find the significance of the coefficients effect (*C*) between the children of schooled and non-schooled fathers. Which means that there is no significant differences between them in terms of the returns to the background characteristics. However, we do find the role of *C* between the children of schooled and non-schooled mothers. In the results for round four we find that *C* explained 0.16

standard deviations (31 percent) of the total difference here. Whereas *C* explained 50 percent (0.28 standard deviations) of the total difference between the children of schooled and non-schooled mothers in round 3 results. This means that if the differential returns to the background characteristics between the children of schooled and non-schooled mothers were to be removed, a large part of the learning difference (one third in round 4 and half in round 3) would have vanished. Looking at the *E* and *C* for round 3 and 4 between the children of schooled and non-schooled mothers we see that the significance of former has become more important (where it was insignificant in round 3) and latter declined. There has been no change in *E* between the children of schooled and non-schooled fathers from round 3 to 4 results. We also find that the interaction of *E* and *C, I* is significant between the children of schooled and non-schooled mothers once in round 3 results when the children were four years younger where it explained 32 percent of total learning difference at this point.

Table 5: The Decomposition Results

	Mothers				Fathers			
	Coefficient (on raw score)		**Coefficients (on Standardized Math Score)**		**Coefficient (on raw score)**		**Coefficients (on Standardized Math Score)**	
	R4	**R3**	**R4**	**R3**	**R4**	**R3**	**R4**	**R3**
Mean Score-Schooled (+)	16.320	12.996	2.308	2.113	15.091	11.871	2.135	1.930
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean Score- Non-Schooled (0)	12.522	9.522	1.771	1.548	13.233	10.174	1.872	1.654
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Difference	3.798	3.475	.537	.565	1.859	1.697	0.263	0.276
	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.003)	(0.005)	(0.003)
Endowment Effect $\{E(X_+)-E(X_0)\}'\beta$	2.472	.643	.349	.105	2.769	3.524	0.392	0.573
	(0.000)	(0.246)	(0.000)	(0.246)	(0.002)	(0.017)	(0.002)	(0.017)
Coefficient Effect $E(X_0)'(\alpha-\beta)$	1.177	1.720	.167	.280	-0.472	-0.752	-0.067	-0.122
	(0.047)	(0.007)	(0.047)	(0.007)	(0.404)	(0.214)	(0.404)	(0.214)
Interaction effect $\{E(X_+)-E(X_0)\}'(\alpha-\beta)$	.1489	1.111	.021	.181	-0.438	-1.075	-0.062	-0.175
	(0.810)	(0.104)	(0.810)	(0.104)	(0.595)	(0.472)	(0.595)	(0.472)

Note*: p-values are in parenthesis.*

Table 6: Detailed Decomposition Results (Round 4)

Variable	Endowments Effect		Coefficient Effect		Interaction Effect	
	Mother	Father	Mother	Father	Mother	Father
Mothers Education	-	.091	-	-.025	-	-.096
		(.090)		(.025)		(.098)
Fathers Education	-.077	-	-.001	-	-.001	-
	(.063)		(0.048)		(0.084)	
Household Size	.004	.004	.201	.172	-.005	-.004
	(0.006)	(.007)	(0.173)	(.186)	(0.008)	(.008)
Assets	-.012	-.005	.032	-.010	.013	-.011
	(0.012)	(.044)	(0.040)	(.043)	(0.017)	(.047)
School Cost	.064*	.091*	-.020	-.048	-.019	-.051
	(0.039)	(.053)	(0.044)	(.051)	(0.043)	(.055)
Expenditure	.001	.006	-.548	-.690	-.011	-.020
	(0.014)	(.023)	(0.902)	(.951)	(0.019)	(.028)
Years	.135***	.079**	.306	-.675	.020	-.020
	(0.036)	(.036)	(0.593)	(.635)	(0.040)	(.021)
Time Allocation	.0220	-.005	.057	.192*	.021	.027
	(0.025)	(.012)	(0.085)	(.100)	(0.032)	(.020)
English Score	.211***	.178***	-.032	.460	-.004	.055
	(0.044)	(.045)	(0.368)	(.367)	(0.044)	(.044)
Gender	-.030*	-.053**	-.098	-.062	.014	.015
	(0.018)	(.025)	(0.077)	(.097)	(0.013)	(.024)
Age in Months	-.018	-.017	-5.105	-3.233	.014	.009
	(0.013)	(0.013)	(3.574)	(3.867)	(0.013)	(.012)
Body Mass Index	0	-.023	-.026	.140	0	.011
	(0.004)	(.038)	(0.215)	(.436)	(0.001)	(.038)
Drinking Water	-.001	-.001	.112	.095	.003	.001
	(0.006)	(.005)	(0.563)	(.461)	(0.014)	(.007)
Household Quality	.029*	.041*	-.006	.0747	-.001	.010
	(0.016)	(.022)	(0.218)	.(206)	(0.019)	(.029)
Type Of School	.023	.005	-.042	.028	-.022	.011
	(0.026)	(.022)	(0.067)	.071	(0.036)	(.028)
Constant	-	-	5.350	3.513	-	-
			(3.775)	(3.961)		
Total	**.349*****	**.392*****	**.167****	**-.067**	**.021**	**-.062**

Note: ***p<0.01, ** p<0.05, *p<0.10: Standard errors are in parentheses

Table 6.1: Detailed Decomposition results (Round 3)

Variable	Endowments Effect		Coefficient Effect		Interaction Effect	
	Mother	Father	Mother	Father	Mother	Father
Mothers Education	-	.105 (.107)	-	.0230 (.030)	-	.090 (.116)
Fathers Education	-.067 (0.075)	-	.075 (0.058)	-	.131 (0.101)	-
Household Size	-.006 (0.020)	.004 (.019)	.542** (0.220)	.467** (.236)	.006 (0.019)	-.003 (.016)
Assets	-.004 (0.007)	.287 (.209)	.003 (0.011)	-.037 (.026)	.002 (0.007)	-.294 (.210)
School Cost	.067 (0.042)	.166* (.093)	-.034 (0.032)	-.063 (.039)	-.051 (0.048)	-.158 (.097)
Expenditure	-.008 (0.020)	-.014 (.044)	2.248** (1.020)	1.519 (1.052)	.062* (0.033)	.077 (.055)
Years	-.000 (0.008)	-.001 (.004)	1.165** (0.518)	1.024* (.559)	.024 (0.019)	.009 (.015)
Time Allocation	.069** (0.033)	.030 (.020)	.016 (0.147)	.034 (.167)	.004 (0.039)	.003 (.016)
English Score	-	-	-	-	-	-
Gender	-.037* (0.023)	-.077** (.032)	-.104 (0.096)	-.099 (.117)	.014 (0.015)	.023 (.028)
Age in Months	-.014 (0.013)	-.011 (.013)	-3.292 (3.470)	-1.310 (3.636)	.014 (0.016)	.005 (.015)
Body Mass Index	-.011 (0.018)	.020 (.024)	.422 (0.514)	-.624 (.581)	.018 (0.023)	-.029 (.029)
Drinking Water	.014 (0.012)	.020 (.016)	-1.166** (0.473)	-1.037*** (.399)	-.055* (0.030)	-.036 (.026)
Household Quality	-.001 (0.006)	-.004 (.014)	.286* (0.166)	.174 (.151)	.018 (0.017)	.026 (.024)
Type Of School	.107** (0.043)	.048 (.071)	.025 (0.156)	-.304 (.216)	-.008 (0.050)	.112 (.081)
Constant	-	-	.092 (3.551)	.111 (3.707)	-	-
Total	**0.105**	**0.573****	**0.280*****	**-0.122**	**0.181***	**-0.175**

Note: ***p<0.01, ** p<0.05, *p<0.10: Standard errors are in parentheses

As we have seen previously that the most consistent contributor to the learning outcome difference between both the groups of children is *E*. We now look into the individual factors that contribute to this effect in Table 6 and 6.1. Our results identify one core factor which is consistent between both the groups (except between the children of schooled and non-schooled mothers in round 3) which is the schooling cost incurred by the child in the last academic year (similar to previous findings such as in Kingdon, 1996). What this means is that better paying capacity towards education entails better learning outcome at two different ages (of children) and between both the groups. What is striking in our results is the sudden importance of influence of schooling cost between the children of schooled and non-schooled mothers. When the children were younger the learning outcome between the children of schooled and non-schooled mothers was not influenced by this cost. However, when the children got four years older in round four we see a sudden significant influence of this variable in explaining the learning gap between these children.[32] Similarly, the gender of a child (similar to findings in Arteaga and Glewwe, 2014; Glewwe *et al.* 2015; Kingdon, 1996) is also found to be a determining factor for the learning outcome gap. This variable stays consistent in terms of its proportional contribution towards *E* at both the rounds. We find no other consistent contributor to the learning outcome gap between these groups of children. We do, however, find several other factors which now contribute to the learning gap between both the groups of children when they have gotten older in round four. The results for round four are worrying since they identify several factors which now explain a significantly (and consistently between both the groups at this point) large part of *E* among both the groups of children. We find that the gap in average years of schooling between the children of schooled and non-schooled mothers, and children of schooled and non-schooled fathers, explain 40 percent (0.135 standard deviation) and 18 percent (.08 standard deviation) of *E* respectively (influence of this is also found in Glewwe *et al.* 2015 in Vietnamese children). The gap in

[32] We must be careful in interpreting this variable since the definition and composition of this variable differs for each round. For details see Table 1.

mean years of schooling was marginal when the children were younger in round three (see table 4). This gap has magnified when the children are older in round four. During this time (from third to fourth round survey) the additional years of schooling obtained by the children of schooled parents is more than the children of non-schooled parents. The latter's average years of schooling too has increased during this time but this increase is lesser than the increase for former. This could mean one of the following two things. The children of non-schooled parents have dropped out from school more often than the children of schooled parents bringing down the overall average in their respective groups. This could also mean that the children of non-schooled parents have repeated a grade more often than the children of schooled parents widening the gap in years of schooling between them.[33] In either case, the average years of schooling obtained/gap is affected against the children of non-schooled parents. As Sanjay *et al.* (2014) points out that the drop out form education in this age group of Indian children is high, we must read these results in that context. Majority of the previous studies have not considered the years of schooling received by children in their analysis which we, in our study, found to have a significant explanation to provide for learning difference. We also find that the English language score has a core explanation to provide for the learning gap between the children of schooled and non-schooled parents in round four results. This stands at 60 percent (0.21 standard deviation) and 46 percent (0.18 standard deviation) of E, between the children of schooled mothers and non-schooled mothers, and schooled and non-schooled fathers respectively. The influence of English language score on the mathematics learning gap between these children should be read in the context of the questionnaire's language. The question in the survey on mathematics were asked both in Telegu (local language) and English. We must also point this out here that the English language score was collected only during the fourth round survey. We could not confirm this result for the groups

[33] The years of schooling variable is obtained by using the information on the highest grade completed. Even if a child has repeated a grade, we do not count that repetition (only completion of grade is counted as one additional year of schooling).

when they were four years younger in round three because of unavailability of information. A marginal contribution by the household quality index is also found between these children in round four results. We also find that the *C* is also significant in round four results between the children of schooled and non-schooled mothers, however, our results do not identify the factors that explain this effect here. On the contrary, the results in round 3 find that *C* is significant and stands at 0.28 standard deviations, and significant factors are also found here. This includes years of schooling gap, household size, household quality index, and household expenditure. Their contribution is found to be more than the total *C* here, but these factors together, they explain 0.28 standard deviations of the total difference in learning.

3.1.4 Conclusion and Policy Implications

Using the third and fourth round of Young Lives Survey for the older cohort of Indian children, we explore the mathematics learning difference between the children of schooled and non-schooled parents. We do this on the basis of each of the parents' schooling status separately. We divide the children of YL into four groups which are children of schooled and non-schooled mothers, followed by children of schooled and non-schooled fathers. We study these children in two stages where the former two are looked together first, and then we also look into the latter two together. We also comment on the findings obtained among these groups to draw patterns of influence. We research the mathematics learning gap between these children with respect to a set of background characteristics which are discussed in Table 1. We see that mathematics learning for the children of schooled parents is better at both the survey points compared to the children of non-schooled parents. We also see that the differences in mathematics learning is more magnified between the children of schooled and non-schooled mothers compared to the children of schooled and non-schooled fathers at both the survey points. The first thing that we notice from the descriptive statistics is that the children of schooled parents have had consistently better average characteristics compared to the children of non-schooled parents. We then move on to explore the learning differences between them with respect to the differences in the background characteristics to find out what explains the former. We do this using the threefold Blinder-Oaxaca decomposition at both of the survey points and based on each parents' schooling status.

In exploring these aspects we learn that the endowments difference between these groups of children at both of the survey points explain a large part of the learning difference between them. The endowments effect is found to be the most significant explainer of the learning differences between these groups of children. This captures the effect of differences in the considered background characteristics (mean of these) on the learning outcome of children. The removal of background differences between these groups of children would have partly (between the children of schooled and non-schooled mothers) and fully[34] (between the children of schooled and non-schooled fathers) removed the gaps in learning. We also find that the differential returns to the background features (captured by *C*) also provides some explanation but this effect is not as consistent as *E*. We find that the differential returns to these background features work in favor of the children of schooled mothers as against the children of non-schooled mothers. This is the reason we saw in our results that the coefficient effect explains almost half and one-third of the learning difference between the children of schooled and non-schooled mothers for round three and round four results respectively. We also find that the interaction of endowments and coefficient effect significantly affecting the learning difference between the children of schooled and non-schooled mothers in the results for round 3 when the children were younger.

The detailed results for the decomposition identifies the factors that contribute to these three aforementioned effects. The most consistent among the factors include the schooling cost difference between the children of schooled and non-schooled parents which is consistent with many of the previous

[34] We have seen in the results that the differences in the features between the children of schooled and non-schooled fathers is not only stopping the children of non-schooled fathers to underperform but to perform better than the children of schooled fathers.

findings. The gender dummy included in our model is also found to have a significant contribution to *E* and *C*. We also find that there are several factors which did not explain these effects when the children were younger but when the children are older they have a significant part to contribute to these. These factors include the gap in years of schooling between the children of schooled and non-schooled parents which has magnified between the rounds. The English language score gap between the children of schooled and non-schooled parents has also a part to explain *E*. We must understand that the score on mathematics test that we have used in our study was asked in both English and Telegu (local language). Our results imply that those children who were better versed in these languages have performed better. A marginal part of *E* is also explained by the household quality index difference between the children of schooled and non-schooled parents. The results for round four do not identify the factors that explain *C* between the children of schooled and non-schooled mothers. Only one-third of *I* (which is significant in round three between the children of schooled and non-schooled mothers) provided by the expenditure variable.

There are several policy implications as well as investigative directions which come out of our study. The foremost policy implication of our study pertains to the schooling cost variable. This aspect must be pursued in a direction to make children avail better quality education. As is known that better spending capacity on education entails better quality of education. Policies must be pursued to level this gap in a way to improve the quality of education which should be reflected in the learning by these children. It should not be the case that the children who could spend better are the only one to perform better. The ability to spend should not be the determining factor in learning. The second important implication pertains to the years of schooling gap. As these children belong to the same age group, sudden magnification of this gap in just four years is worrying. An exploration of this to figure out if there is financial opportunity cost attached to staying in education. If so, financial incentives to be provided to children in order to make them stay in education. The investigative direction in which our study points out is the further exploration of years of schooling gap in order to provide an explanation for widening years of schooling gap given that these children fall in the same age group. Moreover, a more nuanced explanation is needed for the English language score and its relationship with the mathematics score. This needs to be studied in the context of our results.

References

Annual Status of Education Report. (2005). Pratham Resource Centre, New Delhi.

Annual Status of Education Report. (2010). Pratham Resource Centre, New Delhi.

Ammermueller, A. (2007b). Pisa: What Makes the Difference? Explaining the Gap in Pisa Test Scores between Finland and Germany. *Empirical Economics,* 33 (2), 263-287.

Arteaga, I., & Glewwe, P. (2014). Achievement Gap between Indigenous and Non-Indigenous Children in Peru. *Young Lives Working Paper 130*.

Azubuike, Obiageri Bridget, and K. B. (2016). Young Lives' Constructed files. *Oxford Review of Economic Policy*, *23(2)* (7483), 168–195.

Azubuike, O. B., Moore, R., & Iyer, P. (2017). Young Lives School Surveys, 2016–17: The design and development of cross-country Maths and English tests in Ethiopia, India & Vietnam. *Young Lives*.

Baird, K. (2012). Class in the classroom: The relationship between school resources and math performance among low socioeconomic status students in 19 rich countries. *Education Economics*, *20*(5), 484–509.

Barrera-Osorio, F., García-Moreno, V., Patrinos, H. A., & Porta, E. (2011). Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: An application to Indonesia's results in PISA mathematics. *The World Bank*

Blinder, A. S., (1973). Wage Discrimination : Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Brown, L. & Iyengar, S. (2008) Parenting Styles: The Impact on Student Achievement. *Marriage & Family Review*, 43:1-2, 14-38

Burger, R. (2011). School effectiveness in Zambia: The origins of differences between rural and urban outcomes. *Development Southern Africa*, *28*(2), 157–176.

Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review*, *31*(4), 376–390.

Cobb-Clark, D. A., & Gørgens, T. (2014). Parents' economic support of young-adult children: Do socioeconomic circumstances matter? *Journal of Population Economics*, *27*(2), 447–471.

Crane, J. (1996). Effects of home environment, SES, and maternal test scores on mathematics achievement. *Journal of Educational Research*, *89*(5), 305–314.

Crookston, B. T., Forste, R., McClellan, C., Georgiadis, A., & Heaton, T. B. (2014). Factors associated with cognitive achievement in late childhood and adolescence: The Young Lives cohort study of children in Ethiopia, India, Peru, and Vietnam. *BMC Pediatrics*, *14*(1), 1–9.

Cueto, S., and Leon, S. (2012). Psychometric Characteristics of cognitive development and achievement instruments in round 3 of Young Lives (Young Lives Technical no Notes) Young Lives.

Currie, J., & Yelowitz, A. (2000). Are public housing projects good for kids ? *Journal of Public Economics*, *75*(1), 99–124.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304.

Dawes, A. (2020). Measuring the Development of Cognitive Skills Across Time and Context: Reflections from Young Lives. *Young Lives*

Dercon, S., & Krishnan, P. (2009). Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries. *Children, Youth and Environments*, *19*(2), 138–163.

Desai, S., Dubey, A., Vanneman, R., & Banerji, R. (2008). Private Schooling in India: A New Educational Landscape. *India Policy Forum*, *5*(1), 1–58.

Galab, S., Prudhvikar, R., & Singh, R. (2014) Young lives survey design and sampling in Andhra Pradesh, India

Galab, S., P. Prudhvikar Reddy and Renu Singh (2014a). Education and Learning: Round 4 Preliminary Findings, *Round 4 Fact Sheet, Oxford*: Young Lives

Glewwe, P., Chen, Q., & Katare, B. (2015). What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data. *Asia and the Pacific Policy Studies*, *2*(3), 494–516.

Goyal, S. (2009). Inside the house of learning: The relative performance of public and private schools in Orissa. *Education Economics*, *17*(3), 315–327.

Goyal, S., & Pandey, P. (2009). How do government and private schools differ? Findings from two large Indian states. *International Bank for Reconstruction and Development/the World Bank*

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

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*. 8(4), 453-479.

Kingdon, G. G. (1996). The Quality and Efficiency of Public and Private Schooling: A Case Study of India. *Oxford Bulletin of Economics and Statistics*, *58*(1), 57–82.

Keane, M. P., Krutikova, S., & Neal, T. (2018). The impact of child work on cognitive development: results from four Low to Middle Income countries. (No. W18/29) *IFS Working Papers*.

Kingdon, G. G. (2007). The progress of school education in India. *Oxford Review of Economic Policy*, 23(2), 168-19 5.

Leibowitz, A. (1977). Parental Inputs and Children's Achievement. *The Journal of Human Resources*, *12*(2), 242–251.

Li, Z., & Qiu, Z. (2018). How does family background affect children's educational achievement? Evidence from Contemporary China. *The Journal of Chinese Sociology*, *5*(1), 1–21.

Muralidharan, K., & Sundararaman, V. (2015). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics*, *130*(3), 1011–1066.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*, 693–709.

Paxson, C., & Schady, N. (2007). Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. *Journal of Human Resources*, *42*(1), 49–84.

Rolleston, C. (2014). Learning profiles and the "skills gap" in four developing countries: A comparative analysis of schooling and skills development. *Oxford Review of Education*, *40*(1), 132–150.

Sakellariou, C. (2008). Peer effects and the indigenous/non-indigenous early test-score gap in Peru. *Education Economics*, *16*(4), 371–390.

Sanfo, J. B. M. B., & Ogawa, K. (2021). Explaining the rural-urban learning achievements gap in Ethiopian primary education: A re-centered influence function decomposition using Young Lives data. *Education Economics*, 1–29.

Sanjay, S., Gupta, R., Gautam, V., & Meena, S. H. R. (2014). Educational Statistics at a Glance. *MHRD, Bureau of Planning, Monitoring & Statistics, New Delhi*.

Sharma, G. (2014) Study of classroom environment parental education, income and institution as predictor of students' academic achievement. *Indian Association of Health, Research and Welfare*, 4(4), 277-281

Singh, A. (2015). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics, 113*(2015), 16–32.

Singh, R., & Mukherjee, P. (2019). Effect of preschool education on cognitive achievement and subjective wellbeing at age 12: evidence from India. *Compare: A Journal of Comparative and International Education*, *49*(5), 723–741.

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.

Wadhwa, W. (2009). Are private schools really performing better than government schools? *Annual Status of Education Report (Rural) New Delhi*.

Wamalwa, F. M., & Burns, J. (2018). Private schools and student learning achievements in Kenya. *Economics of Education Review, 66*(July), 114–124.

Zhang, L., & Lee, K. A. (2011). Decomposing achievement gaps among OECD countries. *Asia Pacific Education Review, 12*(3), 463–474.

Zhang, Y. (2006). Urban - Rural Literacy Gaps in Sub - Saharan Africa : The Roles of Socioeconomic Status and School Quality. *Comparative Education Review, 50*(4), 581–602.

4.0 Mathematics Achievement of Public and Private School Children: Why is One Better?

Abstract

Departing from the previous two chapters, in this chapter we explore the mathematics learning difference between the children of private schools and public schools. Making use of the fourth round of Young Lives Survey data for the older cohort of Indian children, we categorize the children into those who have attended private school last or attending private schools and public school attended/attending. The first thing that we notice is that there is a clear significant inequality in learning between the children of private schools and public school. Going into their background characteristics we learn that the children of private schools have better background characteristics compared to the children of public schools. After further exploration of the learning difference with respect to these background features difference, using the threefold Blinder-Oaxaca decomposition, we find that the entire difference between these children is because of the background features difference (endowments effect). We find in our results that the differences in the background features of these children is causing the children of public schools to underperform as well as perform better than the children of private schools.

Keywords: Mathematics Achievement; India; Mothers' Education; Blinder-Oaxaca Decomposition; Endowment Effect; Coefficient Effect

4.0.1 Introduction

In India, the demand for secondary education is seeing an increase (a part of which is through private schooling) because of its lucrative nature. The poor management of public schools, particularly the teacher absenteeism may have encouraged the rapid growth of private schooling in India (Kingdon, 2007). This increase is also affected by the common perception that the private schools provide better quality education (Wadhwa, 2009). In India, the enrolment in private schools is associated with better learning outcomes of children after controlling for the background characteristics (Desai *et al.* 2008). Many of the previous literature have captured the quality of education using learning outcomes of children as an indicator (Kingdon, 1996; Goyal, 2009; Singh, 2015; Singh and Mukherjee, 2019). Studies (such as Kingdon, 1996; Desai *et al.,* 2008; Goyal, 2009; Wadhwa, 2009; Chudgar and Quin, 2011; Wamalwa and Burns, 2012; Singh, 2015; Singh and Mukherjee, 2019) have found an association between attending private schools and having better learning/achievement scores among Indian children. There are other studies where this effect is not apparent like the ones mentioned before. Goyal and Pandey (2009 in their study of Indian children have found a private schooling effect but the results suggest that the quality of both the type of schooling is poor. Muralidharan and Sundararaman (2015) in their study of children from the state of Andhra Pradesh (southern state of India) find that there is no public-private difference in learning in some aspects.

The public expenditure on education in India as a percentage of GDP has been increasing over the years and slightly since 2010.[35] The Annual Status of Education Report (ASER) 2005 concludes that the increase in the financial resources towards education does not reflect in the levels of learning in primary education (ASER, 2005). The latter reports conclude that there is a decline in the ability to do the basic

[35] As per the Ministry of Human Resource Development's Educational Statistics at a glance 2014 on Education and other departments.

math nationally, which is visible across all classes (ASER, 2010). In this chapter we explore the mathematics learning inequality between the children of public and private schools. The chapter uses fourth round of Young Lives Survey data for the older cohort of Indian children in the analysis. We delve into the mathematics learning inequality between a set of public and private school children of India at a point. The children that we research in this chapter hail from two southern states of Andhra Pradesh and Telangana. The children surveyed in YL are now spread across these two states. We decompose the gap in mean mathematics learning between the said groups of children using the threefold Blinder Oaxaca decomposition. The YL's preliminary findings for the latter round (fourth) points out that although there is an increasing choice of private schools, the gap is magnifying between the boys attending private schools and girls attending private schools (Galab *et. al.* 2014a).

The previous literature has explored several factors to study the learning outcomes of children in various socio-economic contexts. As we have used the YL dataset in our study, we must mention here that many studies have made use of this dataset to study the learning outcomes of these children. To research the learning aspect of all the four surveyed countries by YL (Ethiopia, India, Peru, and Vietnam), Dercon and Krishnan (2009) and Crookston *et al.* (2014) have investigated these different socio-economic set ups. Using the Peruvian YL dataset, Arteaga and Glewwe (2014) study the learning outcome difference between the indigenous and non-indigenous children. Similarly, Glewwe *et al.* (2015) explore the learning difference based on the ethnicity of Vietnamese children. They study the learning differential between the Kinh community and ethnic minority children using Vietnam's YL dataset. Singh and Mukherjee (2019), using Indian YL dataset, explore the connection between type of school attended and cognitive skills of Indian children. And Sanfo and Ogawa (2021) study the rural-urban learning difference using the Ethiopian YLS data.

Previous literature has studied the learning outcome of children from many countries in depth. Among the lower income countries, several aspects have been explored in connection with the learning outcome of children. For instance, Zhang (2006) looks into the learning disadvantages of children from Sub-Saharan Africa. The learning differences (in Burger, 2011; Das *et al*. 2013) among the Zambian children have been studied as well. Similar to Arteaga and Glewwe (2014), Sakellariou (2008) has studied the mathematics learning difference between the indigenous and non-indigenous Peruvian children. In their study of Ecuadorian children, Paxson and Schady (2007) investigate the connection between the parental education and learning outcomes of children. Learning outcomes of Children from the Asian countries such as China (in Li and Qiu, 2018) and Indonesia (in Barrera-Osorio *et al.* 2011) have also been studied in the previous literature.

Similarly, the learning outcomes of children from high income countries has also been extensively studied in the previous literature. Among these studies, children from the United States of America (in Leibowitz, 1977; Crane, 1996; Currie and Yelowitz, 2000; Davis, 2005; Todd and Wolpin, 2007) have been studied on many similar aspects as done in the studies mentioned before. Similarly, Australian children's (in Cobb *et al.* 2014) learning outcomes with respect to their socio-economic background has been done too. On the same line, the learning outcome of children from Germany and Finland has also been explored in depth (Ammermueller, 2007b). Moreover, learning outcome of children from the OECD countries (in Zhang and Lee, 2011) as well as Swedish children (in Holmlund *et al.* 2011) have been researched too.

The previous literature, although talks about the factors that have an influence on the learning outcomes of children, very few decompose the learning difference between groups of children in relation to their background characteristics. In this chapter, we attempt to add to the growing body of literature on the decomposition of learning outcome difference between public and private school children. Our chapter is an attempt to add to the growing body of literature on the differences in average learning outcome of public and private school children (among children from poorer economic background). Using various features around which these children have grown, varying from socio-economic features, parental education level, and child specific features, the chapter researches the gap in average mathematics learning between the private school and public school children. The idea is also to see whether their respective differences (in the average background features, if any) explain the differences in the average learning outcomes and if so, to what extent. We have done this using the threefold Blinder-Oaxaca decomposition. There have been studies which have used the Blinder-Oaxaca decomposition to study the achievement gaps among children. For instance, Ammermuller (2007b), using the PISA-2000, has studied the achievement gap between the German and Finnish children. Zhang and Lee (2011) have also used this decomposition technique, using PISA-2006, to study the achievement gap between the OECD countries. Burger (2011) too studies the achievement differences between the urban and rural Zambian children using the SAMCEQ-II (South Africa Consortium for Monitoring Educational Quality) dataset. Baird (2012) studies the achievement gaps (using TIMSS for 2003) between the children with high socio-economic status (SES) and lower SES among 19 high income countries. Another interesting study using the Blinder-Oaxaca technique is by Barrera-Osorio *et. al.* (2011) which studies the achievement improvement of the Indonesian children between two points using PISA 2003 and 2006. Using the threefold decomposition, we decompose the average learning outcome difference between the private and public school children and attribute it to the factors responsible for explaining the difference at two points.

The rest of the chapter is organized as follows. Section 2 gives the description of data, source and variables used in our study along with descriptive statistics. Section 3 presents the methodological framework and empirical approach along with the results followed by summary and conclusion in Section 4.

4.0.2 Data and Variables

The study uses the longitudinal survey conducted by the University of Oxford. The survey includes four developing countries Ethiopia, India, Peru, and Vietnam. The 3000 children in the Indian survey were selected from 20 sentinel sites spread now across two southern states of Andhra Pradesh and Telangana. We have made use of the fourth round data for the older cohort on India (united Andhra Pradesh) that was released in 2014. This study uses the data on the older cohort for the fourth round when the children were around19-20 years of age. The round covers 952 children from the older cohort. Because we intend to look into the YL children at two points we need to make sure that the data is consistent between the rounds. Keeping that in mind we ended up with a dataset of 522 children. For more details on the dataset see chapter 2 where we have charted out the sampling method, site selection and everything pertaining to Young Lives Survey. The Table 1 gives the description of the variables used (which is a repetition from chapter 3.0) in the analysis and Table 2 gives the descriptive statistics of these described variables.

Table 1: Description of the Variables

Variables		*Description*
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of the survey out of 30 marks. This score has also been standardized using the standard deviation of the whole sample for comparison and ease of understanding.
Parental Education (PE)[36]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets	Value of the five most valuable assets owned in the YL child's Household
	School Cost	Total expenditure incurred on school in the last academic year.
	Expenditure	The log of per capita monthly expenditure of YL child's household
Child Specific Features (C)	Years	Years of schooling received by YL child at the time of fourth round survey
	Type of School	=1 if attended/attending public school; 0 if private.
	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of fourth round survey
	Body Mass Index	Body mass index of the YL child at the time of fourth round survey
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[37]

In our thesis, the outcome of interest is the mathematics test scores of children at the survey point. The YL has collected the information on child's achievement with the help of language and mathematics tests (Galab *et al.* 2014a). The mathematics test that we use in this study was designed for each round separately and there were certain similarities between the rounds. More details on the test content and the questions that were asked to the surveyed children is available in chapter 2 and 4 in detail.

[36] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[37] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

Table 2: Descriptive Statistics (means) of overall and sub-samples (Round4)

	Overall		Private		Public	
Variables	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Math Score	14.479	7.07	16.38667	6.744	11.90	6.70
Mothers Education	3.632	4.314	4.66	4.514	2.24	3.60
Fathers Education	5.643	5.045	6.75	5.070	4.15	4.62
Household Size	4.753	1.908	4.73	2.036	4.78	1.73
Assets	30018.64	44478.14	35547.63	44685.45	22547.03	43181.86
School Cost	19226.17	26380.52	28759.59	30584.56	6343.17	9218.29
Expenditure	7.170	0.587	7.23	0.580	7.09	0.59
Years	13.448	1.745	13.99	1.376	12.71	1.92
Time Allocation	2.013	1.442	2.3	1.379	1.63	1.44
English Score	16.513	3.306	17.27	2.924	15.50	3.52
Gender	.550	.498	0.58	0.494	0.51	0.50
Age in Months	227.88	4.134	227.76	4.143	228.03	4.13
Body Mass Index	19.933	8.744	19.74	3.489	20.20	12.79
Drinking Water	0.981	0.137	0.99	0.115	0.97	0.16
Household Quality	0.725	.217	0.75	0.204	0.69	0.23
Type of School	.425	.495	-	-	-	-
Observations	**522**		**300**		**222**	

4.0.3 Theoretical Framework

Based on the objective of the research, we employ behavioral framework in order to carry out the empirical analysis. There are various factors that have an influence on the mathematics achievement of the school going children based on the review of literature. The mathematics achievement (EA) can be depicted as a structural relationship between various variables. The input variables in this relationship includes the parental education (PE), household variables (H), child specific variables (C), and **o**thers (O).

$$EA = f(PE, H, C, O) \tag{1}$$

The regression equation following this would take the following form:

$$EA = \alpha_0 + \alpha_1(PE) + \alpha_2(H) + \alpha_3(C) + \alpha_{4(}O) + \epsilon_i \tag{2}$$

The α_is are the row vectors of the coefficients that captures the total influence of the column vectors of set of variables in the bracket. The factors considered in the equation include the ones described in Table 1.

4.0.4 Methodology for Decomposition and Decomposition Analysis

The sub-sample of children that we study are the children who have either attended/attending public school and the children who have either attended/attending private schools. The sub-sample descriptive statistics (differences in the mean background features) of these groups of children is shown in Table 3. We see from this table that the children of private schools have better mathematics learning score, on an average, compared to the children of public schools. We also learn from this table that the children of private school have, on an average, better background features (from Table 1) compared to the children of public schools. These differences in the background features are highly statistically significant.

Table 3 Difference in the means of background features

Variable	Overall Mean	Means of background variables		Difference in the mean of background variables between the private and public school children
	R 4	Private	Public	
Math Score	14.479	16.38667	11.90	4.485^{***}
	(7.074)	(6.744)	(6.70)	
Fathers Education	5.644	4.66	2.24	2.416^{***}
	(5.045)	(4.514)	(3.60)	
Mothers Education	3.632	6.75	4.15	2.601^{***}
	(4.314)	(5.070)	(4.62)	
Household Size	4.753	4.73	4.78	-0.05
	(1.908)	(2.036)	(1.73)	
Assets	30018.64	35547.63	22547.03	13000.61^{***}
	(44478.14)	(44685.45)	(43181.86)	
School Cost	19226.17	28759.59	6343.17	22416.43^{***}
	(26380.52)	(30584.56)	(9218.29)	
Expenditure	7.170	7.23	7.09	$.140^{***}$
	(.588)	(0.580)	(0.59)	
Years	13.448	13.99	12.71	1.281^{***}
	(1.745)	(1.376)	(1.92)	
Time Allocation	2.013	2.3	1.63	$.673^{***}$
	(1.441)	(1.379)	(1.44)	
English Score	16.513	17.27	15.50	1.771^{***}
	(3.306)	(2.924)	(3.52)	
Gender	.550	0.58	0.51	0.07
	(.498)	(0.494)	(0.50)	
Age in Months	227.877	227.76	228.03	-0.27
	(4.134)	(4.143)	(4.13)	
Body Mass Index	19.933	19.74	20.20	-0.46
	(8.743)	(3.489)	(12.79)	
Drinking Water	.980	0.99	0.97	0.01
	(.137)	(0.115)	(0.16)	
Household Quality	.725	0.75	0.69	$.0530^{***}$
	(.217)	(0.204)	(0.23)	
Type Of School	.425			
	(.494)			
Observations	**522**	**300**	**222**	

Note: ***p<0.01, ** p<0.05, * p<0.10: Standard Deviations are in parentheses
The mean values of each background variables for each group is shown in Table 2

The Table 4 summarizes the outcome variable, the mathematics test score, which in this table has been standardized. The normalization of the scores is done by dividing the scores with the standard deviation of overall sample. Even on the standardized test score we learn that the children from the private schools have better mathematics learning outcome compared to the children of public schools.

Table 4: The Mean Math Test score by the groups (Round 4)

	Overall	Private	Public	Difference (Pr)-(Pub)
Math Score (Mean)	2.04	2.317	1.683	$.634^{***}$
S. Dev.	1.00	1.00	.947	
N	522	300	222	

Note: ***p<0.01; Figures in parenthesis are standard errors; The test scores have been normalized by dividing it with the standard deviation of the overall sample for ease of comparison.

As we have seen earlier that the background features are significantly better for the children of private schools, in the following sections, we attempt to learn how much of that difference (in the background) explains the difference in the mathematics score. The primary objective here is to study the achievement scores by the two groups (with respect to a set of background features) and decomposing the differences in the mean test scores into its different components to know about what is causing this difference. We have made use of the Blinder-Oaxaca decomposition in this study.

As mentioned earlier, we have divided the children on the basis of the type of school they have attended and/or attending. The children who have attended/attending private schools (*Pr*) have been put into one category and those children who have attended/attending public schools (*Pub*) have been put into another category. Those children from the dataset who have been to any other type of school have not been taken into account into our research. The difference of the mean test scores between the two groups can be expressed as:

$$D = E(EA_{Pr}) - E(EA_{Pub}) \tag{3}$$

In the decomposition analysis that we deploy in our study we need a reference category to interpret our results. The reference category of children are the children from the private schools and the point of analysis is from the public schools children. Using (2) for each group of children separately (and following Blinder, 1973; Oaxaca, 1973; Jann, 2008) the contribution of the group difference in the predictors to the overall outcome difference can be rearranged as follows:

$$D = \{E(X_{Pr}) - E(X_{Pub})\}'\beta + E(X_{Pub})'(\alpha - \beta) - \{E(X_{Pr}) - E(X_{Pub})\}'(\alpha - \beta)$$

This is referred to as threefold decomposition.

$$D = E + C + I$$

The first component,

$$E = \{E(X_{Pr}) - E(X_{Pub})\}'\beta$$

The part of the learning difference which is due to the group difference in the mean background features and it is called "endowments effect". If the children of public schools were to have same average features as the children of private schools ($\bar{X}_{Pr}$), this part of the equation tells that this is how much their scores would have improved. In other words, the difference in the background features of these two groups of children applied to the impact of the public school children (β).

$$C = E(X_{Pub})'(\alpha - \beta)$$

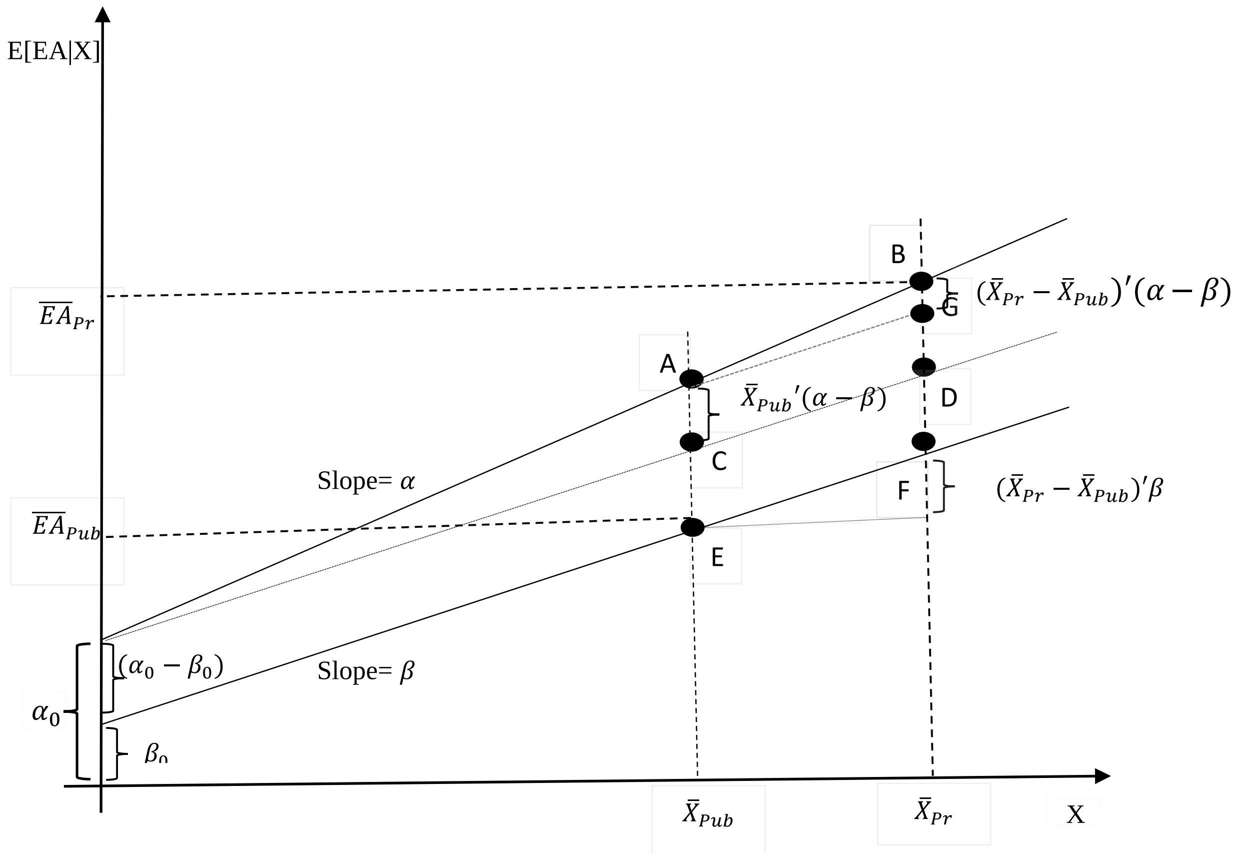
This part measures the contribution to the learning gap by the differences in the coefficients/returns $(\alpha - \beta)$ of the two groups which also includes the intercepts. The part of the gap in learning outcome of the two groups of children explained by the differences in groups' impact of the background features. If the children of public school had same average returns as the children of private schools (α), keeping their features at the current level (X_{Pub}), this is how much the scores of public school children would have improved. In other words, the differences in the average returns to the background features between the two groups of children applied to the average features of the public school children.

$$I = \{E(X_{Pr}) - E(X_{Pub})\}'(\alpha - \beta)$$

And the third component, is an interaction term which accounts for the fact that the differences in endowments and coefficients exist simultaneously between the two groups.

These three effects (from equation 2) are portrayed in figure 1, which is borrowed from Arteaga and Glewwe (2014) and has been modified slightly based on our needs. The decomposition of tests score gap between the groups of children at a point is shown. The first component, endowment effect is shown by the move from point E to F. This shows the increase in the test score of an average child of public school if they had had similar average features as the child of private school. The second term of the equation, the coefficient effect, is shown by the vertical distance between A and C, which is the increase in the test score of an average child of a public school if they had similar returns as a child of private school. The interaction effect, which is the last part of the equation, an interaction of the previous two effects is displayed by the distance between B and G.

Figure 1: Decomposition of Test score gaps between children



4.0.4.1 Empirical Results

We will now talk about the empirical results obtained for the Blinder –Oaxaca decomposition. We can see from table 5 that the difference (on the standardized mathematics score) between the children of private and public school stands at 0.635 standard deviation. The idea behind using the B-O decomposition as our method is to break this difference between them in association with their respective background features. We have already seen earlier that the differences in background features between the children of public and private school is highly significant. The Blinder-Oaxaca decomposition also identifies that if there is a differential returns to the background features and tells us the extent of it between these groups of children.

From our results it is evident that the core and only significant cause behind the mathematics learning difference between the children of public and private school is the differences in their respective average background features. We learn from our results that if the children of public school were to have similar average features as the children of private schools ($\bar{X}_{Pr}$), keeping their returns (β) at the current level, the difference in their mathematics learning (which stands at 0.635 standard deviation) would have been removed (the endowments effect). We must point this out here that the endowments effect explains more than then actual difference between the two groups of children. What this means is that with the features of private school children, the gap in learning would not only have been removed, in fact, the public school children would have performed better than the children of private schools. The differences in the background between the children of private school and public school is not only stopping the latter from performing better, but better than the private school children. We find that the differential returns to background features does not exist between the two groups of children. Moreover, the interaction of endowments and coefficients does not exist between them too.

Table 5: The Decomposition Results

	Coefficient (on raw score)	Coefficients (on Standardized Math Score)
	R4	R4
Private School Children	16.387	2.318
	(0.000)	(0.000)
Public School Children	11.900	1.683
	(0.000)	(0.000)
Difference	4.486	.6345
	(0.000)	(0.000)
Endowment Effect $\{E(X_{Pr}) - E(X_{Pub})\}'\beta$	5.255	.7432
	(0.000)	(0.000)
Coefficient Effect $E(X_{Pub})'(\alpha - \beta)$	.1399	.0198
	(0.810)	(0.810)
Interaction effect $\{E(X_{Pr}) - E(X_{Pub})\}'(\alpha - \beta)$	-.909	-.129
	(0.337)	(0.337)

Note*: p-values are in parenthesis.*

Table 6: Detailed Decomposition Results (Round 4)

	Endowment Effect		Coefficient Effect		Interaction Effect	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Mothers Education	-0.054	.035	0.063	.070	0.039	.044
Fathers Education	0.047	.042	-0.047	.046	-0.051	.050
Household Size	0.003	.008	0.258	.179	-0.003	.009
Assets	-0.009	.015	-0.006	.036	-0.004	.021
School Cost	0.323^{**}	.133	-0.072^{*}	.039	-0.256^{*}	.136
Expenditure	0.009	.014	-1.096	.902	-0.022	.020
Years	0.159^{***}	.045	0.908	.580	0.092	.059
Time Allocation	0.081^{***}	.030	-0.152^{*}	.082	-0.063	.036
English Score	0.149^{***}	.038	1.100^{***}	.359	0.126^{***}	.046
Gender	0.015	.012	0.064	.0678	0.009	.011
Age in Months	-0.005	.007	-0.650	3.520	0.001	.004
Body Mass Index	0.003	.006	0.124	.251	-0.003	.008
Drinking Water	-0.003	.005	0.563	.462	0.008	.001
Household Quality	0.024^{*}	.015	-0.030	.214	-0.002	.016
Constant			-1.007	3.688		
Total	**0.743^{***}**	**0.000**	**.019**	**0.81**	**-.128**	**0.337**

Note: *** p<.01, ** p<.05, * p<.10

The detailed results of the Blinder-Oaxaca decomposition in Table 6 identifies the factors that contribute to the endowments effect. We see that the most important contributor to the endowments effect between the children of private schools and public schools is the schooling cost (average) incurred in the last academic year. As we have seen previously that the children of private schools have spent a significantly large amount on education in the last academic year compared to the children of public schools. We learn that this difference in the schooling expenditure during the last academic year contributes to the learning difference between the children of public and private schools (similar effect of schooling cost on the learning outcome has also been found in the previous studies such as Kingdon, 1996, 2007; Arteaga and Glewwe, 2014). The contribution of the schooling cost variable stands at 43 percent of the total endowments effect. The other factor which contributes to the endowments effect between them is the years of schooling gap between them. We also learn that there is a significant difference between them on the average years of schooling they have received at the time of fourth round survey. This has an effect in terms of its contribution to the total endowments effect, 21 percent (also found in Glewwe *et al.* 2015). This is a worrying sign because these children belong to the same age group and the existence of a gap in years of schooling received points out that there appears to be a drop-out from schooling, more frequently among the children of public schools. The difference in the average time spent by the groups of children on studies apart from that in schools is also a significant contributor to the learning difference on mathematics. We learn from our results that the children of private schools spend an average greater amount of time compared to the children of public schools and that has an influence on the learning on mathematics. One interesting finding is the significance of the English language score in contributing to the endowments effect (20 percent of endowments effect) between the two groups of children. This could be explained with reference to the questionnaire that was used to collect the test scores of these children. The questionnaire that was used to ask questions to the surveyed children asked the questions in both Telegu (local language) as well as English. This is followed by a significant contribution by the time allocation on studies by these children apart from that they spend in school. The contribution of this variable stands at 10 percent of total endowments effect (similar effect is found also in Glewwe *et al.* 2015; Kingdon, 1996). Our results point out that those children who performed well in English have performed better. Moreover, the private school children have this advantage that they were better at English language than the children of public school. There is a marginal contribution by the household quality index to the endowments effect (3 percent). In total, 97 percent of the total endowments effect is explained by these variables. The rest of the endowments effect stands unexplained in our results. As mentioned earlier, the coefficient and interaction effect are not significant between the children of private schools and the children of public schools.

4.0.5 Summary and Conclusions

Just like the previous two chapters, in this chapter we again characterized the children of the Young Lives Survey. Unlike the previous chapter, here the focus of characterization of YL children was the type of school they last attended or attending. They were basically characterized into the children who have a background from private school and those who have in public schools. We look into the mathematics learning of these children as against each other and we noticed that there is a clear significant difference in learning between these children. The children who have a background in private schools have noticeably performed better than the children of the public school. We then looked into a set of their respective background characteristics and we find that the children of private schools have a significantly better background compared to the children of public schools. Moving on, we then tried to study the mathematics learning difference between these children with the actual background differences between them. To do this, we have deployed, the Blinder-Oaxaca decomposition in the empirical analysis. The results of the decomposition tells us that the core and the only cause behind the learning difference between these children is the background differences between them.

In the results, when we associated the background differences of the children of private schools and the children of public schools, we learnt several things. The first thing that come out of the results is that the stark difference in the background features between these children is the core cause behind the learning difference. We also learn that with the features of the private school children, keeping the returns at the current level, the public school children would have not only performed better, but better than the children of private school. The differences in the background characteristics is not only stopping the public school children to perform better, but better than the children of private schools. When we looked deeper into these results we could identify the specific features differences between these children which explain the learning difference. The most important contributor to the endowments effect which is the only contributor is the schooling costs incurred by these children in the last academic year. The children from private schools have spent a significantly large amount on academics as compared to the children of public schools. This difference explains over 40 percent of the endowments effect. This is followed by the gap in the average years of schooling received by these groups of children at the time of fourth round of survey. We must keep this in mind that these children fall in the same age group and an existence of gap in years of schooling between them is a sign of worry. This means that the drop-out of children from education is happening and more frequently among the children of public schools. This gap in schooling explains 21 percent of the endowments effect. A large part of the endowment effect, 20 percent, is also explained by the learning difference between them on the English language score. The time allocation on studies by these children is also significantly different between the two groups of children, the effect of which could also be seen from the results where this variable contributes 11 percent of the endowments effect. A marginal part of the endowments effect is also explained by the household quality index difference between these children.

The children of private schools could perform better than the children of public schools because former could spend more on schooling compared to the latter. A better spending on education entails better educational quality in schools. This must be paid heed to and policies should be framed where the children from public schools should not have to spend as much as the children of private schools to attain a decent level of quality education. Policies should also be framed to keep the children from poorer background, especially those from the public schools in education. There seems to be a monetary opportunity cost attached to staying in schools which affects the children of public schools more than the children of private schools. They should be incentivized to make them stay in schools and should also be provided with quality schooling. The chapter points towards an investigative direction, which is the exploration of the English language score in association with the mathematics score. Our chapter cannot, with certainty, say much about this association of English language score with the mathematics score.

4.1 Mathematics Learning Inequality among the Children of Private and Public Schools

Abstract

In this chapter, we research the gap in average mathematics learning between the children of private and public schools at two points. We have divided them into two groups on the basis of the type of school they are attending or last attended, extending the ideas borrowed from the previous chapter. We first skim through their various background characteristics at these two points. We then explore these background characteristics and using the threefold Blinder-Oaxaca decomposition, we find that it is the difference in the average endowments between the two which consistently explains the gap in average performance between them. We also find the role of differential impact of the background characteristics on the average learning outcome of children on the first point. The most important and consistent contributor to the endowment effect is the schooling cost and the time allocation on studies. One striking result is the now significant contribution of the gap in average years of schooling which is worrying because these children are from the same age group. We conclude that with the average features and returns of the private school children, not only the gap between them would have been removed but, in fact, public school children would have performed better than the private school children.

Keywords: Blinder-Oaxaca Decomposition, Endowment Effect, Mathematics Learning; Public-Private Schooling; Schooling Cost

4.1.1 Introduction

In this chapter, we continue from where we ended the last chapter. We add onto the previous chapter by studying the same set of children at two points on the basis of the type of schools they have attended and/or attending. In India, the demand for secondary education is seeing an increase (a part of which is through private schooling) because of its lucrative nature. The poor management of public schools, particularly the teacher absenteeism may have encouraged the rapid growth of private schooling in India (Kingdon, 2007). This increase is also affected by the common perception that the private schools provide better quality education (Wadhwa, 2009). In India, the enrolment in private schools is associated with better learning outcomes of children after controlling for the background characteristics (Desai *et al.* 2008). Many of the previous literature have captured the quality of education using learning outcomes of children as an indicator (Kingdon, 1996; Goyal, 2009; Singh, 2015; Singh and Mukherjee, 2019). Studies (such as Kingdon, 1996; Desai *et al.*, 2008; Goyal, 2009; Wadhwa, 2009; Chudgar and Quin, 2011; Wamalwa and Burns, 2012; Singh, 2015; Singh and Mukherjee, 2019) have found an association between attending private schools and having better learning/achievement scores among Indian children. Majority of the previous studies have found a private schooling effect in their respective studies. There are other studies where this effect is not apparent like the ones mentioned before. Goyal and Pandey (2009 in their study of Indian children have found a private schooling effect but the results suggest that the quality of both the type of schooling is poor. Muralidharan and Sundararaman (2015) in their study of children from the state of Andhra Pradesh (southern state of India) find that there is no public-private difference in learning in some aspects.

The Trends in Mathematics and Science Study (TIMSS where 46 countries participated) were used on the Indian children from the sates of Rajasthan and Orissa. Their average learning outcome on science and mathematics was not only different but significantly lower than the international average

(Kingdon, 2007). The public expenditure on education in India as a percentage of GDP has been increasing over the years and slightly since 2010.[38] The Annual Status of Education Report (ASER) 2005 concludes that the increase in the financial resources towards education does not reflect in the levels of learning in primary education (ASER, 2005). The later reports conclude that there is a decline in the ability to do the basic math nationally, which is visible across all classes (ASER, 2010). In this chapter we explore the mathematics learning inequality between the children of public and private schools at two points. We use Young Lives Survey (YLS /YL henceforth) data for the older cohort of Indian children in the analysis. Moreover, the data that we have used is at two points, the gap between which is four years. In other words, we delve into the mathematics learning inequality between a set of public and private school children of India at two points. We decompose the gap in mean mathematics learning between the said groups of children using the threefold Blinder Oaxaca decomposition at two survey points. The YL's preliminary findings for the latter round (fourth) points out that although there is an increasing choice of private schools, the gap is magnifying between the boys attending private schools and girls attending private schools (Galab *et. al.* 2014a).

The previous literature (which has been extensively discussed in the previous chapters) although talks about the factors that have an influence on the learning outcomes of children, very few decompose the learning difference between groups of children in relation to their background characteristics. In this chapter, we attempt to add to the growing body of literature on the decomposition of learning outcome difference between public and private school children. These studies have largely focused on the factors influencing the learning outcomes not the extent to which they cause or explain the differences in the learning outcomes between groups of children. This chapter is an attempt to add to the

[38] As per the Ministry of Human Resource Development's Educational Statistics at a glance 2014 on Education and other departments.

growing body of literature on the differences in average learning outcome of public and private school children (among children from poor economic background). We have used the dataset of YL Survey, India, for the older cohort of the third and fourth round. Our study is for a specific population at two specific points in time i.e. when they were 15-16 and 19-20 years old (during the third and fourth round surveys respectively). These children are growing up in the context of poverty and the dataset traces the changes occurring in their lives. Using various features around which these children are growing, varying from socio-economic features, parental education level, and child specific features, the chapter researches the gap in average mathematics learning between the private school and public school children. The idea is also to see whether their respective differences (in the average background features, if any) explain the differences in the average learning outcomes and if so, to what extent. We have done this using the threefold Blinder-Oaxaca decomposition. There have been studies which have used the Blinder-Oaxaca decomposition to study the achievement gaps among children (for instance Ammermueller, 2007b; Zhang and Lee, 2011; Burger, 2011; Baird, 2012; Barrera-Osorio *et al.*, 2011). We decompose the average learning outcome difference between the private and public school children and associate that with the differences in their respective background characteristics.

The rest of the chapter is organized as follows. Section 2 gives the description of data, source and variables used in our study along with descriptive statistics. Section 3 presents the methodological framework and empirical approach along with the results followed by summary and conclusion of this chapter in Section 4.

4.1.2 Data and Variables

Continuing from the previous chapter, this chapter uses two rounds of Young Lives Survey data for India (United Andhra Pradesh). The overall survey includes four developing countries Ethiopia, India, Peru, and Vietnam to trace the nature of changing poverty. The survey is based on a pro-poor sample which has nearly an equal number of boys and girls, based both in rural and urban communities. The 3000 children in the Indian survey were selected from 20 sentinel sites spread now across two southern Indian states of Andhra Pradesh and Telangana. The sentinel site sampling methodology used in the YL is a form of purposive sampling. The observations in the sample represent a certain type of population. As the YL is a longitudinal study, a prolonged contact with the observations had to be maintained. In order to ensure that YL is able to maintain this contact over a long period, this methodology was most suited. We must note here that the YL sample is not strictly statistically representative of the population. Keane *et al.* (2018) mentions that the YLS has a pro-poor sample bias. In our chapter, the eventual number of observation has reduced, primarily because of the missing information on certain variables. So, the analysis and inferences made in our study has to be seen in these light. This is bound to affect the generalizability of our study. The sampling method deployed in the data collection is a form of purposive sampling (Galab *et. al.* 2014). We have made use of the third and fourth round data for the older cohort on India which was released in 2010 and 2014 respectively. The attrition rate in the Indian survey has been kept very low (Singh, 2015). Our chapter had to drop out observations as information on various required variables on many observations were missing. Moreover, we have considered only those children who are studying/studied in purely public or private schools and we had to drop children who have studied in different schools.[39] Eventually we ended up with 522 observations for the analysis, after considering

[39] The other types of school that some of the surveyed children attended included private aided, NGO/charity/Religious, informal or non-formal, charitable trust, bridge schools, and mix of public and private

that the information on each variable is available at both the rounds of survey. The variable of interest here is mathematics test scores of children. In each round YL collected information about children's learning achievement with the help of language and mathematics tests (Galab *et al.* 2014a). A more detailed description of the dataset is available in chapter 2 where the sampling method and the limitations associated with this are discussed.

The tests that were administered to the surveyed children were designed for each round separately and there were certain similarities between them. The test on mathematics in round 3 for the older cohort had two sections. The first section, which includes 20 items, tested these children on addition, subtraction, multiplication, division, and square roots (whole number and fractions). The second section, which had 10 items, involved mathematics problem solving that were developed using PISA (Programme for International Student Assessment) and TIMSS. This section included questions on data interpretation, number problem solving, measurement, and geometry. On the same line, round 4 mathematics test had five sections which asked questions on similar aspects as round 3. As there was a gap of four years between the two rounds of survey, the difficulty level was adjusted (Dawes, 2020). In sum, there were 30 questions in total for each round and the scores were obtained by adding the correct responses out of 30 marks. The detailed information on the surveys is available in the questionnaires for each round. We must point this out here that although there were similarities in these tests between the rounds (Dawes, 2020), the marks obtained cannot be directly compared between the rounds (Dawes, 2020; Rolleston, 2014). In order to make them comparable between the rounds the reliability and the validity of these tests must be checked. There are various studies (such as Cueto and Leon, 2012; Azubuike *et al.* 2017) which have done this using the Classical Test Theory (CTT) and Item Response Theory (IRT). The application of these tests

---

schools. In this study we have considered only those children who have studied in either purely public or private school. See Table 7 for more details.

brought them to a comparable scale between rounds. We do not directly do a comparison of the scores between the rounds hence we have not applied these tests.

Table 1: Description of the Variables

Variables		Description
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of the surveys out of 30 marks. This test score has also been standardized by dividing with the standard deviation of the overall sample.
Parental Education (PE)[40]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets (Round 4)	Value of the five most valuable assets owned, rented or borrowed in the YL child's household
	Assets (Round 3)	Value of assets owned, rented or borrowed in the YL child's household
	School Cost (Round 4)	Total expenditure incurred on schooling in the last academic year.[41]
	School Cost (Round 3)	How much has the YL household spent on school fees and extra tuition for the child per year
	Expenditure (Round 4)	The log of per capita monthly expenditure of YL child's household
	Expenditure (Round 3)	The log of real per capita monthly expenditure of YL child's household base 2006 prices
Child Specific Features (C)	Years	Years of schooling received by YL child at the time of third and fourth round surveys
	Type of School	=1 if attended/attending public school last, 0 if private.
	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey out of 30 marks[42]
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of fourth round survey
	Body Mass Index	Body mass index of the YL child at the time of fourth round survey
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[43]

The Table 1 gives description of the variables selected, based on the literature reviewed, used in our study and Table 2 gives the descriptive statistics of the variables for each round of the survey separately. The idea in this study is to look into the gap in average mathematics learning among private and public school children and try to understand what factors explain this gap between them. The overall

[40] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[41] In the fourth round this cost is the sum total of the tuition fees, education charges, private tuition, accommodation, transportation, uniforms, stationary etc. in the last academic year. The school level heterogeneity is captured by the variance in the sum total of these costs.

[42] This information is available only in the fourth round survey.

[43] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

descriptive statistics of the variables described in Table 1 is shown in Table 2. Table 2 also shows the descriptive statistics by the subgroups of children for each round separately. The test score that we have used in our study has either been standardized or used in its raw form. We cannot directly compare the test averages between rounds because they are not brought to a uniform comparable scale, as discussed before, hence we have not done such comparison.

We must keep this in mind that the pro-poor YL dataset used in our study is for a very specific set of population at two specific point in time. Through this pro-poor sample, the YL is trying to understand the meaning of poverty for Indian children (see Table 4 in chapter 2, which shows a description of whole dataset of YL for each round). We see that the information on various variables is not available on all the observation surveyed in each round. The data that has been used in our study has removed, first, those observations who do not have available information on any of the variables from Table 1. Then, we have taken only those children who have chosen/attended either public or private schools. This is done for both the rounds of data. Lastly, we have studied only those children who have these available information at both the rounds.

Table 2: Descriptive statistics of the Variables

	Round 4						Round 3					
	Overall		Private		Public		Overall		Private		Public	
Variables	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Math Score	14.479	7.07	16.38667	6.744	11.90	6.70	11.312	6.156	14.279	6.009	9.513	5.523
Mothers Education	3.632	4.314	4.66	4.514	2.24	3.60	3.632	4.314	5.532	4.676	2.48	3.629
Fathers Education	5.643	5.045	6.75	5.070	4.15	4.62	5.643	5.045	8.182	4.820	4.104	4.537
Household Size	4.753	1.908	4.73	2.036	4.78	1.73	5.086	1.974	5.162	2.117	5.04	1.884
Assets	30018.64	44478.14	35547.63	44685.45	22547.03	43181.86	9291.452	52007.87	12596.24	60446.55	7288.24	46134.97
School Cost	19226.17	26380.52	28759.59	30584.56	6343.17	9218.29	2723.567	5344.192	6752.056	6904.182	281.683	1114.053
Expenditure	7.170	0.587	7.23	0.580	7.09	0.59	6.776	.561	6.931	.487	6.682	.582
Years	13.448	1.745	13.99	1.376	12.71	1.92	8.641	1.308	8.761	1.494	8.569	1.178
Time Allocation	2.013	1.442	2.3	1.379	1.63	1.44	2.626	1.293	2.766	1.342	2.541	1.257
English Score	16.513	3.306	17.27	2.924	15.50	3.52	-	-	-	-	-	-
Gender	.550	.498	0.58	0.494	0.51	0.50	.547	.498	.583	.494	.526	.500
Age in Months	227.88	4.134	227.76	4.143	228.03	4.13	179.197	4.105	179.376	4.130	179.092	4.092
Body Mass Index	19.933	8.744	19.74	3.489	20.20	12.79	17.608	2.767	18.063	3.111	17.333	2.501
Drinking Water	0.981	0.137	0.99	0.115	0.97	0.16	.963	.187	.974	.157	.956	.203
Household Quality	0.725	.217	0.75	0.204	0.69	0.23	.604	.298	.691	.202	.551	.333
Type of School	.425	.495	-	-	-	-	.622	.485	-	-	-	-
Observations	**522**		**300**		**222**		**522**		**197**		**325**	

Table 3: The Mean Math Test score by the groups

Round 4

	Overall			Private			Public			Difference	
	Obs.	**Mean**	**S.Dev.**	**Obs.**	**Mean**	**S.Dev.**	**Obs.**	**Mean**	**S.Dev**	**(Pr)-(Pub)**	**t-test**
Math Score	522	2.047	.953	300	2.317	1.00	222	1.683	.947	.634	7.53

Round 3

	Overall			Private			Public			Difference	
	Obs.	**Mean**	**S.Dev.**	**Obs.**	**Mean**	**S.Dev.**	**Obs.**	**Mean**	**S.Dev**	**(Pr)-(Pub)**	**t-test**
Math Score	522	1.839	1.00	197	2.321	.977	325	1.546	.898	.774	9.240

*the test scores have been normalized by dividing it with the standard deviation of the overall sample for comparison

Table 4: The Difference in the Means of explanatory variables of the two groups

Variable	Overall		Private		Public		Difference	
	R 4	R3	R 4	R3	R 4	R3	R 4	R3
Math Score	14.48	11.312	16.39	14.279	11.90	9.513	4.485***	4.765***
	(7.07)	(6.156)	(6.74)	(6.009)	(6.70)	(5.523)		
Fathers Education	3.63	3.362	4.66	5.532	2.24	2.48	2.416***	3.052***
	(4.31)	(4.314)	(4.51)	(4.676)	(3.60)	(3.629)		
Mothers Education	5.64	5.643	6.75	8.182	4.15	4.104	2.601***	4.078***
	(5.05)	(5.045)	(5.07)	(4.820)	(4.62)	(4.537)		
Household Size	4.75	5.086	4.73	5.162	4.78	2.117	-0.05	.122
	(1.91)	(1.974)	(2.04)	(5.04)	(1.73)	(1.884)		
Assets	30018.64	9291.452	35547.63	12596.24	22547.03	7288.24	13000.61***	5308.004
	(44478.14)	(52007.87)	(44685.45)	(60446.55)	(43181.86)	(46134.97)		
School Cost	19226.17	2723.567	28759.59	6752.056	6343.17	281.683	22416.43***	6470.373***
	(26380.52)	(5344.192)	(30584.56)	(6904.182)	(9218.29)	(1114.053)		
Expenditure	7.17	6.776	7.23	6.931	7.09	6.682	.140***	.249***
	(0.59)	(.561)	(0.58)	(.487)	(0.59)	(.582)		
Years	13.45	8.641	13.99	8.761	12.71	8.569	1.281***	.192**
	(1.75)	(1.308)	(1.38)	(1.494)	(1.92)	(1.178)		
Time Allocation	2.01	2.626	2.30	2.766	1.63	2.541	.673***	.224**
	(1.44)	(1.293)	(1.38)	(1.342)	(1.44)	(1.257)		
English Score	16.51	-	17.27	-	15.50	-	1.771***	-
	(3.31)	-	(2.92)	-	(3.52)	-		
Gender	0.55	.547	0.58	.583	0.51	.526	0.07	.058*
	(0.50)	(.498)	(0.49)	(.494)	(0.50)	(.500)		
Age in Months	227.88	179.197	227.76	179.370	228.03	179.092	-0.27	.278
	(4.13)	(4.105)	(4.14)	(4.130)	(4.13)	(4.092)		
Body Mass Index	19.93	17.608	19.74	18.063	20.20	17.333	-0.46	.730***
	(8.74)	(2.767)	(3.49)	(3.111)	(12.79)	(2.501)		
Drinking Water	0.98	.963	0.99	.974	0.97	.956	0.01	.0176
	(0.14)	(.187)	(0.11)	(.157)	(0.16)	(.203)		
Household Quality	0.73	.604	0.75	.691	0.69	.551	.0530***	.140***
	(0.22)	(.298)	(0.20)	(.202)	(0.23)	(.333)		
Observations	**522**		**300**	**325**	**222**	**197**		

Note: ***p<0.01, ** p<0.05: Standard Deviations are in parentheses

In Table 3 the standardized mathematics test scores is shown for both the rounds as well as by the type of school. We learn from table 3 that the difference in mathematics learning between the children of private school and public school is 0.63 standard deviation in round 4 and it stood at 0.77 standard deviation in round 3. Table 4 summarizes the mean values of each of the background variables by round and subgroup of children. We see that majority of the variables, on an average, are different, significant, and better for the private school children for each round. Almost all the similar variables are significantly better for private school children in both the rounds. The aim now is to see which of these background characteristics contribute to the difference in the average performance of these two groups at these two survey points. In rest of the paper, using the threefold Blinder-Oaxaca decomposition methodology, we study the learning difference with respect to a set of their background characteristics.

4.1.3 Methodological Framework and Empirical Approach

Continuing with the ideas discussed and used in the previous chapters, the mathematics learning (EA) could be depicted as a function of input variables like parental education (PE), household factors (H), Child specific features (C), and other features (O) (as discussed in Table 1).

$$EA = f(PE, H, C, O) \qquad (1)$$

Yet even when an educational production function exists, there is no guarantee that one can estimate it (Artega & Glewwe, 2014). In this chapter we have used the threefold Blinder Oaxaca decomposition for the analysis of the gap in mean mathematics learning between private and public school children. In this empirical analysis we need a reference category from the two groups so we have taken the first group of children ($Private$) to be that and the point of analysis is the public school children.

The reduced form equation for Blinder-Oaxaca decomposition in our analysis, (following Blinder, 1973; Oaxaca, 1973; Jann, 2008) the contribution of the group difference in the predictors to the overall outcome difference can be rearranged as follows:

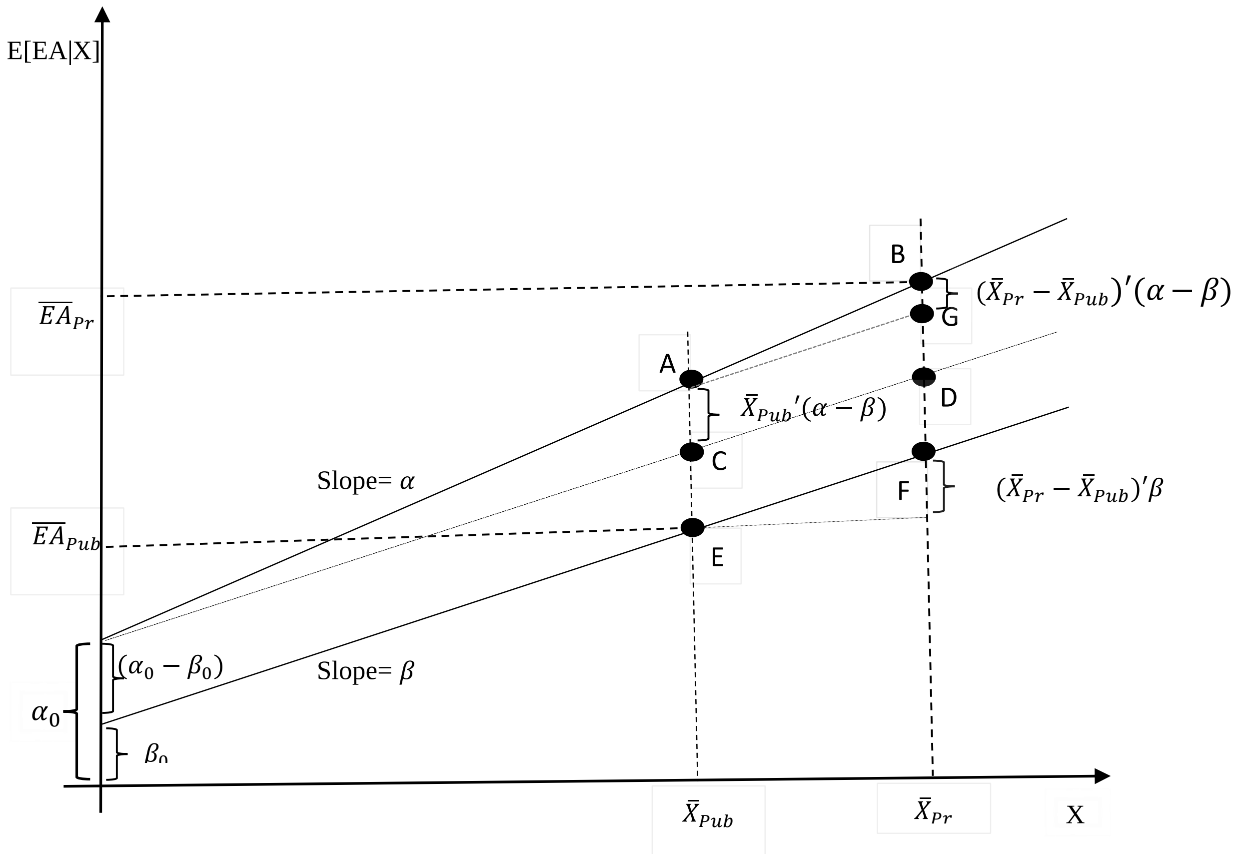
$$D = \{E(X_{Pr}) - E(X_{Pub})\}'\beta + E(X_{Pub})'(\alpha - \beta) + \{E(X_{Pr}) - E(X_{Pub})\}'(\alpha - \beta) \tag{2}$$

Where, D is the difference in the mean standardized mathematics score, Xs are the set (vector) of background characteristics, β is the vector of slope coefficients for the public school children and α is for the private school children. These coefficients include the intercepts as well. This is called the threefold decomposition.

$$D = E + C + I$$

The three effects have been described in detail in the previous chapters already. These three effects (from equation 2) are portrayed in figure 1, which has been borrowed from Arteaga and Glewwe (2014) and has been modified slightly based on our needs. The figure 1 portrays these three effects, borrowing from Arteaga and Glewwe (2014) which has been modified with respect to our study. The decomposition of the test score gap between the groups at a point is shown in figure 1. The first composition of equation (2), the endowment effect, is shown by the move from E to F. This shows the increase in the score of an average public school child if they had similar characteristics as an average private school child. The second term of the equation, the coefficient effect is shown by the vertical distance between A and C which is the increase in the test scores of an average public school child if they had similar returns as an average private school child. The last part of the equation, the interaction of the previous two effects, is shown by the distance between B and G.

Figure 1: Decomposition of Test score gaps between children



4.1.3.1 Empirical Results

Table 5 presents the decomposition of the gap in raw (and standardized) average mathematics test scores between the private school and public school children at the two survey points. The difference in the mean scores stands approximately at 4.486 points (0.63 on standardized scores) in round four where the same in round three stood at 4.77 points (0.77 on standardized score). In the second part of the table we have shown the decomposition of the gaps in the mean scores between the groups of children for each round. The results indicate that the difference in the learning outcome is primarily due to the endowments effect which captures the differences in mean values of the background characteristics (from Table 1) between the two groups of children. We also see the role of coefficient effect, or differences between the groups of children in the impacts of background characteristics in explaining the difference in mean learning in the third round results. In our results, the interaction of endowments and coefficients does not significantly affect the learning outcome difference in either round.

Table 5: Difference in Scores and the Blinder Oaxaca Decomposition

	Coefficient (on raw score)		Coefficients (on Standardized Math Score[44])		p-value	
	Round 4	**Round 3**	**Round 4**	**Round 3**	**Round 4**	**Round 3**
Private	16.387	14.279	2.318	2.321	0.000	0.000
Public	11.900	9.513	1.683	1.547	0.000	0.000
Difference	4.486	4.765	.6345	.774	0.000	0.000
Endowment Effect $\{E(X_{Pr}) - E(X_{Pub})\}'\beta$	5.255	4.719	.7432	.767	0.000	0.021
Coefficient Effect $E(X_{Pub})'(\alpha - \beta)$	.1399	2.425	.0198	.394	0.810	0.001
Interaction Effect $\{E(X_{Pr}) - E(X_{Pub})\}'(\alpha - \beta)$	-.909	-2.379	-.129	-.387	0.337	0.258

The results in third round here mean that if the public school children had had similar average features as their private school counterparts (X_{Pr}), their scores would have improved by 0.76 standard deviations. This is the exact amount of the difference in average learning between the two groups of children. This, assuming that the returns to the background characteristics of the public school children stayed the same (β). The difference in the features applied to the impact of public school children would have exactly removed the gap in learning. However, we also find in our results that *C* also explains a part of the learning difference individually at this point. *C* explains almost half of the learning difference in the third round results (0.39 standard deviations). If the average features of the public school children (X_{Pub}) stayed at the current level, with the impact/returns of the private school children (α), the gap in learning would have been removed by fifty percent of the total difference. The results basically indicate that not only do the private school children have better characteristics than the public school children they also have better returns to those characteristics. Summing *E* and *C* of our results here, we find that these two differences is not only stopping public school children to perform better, in fact, it is acting as an hurdle for them to perform better than the private school children. On the contrary, in the fourth round

[44] By dividing with the Standard deviation of math score for the overall sample.

results we find that these differences in returns $(\alpha - \beta)$ have vanished, but E still exists. However, E explains more than the actual difference in learning at this point of results. Here, the difference in the background characteristics between the two groups of children (unlike round 3 results) alone is stopping the public school children from outperforming their private school counterparts. The actual difference here stood at 0.63 standard deviations whereas E explains 0.74 standard deviations. When we look closely at the results of both the rounds together we see that the consistent contributor to the mathematics learning difference is the endowments effect. However, the contribution of the coefficient effect in explaining the learning difference has withered from round three to round four. This suggests that the contribution of C to the learning outcome difference has vanished and/or the differential returns to the background characteristics operated only when the children were four years younger. As the children got older, it is not the difference in the impact of the background characteristics, rather the difference in the absolute average background characteristics which explains the learning outcome difference.

Table 6: The Decomposition of Gaps for the Standardized Math Score

Round 4

	Endowment Effect		Coefficient Effect		Interaction Effect	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Mothers Education	0.047	.042	-0.047	.046	-0.051	.050
Fathers Education	-0.054	.035	0.063	.070	0.039	.044
Household Size	0.003	.008	0.258	.179	-0.003	.009
Assets	-0.009	.015	-0.006	.036	-0.004	.021
School Cost	0.323**	.133	-0.072*	.039	-0.256*	.136
Expenditure	0.009	.014	-1.096	.902	-0.022	.020
Years	0.159***	.045	0.908	.580	0.092	.059
Time Allocation	0.081***	.030	-0.152*	.082	-0.063	.036
English Score	0.149***	.038	1.100***	.359	0.126***	.046
Gender	0.015	.012	0.064	.0678	0.009	.011
Age in Months	-0.005	.007	-0.650	3.520	0.001	.004
Body Mass Index	0.003	.006	0.124	.251	-0.003	.008
Drinking Water	-0.003	.005	0.563	.462	0.008	.001
Household Quality	0.024*	.015	-0.030	.214	-0.002	.016
Constant			-1.007	3.688		
Total	**0.743*****	**0.000**	**.019**	**0.81**	**-.128**	**0.337**

Round 3

	Endowment Effect		Coefficient Effect		Interaction Effect	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Mothers Education	.148***	.0493	-.048	.062	-.060	.076
Fathers Education	-.068	.0492	.113	.090	.113	.090
Household Size	-.006	.010	.151	.229	.004	.008
Assets	-.004	.007	.004	.012	.003	.009
School Cost	.596*	.341	-.023	.016	-.533	.348
Expenditure	.028	.023	.885	1.182	.033	.045
Years	.008	.009	.358	.521	.008	.0128
Time Allocation	.048*	.027	-.469***	.163	-.041	.026
Gender	.021	.018	.005	.093	.001	.010
Age in Months	.002	.004	1.454	3.661	.002	.006
Body Mass Index	-.003	.014	-.104	.523	-.004	.022
Drinking Water	.003	.005	-.355	.479	-.007	.011
Household Quality	-.006	.020	0.375*	.204	.0956	.054
constant	-	-	-1.952	3.753	-	-
Total	**.767****	**0.021**	**.394*****	**0.001**	**-.387**	**0.258**

Note: *** p<.01, ** p<.05, * p<.10

In Table 6 we look closely at the detailed contribution of various background characteristics in both of the above discussed effects. As we have found in the fourth round results that it is just the endowments effect which explains the gap in mean learning, we shall look into the contributors to this effect. Our results identify the contributors to the above mentioned effects. In round four results, the major contributor to *E* is the schooling cost incurred in the last academic year which explains 43 percent of total *E*. This is followed by years of schooling received by the child at the time of survey and the English language score, each of which explain 22 percent of total *E*. The rest of the *E* is explained by time allocation on studies (11 percent of *E*) and the household quality (2 percent of *E*). In round 3, 78 percent (which is higher compared to round 4) of the total *E* is explained by the schooling cost incurred in the last academic year. This is followed by an 18 percent contribution by the mothers' education years and 7 percent by the time allocation on studies. In the coefficient effect we find that there is just one contributor, the household quality index.

When we look at the results of these two rounds together, we notice that the consistent effect is the endowments effect. And the consistent contributor to *E* includes the schooling cost (similar effect is also found in Kingdon, 1996; Arteaga and Glewwe, 2014) and time allocation on studies (also found in Kingdon, 1996; Glewwe *et al.*, 2015). Attention has to be paid while interpreting the schooling cost with reference to the learning differences here. The composition of the schooling cost is different in each of the rounds because of the data limitations. Moreover, schooling cost here, not only captures the household's ability to spend on schools but school level heterogeneity as well. Those differences has to be kept in mind while reading the schooling cost in our results. The contribution of mothers' education years has vanished between the rounds. However, a large part now in the later round is explained by the years of schooling gap (as discussed in Glewwe *et al.* 2015). When the children are older, one of the core contributor to the mathematics learning difference between the public and private school children is the years of schooling. We must take note of the fact that years of schooling gap (as shown in table 4) has magnified between the rounds where it has increased even further in round 4. The gap in mean years of schooling between these children was marginal when these children were younger (15-16 years old). However, this gap has magnified in four years where the private school children have gained more additional years of schooling compared to the public school children. The latter's years of schooling has increased as well in the following four years but this increase is less than the former's increase. This could mean one of the following two things. The public school children have dropped out from school more often than the private school, which has eventually widened the gap in average schooling years. Or, the public school children have repeated a course/grade more often than the private school children.[45] Meaning, the private school children have stayed in school for a longer period of time compared to their public school counterparts. Additionally, this could also means that the drop out/discontinuation is greater for the public school children because in the same four years gap they have completed lesser additional years of schooling. As Sanjay *et al.* (2014) points out that the dropout rate among Indian children in this age group is high, these results should be understood in that context. In either case, the gap in average years of schooling between the public and private school children would increase. Both of these situation work in favor of the children of private school. Moreover, in round 3 results, *C* explained the gap in learning only when children were younger and the sole contributor to this effect was differential returns to index of household quality. The whole of coefficients effect has vanished between the rounds which is promising as it indicates that the differential returns to background characteristics do not operate when these children are older. It is only the differences in the background features which consistently explain the learning outcome gap between the two groups of children.

4.1.4 Conclusion and Policy Implications

Using the third and fourth round of Young Lives Survey (YLS) data for the older cohort of Indian children, we study the mathematics learning inequality between the children of public and private school. The children in this chapter have been approached from the direction of the type of school they are attending/last attended. Skimming through their background characteristics we notice that they are different and significantly better for the children of private school. In the further analysis, we study the difference in average mathematics learning with respect to these differences in background characteristics

[45] The years of schooling variable is obtained from the information provided by the children on the highest grade they have completed at each round of survey. Each grade completed (even if there is a repetition of grade) is treated as having obtained one year of schooling. There is no pre-school attendance years for either choice of schooling.

using the threefold Blinder-Oaxaca decomposition. In the broad decomposition results (for each round separately, where the gap between rounds is four years) the mathematics learning difference is consistently explained by their endowments difference/effect. The contributors to this effect have also been found to be consistent, which are the schooling cost and time allocation to studies. The contribution to *E* of former round, though has proportionately declined between the rounds. However, the time allocation on studies' contribution has increased between the rounds. As we have used the English language score in the analysis for the fourth round, we have found that this too has a significant role to play in the learning gap between these groups of children. The test questionnaire on mathematics was framed/asked in the local language, Telugu as well as English. Our results point out that those children who better performed in both have had better learning outcomes in the fourth round results and we have seen that the children of private schools have performed better on English. We do not have this score available in the third round survey, so we were unable to use the information in the chapter for the third round. Additionally, we find that the significance of *C* is present only in the results of third round which is when the children were four years younger. Unlike the consistent contribution of the differences in the average background features, the differential returns to these features has withered from round 3 to round 4 results. In sum, the results tell us that the major part of the contribution to the learning difference between the two groups of children is explained by the mean differences in the background characteristics. Where at both of these survey points, the schooling cost explains a larger chunk of the total endowments effect followed by the time allocation on studies. The significance of schooling cost indicates that the children who were able to spend a larger amount on schooling were able to perform better on mathematics. The schooling cost, apart from household's ability to spend on education, captures the quality and services a child is able to avail from schools. This cost also captures the school level heterogeneity. The years of schooling gap has magnified between the rounds and now explains a large proportion of the endowments effect.

The results in our chapter point towards both policy as well as investigative direction. The foremost policy implication comes from the consistent contribution of the schooling cost to the learning difference between the two groups of children. Attention needs to be paid to this as higher paying ability to schooling (towards private schools) entails better schooling facilities. This has to be dealt with in terms of providing better facilities in public schools (the cost there is already lesser) which should add to the environment of learning. Policies are needed to address this dilapidated state of public education system in terms of making and creating a better atmosphere of learning for these children. Secondly, the sudden significance of years of schooling gap in explaining the learning gap is worrying because the surveyed children are almost of the similar age. The increase in the years of schooling gap within four years warrants policy intervention in terms of incentivizing (monetary or kind) staying in education. Lastly, the consistent significance of time allocation on studies needs exploratory attention as we see that the public school children spend a lesser average amount of time on studies. Moreover, this gap has magnified within four years which raises question pertaining to their time allocation pattern. Is it that they spend their time various other activities (economic?) because there is an opportunity cost attached to staying in education. The time allocation on studies and the years of schooling has to be re-looked as they both seem to be connected.

References

Annual Status of Education Report. (2005). Pratham Resource Centre, New Delhi.

Annual Status of Education Report. (2010). Pratham Resource Centre, New Delhi.

Ammermueller, A. (2007b). Pisa: What Makes the Difference? Explaining the Gap in Pisa Test Scores between Finland and Germany. *Empirical Economics,* 33 (2), 263-287.

Arteaga, I., & Glewwe, P. (2014). Achievement Gap between Indigenous and Non-Indigenous Children in Peru. *Young Lives Working Paper 130*.

Azubuike, Obiageri Bridget, and K. B. (2016). Young Lives' Constructed files. *Oxford Review of Economic Policy*, *23(2)* (7483), 168–195.

Azubuike, O. B., Moore, R., & Iyer, P. (2017). Young Lives School Surveys, 2016–17: The design and development of cross-country Maths and English tests in Ethiopia, India & Vietnam. *Young Lives*.

Baird, K. (2012). Class in the classroom: The relationship between school resources and math performance among low socioeconomic status students in 19 rich countries. *Education Economics, 20*(5), 484–509.

Barrera-Osorio, F., García-Moreno, V., Patrinos, H. A., & Porta, E. (2011). Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: An application to Indonesia's results in PISA mathematics. *The World Bank*

Blinder, A. S., (1973). Wage Discrimination : Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Burger, R. (2011). School effectiveness in Zambia: The origins of differences between rural and urban outcomes. *Development Southern Africa*, *28*(2), 157–176.

Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review*, *31*(4), 376–390.

Cobb-Clark, D. A., & Gørgens, T. (2014). Parents' economic support of young-adult children: Do socioeconomic circumstances matter? *Journal of Population Economics*, *27*(2), 447–471.

Crane, J. (1996). Effects of home environment, SES, and maternal test scores on mathematics achievement. *Journal of Educational Research*, *89*(5), 305–314.

Crookston, B. T., Forste, R., McClellan, C., Georgiadis, A., & Heaton, T. B. (2014). Factors associated with cognitive achievement in late childhood and adolescence: The Young Lives cohort study of children in Ethiopia, India, Peru, and Vietnam. *BMC Pediatrics*, *14*(1), 1–9.

Cueto, S., and Leon, S. (2012). Psychometric Characteristics of cognitive development and achievement instruments in round 3 of Young Lives (Young Lives Technical no Notes) Young Lives.

Currie, J., & Yelowitz, A. (2000). Are public housing projects good for kids ? *Journal of Public Economics*, *75*(1), 99–124.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Score. *American Economic Journal: Applied Economics*, *5*(2), 29–57.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304.

Dawes, A. (2020). Measuring the Development of Cognitive Skills Across Time and Context: Reflections from Young Lives. *Young Lives*

Dercon, S., & Krishnan, P. (2009). Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries. *Children, Youth and Environments*, *19*(2), 138–163.

Desai, S., Dubey, A., Vanneman, R., & Banerji, R. (2008). Private Schooling in India: A New Educational Landscape. *India Policy Forum*, *5*(1), 1–58.

Galab, S., Prudhvikar, R., & Singh, R. (2014) Young lives survey design and sampling in Andhra Pradesh, India

Galab, S., P. Prudhvikar Reddy and Renu Singh (2014a). Education and Learning: Round 4 Preliminary Findings, *Round 4 Fact Sheet*, *Oxford*: Young Lives

Glewwe, P., Chen, Q., & Katare, B. (2015). What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data. *Asia and the Pacific Policy Studies*, *2*(3), 494–516.

Goyal, S. (2009). Inside the house of learning: The relative performance of public and private schools in Orissa. *Education Economics*, *17*(3), 315–327.

Goyal, S., & Pandey, P. (2009). How do government and private schools differ? Findings from two large Indian states. *International Bank for Reconstruction and Development/the World Bank*

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

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*. 8(4), 453-479.

Kingdon, G. G. (1996). The Quality and Efficiency of Public and Private Schooling: A Case Study of India. *Oxford Bulletin of Economics and Statistics*, *58*(1), 57–82.

Keane, M. P., Krutikova, S., & Neal, T. (2018). The impact of child work on cognitive development: results from four Low to Middle Income countries. (No. W18/29) *IFS Working Papers*.

Kingdon, G. G. (2007). The progress of school education in India. *Oxford Review of Economic Policy*, 23(2), 168-19 5.

Leibowitz, A. (1977). Parental Inputs and Children's Achievement. *The Journal of Human Resources*, *12*(2), 242–251.

Li, Z., & Qiu, Z. (2018). How does family background affect children's educational achievement? Evidence from Contemporary China. *The Journal of Chinese Sociology*, *5*(1), 1–21.

Muralidharan, K., & Sundararaman, V. (2015). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics*, *130*(3), 1011–1066.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*, 693–709.

Paxson, C., & Schady, N. (2007). Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. *Journal of Human Resources*, *42*(1), 49–84.

Rolleston, C. (2014). Learning profiles and the "skills gap" in four developing countries: A comparative analysis of schooling and skills development. *Oxford Review of Education*, *40*(1), 132–150.

Sakellariou, C. (2008). Peer effects and the indigenous/non-indigenous early test-score gap in Peru. *Education Economics*, *16*(4), 371–390.

Sanfo, J. B. M. B., & Ogawa, K. (2021). Explaining the rural-urban learning achievements gap in Ethiopian primary education: A re-centered influence function decomposition using Young Lives data. *Education Economics*, 1–29.

Sanjay, S., Gupta, R., Gautam, V., & Meena, S. H. R. (2014). Educational Statistics at a Glance. *MHRD, Bureau of Planning, Monitoring & Statistics, New Delhi*.

Singh, A. (2015). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics*, *113*(2015), 16–32.

Singh, R., & Mukherjee, P. (2019). Effect of preschool education on cognitive achievement and subjective wellbeing at age 12: evidence from India. *Compare: A Journal of Comparative and International Education*, *49*(5), 723–741.

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.

Wadhwa, W. (2009). Are private schools really performing better than government schools? *Annual Status of Education Report (Rural) New Delhi*.

Wamalwa, F. M., & Burns, J. (2018). Private schools and student learning achievements in Kenya. *Economics of Education Review*, *66*(July), 114–124.

Zhang, Y. (2006). Urban - Rural Literacy Gaps in Sub - Saharan Africa : The Roles of Socioeconomic Status and School Quality. *Comparative Education Review*, *50*(4), 581–602.

Zhang, L., & Lee, K. A. (2011). Decomposing achievement gaps among OECD countries. *Asia Pacific Education Review*, *12*(3), 463–47

5.0 Inside the Mathematics Learning Inequality at a Point

Abstract

In this chapter, unlike the previous chapters, we approach the children directly from the direction of their mathematics learning outcome. Using the 2014 Young Lives Survey data for India, we divide the children into two groups, better-performing children and rest of the children. The former have better background characteristics on an average as compared to the latter. Using the Blinder-Oaxaca decomposition, we find that a large part of learning difference is explained by the differences in their characteristics. Moreover, the impact of these differences in the form of average returns they reap for both the groups is significantly different where it is better for the former.

Keywords: Mathematics Learning; Blinder-Oaxaca Decomposition; Endowment Effect; Coefficient Effect

5.0.1 Introduction

We have pointed out earlier in the previous chapters that the learning outcomes of children from India is significantly lower than the international average (Kingdon, 2007).[46] However, there are states in India that perform better as compared to the national average. The National Achievement Survey (NAS, 2014) class III (cycle 3) finds no significant difference in the performance of students form Andhra Pradesh and the national average in language while the former's performance is significantly better than the national average on mathematics. It is here we investigate this question that who are the children that perform better than the others on the same mathematics test using Young Lives Survey for India. Is there a role of characteristics difference that result in the better performance of few children as compared to the others?

A number of studies have established factors that have a key role to play in influencing the educational outcomes of children. Blau and Duncan (1967) and Duncan *et. al.* (1972) using 1962 survey data by the United States Bureau of the census obtained for five year birth cohorts of adult males talk about the relationship between their socio-economic characteristics and educational achievement. A similar effect of household facilities is captured in Currie and Yelowitz, (2000) using the Survey of Income and Program Participation (SIPP), USA. The academic achievement is affected indirectly by the household expenditure (Blau and Duncan, 1967; Duncan *et. al.* 1972) for USA, and similarly for Australian children in Cobb *et al*. (2014) using Youth in Focus survey. It is also affected by the whole Socio-economic status of the family in Davis, (2005) using national cross sectional study of children (8-12 years old) and in Jeynes', (2007) meta-analysis of 52 studies. Moreover, the influence of household wealth on the language score in the study of poor preschool-aged children from Ecuador (in Paxson and

[46] The Trends in Mathematics and Science Study (TIMSS) were applied where questions were asked to the secondary school going children from the states of Rajasthan and Orissa (Indian States).

Schady, 2007) and overall advantages and disadvantages of family (in Cobb *et al*. 2014) affects the choice of either education and employment of young adults.. The household and child characteristics (Witte, 1992; Kingdon, 2007), cost effectiveness of private schools (Kingdon, 1996; Goyal, 2009) have an influence on the learning outcomes of children from India. The parental level of education exerts an influence on the educational outcomes of children·.[47] The parental education influences the educational outcome of children directly, as discussed in Mare (1980, 1981), where the influence to the educational continuation probability of American male children to further grades is discussed. The education of parents specifically influences the educational performance of children directly (as found in Paxson and Schady, 2007) and similar results are found in the study of Peruvian children by Artega and Glewwe (2014) and indirectly in (Davis, 2005). There are studies that found an influence of just one parents' education on the educational outcomes of children. Family head's education influences the educational achievement of children (Blau and Duncan, 1967) and same is influenced by just fathers' education (Duncan *et. al.* 1972). Similarly, Sakellariou (2008) finds that fathers' education influences a significant part of mathematics and language test scores of children from Peru. Fathers' education is found to be more important than mothers' education in influencing the child's academic performance in Holmlund *et al*. (2011) using Swedish register data. On the other hand, there are studies that have results where it is just mothers' education that has an influence on the educational performance of children. Leibowitz (1977) in a study of children from California finds that there is a significant positive relationship between mothers' schooling and children's test scores. Similarly, mothers' cognitive test scores has significant effects on the mathematics test scores of children using National Longitudinal Survey of Youth in (Crane, 1996; Todd and Wolpin, 2007).

[47] The educational outcome discussed includes the achievement scores, educational continuation decision, returns to education, etc.

There are studies that talk, unlike the ones discussed so far, about parental involvement and academic performance (such as Reynolds, 1992; Glewwe *et. al.*, 2017) in Chicago, Vietnam and Peru. There are also meta-analysis (for instance Fan and Chen, 2001; Jeynes, 2005; Jeynes; 2007) which have talked about the parental involvement and the relationship it has with different educational outcomes of children.[48] The educational expectations of parents in a study of American children by Mau (1997), and parental expectations, supervision, acceptance and psychological autonomy of Canadian children in Deslandes *et al*. (1997), parenting style (in Paxson and Schady, 2007), have discussed about its influence on the academic performance of children.

The child features too are responsible for how he/she performs in school and/or continues in school. The schooling continuation decision of American, Swedish, and Indian children is greatly influenced by how a child performs academically (Mare, 1979; 1980; Breen and Jonsson, 2000; Mukhejee and Pal, 2016). The health of a child has a crucial and a very important association with the educational achievement of children. For instance, studies (such as Behram, 1996; Glewwe *et. al.*, 2001; Paxson and Schady, 2007; Frisvold, 2015; Glewwe *et. al.*, 2017) have pointed out the connection between health/nutrition and the learning outcome of children from Philippines, Ecuador, and America. The educational outcome in general (Belot and James, 2011; Mukherjee and Pal, 2016) is affected by the child's health of English and Indian children.

The learning outcomes of children is also influenced by the type of schools they are in, where there are studies that have found the existence of a private schooling effect in learning outcomes. The private enrolment is associated with better child outcomes after controlling for family background characteristics in India (Desai *et. al.* 2008). Similarly, (Chudgar and Quin, 2012; Wamlawa and Burns, 2012; Singh,

[48] The involvement is broadly defined as: parental aspiration/expectation for children's education achievement and parental home supervision.

2015) have also found an association between attending private schools and having better achievement outcomes in Indian children. Wadhwa (2009) points out that the supplemental helps provided by Indian parents at home like tuitions etc. also explains a part of this difference in the performance of children from these two types of schools. There are studies that have found different results (for instance, Goldhaber, 1996) where the overall private schools have no significant advantages in education on mathematics and reading. Goyal and Pandey (2009) finds that the private performance in India is better than the public schools however the quality of both the type of schooling is low in both the cases.

These studies have largely focused on the factors influencing the learning outcomes not on the extent to which they cause or explain the differences in the learning outcomes between groups of children. This chapter is an attempt to add to the growing body of literature on the learning outcomes of school going children. We have used the dataset of Young Lives (YL henceforth) Survey, India, for the older cohort of the fourth round. Our chapter is for a specific population at a specific point in time. These children have grown up in the context of poverty and the dataset traces the changes occurring in their lives as they are growing up. Using various features around which these children are growing, varying from socio-economic features, parental education and childhood features, the study shall research the mathematics learning gaps between the first division scoring children and rest of the children.[49] The primary focus of this study is to look into the background of the first division scoring children as against rest of the children at the time of fourth round YL survey. The idea is to see whether their respective differences (if any) explain the differences in the learning outcomes and to what extent they explain the differences in the performance of these children. We slightly depart in this chapter from the previous chapters and studies that have been done so far, especially in India. Here, instead of studying the factors explaining the learning

[49] Those who have scored 60 percent and above marks in the mathematics test conducted during the fourth round of YLS. Are called the better performing or first division children and remaining as rest.

outcomes of children, we, by using the major factors that are established to have an impact on the learning outcomes, study how they contribute to the differences in the learning outcome of children where few perform better than the rest. The studies of similar sort are minimal that try to explain the learning outcome gaps between Indian children and attribute it to the factors that contribute to the difference in the performance.

The rest of the chapter is organized as follows. Section 2 gives the description of data, source and variables used in our study along with descriptive statistics. Section 3 presents the methodological framework and empirical approach with the results followed by summary and conclusion of this chapter in Section 4.

5.0.2 Data and Variables

This chapter uses the Young Lives data for India (United Andhra Pradesh) which is a collaborative longitudinal research project coordinated by a team based at the University of Oxford. The survey includes four developing countries Ethiopia, India, Peru, and Vietnam to trace the nature of changing poverty. The survey has so far done five rounds across these countries tracing the lives of 12000 children growing up in the context of poverty. The survey is based on a pro-poor sample which has nearly an equal number of boys and girls, based both in rural and urban communities. These 3000 children in the Indian survey were selected from 20 sentinel sites spread now across two southern states of Andhra Pradesh and Telangana. The sites have been defined specifically for each country. The sampling method deployed in the data collection is a form of purposive sampling (Galab *et. al.* 2014). We have made use of the fourth round data, just like chapter 3.0 and 4.0, for the older cohort on India (united Andhra Pradesh) that was released in 2014. The survey in India was conducted on two age cohorts, young and old, where the younger cohort was 8 months of age when the survey was first conducted in the year 2002 and the older cohort was 8 years of age at that time. The round covers 952 children from the older cohort. Our chapter had to drop out observations as many of the information on various variables were missing. Majority of the observations were missing on the parental education variable. Moreover, we have considered only those children who are studying/studied in purely public, private schools. Those who have studied in a school different form these are not considered. Eventually we ended up with 522 observations for the analysis. We also had to consider the consistency of variable information both for round 3 and round 4 as the next chapter looks into both the rounds together. The Young Lives Survey provides information on various background characteristics of children and detailed record of necessary information needed for this research. The Table 1 gives description of the variables selected based on the literature reviewed used in our study and Table 2 gives the descriptive statistics of the variables.

Table 1: Description of the Variables

Variables		*Description*
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of the survey out of 30 marks
Parental Education (PE)[50]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets	Value of the five most valuable assets owned, rented or borrowed in the YL child's household
	School Cost	Total expenditure incurred on school in the last academic year.[51]
	Expenditure	The log of per capita monthly expenditure of YL child's household
	Years	Years of schooling received by YL child at the time of fourth round survey
Child Specific Features (C)	Type of School	=1 if attended/attending public school last, 0 if private.
	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of fourth round survey
	Body Mass Index	Body mass index of the YL child at the time of fourth round survey
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[52]

Table 2: Descriptive statistics of the Variables

Variables	Observations	Mean	Median	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis
Math Score	522	14.479	15	0	29	7.074	-0.105	1.958
Mothers Education	522	3.632	1	0	14	4.314	0.777	2.287
Fathers Education	522	5.644	5	0	14	5.046	0.220	1.586
Household Size	522	4.753	4	1	26	1.908	3.607	33.576
Assets	522	30018.640	16000	100	563900	44478.140	6.573	69.525
School Cost	522	19226.170	8425	0	189000	26380.520	2.486	10.487
Expenditure	522	7.170	7.200	5.193	8.739	0.588	-0.396	3.277
Years	522	13.448	14	10	15	1.745	-0.617	1.922
Time Allocation	522	2.013	2	0	8	1.442	0.581	3.286
English Score	522	16.513	17	4	22	3.306	-0.570	3.078
Gender	522	0.550	1	0	1	0.498	-0.200	1.040
Age in Months	522	227.877	228	219	238	4.134	-0.069	1.958
Body Mass Index	522	19.934	18.900	0.196	200.4008	8.744	17.119	350.045
Drinking Water	522	0.981	1	0	1	0.137	-7.016	50.220
Household Quality	522	0.725	0.821	0.021	1	0.217	-1.256	4.009
Type of School	522	0.57	1	0	1	.49	-.30	1.091

[50] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[51] This cost is the sum total of the tuition fees, education charges, private tuition, accommodation, transportation, uniforms, stationary etc. The school level heterogeneity will be captured by the variance in the sum total of these costs.

[52] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

As discussed previously, we have divided the children into two groups, namely, the better performing children, *F*, (first division scoring: who have scored sixty percent or above marks in mathematics) and the rest of the children as *R*. In Table 3 we present the descriptive statistics of all the variables described in Table 1 by the sub-group of children. The first part of the table describes the statistics of the first division scoring children and the second part describes the same for rest of the children.

Table 3 Sub-sample Descriptive statistics

Variables	Observations	Mean	Median	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis
First Division Children								
Math Score	180	22.289	22	19	29	2.482	0.463	2.318
Mothers Education	180	5.161	5	0	14	4.479	0.291	1.945
Fathers Education	180	6.872	7	0	14	5.302	-0.083	1.536
Household Size	180	4.600	4	2	26	2.116	6.131	59.961
Assets	180	35562.420	20650	1100	515000	48508.220	5.980	55.339
School Cost	180	31147.950	16300	0	189000	33431.980	1.643	5.922
Expenditure	180	7.193	7.196	5.263	8.728	0.570	-0.480	3.473
Years	180	14.411	15	11	15	1.056	-1.592	4.053
Time Allocation	180	2.461	2	0	8	1.283	0.683	3.833
English Score	180	18.700	19	11	22	2.347	-0.695	3.111
Gender	180	0.650	1	0	1	0.478	-0.629	1.396
Age in Months	180	228.289	229	219	237	4.170	-0.133	1.973
Body Mass Index	180	19.583	18.964	0.196	30.963	3.272	-0.213	10.230
Drinking Water	180	0.994	1	0	1	0.075	-13.304	178.006
Household Quality	180	0.782	0.848	0.125	1	0.182	-1.493	4.651
Type of School	180	0.255	1	0	1	0.437	-1.120	2.256
Rest of the Children								
Math Score	342	10.368	10	0	18	4.908	-0.177	1.896
Mothers Education	342	2.827	0	0	14	4.003	1.101	2.851
Fathers Education	342	4.997	5	0	14	4.788	0.356	1.668
Household Size	342	4.833	4	1	13	1.787	1.433	6.612
Assets	342	27100.860	13450	100	563900	41983.730	7.042	81.612
School Cost	342	12951.550	6025	0	155850	19028.050	3.300	17.766
Expenditure	342	7.158	7.200	5.193	8.739	0.598	-0.353	3.188
Years	342	12.942	12	10	15	1.823	-0.166	1.583
Time Allocation	342	1.778	2	0	7	1.466	0.707	3.295
English Score	342	15.363	16	4	22	3.155	-0.488	3.099
Gender	342	0.497	0	0	1	0.501	0.012	1.000
Age in Months	342	227.661	228	219	238	4.104	-0.040	1.953
Body Mass Index	342	20.118	18.851	14.076	200.401	10.540	14.896	252.459
Drinking Water	342	0.974	1	0	1	0.160	-5.918	36.027
Household Quality	342	0.695	0.806	0.021	1	0.228	-1.122	3.631
Type of School	342	0.515	0	0	1	0.500	0.585	1.003

The Table 4 portrays the standardized outcome variable (mathematics score) for the whole dataset as well as the sub-sample dataset. The mathematics performance gap as we see here is 1.686 standard deviations, which is statistically significant. We see a big difference between the performances of both of these group of children. The mathematics scores here have been standardized using the standard deviation of the whole dataset for ease of comparison.

Table 4: The Mean Math Test score by the groups

	For Whole Sample			First Division			Rest			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(F)-(R)	t-test
Math Score	522	2.047	1.000	180	3.152	.351	342	1.466	.694	1.686	30.58

*the test score has been normalized by dividing it with the standard deviation of the overall sample for comparison

Table 5 summarizes the mean of each variable for the whole dataset and by the division they have scored on mathematics test. The differences in the means of each variable is also presented to see whether they are significant. We see that majority of the variables are different, significant, and better for the first division scoring children. For instance, the first division scoring children have better schooled parents on an average. The mean years of schooling of both fathers and mothers is greater for them. The average asset holding (owned, rented, or borrowed) is 31 percent greater for the first division scoring children as compared to the rest. The average schooling expenditure in the last academic year for the first division scoring children is more than double the average expenditure for the rest of children. Similarly, the first division children have spent an average greater number of years on schooling than rest of the children. They have also spent a greater average amount of time on studies on a typical day apart from that in school. They also have better average score on English language. Moreover, 65 percent of the better performing children are males as against 49 percent for rest of the children. They have better household quality as measured by the household quality index than the rest of the children. Lastly, around 74 percent of the first division scoring children have studied/studying in the private school as against 48 percent for the rest of the children. Whether these differences in the features of these two groups of children contribute to the differences in the mathematics learning of the children is what we attempt to answer in the remaining section of this chapter.

5.0.3 Methodological Framework and Empirical Approach

The learning outcome of children depends on various background characteristics. This relationship could be depicted in a mathematical relationship. A simple production function can be used to express this relationship between mathematics learning and a set of background variables. This relationship has been discussed in the previous chapters too, albeit differently. We can see from table 5 that the better performing children (*F*) have better average characteristics compared to the rest of the children (*R*).

Table 5: The Difference in the Means of explanatory variables of the two groups

Variable	Overall	First Division (F)	Rest of the children (R)	Difference
Fathers Education	5.644	6.872	4.997	1.875146***
	(5.046)	(5.302)	(4.788)	
Mothers Education	3.632	5.161	2.827	2.333626***
	(4.314)	(4.479)	(4.003)	
Household Size	4.753	4.600	4.833	-0.233
	(1.908)	(2.116)	(1.787)	
Assets	30018.640	35562.420	27100.860	8461.554***
	(44478.140)	(48508.220)	(41983.730)	
School Cost	19226.170	31147.950	12951.550	18196.4***
	(26380.520)	(33431.980)	(19028.050)	
Expenditure	7.170	7.193	7.158	0.035
	(0.588)	(0.570)	(0.598)	
Years	13.448	14.411	12.942	1.469591***
	(1.745)	(1.056)	(1.823)	
Time Allocation	2.013	2.461	1.778	.6833333***
	(1.442)	(1.283)	(1.466)	
English Score	16.513	18.700	15.363	3.337427***
	(3.306)	(2.347)	(3.155)	
Gender	0.549	0.65	0.49	0.15***
	(0.497)	(0.478)	(0.500)	
Age in Months	227.877	228.289	227.661	0.628
	(4.134)	(4.170)	(4.104)	
Body Mass Index	19.934	19.583	20.118	-0.536
	(8.744)	(3.272)	(0.540)	
Drinking Water	0.981	0.994	0.974	0.021
	(0.137)	(0.075)	(0.160)	
Household Quality	0.725	0.782	0.695	.086744***
	(0.217)	(0.182)	(0.228)	
Type Of School	0.425	0.255	0.515	-0.259***
	(0.494)	(0.437)	(0.500)	
Observations	**522**	**180**	**342**	

Note: ***p<0.01, ** p<0.05: Standard Deviations are in parentheses

The mathematics learning (EA) is a function of input variables like parental education (PE), household factors (H), Child specific features (C), and other features (O).

$$EA = f(PE, H, C, O) \qquad (1)$$

Yet even when an educational production function exists, there is no guarantee that one can estimate it (Artega & Glewwe, 2014). The linear regression equation for the whole dataset is shown below:

$$EA = \theta_0 + \theta_1(PE) + \theta_2(H) + \theta_3(C) + \theta_4(O) + \epsilon_i \qquad (2)$$

Following (Blinder, 1973; Oaxaca, 1973; Jann, 2008) the contribution of the group difference in the predictors to the overall outcome difference can be rearranged as follows:

$$D = \{E(X_F) - E(X_R)\}'\beta + E(X_R)'(\alpha - \beta) + \{E(X_F) - E(X_R)\}'(\alpha - \beta)$$

This is referred to as threefold decomposition. We need a reference category in the empirical analysis of our research hence we have taken the *F* children as that category and the results are interpreted from the point of *R* children.

$$D = E + C + I$$

The first component,

$$E = \{E(X_F) - E(X_R)\}'\beta$$

Amounts to the part of the difference that is due to the group difference in the predictors which is called the "endowment effects". That part of the difference in the learning outcome of (F) and (R) children's mathematics score which is explained because of the differences in the mean values of the background characteristics. The difference in their average background characteristics applied to the impact (returns) of the rest of the children. If the rest of the children were to have similar average background features ($\bar{X}_F$) as F children, this is how much their scores would have improved. This, assuming, the returns of the *R* children (β) at the current level.

The second component,

$$C = E(X_R)'(\alpha - \beta)$$

Measures the contribution of the differences in the coefficients which includes also the intercept. That part of the difference in the learning outcome of these two groups of children explained by the difference between the groups' impact of the background characteristics. To put it differently, the impact on the outcome variable is different for each group and that is causing/explaining the difference in the outcome of each group. Keeping the average features of the *R* children at the current level ($\bar{X}_R$), if they were to have similar average returns as the better performing children (α), this is how much their scores would have improved. Differences in the returns to the background features between the two groups applied to the average features of *R* children.

And the third component,

$$I = \{E(X_F) - E(X_R)\}'(\alpha - \beta)$$

This is an interaction term which accounts for the fact that differences in endowments and coefficients exist simultaneously between the two groups and is causing the difference in the mean mathematics learning simultaneously.

5.0.3.1 Empirical Results

The Table 5 presents first the breakdown of the raw (and standardized) mathematics test scores for both the first division scoring and rest of the children. The first row shows the mean test scores for the first division children, followed by for the rest of the children in second row. Third row shows the difference between mean test scores of both the groups of children. The difference in the mean scores stands approximately at 12 points (1.69 on standardized scores). The results indicate that the difference in the learning outcomes is primarily due to the coefficient effect, or differences between first division and rest of the children in the impacts of considered background characteristics (from Table 1). The learning outcome gap is also explained by the endowment effect (which is the difference between two groups in the mean values of background characteristics). In our results, the interaction of endowments and coefficients, though small, exists too which is negative in value. On the standardized mathematics score, out of the total 1.68 standard deviations, the coefficient effect contributes 1.42 standard deviation, endowment effect contributes 0.48 standard deviation, and interaction effect stands at -0.22. The value of the interaction effect is -0.22 which offsets a part of the learning gap explained more by the combined endowment and coefficient effects. All of these effects are significant at 1 percent level. With the average features of the *F* children, keeping the returns of *R* children at the current level, the gap in learning would have been removed by 84 percent. Similarly, with the returns of *F* children, keeping the features of *R* children at the current level, the gap in learning would have been reduced by 29 percent. Since the value of the interaction effect is -0.22 standard deviation or 13 percent, the total of these three effects stands at 100 percent or the total actual difference in learning between these children.

Table 6: Difference in Scores and the Blinder Oaxaca Decomposition

	Coefficient	**Standardized Math Score**[53]	**p-value**
First Division Children	22.289	3.153	0.000
Rest of the children	10.368	1.467	0.000
Difference	11.920	1.686	0.000
Endowment Effect $\{E(X_+) - E(X_0)\}\alpha$	3.457	0.489	0.000
Coefficient Effect $E(X_+)'(\alpha - \beta)$	10.052	1.421	0.000
Interaction Effect $E(X_0)\}'(\alpha - \beta)$	-1.590	-0.224	0.001

Because the coefficient effect is so important in our results, we look more closely at the individual factors that contribute to it. The Table 7 shows the detailed decomposition results by individual variables. The greatest contributor to this is the schooling years received by the child at the time of fourth

[53] By dividing with the Standard deviation of math score for the overall sample.

round survey which contributes to almost two-third of the total coefficient effect and one-third of the C is unexplained in our results. Similar effect of years of schooling is also found in the study of Vietnamese children by Glewwe *et al.* (2015). The result is a worrying sign as we have mentioned before that the children surveyed belong to the same age group and the existence (and further influence of it on the learning difference) points out that there is drop-out happening from schooling, more so for the children who cannot/did not perform well. Similarly, for the endowments effect, English language is the greatest contributor (a little over one-third or 39 percent of the total endowments effect) followed by years of schooling (contributing almost one-third or 33 percent of the total *E*). As we have pointed out in the earlier chapters that the significance of English language score in determining the mathematics score has to be understood in the context of the test questionnaire which was asked in both English and Telegu. Around 14 per cent of the endowment effects is explained by the schooling cost incurred in the last academic year by the young lives child. 10 percent of the same is explained by the time allocation on studies on a typical day by the YL child apart from that in school. The gender of the child also explains a marginal part (4 percent) of the endowment effects.

The interaction effect stands out to be -0.22 (contributed almost two third by the years of schooling) and rest is unexplained in the results. The negative value of *I* has a different meaning at the point. The interpretation of *E* and *C* is premised upon isolation from one another. In order to get a negative value of *I*,$(\bar{X}_F - \bar{X}_R)'(\alpha - \beta)$, we must have to have at least one of the two parts to be negative. The first part, $(\bar{X}_F - \bar{X}_R)$ is positive as we see from table 4 where all the average features are better for *F* children. However, when we look at the detailed composition of *C* (in table 7) the individual returns are better for *R* children (for years of schooling in round 4). The better individual returns of these factors for *R* children results in the negative value of *I* (through negative$[\alpha - \beta]$). We must also notice that *C* captures the average returns to all the background features taken together, in isolation from *E*. This is what is captured by *I* and in our results these negative values at the point is bringing down the total contribution of *E* and *C*. In sum the total difference is explained by the contributions of these three effects together.[54]

Table 7: The Decomposition of Gaps for the Standardized Math Score

	Endowment Effect		Coefficient Effect		Interaction Effect	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Mothers Education	-0.003	0.022	-0.035	0.036	-0.024	0.030
Fathers Education	-0.008	0.016	0.052	0.054	0.019	0.021
Household Size	0.006	0.007	0.171	0.110	-0.008	0.008
Assets	-0.006	0.007	0.021	0.028	0.006	0.009
School Cost	0.067*	0.037	-0.038	0.028	-0.054	0.041
Expenditure	0.000	0.002	0.480	0.572	0.002	0.005
Years	0.159***	0.033	-1.083***	0.418	-0.122***	0.049
Time Allocation	0.045**	0.019	-0.085	0.057	-0.033	0.023
English Score	0.188***	0.041	-0.030	0.254	-0.006	0.055
Gender	0.022*	0.012	-0.033	0.041	-0.010	0.013
Age in Months	0.007	0.006	-1.720	2.257	-0.004	0.007
Body Mass Index	0.001	0.002	0.088	0.165	-0.002	0.005
Drinking Water	0.000	0.004	-0.001	0.387	0.000	0.008
Household Quality	0.007	0.013	0.108	0.144	0.013	0.018
Type of School	0.003	0.019	0.000	0.097	0.000	0.025
			3.520	2.397		
Total	**0.489*** **	**0.000**	**1.421*** **	**0.000**	**-0.224*** **	**0.001**

5.0.4 Conclusion and Policy Implications

Just like chapter 3.0 and 4.0, chapter 5.0 used the fourth round of Young Lives survey data for the older cohort. Approaching the children from the direction of their mathematics score, we characterized them into two groups. The first group of children included those children who scored 60 percent and above marks and calling them first division children. The rest of the children are those who have less than 60 percent score on mathematics test. The significant difference in learning between them was pointed out. We then moved on to their respective background characteristics where we saw that the better performing children have significantly better average characteristics compared to the rest of the children. We then put the learning difference and the differences in the background features together and tried to associate the latter with former. To do that, we have used the threefold Blinder-Oaxaca decomposition to break the difference in mean mathematics learning in order to attribute it to various factors.

[54] 6. For more details on the interpretation and explanation of the negative value of interaction effect see Biewen (2012).

The results tell us that the most important contributor to the mathematics learning difference between the *F* and *R* children is the differential returns to the background features. We learnt from our results that the *F* children have better average returns to background features compared to *R* children. The differential returns to background features which is called the coefficient effect explains 85 percent of the total standardized learning difference between them. Of the total coefficient effect here, 76 percent of this effect is explained by the gap in average years of schooling between the better performing and rest of the children. Similarly, the actual difference between them in terms of the average background features contributes to around 29 percent of the total standardized learning difference between them. There are several factors which contribute to the endowments effect, which is what it is called, where the major part is contributed by the difference in the average English scores between them. This is followed by the contribution of the years of schooling gap between these groups of children which, just like English, contributed to one-third of the total *E*. The schooling cost which captures the spending in the last academic year also contributes to the endowments effect between them. The time allocation on studies by these children apart from the hours they spend in schools has a significant contribution too. And a marginal part is also contributed by the gender of the child in favor of the male children.

The results in our study point lucidly in a policy direction: increasing the years of education (similar to Artega and Glewwe, 2014). The number of years a child has been in education has the single biggest positive impact on the mathematics learning outcomes (in both the endowments and coefficients effect). This is an issue which needs further exploration to find out why, the children who belong to the same age group have varied levels of education years. Is this because of a drop-out? As in Wadhwa (2009), there is an association between tuition cost and achievement outcomes, we find similar influence of school cost (which also includes the tuition cost) on the mathematics learning outcomes of children. Policies must be made to ensure that children from poorer background should not be left behind in availing better quality education simply because they cannot pay a higher cost. The time allocation on studies and the gender of the child show similar results as found in Kingdon (1996). The time allocation variable needs to be explored as well. In order to understand who are the children that spend greater time on studies apart from that in schools, this exploration is necessary. This would make it clearer as to what motivates child to spend time on education. One interesting finding of our study is a large contribution of the English score in the endowments effect. Those who are good in one are good in the other and former is influencing the score on the latter. The mathematics test questionnaire were asked in both English and Telegu (the local language) so the results suggest that the knowledge of English catered to the understanding of these questions asked at the time of survey. It should also be noted (as in Artega and Glewwe, 2014) that enrolment does not mean attendance, focus must also be paid on the latter and to the quality of the attendance. Unlike this study, where they have not considered years of schooling of the child at the time of survey, we have included this as well and have found its significant impact on the mathematics learning outcome of children.

5.1 Inside the Mathematics Learning Inequality: Analysis of Young Lives Survey Data, India

Abstract

This chapter studies mathematics learning gaps within Indian children at two points in time. Dividing them into two groups, better performing and rest, we investigate causes of the difference in the average learning between them at those two points. We explore this question using the threefold Blinder-Oaxaca decomposition at these survey points (collected over a gap of four years). We find that when the children were younger the private schooling effect was the core contributor towards this learning gap. When these children got older, the effect vanished and the gap in average years of schooling, which has magnified during this time between these groups of children, contributes most to this learning gap.

Keywords: Mathematics Learning; Blinder-Oaxaca Decomposition; Endowment Effect; Coefficient Effect; Schooling

5.1.1 Introduction

Continuing from the last chapter, in this chapter, we look into the mathematics learning inequality among Indian children at two points in their lives. These children are from the southern states of Andhra Pradesh and Telangana. The state was bifurcated into these two in 2014 and surveyed children are now spread across these two states. We investigate learning inequality among the children who perform better as against those children who have not. Using YLS data for India we attempt to study two groups of children (better performing and rest) at two points in their lives as an attempt to understand the reason for better performance of few as against the rest.

The World Bank in 2006 applied the Trends in Mathematics and Science Study (TIMSS-2003) on the secondary school going Indian children (from the states of Rajasthan and Orissa). Their Mathematics learning was found to be significantly poorer compared to the international average (Kingdon, 2007). The public expenditure on education in India as a percentage of GDP has been increasing over the years and marginally since 2010.[55] The Annual Status of Education Report (ASER) 2005 concludes that the increase in the financial resources towards education does not reflect in the levels of learning in primary education (ASER, 2005). There is a decline in the ability to do basic math nationally, which is visible across all classes (ASER, 2010). ASER (2017) further reports that in the age group of 14-18, a quarter of the surveyed children could not read basic text in their own language and they also struggled with doing basic division exercises.

In this study, using the novel dataset of Young Lives Survey (YLS henceforth), we look into the mathematics learning inequality among Indian children. These children are from the southern states of Andhra Pradesh and Telangana. The state was bifurcated into these two in 2014 and surveyed children are

[55] As per the Ministry of Human Resource Development's Educational Statistics at a glance, 2014 on Education and other departments.

now spread across these two states. We investigate the children who perform better as against those children who have not. We attempt to study two groups of children (better performing and rest) at two points in their lives as an attempt to understand the reason for better performance of few as against the rest.

Several previous studies (for instance Kingdon, 1996; Desai *et al.*, 2008; Goyal, 2009; Wadhwa, 2009; Chudgar and Quin, 2011; Wamalwa and Burns, 2012; Singh, 2015; Singh and Mukherjee, 2019) researched the connection between type of school and child achievement in the Indian context. Majority of these studies have found a private schooling effect on learning outcome. On the contrary, there are studies that have found that in different set ups this effect is not apparent. Goyal and Pandey (2009) have found a private schooling effect, however, their results suggest that despite this the quality of both public and private schools is low. Moreover, Muralidharan and Sundararaman (2015) in their study of the children from Andhra Pradesh point out that there is no public-private difference in learning in some aspects.

In this research, we have used YLS to explore the mathematics learning of Indian children. This YLS dataset has been used previously to research the learning aspect of children across all the four surveyed countries. The resources around which a child grows, has an association with the cognitive and psycho-social well-being of children. The material circumstances in general (in Dercon and Krishnan, 2009) and household features in particular (in Crookston *et. al.* 2014) have strong association with the outcome. This association is strongly positive with the child cognitive scores (Crookston *et. al.* 2014). Similarly, the influence of the parents' education and child's health is strong with learning outcomes (in Arteaga and Glewwe, 2014 and Glewwe *et al* 2015 for Peruvian and Vietnamese children respectively). Furthermore, the urban rural-learning gap among the Ethiopian children (in Sanfo and Ogawa, 2021) is found to have been explained in large part by the differences in the child and family level characteristics.

Singh and Mukherjee (2019) have found a connection between private pre-school attendance and its positive influence on mathematics learning among Indian children.

Past research has looked into these aspects among children from lower and middle income countries as well. Among the Sub-Saharan African children, Zhang (2006) finds that the inferior socio-economic conditions of rural children is a core reason for their poor performance compared to their urban counterparts. Similarly, in Zambia (in Burger, 2011) the urban-rural learning gap is shaped by the resource differences between them. Sakellariou (2008), just like Arteaga and Glewwe (2014), looked into the role of family level, child level, and school level characteristics with children's test scores (for Peruvian children). And Paxson and Schady (2007) also talks about the connection between household features and learning among children from Ecuador. In both of these studies the association is found to be positive. One improvement in this area of literature is by Barrera-Osorio *et al* (2011) where they study the children from Indonesia using the PISA dataset (2003 and 2006). They study the learning difference between the two points in the context of family, school, and student level inputs. The findings are again similar in the study of Chinese children's (in Li and Qiu, 2018) where the association of family's SES is strong, more so for the urban children, with the academic performance. Irrespective of the context, we learn from the previous literature that there is a clear connection, specifically between household characteristics and learning outcomes of children.

On similar line, the high income countries have been extensively researched on several aspects of child learning. Among these studies, children of USA (in Leibowitz, 1977; Crane, 1996; Currie and Yelowitz, 2000; Davis, 2005; Todd and Wolpin, 2007) have been frequent. Cobb *et al.* (2014) have studied the Australian children's learning outcome. German and Finnish children's learning outcome has also been studied (Ammermueller, 2007b). The learning outcomes of children from OECD countries is also done (Zhang and Lee, 2011). Holmlund *et al.* (2011) have looked into the Swedish children's learning outcomes.

The need to study the Indian children arise from growing reports from ASER suggesting a decline in the ability to do basic mathematics of Indian children. We have divided the YL Indian children into two groups. The first division scoring children (*F*; 60 percent and above marks) as better performers and remaining as rest(*R*; below 60 percent marks).[56] Previous studies have grouped children based on their background characteristics (parental education, gender, ethnicity etc.). There are researches that have chosen the 60 percent cutoff (as an independent/dummy variable) in the literature. As a measure of teachers' quality, Kingdon and Teal (2007) and Atherton and Kingdon (2010) have considered a cutoff (60 percent threshold) of teachers in their degrees. In both of these studies, the cutoff has been used as a dummy variable, just like any other dummy in the previous literature. Furthermore, Singh (2011b; 2012) and Asif *et al.* (2017) too have a detailed mention of this cutoff in their respective studies. Borrowing from these, division of children in our study was decided based upon their performance on mathematics. There are better methods available to study the outcome variable in this fashion. Quantile regression, from the wage literature (Machado and Mata, 2005; Melly, 2005; Rajas *et al.*, 2017) has been applied in the education/performance literature, to study the influence of background features on student achievement across different distribution of outcome/performance score (in Koenker and Hallock, 2001; Tian, 2006; Reeves and Lowe, 2014; Elizabeth and Schatschneider, 2014; Le and Nguyen, 2018). This research could not use the quantile regression despite the probable suitability because, as Reeves and Lowe (2009) point out, this method is more suitable for large sample sizes. As the sample size in our research is small, the robustness and generalizability of the results obtained using *F* and *R* categorization is expected to be

[56] The definition of first division children and rest of the children is same in both the rounds of this study. However, the children falling into these categories at both of these points are not necessarily the same. The total number of children are the same at both of these points.

affected. Despite this limitation, this research does provide some valuable insights in the dynamics of learning inequality. Lipovetsky (2012) discusses this method of classifying the dataset based on the outcome which entails a loss of information for all the groups created. We will see that when the children were younger the factors that explained the difference were largely different compared to when they are four years older. The core factors to understand learning differences include the private schooling effect, time allocation on studies and years of schooling. Our study is for a specific population at two specific points in time i.e. when they were 15-16 and 19-20 years old. To our knowledge, there is no study for the Indian children where this method was used to explore learning inequality among children over time. Such learning difference exploration has not been done previously, especially in the context of India. However, there are few studies (such as Barrera-Osorio *et al.*, 2011; Arteaga and Glewwe, 2014; Parvez and Laxminarayana, 2021) that have identified that the predictors of learning outcome change over time. Our research also finds similar results to these as we shall discuss later. Using the threefold Blinder-Oaxaca decomposition we break down the mean mathematics learning gap between them at two points. The Blinder-Oaxaca decomposition has been previously used (in Ammermueller, 2007b; Zhang and Lee, 2011; Barrera-Osorio *et al.*, 2011; Burger, 2011; Baird, 2012; Arteaga and Glewwe, 2014; Glewwe *et al.*, 2015; Sanfo and Ogawa, 2021) to study the learning outcome difference.

The rest of the chapter is organized as follows. Section 2 gives the description of data, source and variables used in our chapter along with descriptive statistics. Section 3 presents the methodological framework and empirical approach along with the results followed by summary and conclusion of this chapter in Section 4.

5.1.2 Data and Variables

This study uses two rounds of Young Lives Survey data for India (United Andhra Pradesh) which is a collaborative longitudinal research project coordinated by a team based at the University of Oxford.[57] The overall survey includes four developing countries Ethiopia, India, Peru, and Vietnam to trace the nature of changing poverty. The survey has so far done five rounds across these countries tracing the lives of 12000 children growing up in the context of poverty. The survey is based on a pro-poor sample (a pro-poor sample bias as mentioned in Keane *et al.* (2018)) which has nearly an equal number of boys and girls, based both in rural and urban communities. The 3000 children in the Indian survey were selected from 20 sentinel sites spread now across two southern Indian states of Andhra Pradesh and Telangana. The sites have been defined specifically for each country. The sampling method deployed in the data collection is a form of purposive sampling (Galab *et. al.* 2014).[58] This survey is repeated on the same set of children every four years. We have made use of the third and fourth round data for the older cohort on India which was released in 2010 and 2014 respectively. The older cohort which makes the Indian sample one-third of the total survey children, were born between Jan'94 and June'95. The remaining children, younger cohort, makes two-third of the Indian sample and they were born between Jan'01 and June'02. The third round covered 977 and fourth round covers 952 children from the older cohort. The attrition rate in the Indian survey has been kept really low (Singh, 2015). Our study had to drop out observations as information on various variables for many observations were missing. Majority of the observations were missing on the parental education. Moreover, we have considered only those children who are studying/studied in purely public or private schools and we had to drop children who have studied in different schools (54 observations). We are only studying those children who have participated and on

[57] The Indian state selected for this survey was Andhra Pradesh when the survey began in 2002. The state was bifurcated into two separate states in 2014, now the surveyed children belong to either one of the two states.

[58] Or 'cluster' in the language of sampling.

whom the information on the required variables are available at both the rounds and we ended up with 522 observations for the analysis. The Young Lives Survey provides information on various background characteristics of children and detailed record of necessary information needed for this research. The variable of interest here is mathematics test scores of children. In each round YL collected information about children's learning achievement with the help of language and mathematics tests (Galab *et al.* 2014a). The tests that were administered to YL children were different for each round, though there were few similarities. The detailed information on these tests are available in YLS questionnaires. In round 3, the test for the older cohort had two sections. The first section (which included 20 items) asked questions pertaining to addition, subtraction, multiplication, division, square roots (which included both whole numbers as well as fractions). The second section (10 items) included questions on mathematics problem solving which were developed using PISA and TIMSS. This section's questions were on data interpretation, number problem solving, measurement, and geometry. Similarly, the round 4 test had five sections which asked problems on similar aspects as round 3. The difficulty level was adjusted, as there was a gap of four years between the two rounds (Dawes, 2020). In total there were 30 questions in each round. The scores were obtained by adding the correct responses out of 30.

We must note here that there were similarities in the questions between the rounds (Dawes, 2020), they cannot be compared directly (Dawes, 2020; Rolleston, 2014). In order to make these tests comparable, especially mathematics, between the rounds, the reliability and validity has to be checked. There are several studies (Cueto and Leon, 2012; Azubuike *et al.* 2017) which have done that using the Classical Test Theory (CTT) and Item Response Theory (IRT). The application of these tests brought them to a uniform comparable scale, which allowed for a comparable study between the rounds. Our study does not directly do a comparison between the rounds hence we have not used this in our study.

The Table 1 gives description of the variables selected based on the literature reviewed used in our study. The Table 2 presents the overall descriptive statistics of all the variables described in Table 1 for each round of the survey, and by the sub-group of children by their performance. What is visible is that the absolute number of better performing children has doubled from round three to round four. The Table 3 provides information on the mathematics score for the children in both the rounds. For the ease of understanding at a point, the scores have been normalized by dividing with the standard deviation of the overall sample from respective rounds. We see that there is a reduction in the gap in learning from round three to round four. One reason this could have happened is because of the item composition in each round. The questions were different for each round which could also have affected their performance and in turn, may have caused the gap between the groups to come down. This test effect must be kept in mind while reading the results. Also, the broad areas that were covered in the questionnaire were same and children might have learnt in the four years gap. Furthermore, the idea of this research is not to look into the movement of this gap or do an inter-round comparison. Since the children were four years older in round 4 survey, we cannot, with certainty, say that the improvement in the average score (although incomparable with standardization as well) is an improvement. Glewwe *et al.* (2015) and Barrera-Osorio *et al.* (2021), for instance, points out that there is a connection between age of the child and their learning outcome. We have to be careful while interpreting the averages at these two points as we do expect a change in score simply because the children are older. We considered the age of child at the time of both of these surveys (and analysis) but since there is no significant variation in their age, we do not find the significant role of this variable at these points. In Table 4 the mean values of each variable for the whole dataset and by the sub-groups is summarized. We see that majority of the variables are different, significant, and better for *F* children at each round. Almost all the similar variables are significantly better for them at both the points.

Table 1: Description of the Variables

Variables		*Description*
Mathematics test Score		Raw test score on mathematics of YL child collected at the time of surveys out of 30 marks. This has also been standardized using the standard deviation of whole sample in each round for ease of comparison.
Parental Education (PE)[59]	Mothers Education	YL child's mothers' education in years
	Fathers Education	YL child's fathers' education in years
Household Features (H)	Household Size	Number of family members in YL child's household
	Assets (Round 4)	Value of the five most valuable assets owned, rented or borrowed in the YL child's household
	Assets (Round 3)	Value of assets owned, rented or borrowed in the YL child's household
	School Cost (Round 4)	Total expenditure incurred on school in the last academic year.[60]
	School Cost (Round 3)	How much has the YL household spent on school fees and extra tuition for the child per year
	Expenditure (Round 4)	The log of per capita monthly expenditure of YL child's household
	Expenditure (Round 3)	The log of real per capita monthly expenditure of YL child's household; base 2006 prices
Child Specific Features (C)	Years	Years of schooling received by YL child at the time of surveys
	Type of School	=1 if attended/attending public school last, 0 if private.
	Time Allocation	Time spent by YL child on studies apart from that in school on a typical day
	English Score	Raw test score on English of YL child collected at the time of survey out of 30 marks[61]
	Gender	=1 if male; 0 otherwise
	Age in Months	Age of the YL child in months at the time of surveys
	Body Mass Index	Body mass index of the YL child at the time of surveys
Others (O)	Drinking Water	Index constructed for whether or not the YL household has safe drinking water facility
	Household Quality	A simple average of the following:[62]

[59] The parents in our dataset have received school education (up to Grade 12), post-secondary/vocational education, adult literacy, and university education. Those who have received post-secondary education, they have been treated as having obtained 13 years of education. The university graduates have been treated as having received 15 years of education since university education lasts for 3 years after finishing Grade 12. There are no parents who have received education beyond university education. The parents who have received just the adult literacy, they have been taken as having received just 1 year of education.

[60] In the fourth round this cost is the sum total of the tuition fees, education charges, private tuition, accommodation, transportation, uniforms, stationary etc. in the last academic year. The school level heterogeneity is captured by the variance in the sum total of these costs.

[61] This information is available only in the fourth round survey.

[62] Crowding (scaled sleeping rooms per person), main materials of walls-dummy variable that takes the value 1 if main materials of the walls satisfied the basic norms of quality, main materials of roof, and main materials of floor. (Azubuike and Briones, 2016).

Table 2: Descriptive Statistics of the Variables

	Round 4						Round 3					
	Overall		First Division		Rest		Overall		First Division		Rest	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Math Score	14.479	7.074	22.289	2.482	10.368	4.908	11.312	6.156	21.044	2.546	9.258	4.504
Mothers Education	3.632	4.314	5.161	4.479	2.827	4.003	3.632	4.314	6.043	4.730	3.122	4.047
Fathers Education	5.644	5.046	6.872	5.302	4.997	4.788	5.643	5.045	7.736	5.305	5.201	4.882
Household Size	4.753	1.908	4.600	2.116	4.833	1.787	5.086	1.974	5.065	2.421	5.090	1.869
Assets	30018.640	44478.140	35562.420	48508.220	27100.860	41983.730	9291.452	52007.87	8172.802	33687.21	9527.64	55130.67
School Cost	19226.170	26380.520	31147.950	33431.980	12951.550	19028.050	2723.567	5344.192	5979.121	7369.401	2036.2	4530.036
Expenditure	7.170	0.588	7.193	0.570	7.158	0.598	6.776	.561	6.977	.455	6.734	.572
Years	13.448	1.745	14.411	1.056	12.942	1.823	8.641	1.308	8.978	1.282	8.571	1.304
Time Allocation	2.013	1.442	2.461	1.283	1.778	1.466	2.626	1.293	3.066	1.459	2.534	1.238
English Score	16.513	3.306	18.700	2.347	15.363	3.155	-	-	-	-	-	-
Gender	0.550	0.498	0.650	0.478	0.497	0.501	.550	.498	.703	.459	.515	.500
Age in Months	227.877	4.134	228.289	4.170	227.661	4.104	179.197	4.105	179.376	3.784	179.160	4.172
Body Mass Index	19.934	8.744	19.583	3.272	20.118	10.540	17.608	2.767	17.709	2.905	17.588	2.741
Drinking Water	0.981	0.137	0.994	0.075	0.974	0.160	.963	.187	.989	.105	.958	.200
Household Quality	0.725	0.217	0.782	0.182	0.695	0.228	.604	.298	.712	.213	.582	.309
Type of School	0.425	.494	0.255	0.437	0.515	0.500	.622	.485	.297	.459	.691	.462
Observations	**522**		**180**		**342**		**522**		**91**		**431**	

Table 3: The Mean Math Test score by the groups

Round 4

	Overall			First Division			Rest			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(F)-®	t-test
Math Score	522	2.047	1.00	180	3.152	.351	342	1.466	.694	1.686	30.58

Round 3

	Overall			First Division			Rest			Difference	
	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev.	Obs.	Mean	S.Dev	(F)-®	t-test
Math Score	522	1.839	1.00	91	3.422	.414	431	1.505	.732	1.916	24.148

*the test scores have been normalized by dividing it with the standard deviation of the overall sample for ease

Table 4: The Difference in the Means of explanatory variables of the two groups

Variable	Overall		First Division (F)		Rest of the children (R)		Difference	
	R 4	R3	R 4	R3	R 4	R3	R 4	R3
Math Score	14.479	11.312	22.289	21.042	10.368	9.257	11.920***	11.786***
	(7.074)	(6.156)	(2.482)	(2.547)	(4.908)	(4.504)		
Fathers Education	5.644	3.632	6.872	6.044	4.997	3.123	1.875***	2.921***
	(5.046)	4.31428	(5.302)	(4.730)	(4.788)	(4.047)		
Mothers Education	3.632	5.643	5.161	7.736	2.827	5.202	2.333***	2.534***
	(4.314)	(5.045)	(4.479)	(5.306)	(4.003)	(4.882)		
Household Size	4.753	5.086	4.600	5.065	4.833	5.090	-0.233	-.025
	(1.908)	(1.974)	(2.116)	(2.421)	(1.787)	(1.869)		
Assets[63]	30018.640	9291.452	35562.420	8172.802	27100.860	9527.64	8461.554***	-1354.838
	(44478.140)	(52007.87)	(48508.220)	(33687.21)	(41983.730)	(55130.67)		
School Cost	19226.170	2723.567	31147.950	5979.121	12951.550	2036.2	18196.4***	3942.921***
	(26380.520)	(5344.192)	(33431.980)	(7369.401)	(19028.050)	(4530.036)		
Expenditure	7.170	6.776	7.193	6.976	7.158	6.734	0.035	.242***
	(0.588)	(.561)	(0.570)	(.455)	(0.598)	(.573)		
Years	13.448	8.641	14.411	8.978	12.942	8.571	1.470***	.407***
	(1.745)	(1.308)	(1.056)	(1.282)	(1.823)	(1.304)		
Time Allocation	2.013	2.626	2.461	3.065	1.778	2.533	.683***	.532***
	(1.442)	(1.293)	(1.283)	(1.459)	(1.466)	(1.238)		
English Score	16.513	-	18.700	-	15.363	-	3.337***	-
	(3.306)	-	(2.347)	-	(3.155)	-		
Gender	0.549	.547	0.65	.703	0.49	.515	0.15***	.188***
	(0.497)	(.498)	(0.478)	(.459)	(0.500)	(.500)		
Age in Months	227.877	179.197	228.289	179.373	227.661	179.160	0.628	.213
	(4.134)	(4.105)	(4.170)	(3.785)	(4.104)	(4.173)		
Body Mass Index	19.934	17.608	19.583	17.708	20.118	17.587	-0.536	.121
	(8.744)	(2.767)	(3.272)	(2.906)	(0.540)	(2.741)		
Drinking Water	0.981	.963	0.994	.989	0.974	.958	0.021	.031*
	(0.137)	(.187)	(0.075)	(.105)	(0.160)	(.200)		
Household Quality	0.725	.604	0.782	.712	0.695	.582	.086744***	.131***
	(0.217)	(.298)	(0.182)	(.213)	(0.228)	(.309)		
Type Of School	0.425	.622	0.255	.297	0.514	.691	-0.259***	-.395***
	(0.494)	(.485)	(0.437)	(.459)	(0.500)	(.462)		
Observations	**522**		**180**	**91**	**342**	**431**		

Note: ***p<0.01, ** p<0.05, *p<.10
Standard Deviations are in parentheses

In Table 4 the mean values of each variable for the whole dataset and by the division they have scored on the mathematics test for each round separately is summarized. We see that majority of the variables are different, significant, and better for *F* children at each round. Almost all the similar variables are significantly better for them at both the points. The aim now is to see which of these background characteristics contribute to the difference in the average performance of these two groups of children at these two survey points. To what extent does these differences in the average background features

[63]We have to be careful while reading this as the definition of this variable is slightly different for each round due to data limitation.

contribute to the difference in the average learning outcome gap is what we attempt to answer in the remaining section of this paper.

5.1.3 Methodological Framework and Empirical Approach

A simple cognitive production function can be used to express the relationship between mathematics learning and a set of background characteristics. The mathematics learning (EA) is a function of input variables like parental education (PE), household factors (H), Child specific features (C), and other features (O) which are discussed in Table 1.

$$EA = f(PE, H, C, O) \qquad (1)$$

Yet even when an educational production function exists, there is no guarantee that one can estimate it (Artega & Glewwe, 2014). The reduced form equation for the decomposition in our analysis (following Blinder, 1973; Oaxaca, 1973; Jann, 2008) can be rearranged as shown in equation 2. In the analysis, for the application of Blinder-Oaxaca decomposition, we need a reference category from the two groups and we have taken *F* children as such. Moreover, the point of analysis is *R* children, meaning, the results are interpreted from their point.

$$D = \{E(X_F) - E(X_R)\}'\beta + E(X_R)'(\alpha - \beta) + \{E(X_F) - E(X_R)\}'(\alpha - \beta) \qquad (2)$$

This is called the threefold decomposition. The α and β are the coefficient vectors of *F* and *R* children, respectively, including the intercept.

$$D = E + C + I$$

The first component,

$$E = \{E(X_F) - E(X_R)\}'\beta$$

Amounts to the part of the difference that is due to the group difference in the predictors and is called the "endowment effects". The gap in the average background features applied to the impact (returns) of *R* children.

The second component,

$$C = E(X_R)'(\alpha - \beta)$$

Measures the contribution of the differences in the coefficients which includes also the intercept. It captures the differential returns to average features for each group. Gap in the average returns applied to the average features of *R* children.

And the third component,

$$I = \{E(X_F) - E(X_R)\}'(\alpha - \beta)$$

This is an interaction term which accounts for the fact that *E* and *R* effects between the two groups exist simultaneously. Our study estimates this decomposition and discusses about three effects in detail at both the survey points.

5.1.3.1 Empirical Results

On the standardized mathematics score, our results explain the entirety of the learning gap for each round (Table 5). Out of the total difference of 1.69 standard deviations (s.d.) in round 4, *E* explains 28 percent where the corresponding figure for round 3 stood at 16 percent (out of total 1.92 s.d.). If the *R* children had had similar average features as the *F* children ($\bar{X}_F$), their scores would have improved by 0.48 s.d. in round 4 (and 0.31 s.d. in round 3) given their returns stayed at the current level (β). A large part of the learning difference between these children is because of the fact that *F* children have had better average features as against *R* children. However, the *E's* influence in explaining the learning difference between the rounds has magnified proportionately. We see from Table 4 that several important features (such as assets, school cost, time allocation) are significant at both the points but these differences have magnified in favor of the *F* children between the rounds. This is the reason we find a proportionately larger contribution of *E* in round 4 than in round 3.

Similarly, *C's* contribution is 84 percent of the total difference in round 4 and the corresponding figure is 92 percent for round 3. If the *R* children's features stayed at the current level ($\bar{X}_R$), with the average returns of the *F* children (α) the improvement in their scores would have been 1.42 s.d. in round 4 (and 1.78 s.d. in round 3). We learn from here that, not only do *F* children have consistently better features, they have better and more efficient utilization of those resources compared to *R* children. This is true at both of the studied points. *F* children have been reaping better returns to already better average background features when compared to *R* children. One promising element, however, is the proportional decline in *C* between the rounds.

Table 5: Difference in Scores and the Blinder Oaxaca Decomposition

	Coefficient (on raw score)		**Coefficients (on Standardized Math Score[64])**		**p-value**	
	Round 4	**Round 3**	**Round 4**	**Round 3**	**Round 4**	**Round 3**
First Division	22.289	21.044	3.153	3.421	0.000	0.000
Rest of the children	10.368	9.258	1.467	1.505	0.000	0.000
Difference	11.920	11.786	1.686	1.916	0.000	0.000
Endowment Effect $\{E(X_F) - E(X_R)\}'\beta$	3.457	1.886	0.489	.306	0.000	0.000
Coefficient Effect $E(X_R)'(\alpha - \beta)$	10.052	10.913	1.421	1.775	0.000	0.000
Interaction effect $\{E(X_F) - E(X_R)\}'(\alpha - \beta)$	-1.590	-1.013	-0.224	-.165	0.001	0.055

[64] By dividing with the Standard deviation of math score for the overall sample.

Table 6: The Decomposition of Gaps for the Standardized Math Score

Round 4

	Endowment Effect		Coefficient Effect		Interaction Effect	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Mothers Education	-0.003	0.022	-0.035	0.036	-0.024	0.030
Fathers Education	-0.008	0.016	0.052	0.054	0.019	0.021
Household Size	0.006	0.007	0.171	0.110	-0.008	0.008
Assets	-0.006	0.007	0.021	0.028	0.006	0.009
School Cost	0.067*	0.037	-0.038	0.028	-0.054	0.041
Expenditure	0.000	0.002	0.480	0.572	0.002	0.005
Years	0.159***	0.033	-1.083***	0.418	-0.122***	0.049
Time Allocation	0.045**	0.019	-0.085	0.057	-0.033	0.023
English Score	0.188***	0.041	-0.030	0.254	-0.006	0.055
Gender	0.022*	0.012	-0.033	0.041	-0.010	0.013
Age in Months	0.007	0.006	-1.720	2.257	-0.004	0.007
Body Mass Index	0.001	0.002	0.088	0.165	-0.002	0.005
Drinking Water	0.000	0.004	-0.001	0.387	0.000	0.008
Household Quality	0.007	0.013	0.108	0.144	0.013	0.018
Type of School	0.003	0.019	0.000	0.097	0.000	0.025
			3.520	2.397		
Total	**0.489*****	**0.000**	**1.421*****	**0.000**	**-0.224*****	**0.001**

Round 3

	Endowment Effect		Coefficient Effect		Interaction Effect	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Mothers Education	.076**	0.033	-.089	0.056	-.083	0.055
Fathers Education	-.010	0.022	.109	0.082	.053	0.042
Household Size	.001	0.007	.091	0.137	-.0004	0.005
Assets	.0001	0.001	.007	0.015	-.001	0.004
School Cost	.006	0.038	.001	0.025	.002	0.048
Expenditure	.010	0.016	.710	0.863	.0255	0.032
Years	.017	0.012	-.008	0.378	-.0004	0.017
Time Allocation	.064**	0.025	-.363***	0.107	-.076**	0.033
Gender	.036**	0.017	.040	0.066	.0144	0.024
Age in Months	.002	0.004	-1.198	2.814	-.001	0.004
Body Mass Index	-.0001	0.002	.239	0.376	.002	0.005
Drinking Water	.002	0.005	-.418	0.435	-.013	0.015
Household Quality	-.012	0.015	.1767	0.142	.0397	0.033
Type of School	.115***	0.040	.219**	0.105	-.125**	0.063
constant			2.259	2.791		
Total	**.306*****	**0.000**	**1.775*****	**0.000**	**-.17****	**0.055**

Note: *** p<.01, ** p<.05, * p<.10

The contributors to *E* (in round 4 in order of importance) are the English score, years of schooling, schooling cost, time allocation and gender of the child (Table 6). For round 3 they were type of school, mothers' education, time allocation and gender. The common factor between round were the time allocation on studies (similar influence is found in Kingdon, 1996; Glewwe *et al.* 2015) and gender (marginal contribution). We learn that the influence of mothers' education and type of school, through *E*, existed only in round 3 and they have withered in round 4. The withering influence of mothers' education indicate that its impact does not operate at all stages of a child's learning (also found in Parvez and Laxminarayana, 2021). The type of school is an important factor in influencing learning outcome as we find in round three results (similar to Desai *et al.,* 2008; Goyal, 2009; Wadhwa, 2009; Chudgar and Quin, 2011; Wamalwa and Burns, 2012; Singh, 2015). Table 4 also tells that the gap in average *F* children attending private schools and *R* children attending the same has reduced between the rounds (although more proportion of each of them is attending private schools). This is the why type of school does not have a contribution to *E* in round 4 results (Chetty *et al.* 2014 points out that private schooling effect vanishes, though in a different context). Furthermore, as the children have gotten older, years of schooling is the single most (after English score) important contributor to *E*. When the children were younger a gap in average years of schooling existed but not as significantly as it does in round 4. Table 4 shows that the gap in average years of schooling between the groups has magnified too. Which means that *F* children have, on an average, stayed in schooling longer than the *R* children. This is precisely the reason we find the significant contribution to *E* (similar influence is also found in Glewwe *et al.* 2015; Parvez and Laxminarayana, 2021). Similarly, the schooling cost gap between these children has magnified (note that the composition of this cost is different for each round) between the rounds which now, in round 4 results, has a significant role to play. We learn that more spending capacity of *F* children and further, their ability to stay longer in education, compared to *R* children, during these four years explains their better scores. One interesting finding is that the large contribution of the English score in the endowments effect in the fourth round results. Those who are good in one are good in the other and former is influencing the gap in learning on latter. The mathematics test questionnaire were asked in both English and Telegu (the local language) so the results suggest that the knowledge of English catered to the understanding of these questions asked at the time of survey. We could not confirm this result for the third round because of lack of scores on English test in the third round.[6]

Similarly, the contributors to *C* in round 3 were time allocation and the type of school. The contribution of both of these variables has withered between the rounds. Years of schooling is now explaining a large part of this effect. We must note here that the children in this dataset were almost of the same age and an increase in this gap (tripled) is worrisome. This means that a dropout/discontinuation from education is happening and more so among *R* children. Sanjay *et al.,* (2014) points that dropout rates are high in this age group. ASER (2017) further reports that the probability of dropping out from education is more for the children who hail from the poorer and disadvantaged background. Also, in our dataset the years of schooling is obtained from the information on the highest grade completed by a child. Only after a grade is finished (even if there a repetition) that that child has been treated as having obtained one additional years of schooling. This could also be the case that, if there is a repetition of grade, it is more so for the *R* children. Lastly, a part of *C* is unexplained at both of these points.

The negative value of *I* has a different meaning at these points. The interpretation of *E* and *C* is premised upon isolation from one another. In order to get a negative value of I,$(\bar{X}_F - \bar{X}_R)'(\alpha - \beta)$, we must have to have at least one of the two parts to be negative. The first part, $(\bar{X}_F - \bar{X}_R)$ is positive as we see from table 4 where all the average features are better for *F* children. However, when we look at the detailed composition of *C* (in table 6) the individual returns are better for *R* children (for type of school and time allocation in round 3, and for years of schooling in round 4). The better individual returns of these factors for *R* children results in the negative value of *I* (through negative$[\alpha - \beta]$) at both of these points. We must also notice that *C* captures the average returns to all the background features taken together, in isolation from *E.* This is what is captured by *I* and in our results these negative values at these points is bringing down the total contribution of *E* and *C.* In sum the total difference is explained by the contributions of these three effects together.[65]

[65] For more details on the interpretation and explanation of the negative value of interaction effect see Biewen (2012).

5.1.4 Conclusion and Policy Implications

Using the Young Lives Survey (Older cohort), India, dataset for the third and fourth ~~two~~ rounds, we have explored the mathematics learning inequality among Indian children. Approaching these children directly from the outcome of interest (mathematics score), we have divided them into better performing and rest. We have used the Blinder-Oaxaca decomposition, a counterfactual decomposition methodology to study this learning gap. These children have been studied at two points where the gap is 4 years between the rounds. The *F* children have had significantly better average background characteristics at both of these points which consistently explains a large part of the learning differences between the groups. Moreover, the gap in average returns to those features, in favor of the *F* children, is also consistently explaining the learning inequalities. The Interaction of these two effects is also consistent, though proportionately small.

We learn from this research that the mathematics learning inequality between *F* and *R* children is in large part because of better endowments to *F* children. *F* children are found to be more efficient in putting these resources to use as compared to *R* children. This research finds that when *F* children were younger in round 3, the better endowments (which were significant) included mothers' education, time allocation, type of school. When they were older the corresponding features included school cost, years of schooling, time allocation, English score. These better features alongside better returns to them, at both of these points, helped shape the better mathematics performance of *F* children compared to *R* children (through *E* and *C*).

The research attempted to identify factors that shape the learning inequality among Indian children, in the context of their background resources. The investigative direction of this research is the exploration of the gap in average years of schooling (and time allocation on studies in this context) which has magnified during this time since all the children are in the same age group. The number of years a child has been in education has the single biggest positive impact on the mathematics learning outcomes (in both the endowments and coefficients effect) in the fourth round results. On the policy front, this research has some valuable insights to provide. Bridging the resource gap in the early lives of children should be a priority. However, we also learn from that there is an efficiency gap as well (although reducing), meaning, bridging this gap is not enough to deal with the issue. The gap in the average years of schooling (and time allocation on studies) which has magnified during this time needs to be further explored. The young adults, that we research here are discontinuing (given their age) from education or not continuing to higher education. Incentivizing staying in education without making the children compromise on the opportunity cost staying away from education entails, is one way to go about it.

References

Annual Status of Education Report. (2005). Pratham Resource Centre, New Delhi.

Annual Status of Education Report. (2010). Pratham Resource Centre, New Delhi.

Annual Status of Education Report. (2017). Pratham Resource Centre, New Delhi.

Ammermueller, A. (2011). PISA: What Makes the Difference? Explaining the Gap in Pisa Test Scores between Finland and Germany. *Empirical Economics*, *33*(2), 263–287.

Arteaga, I., & Glewwe, P. (2014). Achievement Gap between Indigenous and Non-Indigenous Children in Peru. *Young Lives Working Paper 130*.

Asif, R., Merceron, A., Abbas, S., & Ghani, N. (2017). Computers & Education Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194.

Atherton, P., & Kingdon, G. G. (2010). The relative effectiveness and costs of contract and regular teachers in India by. *Institute of Education, University of London*, 256-276.

Azubuike, Obiageri Bridget, and K. B. (2016). Young Lives' Constructed files. *Oxford Review of Economic Policy, 23(2)* (7483), 168–195.

Azubuike, O. B., Moore, R., & Iyer, P. (2017). Young Lives School Surveys, 2016–17: The design and development of cross-country Maths and English tests in Ethiopia, India & Vietnam. *Young Lives*.

Baird, K. (2012). Class in the classroom: The relationship between school resources and math performance among low socioeconomic status students in 19 rich countries. *Education Economics*, *20*(5), 484–509.

Barrera-Osorio, F., García-Moreno, V., Patrinos, H. A., & Porta, E. (2011). Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: An application to Indonesia's results in PISA mathematics. *World Bank Policy Research Working Paper*, (5584).

Biewen, M. (2012). Additive Decompositions with Interaction Effects. *IZA Discussion Paper*, (6730).

Blinder, A. S., (1973). Wage Discrimination : Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Burger, R. (2011). School effectiveness in Zambia: The origins of differences between rural and urban outcomes. *Development Southern Africa*, *28*(2), 157–176.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *National Bureau of Economic Research Working Paper*, (17699).

Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review*, *31*(4), 376–390.

Cobb-Clark, D. A., & Gørgens, T. (2014). Parents' economic support of young-adult children: Do socioeconomic circumstances matter? *Journal of Population Economics*, *27*(2), 447–471.

Crane, J. (1996). Effects of home environment, SES, and maternal test scores on mathematics achievement. *Journal of Educational Research*, *89*(5), 305–314.

Crookston, B. T., Forste, R., McClellan, C., Georgiadis, A., & Heaton, T. B. (2014). Factors associated with cognitive achievement in late childhood and adolescence: The Young Lives cohort study of children in Ethiopia, India, Peru, and Vietnam. *BMC Pediatrics*, *14*(1), 1–9.

Cueto, S., and Leon, S. (2012). Psychometric Characteristics of cognitive development and achievement instruments in round 3 of Young Lives (Young Lives Technical no Notes) Young Lives.

Currie, J., & Yelowitz, A. (2000). Are public housing projects good for kids ? *Journal of Public Economics*, *75*(1), 99–124.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304.

Dawes, A. (2020). Measuring the Development of Cognitive Skills Across Time and Context: Reflections from Young Lives. *Young Lives*

Dercon, S., & Krishnan, P. (2009). Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries. *Children, Youth and Environments*, *19*(2), 138–163.

Desai, S., Dubey, A., Vanneman, R., & Banerji, R. (2008). Private Schooling in India: A New Educational Landscape. *India Policy Forum*, *5*(1), 1–58.

Galab, S., Prudhvikar, R., & Singh, R. (2014) Young lives survey design and sampling in Andhra Pradesh, India

Galab, S., P. Prudhvikar Reddy and Renu Singh (2014a). Education and Learning: Round 4 Preliminary Findings, *Round 4 Fact Sheet, Oxford*: Young Lives

Glewwe, P., Chen, Q., & Katare, B. (2015). What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data. *Asia and the Pacific Policy Studies*, *2*(3), 494–516.

Goyal, S. (2009). Inside the house of learning: The relative performance of public and private schools in Orissa. *Education Economics*, *17*(3), 315–327.

Goyal, S., & Pandey, P. (2009). How do government and private schools differ? Findings from two large Indian states. *International Bank for Reconstruction and Development/the World Bank*

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

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*. 8(4), 453-479.

Keane, M. P., Krutikova, S., & Neal, T. (2018). The impact of child work on cognitive development: results from four Low to Middle Income countries. (No. W18/29) *IFS Working Papers*.

Kingdon, G. G. (1996). The Quality and Efficiency of Public and Private Schooling: A Case Study of India. *Oxford Bulletin of Economics and Statistics*, *58*(1), 57–82.

Kingdon, G. G. (2007). The progress of school education in India. *Oxford Review of Economic Policy*, 23(2), 168-19 5.

Kingdon, G. G., & Teal, F. (2007). Does performance related pay for teachers improve student performance ? Some evidence from India. *Economics of Education Review*, *26*, 473–486.

Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, *15*(4), 143–156.

Le, H. T., & Nguyen, H. T. (2018). The evolution of the gender test score gap through seventh grade : new insights from Australia using unconditional quantile regression and decomposition. *IZA Journal of Labor Economics*, *7*(2).

Leibowitz, A. (1977). Parental Inputs and Children's Achievement. *The Journal of Human Resources*, *12*(2), 242–251.

Li, Z., & Qiu, Z. (2018). How does family background affect children's educational achievement? Evidence from Contemporary China. *The Journal of Chinese Sociology*, *5*(1), 1–21.

Lipovetsky, S. (2012). Regression split by levels of the dependent variable. *Journal of Modern Applied Statistical Methods*, *11*(2), 319–324.

Machado, J. A. F., & Mata, J. (2005). Counterfactual Decomposition Changes in Wage Distributions Using Quantile Regression. *Journal of Applied Econometrics*, *20*, 445–465.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, *12*, 577–590.

Montes-Rojas, G., Siga, L., & Mainali, R. (2017). Mean and quantile regression Oaxaca-Blinder decompositions with an application to caste discrimination. *Journal of Economic Inequality*, *15*, 245–255.

Muralidharan, K., & Sundararaman, V. (2015). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics*, *130*(3), 1011–1066.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*, 693–709.

Parvez, A., & Laxminarayana, K. (2021). Mathematics learning inequality among children of private and public schools. *Asia Pacific Education Review*.

Paxson, C., & Schady, N. (2007). Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. *Journal of Human Resources*, *42*(1), 49–84.

Reeves, E. B., & Lowe, J. (2009). Quantile Regression : An Education Policy Research Tool. *Journal of Rural Social Sciences*, *24*(1).

Rolleston, C. (2014). Learning profiles and the "skills gap" in four developing countries: A comparative analysis of schooling and skills development. *Oxford Review of Education*, *40*(1), 132–150.

Sakellariou, C. (2008). Peer effects and the indigenous/non-indigenous early test-score gap in Peru. *Education Economics*, *16*(4), 371–390.

Sanfo, J. B. M. B., & Ogawa, K. (2021). Explaining the rural-urban learning achievements gap in Ethiopian primary education: A re-centered influence function decomposition using Young Lives data. *Education Economics*, 1–29.

Sanjay, S., Gupta, R., Gautam, V., & Meena, S. H. R. (2014). Educational Statistics at a Glance. *MHRD, Bureau of Planning, Monitoring & Statistics, New Delhi*.

Singh, A. (2015). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics*, *113*(2015), 16–32.

Singh, A. (2011). Family background, academic ability and associated inequality of opportunity in India. *Economics Bulletin*, *31*(2), 1463–1473.

Singh, A. (2012). Reducing Gender Based Inequalities In Employment And Wages: Is India On Right Track? *Indira Gandhi Institute of Development Research*, (June).

Singh, R., & Mukherjee, P. (2019). Effect of preschool education on cognitive achievement and subjective wellbeing at age 12: evidence from India. *Compare: A Journal of Comparative and International Education*, *49*(5), 723–741.

Tian, M. (2006). A Quantile Regression Analysis of Family Background Factor Effects on Mathematical Achievement. *Journal of Data Science*, *4*, 461–478.

Tighe, E. L., & Schatschneider, C. (2014). A Quantile Regression Approach to Understanding the Relations among Morphological Awareness, Vocabulary, and Reading Comprehension in Adult Basic Education Students. *Journal of Learning Disabilities*.1-13

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.

Wadhwa, W. (2009). Are private schools really performing better than government schools? *Annual Status of Education Report (Rural) New Delhi*.

Wamalwa, F. M., & Burns, J. (2018). Private schools and student learning achievements in Kenya. *Economics of Education Review*, *66*(July), 114–124.

Zhang, L., & Lee, K. A. (2011). Decomposing achievement gaps among OECD countries. *Asia Pacific Education Review*, *12*(3), 463–474.

Zhang, Y. (2006). Urban - Rural Literacy Gaps in Sub - Saharan Africa : The Roles of Socioeconomic Status and School Quality. *Comparative Education Review*, *50*(4), 581–602.

6. Summary, Conclusions, and Policy Implications

In this thesis we explored the mathematics learning inequality among a set of Indian children who hail from poorer socio-economic background. In the literature on learning, linkages of child learning have been established with a lot of background characteristics. The socio-economic aspects, community aspects, school level aspects have all been researched in depth to see how the learning of children from various background is being shaped. There are researches in the context of India as well where these aspects are looked into too. The gap that our thesis fills pertains to the literature on the learning inequality among Indian children in general, and from poorer background in particular. Keeping that in mind, the thesis has explored the mathematics learning inequality among the Indian children based on three benchmarks. The foremost benchmark that we studied in this thesis is the schooling status of the parents (both separately) with respect to a bunch of background features of these children. We studied the learning inequality on mathematics between the children of schooled and non-schooled parents. Then we looked into the mathematics learning inequality between the children of public schools and private schools. Again we explore this with respect to a set of background features of these children. Lastly, departing from the conventional approaches to exploring learning inequality in children based on background features, we approached the children directly from the side of their mathematics learning scores. We divided them into groups where who scored 60 percent and above were categorized as better performing and the remaining as rest. We then looked into their background features to understand the reasons behind better performance of few as against others with respect to a set of background features.

The introductory chapter of the thesis problematizes the research that we have done in this thesis. The chapter detailed the reviews on learning of Indian children from international agencies such as the World Bank. The World Bank study puts the Indian children in the limelight in terms of their learning on the international stage. The reports point out how the Indian children perform significantly poorly. The Trends in Mathematics and Science Study was used in this study where 46 countries had participated and it was found that the score of Indian children was significantly below the international average. Similarly, the National Achievement Survey (NAS) and Annual Statistics of Education Report (ASER) both point out how different states in India perform at different levels of education. This chapter pointed out how the scores of Andhra Pradesh is different on mathematics as compared to the national average. It was observed from these reports that the children from this state have performed better on mathematics compared to the national average. It is here that we decided to explore the mathematics learning of children from united Andhra Pradesh. We used the rich dataset which was collected by the Department of Foreign and International Development (DFID), University of Oxford. The longitudinal survey, Young Lives, on the Indian children from the states of Andhra Pradesh and Telangana was found to be rich and useful enough to be researched in depth on these aspects. The chapter after the problematization of the research gives a detailed account of previous research on educational outcome and socio-economic background in general, and socio-economic background and learning outcome in particular. We have started by pointing out how previous literature has discussed the background characteristics in relation to the educational outcomes such as continuation, drop-out etc. in several contexts. We have also pointed out how these literature have discussed specific aspects of these characteristics such as parental education, parental involvement, type of schooling, social background, social/economic class etc. in relation to the educational outcome of children. After this we moved on to pointing those studies that have used the YL dataset to study the learning outcomes of children from the four surveyed YL countries, Ethiopia, India, Peru, and Vietnam. The studies have looked into several aspects of learning among these countries using the dataset of YL. We also charted out the literature where the learning outcome of children from the lower income countries has been looked into. Countries like Indonesia, Philippines, Mexico, and Ecuador etc. have all been a matter of discussion in the previous literature on learning outcomes of children. We then discussed the children who hail from high income countries where their learning outcomes have been researched in relation to the socio-economic background specific to their contexts. These studies have talked about children from United States, Australia, Sweden, Germany, Finland, and OECD countries etc. The purpose of reviewing literature on different socio-economic contexts was to identify the pattern and dissimilarities alike among these studies. As we have categorized our children into two groups based on each benchmark characteristics and used the threefold Blinder-Oaxaca decomposition to study their learning gaps, we have pointed out several literature where this method has been deployed. After this, we charted out the broad objectives of this thesis which we attempted to explore in the following chapters.

In the next chapter we described the dataset that we have used for the purpose of our research. The source of the data along with the sampling methodology was described. We also detailed the limitations of the dataset and also the limits of inference that is to be kept in mind while reading the results from the upcoming chapters. This chapter also details the variables that we have used in the thesis based on the literature and the context of India. We give the descriptive statistics of the dataset as well as the preliminary observations from the statistics. Moving on, we give a detailed information on the decomposition methodology that we have used in all our main chapters where we derived the methodology and explained how the results obtained from this are to be read.

Like we said earlier that we have explored the mathematics learning inequality among the children of united Andhra Pradesh based on three benchmarks, the next three chapters (divided into two parts each) explored those. In chapter 3.0 and 3.1 the inequality in mathematics learning between the children of schooled and non-schooled parents (individually for each parent) was explored. In chapter 3.0, we began by pointing out how the children of schooled parents have significantly better learning compared to the children of non-schooled parents. We also point out here that they have better average background characteristics compared to the children of non-schooled parents. The chapter then places the mathematics learning inequality in the context of their background differences. The Blinder-Oaxaca decomposition was used to break the difference in mean learning between the two groups into various components that explain this. The chapter does this for a specific point of time using the fourth round of YLS for the older cohort of the Indian children. Similarly, in chapter 3.1, we repeat the same exercise adding third round of YLS in the analysis. Successively the results that we obtain from both of the rounds were placed together to report the observations and findings. It must be pointed out here that the gap between the two points of dataset that we have used is four years. However, the children that have been studied in the thesis are same. We report our results from the two point analysis of the mathematics learning difference between the children of schooled and non-schooled parents in chapter 3.1.

In the next chapter, divided again in two parts, 4.0 and 4.1, we explore the mathematics learning inequality between the children of private and public schools, the second benchmark for the thesis. In chapter 4.0 we begin by pointing out how the children of private schools have better mathematics learning score compared to the children of public schools. The chapter also points out how the children of private schools have better average background features as well compared to the children of public schools. Using again the Blinder-Oaxaca decomposition we study the difference in mean mathematics learning between them in the context of their respective backgrounds. We use the round four YLS data for the older cohort of Indian children in this chapter. Similarly, in chapter 4.1 we continue from chapter 4.0 where we use the third round of YLS as well where the gap between the two rounds is four years, however, the children in both the rounds that we study are same. Following the same steps we obtain the back ground differences between them alongside their mathematics learning difference and decompose that difference in the context of former. The chapter reports the results obtained from the two point study of mathematics learning difference between the children of private and public schools. The chapter also reports the crucial observations and findings from the empirical results.

Departing from the previous two benchmarks, the next chapter uses an unconventional approach to categorize the children that we study. In the previous two chapters the children were categorized on the basis of their background features (parental schooling status and type of school attended). However, in the third benchmark the children were categorized by approaching them directly on the basis of marks that they have obtained on mathematics. The children who scored sixty percent and above were called better performing and remaining as rest. In chapter 5.0 and 5.1 we explore these groups of children. In chapter 5.0, using the fourth round of YLS we see that the better performing children have better average background features compared to rest of the children. The chapter then uses the Blinder-Oaxaca decomposition methodology associate the differences in background to the differences in mathematics score obtained. Similarly, in chapter 5.1, we repeat the same effort using the third round of YLS for the same set of children and report our findings. The chapter talks about the pattern of influence and crucial findings from the two point study of a set of children who were either better performing and who could not perform better.

The thesis charts out some major findings that identify some core factors which need policy level as well as research level attention. The summary of results obtained from the thesis on the basis of each of the benchmarks (at both of the studied points) have been summarized in Table *I.* The thesis researches eight sub-groups of children at two points which are four years apart. These groups are created on the basis of three benchmarks. The first four groups were described on the basis of the schooling status of the YL child's parents (separately for mothers' and fathers' schooling status). The second classification was done on the basis of the type of school the YL child is attending/last attended. Lastly, on the basis of their performance itself on mathematics. The first part of table *I* details the three effects of the threefold

Blinder-Oaxaca decomposition for all of these groups at both of the points. In total there are eight instances where the three effects have been obtained in the empirical exercise and they are charted out in this table. The detailed description and derivation of these effects is available in chapter 2 and subsequently in chapter 3, 4, and 5.

Table I: Summary of Thesis Results

Three Effects	**Round 3 (2010)-15/16 years old**				**Round 4 (2014)-19/20 years old**			
	Parental Schooling Status		*Type of School*	*Division (F/R)*	*Parental Schooling Status*		*Type of School*	*Division (F/R)*
	Mother	*Father*			*Mother*	*Father*		
$E = \{E(X_A) - E(X_B)\}'\beta$	N	Y	Y	Y	Y	Y	Y	Y
$C = E(X_B)'(\alpha - \beta)$	Y	N	Y	Y	Y	N	N	Y
$I = \{E(X_+) - E(X_0)\}'(\alpha - \beta)$	Y	N	N	Y	N	N	N	Y
Variables	**R3**				**R4**			
Mothers Education			Y	Y				
Fathers Education								
Household Size	Y							
Assets								
School Cost		Y	Y		Y	Y	Y	Y
Expenditure	YY							
Years	Y				Y	Y	Y	YYY
Time Allocation			Y Y	YYY			Y	Y
English Score					Y	Y	Y	Y
Gender		Y		Y	Y	Y		Y
Age in Months								
Body Mass Index								
Drinking Water	YY							
Household Quality	Y		Y		Y	Y	Y	
Type of School				YYY				

The Table *I* summarizes the results of this thesis. The three benchmarks that were taken up to explore the mathematics learning inequalities, were done using the threefold Blinder-Oaxaca decomposition. The results to this decomposition gave three effects, namely, Endowments, Coefficient, and Interaction. The Table has color coded these effects, green, yellow, and blue respectively. There were eight groups that were looked into, four at each point. In each of these groups, the thesis has estimated these three effects. To reiterate, the groups were; children of schooled and non-schooled parents (separately for mothers' and fathers' schooling status), children of public and private schools, and better performing and rest of the children. The thesis looks into the same set of children at two points where the gap between those points is four years. At the first point the children were 15/16 years of age, whereas on the second point they were 19/20 years old.

As can be seen from the Table that the most frequent significant color is green. Out of eight instances/groups that we looked into, this effect came out to be most significant/consistent explainer to the mathematics learning difference. The only time/group in which this did not explain the mathematics learning difference is between the children of schooled and non-schooled mothers at the first point. The endowments effect talks about the differences in the average background features between the two groups as the contributor to the learning difference. The second most significant/consistent contributor to the learning difference between the groups of children is the coefficient effect which is significant five out of eight instances/groups. This effect explains the learning difference in the context of impact/returns to the background features between the groups. It could also be interpreted as the efficiency with which the resources at disposal are put to use. And interaction effect is the least consistent, three out of eight groups. This effect captures the effect of both the previously discussed effects simultaneously. The results also identify some of the consistent variables which are core contributors to these three effects, among several groups, and at both points. These variables have been shown in the lower part of the table.

The first and the foremost finding at one point (round four results) is that the major part of the mathematics learning difference between the children of schooled and non-schooled mothers is because of the fact that the children of schooled mothers have better resources available to them. On the contrary, the children of non-schooled mothers are significantly poorer endowed in terms of the same resources. Moreover, the children of schooled mothers have better returns to those resources as compared to the children of non-schooled mothers. However, the differences in mathematics learning between the children of schooled and non-schooled fathers is entirely because of the resource differences between them. One crucial result that we have found is that if the children of non-schooled fathers had had similar average features as the children of schooled fathers, they would have performed better than the children of schooled fathers. The core factors that explain the learning difference at this point includes the schooling cost (spent in the last academic year) difference between the children of schooled and non-schooled parents. The results also point out that the children of schooled parents have stayed in education for longer number of years compared to the children of non-schooled parents. This difference in the average years of schooling received by these children, given they fall in the same age group, is worrying, more so, because it is an important contributor to the learning difference between them. Another core contributor here is the English language score obtained by these children at the time of fourth round survey. We have found a clear connection between better performance on English by the children of schooled parents and not so better performance of the children of non-schooled parents. The role of English language needs to be further explored.

While looking at the results of round four and round three together between the children of schooled and non-schooled parents, we could identify some consistent factors that have a role in determining the learning gap at both the survey points. Moreover, the results also identify the factors which have a role to play when they were younger but the effect of which has vanished in round four results and vice-versa. Between the children of schooled and non-schooled mothers in round three results we see that the differences in returns to the resources was the core contributor towards the learning difference between them. Unlike the results of round four, we see that when the children have gotten older it is precisely the differences between the resource endowments that has an impact on the learning difference between the children of schooled and non-schooled mothers. However, the differences in the resources is consistently determining the learning difference between the children of schooled and non-schooled fathers at both the survey points. Looking at the core factors that consistently contribute to the mathematics learning difference between the children of schooled and non-schooled parents, the results identify few variables that have more important role to play at both the ages in the lives of these children. The difference in the average schooling cost incurred by the child in the last academic year between the children of schooled and non-schooled parents is one core consistent variable that explain the learning difference between them among both groups and at both the points. Another core factor that our results identify is the sudden significant role played by the years of schooling received by the child at the time of fourth round survey. When the children were younger this variable was not the core determinant of the mathematics learning gap. However, when the children got older we find that this is one of the most important contributor to the learning gap. This result between the children of schooled and non-schooled parents is worrying because these children belong to the same age group and increase in the years of schooling for a group, and for another not that much, points out towards dropping out.

Between the children of public and private schools we find that the core determinant of the mathematics learning difference is the resource endowments between them. We find from our results that if the children of public school were to have similar average features as the children of private school, they would have performed better than the children of private schools. The resource difference between these children is not only stopping them from performing better but also to perform better than the children of private schools. In the further analysis, the results identify some important variables that work behind the type of school a child is in in determining the mathematics learning gap. The most important among those variables is time allocation on studies by the child and the schooling cost incurred in the last academic year. We learn from descriptive statistics that the children who have studied in the private schools have spent a greater amount on schooling in the last academic year compared to the children of public schools. This result is worrying as this implies that the children who are able to spend more are able to perform better which should not be the case. Also the children from private schools spend greater amount of time on studies apart from that in school compared to the children of public schools. This needs further exploration as there seems to be a role of certain other activities (financial?) involved in this for public school children.

In the last part of the thesis we have departed from the conventional approach of dividing the children into groups on the basis of their background characteristics. We find that the consistent explanation to the gap in average learning between them is provided by these gaps in average features. Moreover, the gap in average returns to those features, in favor of the *F* children, is also consistently explaining the learning inequalities. The Interaction of these two effects is also consistent, though proportionately small. Our results also identify the factors in terms of their proportional contribution to these effects at both of these points. The consistent contributor to *E* are the time allocation to studies (similar to Kingdon, 1996 and Glewwe *et al.* 2015) and gender (marginal and declining contribution). The contributors to *E* which have withered from round 3 to round 4 includes mothers' education and type of school. However, the private schooling effect is significant when the children were younger (Desai *et al.*, 2008; Goyal, 2009; Wadhwa, 2009; Chudgar and Quin, 2011; Wamalwa and Burns, 2012; Singh, 2015), similar to other findings in India. The contributors to *C* were similar to *E* for the respective rounds. In round 3, private schooling effect and time allocation, just like the contributors to *E* here, explain the learning gap. However, these effects have vanished for *C* in round 4 where the sole contributor is years of schooling. The factors to keep in mind is that these two variables, mothers' education level and type of school were significant contributor to both *E* and *C*. However, as the children have grown four years older, these effects no longer influence the learning outcome. This disappearing effect of these variables points out that some of the variables have an influence only at a particular age of child. Same is true for the contributors to *I*. One interesting finding is that the large contribution of the English score in the endowments effect in the fourth round results. Those who are good in one are good in the other and former is influencing the gap in learning on latter. The mathematics test questionnaire were asked in both English and Telegu (the local language) so the results suggest that the knowledge of English catered to the understanding of these questions asked at the time of survey. We could not confirm this result for the third round because of lack of scores on English test in the third round. When the children were younger the type of schooling in favor of the private schools contributed to a large part of the learning outcome difference (both endowment and coefficient effects). The older age of the child in round four is significant in understanding our findings, especially the now significant contribution of years of schooling variable. When the children were younger a gap in average years of schooling existed but not as significantly as it does in round 4. The average years of schooling has marginally increased for the whole sample of children but at the same time the gap has almost tripled during this time. This is because *F* children, on an average, have continued with schooling longer than *R* children. This increasing gap pulls down the average years of schooling for *R* children. The core reason behind this occurrence is the discontinuation by many of these children after round 3. Sanjay *et al.*, (2014) points that dropout rates are high in this age group.

Table II: Number of Schools by School Management and School Category (2014-15)

School Management and School Category	Telangana			Andhra Pradesh			
	Hyderabad	Karimn--agar	Mahboob-nagar	Anantapur	YSR Kadapa	West Godavari	Srikakulam
Department of Education	529	105	111	32	19	3	42
Tribal Welfare Department	1	28	61	7	11	129	247
Government Aided	168	25	36	30	104	236	23
Private Unaided (Recognized)	1248	984	824	761	602	713	364
Central Government	6	3	1	6	1	2	3
Unrecognized	20	5	12	10	3	12	18
Madarsa Unrecognized	152	22	38	11	40		
Local Body		2243	3040	3333	2971	2570	2782
Other Government Managed		17					

Telangana

School Management and School Category	Adilabad	Khammam	Medak	Nalgonda	Nizamabad	Rangareddy	Warangal
Department of Education	121	52	84	128	131	51	160
Tribal Welfare Department	940	344	22	28	15	17	230
Government Aided	19	41	6	59	29	31	75
Private Unaided (Recognized)	561	401	536	768	579	1725	848
Central Government	1	2	2	1	1	6	3
Unrecognized	29	15	18	27	3	30	19
Madarsa Unrecognized		2	29	21	48	25	32
Local Body	2346	2086	2332	2476	1641	1814	2279
Other Government Managed	7	26			22	2	

Andhra Pradesh

School Management and School Category	Chittor	East Godavari	Guntur	Krishna	Kurnool	Nellore	Prakasam	Vishakapatnam	Vizianagram
Department of Education	38	7	35	5	88	23	20	140	89
Tribal Welfare Department	45	427	28	13	33	33	62	692	372
Government Aided	35	108	269	480	130	99	195	60	69
Private Unaided (Recognized)	736	1086	500	738	796	528	458	734	371
Central Government	3	2	3	2	2	3	3	9	3
Unrecognized	23	73	174	46	2	7	68	11	
Madarsa Unrecognized	6	1	18	6	61	30	16	10	
Local Body	4392	3343	2921	2308	2227	3009	2957	2816	2065
Other Government Managed			16	1	4	4			

Source: Unified District Information System for Education (*UDISE*)

Note: The number of schools specified includes schools that offer various levels/Grade such as (I-V), (I-VIII), (I-XII), (VI-VIII), and (VIII-XII)

There are several policy implications that come out of this research. The foremost implication of the research pertains to the schooling cost variable. As our findings suggest that the most important and core contributor to the mathematics learning difference among all the groups of children at both of the surveyed points is the schooling cost incurred in the last academic year. From *Table II* we see that in all of the surveyed regions by the Young Lives dataset, the major education provider to these children is the private schools (apart from the local body schools). It must be pointed out here that the private schools are more expensive alongside providing better quality education. We see that a better spending capacity entails better learning outcomes. Policies must be pursued to level this gap in a way to improve the quality of education which should be reflected in the learning by these children. It should not be the case that the children who could spend better are the only one to perform better. The ability to spend should not be the determining factor of learning. The findings suggest that the children of schooled parents (and private school children and better performing children) could perform better mainly because they have spent an average greater amount on schooling in the last academic year at both of these points. The children of non-schooled parents (and public schools and rest of the children) could not spend that much is one reason they could not perform better than their respective counterparts. Policies must be pursued to fill this gap in spending or at least ensuring that the children receive similar quality of schooling irrespective of whether or not they are able to spend a lucrative amount.

The second important implication pertains to the years of schooling gap. As these children belong to the same age group, sudden increase in this gap in just four years is worrying. An exploration of this to figure out if there is financial opportunity cost attached to staying in education. If so, financial incentives to be provided to children in order to make them stay in education. The investigative direction in which our study points out is the further exploration of years of schooling gap in order to provide an explanation for widening years of schooling gap given that these children fall in the same age group. Moreover, a more nuanced explanation is needed for the English language score and its relationship with the mathematics score. This needs to be studied in the context of our results.

The results in our chapters on the basis of type of schooling point towards both policy as well as investigative direction. The foremost policy implication comes from the consistent contribution of the schooling cost to the learning difference between the two groups of children. Attention needs to be paid to this as higher paying ability to schooling (towards private schools) entails better schooling facilities. This has to be dealt with in terms of providing better facilities in public schools (the cost there is already lesser) which should add to the environment of learning. Policies are needed to address this dilapidated state of public education system in terms of making and creating a better atmosphere of learning for these children. Secondly, the sudden significance of years of schooling gap in explaining the learning gap is worrying because the surveyed children are almost of the similar age. The increase in the years of schooling gap within four years warrants policy intervention in terms of incentivizing (reiterating what we pointed out earlier with respect to providing constant monetary benefit to school going children) staying in education. Lastly, the consistent significance of time allocation on studies needs exploratory attention as we see that the public school children spend a lesser average amount of time on studies. Moreover, this gap has magnified within four years which raises question pertaining to their time allocation pattern. Is it that they spend their time various other activities (economic?) because there is an opportunity cost attached to staying in education. The time allocation on studies and the years of schooling has to be re-looked as they both seem to be connected.

The results in the last two chapters point again towards further investigative as well as policy direction. The investigative direction is the exploration of the gap in the average years of schooling (and time allocation on studies in this context) which has magnified during this time since all the children are in the same age group. The number of years a child has been in education has the single biggest positive impact on the mathematics learning outcomes (in both the endowments and coefficients effect) in the fourth round results. The reduction of this gap through policy initiatives is one way to deal with this. Incentivizing staying in school without making the children compromise on the financial earns staying away from school entails, is one way to go about it.

We must also look into the policy implication in light of existing education/school level policies of the state/central governments. There are several policies that are in place in both Telangana and Andhra Pradesh. In order to reduce the drop outs among the children from these states, few policies are in place to tackle the issue. To provide residential arrangements to the school going children (from minority communities) of Telangana state. The name of the scheme is Telangana Minorities Residential Junior College. The objective of this scheme is to come up with a full residential schools and help develop the overall child's development. The idea is to have at least one residential school in each of the constituency of the Telangana state. The scheme falls under the scholarship scheme where fellowship as well as living arrangement is to be provided. For minority students, there are other schemes by the state government that aim to provide for living arrangements. There are other scheme run by the minority welfare department which has two parts where the first part is for the children from the minority communities and the second one is for the girl child from the minority communities. Similarly, there are educational policies specifically meant for the children from the Scheduled Caste (SC) and Scheduled Tribe (ST) background. Scheme from the ministry of social justice and empowerment which are in place to provide fellowships to the children who come from these backgrounds. Some of these schemes are meant to provide subsidized hostel facility to these students. There is pre matric scholarship for the children with these background too which is aimed to provide for fellowships to these children. Along with SC and ST students, this scheme also covers the children from Other Backward Classes (OBC). In order to minimize the incidence of dropout in these communities the scheme from the department of backward class welfare gives variable awards to the children of class 9 and 10. On the same line the same ministry has fellowship meant for the same children in the grade range of 11-PhD. The idea is to give them fellowship in the form of reimbursement of fees that they pay while they are in the education system. There are various other scholarships such as pre matric national scholarship, post matric national scholarship, merit cum means scholarship, and Telangana post matric scholarship. All of these scholarships are meant for the children from the minority communities. There is another pre matric scholarship specifically meant for the children of those engaged in occupations involving cleaning and are prone to health hazards. Under this scheme, the children in the grades between (*I-X)* receive a monthly amount for 10 months. The amount ranges from rupees 225-750 per month depending upon whether the children are day scholars or hostellers respectively.

Similar to the state of Telangana, the state of Andhra Pradesh also has several education related policies in place or are to be put in place. Jagananna Vidya Kanuka Kit Scheme which was launched in October 2020 is intended to reach 43 lakh beneficiaries. This scheme is meant to provide for school uniforms, notebooks, shoes, socks, and school bags. The target students under this scheme includes the children who are in the grade between *(I-X)* in the government schools. There are schemes by the state government which are meant to provide career guidance as well. AP career guidance portal for which the target audience/students are the secondary and higher secondary students. The scheme was rolled out last in the year 2021. This scheme was launched by the department of school education in collaboration with the UNICEF. This is aimed to provide access to information on different careers to the students between the grades (*IX-XII*). In order to provide financial assistance to the children who hail from the poorer economic background there is another scheme by the state government, namely, AP Jagananna Vasathi Deevena Scheme (2020). The aim of this scheme is to help the students who are in IITs/Polytechnic/degree pay their hostel and mess charges in the institutions they are in. For the IIT students the scheme provides an assistance of 10000/ year. Similarly, they provide 15000 rupees in two equal instalments to the students of polytechnic. And to the students of degree, they provide 20000 in two instalments. Similar to this scheme there is another scheme, namely, AP Jagananna Vidya Deevena Scheme (2022). The aim is to pay the full fee reimbursement for the children from the marginalized background. The SC, ST, OBC, minorities, Kapus, EWS, PD students are eligible to avail the benefits of the scheme. The students who are pursuing B.Tech, B.Pharma, M.Tech, M.Pharma, MBA, MCA, B.eD are all eligible to benefit from the scheme in the form of receiving the full reimbursement of the fee. On a different line, the Mid-Day meal which has now been renamed as, namely, AP Jagananna Gorumudda Scheme is also in place to maintain the nutritional standing of school age children. The scheme provides free lunches to the children of primary and upper primary grade in the government and government aided schools. Other schemes such as the AP free laptop scheme (2022) for the students above class (*IX)* and AP pre matric scholarship for the children who hail from the poorer socio-economic background, are aimed to provide assistance in some form (not regular) to the school going children. On a completely different line, AP Mana Badi Nadu Nedu Scheme (2022) which is aimed at developing government schools' infrastructure for the benefit of the children who go to these schools. The target is to construct new anganwadis, renovation of the exiting anganwadis. Moreover, revamping of primary, upper primary, high schools, junior colleges etcetera. Electrification of schools, construction of toilets, maintaining drinking water supply, and other school level repairs are the target of this scheme.

Browsing through all the education related schemes that are in place in both of these states there are several things that are to be pointed out here in the context of our results from this thesis. Foremost among these pertains to the schooling cost. As we have seen from Table *II* that the major education provider across the surveyed regions is the private schools. The private schools charge a greater chunk of money for education compared to the public school counterparts. There is no scheme by these state governments or the central government to improve the condition of the schools, the other providers of education. Except for one scheme (AP Mana Nada Nedu Scheme, 2022) which has recently come up in the state of Andhra Pradesh there is none to improve the dilapidated state the government schools and other providers of education (non-private schools) are in. There has to be an institutionalized mechanism to improve these schools to make the condition of these schools better. The fundamental reason is that, as we have found from our research that public school children are lacking behind because they do not have better background features compared to the children of private schools. This gap needs to be filled by improving drastically the quality of infrastructure of non-private schools. Furthermore, we have seen that the children from the poorer background (children of non-schooled parents, children of public schools, and rest of the children) they discontinue their schooling early on compared to their respective counterparts. We could not locate a scheme by either of the state which incentivize the stay in schooling, especially by the children of poorer background. A constant financial support is needed to them to stay in education and schemes are needed to target that children for their eventual stay and reaping the benefit in terms of learning. We learn from the research that the mathematics learning inequality among the groups of children is because of several factors. The inequality in learning between the children of schooled and non-schooled parents, between private and public school children, and better performing and rest of the children, is in large part because of their respective differences in endowments. The children of schooled parents, private schools, and better performing children, could perform better in large part because of their respective better background features compared to their counterparts.

On the policy front, this research has some valuable insights to provide. Bridging the resource gap in the early lives of children should be a priority. However, we also learn from that there is an efficiency gap as well (although reducing), meaning, bridging this gap is not enough to deal with the issue. The gap in the average years of schooling (and time allocation on studies) which has magnified during this time needs to be further explored. The young adults, that we research here are discontinuing (given their age) from education or not continuing to higher education. Incentivizing staying in education without making the children compromise on the opportunity cost staying away from education entails, is one way to go about it.

Discussions

This thesis, just like any research, has certain limitations on certain grounds. The foremost among those pertains to the dataset of Young Lives Survey. The sampling methodology used for the survey was borrowed from the health surveillance studies, sentinel site methodology. This methodology for the collection of information on these children was found to be most suitable because of the nature of the information required over a period of time i.e. longitudinal. The foremost criteria that was chosen by YL was to select the children specifically from the poorer background as they wanted to study the changing nature of poverty. Hence the data that they could collect is not strictly statistically representative of the population. Moreover, given the nature and objectives of our thesis we lost a lot of information from the available dataset which cut the sample size of children we could actually study. In that context, the results and inferences drawn from our thesis should be seen in that light. Especially the limits pertaining to the generalizability of our results has been affected.

Elaborating further on the previous point, the information collected in the survey relates mostly to the background in which the children are growing/have grown. Indeed, the YL dataset has an elaborate number of variables on which information is available to track the changes occurring in the lives of these children. However, the school level information is not available in as much detail as it is for other background variables. We could not include school level features to capture the effect of those in explaining the learning differences between these groups of children. Given this limitation of the data we used the schooling cost incurred by these children in the last academic year as a proxy to capture the school level heterogeneity. This variable too has issues as the composition of this variable (in terms of different costs it captured) was different between the rounds. Hence, we could comment on the role of schooling cost in determining the learning gap between the groups of children, however with the disclaimer that this difference in the composition of this cost has to be kept in mind while reading the results.

The methodology that we have deployed in the study of learning difference between certain groups of children is indeed a suitable one. However, the Blinder-Oaxaca methodology for decomposition of mean difference between groups has been derived from linear regression. An existence of non-linear relationship could not be explored in our thesis. The non-linear B-O decomposition is available for use but we could not do it in our research. Furthermore, the decomposition that we have done between each set of groups has divided them only in two groups for each benchmark feature. Take for example the first benchmark feature, parental schooling status. The thesis has divided the children into two groups which are children of schooled parents and non-schooled parents. There is no issue in bracketing the children non-schooled parents as one as they all have zero year in schooling. However, bracketing the children of schooled parents into one has an issue where we treat all the levels of education (say one year or ten years of schooling) as the same. The reading of results obtained from this analysis should consider that. As the sample size was small, we could not do anything about it and went ahead with classifying them as such. Under different circumstances, it would have been interesting to look into the role of background features in determining the learning outcome among different classes of children whose parents have varying level of schooling years. Similarly, for the third classification of children directly on the basis of their marks has similar issue where it would have been interesting to look into the role of background features in determining the learning outcome for different classes of performance. On a different line, in the second classification, we have only looked into the learning difference between the children of private and public schools with respect to background characteristics. As these choices were the most often taken we chose to study them leaning children from other schools. This cut our sample size smaller and hence the inference limited. It would have been more holistic to have a look at the learning variations among various choices of schooling. However, given our sample size we could not go about with that exercise.

Another important thing that has to be taken into consideration is the usage of just mathematics score as an indicator of learning. A more nuanced, where more subjects like sciences, social science etcetera should also have been taken into account. A more holistic understanding of learning in terms of subjects must include an all-round educational indicator when making a comment, especially empirical. In sum, in further studies on learning outcomes must include a more heterogeneous set of subjects. Moreover, the scores on mathematics that were used in this study were designed for each round separately using TIMSS and PISA. There were certain similarities in the questions that were asked between the rounds but we could not comment on the direction of the learning outcome difference. In order for us to do that these tests needed to have been brought to a uniform comparable scale using IRT and CTT. We could not comment much on the differences and the direction those differences are taking between the rounds where the children have gotten older for precisely the same reason. All of these limitations restrict the limit to which these results are to be used for policy implications. The policy implications obtained from this research are very specific and cannot be generalized to a great extent because of the issues raised above.

References

Annual Status of Education Report. (2005). Pratham Resource Centre, New Delhi.

Annual Status of Education Report. (2010). Pratham Resource Centre, New Delhi.

Annual Status of Education Report. (2017). Pratham Resource Centre, New Delhi.

Ammermueller, A. (2011). PISA: What Makes the Difference? Explaining the Gap in Pisa Test Scores between Finland and Germany. *Empirical Economics*, *33*(2), 263–287.

Arteaga, I., & Glewwe, P. (2014). Achievement Gap between Indigenous and Non-Indigenous Children in Peru. *Young Lives Working Paper 130*.

Asif, R., Merceron, A., Abbas, S., & Ghani, N. (2017). Computers & Education Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194.

Atherton, P., & Kingdon, G. G. (2010). The relative effectiveness and costs of contract and regular teachers in India by. *Institute of Education, University of London*, 256-276.

Azubuike, Obiageri Bridget, and K. B. (2016). Young Lives' Constructed files. *Oxford Review of Economic Policy*, *23(2)* (7483), 168–195.

Azubuike, O. B., Moore, R., & Iyer, P. (2017). Young Lives School Surveys, 2016–17: The design and development of cross-country Maths and English tests in Ethiopia, India & Vietnam. *Young Lives*.

Baird, K. (2012). Class in the classroom: The relationship between school resources and math performance among low socioeconomic status students in 19 rich countries. *Education Economics*, *20*(5), 484–509.

Barrera-Osorio, F., García-Moreno, V., Patrinos, H. A., & Porta, E. (2011). Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: An application to Indonesia's results in PISA mathematics. *World Bank Policy Research Working Paper*, (5584).

Behrman.]ere R. (1996). The impact of health and nutrition on education. *The World Bank Research Observer*, *11*(1), 23–37.

Belot, M., & James, J. (2011). Healthy school meals and educational outcomes. *Journal of Health Economics*, *30*(3), 489–504.

Blau, P.M., Duncan, O. (1967). The American Occupational Structure. *In New York: John Wiley & Sons, 1967*.

Biewen, M. (2012). Additive Decompositions with Interaction Effects. *IZA Discussion Paper*, (6730).

Blinder, A. S., (1973). Wage Discrimination : Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Breen, R., & Jonsson, J. A. N. (2000). Analyzing Educational Careers : A Multinomial Transition Model. *American Sociological Review*, *65*(5), 754–772.

Brown, L. & Iyengar, S. (2008) Parenting Styles: The Impact on Student Achievement. *Marriage & Family Review*, 43:1-2, 14-38

Burger, R. (2011). School effectiveness in Zambia: The origins of differences between rural and urban outcomes. *Development Southern Africa*, *28*(2), 157–176.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *National Bureau of Economic Research Working Paper*, (17699).

Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review*, *31*(4), 376–390.

Cobb-Clark, D. A., & Gørgens, T. (2014). Parents' economic support of young-adult children: Do socioeconomic circumstances matter? *Journal of Population Economics*, *27*(2), 447–471.

Crane, J. (1996). Effects of home environment, SES, and maternal test scores on mathematics achievement. *Journal of Educational Research*, *89*(5), 305–314.

Crookston, B. T., Forste, R., McClellan, C., Georgiadis, A., & Heaton, T. B. (2014). Factors associated with cognitive achievement in late childhood and adolescence: The Young Lives cohort study of children in Ethiopia, India, Peru, and Vietnam. *BMC Pediatrics*, *14*(1), 1–9.

Cueto, S., and Leon, S. (2012). Psychometric Characteristics of cognitive development and achievement instruments in round 3 of Young Lives (Young Lives Technical no Notes) Young Lives.

Currie, J., & Yelowitz, A. (2000). Are public housing projects good for kids ? *Journal of Public Economics*, *75*(1), 99–124.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Score. *American Economic Journal: Applied Economics*, *5*(2), 29–57.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304.

Dawes, A. (2020). Measuring the Development of Cognitive Skills Across Time and Context: Reflections from Young Lives. *Young Lives*

Dercon, S., & Krishnan, P. (2009). Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries. *Children, Youth and Environments*, *19*(2), 138–163.

Desai, S., Dubey, A., Vanneman, R., & Banerji, R. (2008). Private Schooling in India: A New Educational Landscape. *India Policy Forum*, *5*(1), 1–58.

Deslandes, R., E. Royer, D. Turcotte, and R. B. (1997). School Achievement at the Secondary Level: Influence of Parenting Style ant Parent involvement in Schooling. *McGill Journal of Education/Revue Des Sciences de l'éducation de McGill*, *32*(3), 191–207.

Duncan, O.D., Featherman, D. L., and Duncan B. (1972). Socioeconomic background and achievement. *In New York: Seminar*.

Dutta Chowdhury, S., & Ghosh, T. (2011). Nutritional and socioeconomic status in cognitive development of Santal children of Purulia district, India. *Annals of Human Biology*, *38*(2), 188–193.

Fan, X., & Chen, M. (2001). Parental involvement and student's academic achievement: a meta-analysis. *Educational Psychology Review*, *13*(1), 1–22.

Frisvold, D. E. (2015). Nutrition and cognitive achievement: An evaluation of the School breakfast program. *Journal of Public Economics*, *124*(2015), 91–104.

Galab, S., Prudhvikar, R., & Singh, R. (2014) Young lives survey design and sampling in Andhra Pradesh, India

Galab, S., P. Prudhvikar Reddy and Renu Singh (2014a). Education and Learning: Round 4 Preliminary Findings, *Round 4 Fact Sheet, Oxford*: Young Lives

Glewwe, P., Jacoby, H. G., & King, E. M. (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics*, *81*(3), 345–368.

Glewwe, P., Krutikova, S., & Rolleston, C. (2017). Do Schools Reinforce or Reduce Learning Gaps between Advantaged and Disadvantaged Students? Evidence from Vietnam and Peru. *Economic Development and Cultural Change*, 65(4), 699-739.

Glewwe, P., Chen, Q., & Katare, B. (2015). What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data. *Asia and the Pacific Policy Studies*, *2*(3), 494–516.

Goldhaber, D. D. (1996). Public and private high schools: Is school choice an answer to the productivity problem? *Economics of Education Review*, *15*(2), 93–109.

Goyal, S. (2009). Inside the house of learning: The relative performance of public and private schools in Orissa. *Education Economics*, *17*(3), 315–327.

Goyal, S., & Pandey, P. (2009). How do government and private schools differ? Findings from two large Indian states. *International Bank for Reconstruction and Development/the World Bank*

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

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*. 8(4), 453-479.

Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. *Urban Education*, *40*(3), 237–269.

Jeynes, W. H. (2007). The Relationship between Parental Involvement and Urban Secondary School Student Academic Achievement. *Urban Education*, *42*(1), 82–110.

Keane, M. P., Krutikova, S., & Neal, T. (2018). The impact of child work on cognitive development: results from four Low to Middle Income countries. (No. W18/29) *IFS Working Papers*.

Kingdon, G. G. (1996). The Quality and Efficiency of Public and Private Schooling: A Case Study of India. *Oxford Bulletin of Economics and Statistics*, *58*(1), 57–82.

Kingdon, G. G. (2007). The progress of school education in India. *Oxford Review of Economic Policy*, 23(2), 168-19 5.

Kingdon, G. G., & Teal, F. (2007). Does performance related pay for teachers improve student performance ? Some evidence from India. *Economics of Education Review*, *26*, 473–486.

Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, *15*(4), 143–156.

Le, H. T., & Nguyen, H. T. (2018). The evolution of the gender test score gap through seventh grade : new insights from Australia using unconditional quantile regression and decomposition. *IZA Journal of Labor Economics*, *7*(2).

Leibowitz, A. (1977). Parental Inputs and Children's Achievement. *The Journal of Human Resources*, *12*(2), 242–251.

Li, Z., & Qiu, Z. (2018). How does family background affect children's educational achievement? Evidence from Contemporary China. *The Journal of Chinese Sociology*, *5*(1), 1–21.

Lipovetsky, S. (2012). Regression split by levels of the dependent variable. *Journal of Modern Applied Statistical Methods*, *11*(2), 319–324.

Machado, J. A. F., & Mata, J. (2005). Counterfactual Decomposition Changes in Wage Distributions Using Quantile Regression. *Journal of Applied Econometrics*, *20,* 445–465.

Mare, R. D. (1981). Change and Stability in Educational Stratification. *American Sociological Review*, *46*(1), 72–87.

Mare, R. D. (1980). Social Background and School Continuation Decisions. *Journal of American Statistical Association*, *75*(370), 295–305.

Mare, R. D. (1979). Social Background: Composition and Educational Growth. *The Population Association of America*, *16*(1), 55–71.

Mau, W.-C. (1997). Parental Influences on the High School Students' Academic Achievement: A Comparison of Asian Immigrants, Asian Americans, and White Americans. *Psychology in the Schools*, *34*(3), 267–277.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, *12*, 577–590.

Montes-Rojas, G., Siga, L., & Mainali, R. (2017). Mean and quantile regression Oaxaca-Blinder decompositions with an application to caste discrimination. *Journal of Economic Inequality*, *15*, 245–255.

Mukherjee, C., & Pal, R. (2016). Role of Parental Expectations in Determining Child Labour and Schooling. *Working Paper 2016:6*

Muralidharan, K., & Sundararaman, V. (2015). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics*, *130*(3), 1011–1066.

National Achievement Survey (2012) Ministry of Human Resource Development, New Delhi.

National Achievement Survey (2014) Ministry of Human Resource Development, New Delhi.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*, 693–709.

Parvez, A., & Laxminarayana, K. (2021). Mathematics learning inequality among children of private and public schools. *Asia Pacific Education Review*.

Paxson, C., & Schady, N. (2007). Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. *Journal of Human Resources*, *42*(1), 49–84.

Reeves, E. B., & Lowe, J. (2009). Quantile Regression : An Education Policy Research Tool. *Journal of Rural Social Sciences*, *24*(1).

Reynolds, J. (1992). Comparing Measures of Parental Involvement and Their Effects on Academic Achievement. *Early Childhood Research Quarterly, 7*(3), 441–462.

Rolleston, C. (2014). Learning profiles and the "skills gap" in four developing countries: A comparative analysis of schooling and skills development. *Oxford Review of Education*, *40*(1), 132–150.

Sakellariou, C. (2008). Peer effects and the indigenous/non-indigenous early test-score gap in Peru. *Education Economics*, *16*(4), 371–390.

Sanfo, J. B. M. B., & Ogawa, K. (2021). Explaining the rural-urban learning achievements gap in Ethiopian primary education: A re-centered influence function decomposition using Young Lives data. *Education Economics*, 1–29.

Sanjay, S., Gupta, R., Gautam, V., & Meena, S. H. R. (2014). Educational Statistics at a Glance. *MHRD, Bureau of Planning, Monitoring & Statistics, New Delhi*.

Sharma, G. (2014) Study of classroom environment parental education, income and institution as predictor of students' academic achievement. *Indian Association of Health, Research and Welfare*, 4(4), 277-281

Singh, A. (2015). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics*, *113*(2015), 16–32.

Singh, A. (2011). Family background, academic ability and associated inequality of opportunity in India. *Economics Bulletin*, *31*(2), 1463–1473.

Singh, A. (2012). Reducing Gender Based Inequalities In Employment And Wages: Is India On Right Track? *Indira Gandhi Institute of Development Research*, (June).

Singh, R., & Mukherjee, P. (2019). Effect of preschool education on cognitive achievement and subjective wellbeing at age 12: evidence from India. *Compare: A Journal of Comparative and International Education*, *49*(5), 723–741.

Tian, M. (2006). A Quantile Regression Analysis of Family Background Factor Effects on Mathematical Achievement. *Journal of Data Science*, *4*, 461–478.

Tighe, E. L., & Schatschneider, C. (2014). A Quantile Regression Approach to Understanding the Relations among Morphological Awareness, Vocabulary, and Reading Comprehension in Adult Basic Education Students. *Journal of Learning Disabilities*.1-13

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.

Wadhwa, W. (2009). Are private schools really performing better than government schools? *Annual Status of Education Report (Rural) New Delhi*.

Wamalwa, F. M., & Burns, J. (2018). Private schools and student learning achievements in Kenya. *Economics of Education Review*, *66*(July), 114–124.

Witte, J. F. (1992). Private school versus public school achievement: Are there findings that should affect the educational choice debate? *Economics of Education Review*, *11*(4), 371–394.

Zhang, L., & Lee, K. A. (2011). Decomposing achievement gaps among OECD countries. *Asia Pacific Education Review*, *12*(3), 463–474.

Zhang, Y. (2006). Urban - Rural Literacy Gaps in Sub - Saharan Africa : The Roles of Socioeconomic Status and School Quality. *Comparative Education Review*, *50*(4), 581–602.

Asia Pacific Education Review
https://doi.org/10.1007/s12564-021-09733-6

Mathematics learning inequality among children of private and public schools

Aquib Parvez[1] · K. Laxminarayana[2]

Received: 3 April 2021 / Revised: 8 November 2021 / Accepted: 10 November 2021

Abstract
In this study, we research the gap in average mathematics learning between the children of private and public schools at two points. We have divided them into two groups on the basis of the type of school they are attending or last attended. We first skim through their various background characteristics at these two points. We then explore these background characteristics and using the threefold Blinder–Oaxaca decomposition, we find that it is the difference in the average endowments between the two which consistently explains the gap in average performance between them. We also find the role of differential impact of the background characteristics on the average learning outcome of children on the first point. The most important and consistent contributor to the endowment effect is the schooling cost and the time allocation on studies. One striking result is the now significant contribution of the gap in average years of schooling which is worrying because these children are from the same age group. We conclude that with the average features and returns of the private school children, the gap in learning between them would have been removed. Moreover, the public school children would have performed better than the private school children.

Keywords Blinder–Oaxaca decomposition · Endowment effect · Mathematics learning · Public–private schooling · Schooling cost

JEL Classification I20 · I21 · I24 · I25 · I29

Introduction

In India, there is an increasing demand for secondary education. Because of its lucrative nature, private schooling meets some of this demand. Furthermore, the rapid growth of private schooling in India is due to the poor management of public schools and in particular, the high absenteeism among teachers (Kingdon, 2007). Moreover, this increase is affected by the common perception that private schools provide better quality education (Wadhwa, 2009). In India, enrollment in private schools is associated with children's enhanced learning outcomes, after controlling for background characteristics (Desai et al., 2008). A plethora of literature has captured the quality of education by employing children's learning outcomes as an indicator (Goyal, 2009; Kingdon, 1996; Singh, 2015; Singh & Mukherjee, 2019). Studies (e.g., Kingdon, 1996; Desai et al., 2008; Goyal, 2009; Wadhwa, 2009; Chudgar and Quin, 2012; Wamalwa and Burns, 2018; Singh, 2015; Singh & Mukherjee, 2019) have revealed an association between Indian children attending private schools and enjoying more enhanced learning and achievement scores. While the majority of studies have demonstrated a private schooling effect, this effect has not been apparent in some studies. Although Goyal and Pandey (2009 revealed a private schooling effect among Indian children, their results suggest that the quality of both types of schooling were poor. Muralidharan and Sundararaman (2015) found no public–private differences in various aspects of learning in children from the southern state of Andhra Pradesh in India.

✉ Aquib Parvez
aquibparvez78@gmail.com

K. Laxminarayana
klnhcueco@gmail.com

[1] School of Economics, University of Hyderabad, Hyderabad, Telangana 500046, India

[2] School of Economics, University of Hyderabad, Gachibowli, Hyderabad, Telangana 500046, India

Published online: 23 November 2021

Certificate

NATIONAL INSTITUTE OF SECURITIES MARKETS
An Educational Initiative of SEBI

The Indian Econometric Society
1960

University of Mumbai

55th Annual Conference of The Indian Econometric Society (TIES)

Mumbai School of Economics & Public Policy, University of Mumbai & National Institute of Securities Markets

8th, 9th and 10th January, 2019

This is to certify that Aquib Parvez

has presented a paper titled Educational Achievement and Mothers' education: Evidence from India (Andhra Pradesh)

Co-authored with

at the 55th Annual Conference of The Indian Econometric Society (TIES) held during 8-10 January, 2019 at the NISM Campus, Patalganga.

Dr. Neeraj Hatekar	Dr. Latha Chari	Dr. K. Shanmugan
Local Organizing Secretary	Local Organizing Secretary	Convener, Programme Committee, TIES

56th ANNUAL CONFERENCE OF THE INDIAN ECONOMETRIC SOCIETY (TIES)

The Indian Econometric Society 1960

Madurai Kamaraj University

8 - 10 January 2020

CERTIFICATE

This is to certify that Mr/Ms/Dr. Aquib Parvez .... **of** University of Hyderabad, Hyderabad, Telangana .... **presented a research paper entitled** Mathematics Achievement of Public and Private School Children: Why is One Better? ....

in the 56th Annual Conference of The Indian Econometric Society (TIES) held during 08 – 10 January 2020, organized by School of Economics, Madurai Kamaraj University, Madurai, Tamil Nadu, India.

மெய்ப்பொருள் காண்பதறிவு

Prof. N.R. Bhanumurthy	**Prof. D. Swamikannan**	**Dr. S. Pushparaj**
Secretary, TIES	**Local President**	**Local Organising Secretary**

CERTIFICATE OF PLAGIARISM

This is to declare that the research conducted in the present thesis entitled *"Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks"* by *Aquib Parvez* (registration no:17SEPH16) is an original work, carried out by him under my supervision, in the School of Economics, for the award of the Doctor of Philosophy degree from the University of Hyderabad. This is also to certify that the plagiarism shown in the plagiarism report is mainly from his own published work, which is a chapter (4.1) from this thesis.

K. Laxminarayana

Place: Hyderabad
Date: 10/03/2022

Supervisor
Prof. K. Laxminarayana

SUPERVISOR
SCHOOL OF ECONOMICS
UNIVERSITY OF HYDERABAD
Hyderabad-500 046 (INDIA

Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks

by Aquib Parvez

Submission date: 11-Mar-2022 11:59AM (UTC+0530)
Submission ID: 1781746435
File name: AQUIB_PARVEZ.docx (723.81K)
Word count: 55337
Character count: 293289

Mathematics Learning Inequality among Indian Children: A Look Inside with respect to Three Benchmarks

ORIGINALITY REPORT

21%	5%	20%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Aquib Parvez, K. Laxminarayana. "Mathematics learning inequality among children of private and public schools", Asia Pacific Education Review, 2021 Publication	16%
2	www.journaltocs.ac.uk Internet Source	1%
3	www.younglives-india.org Internet Source	<1%
4	Submitted to University of Hyderabad, Hyderabad Student Paper	<1%
5	www.edubloxtutor.com Internet Source	<1%
6	academic.oup.com Internet Source	<1%
7	fr.scribd.com Internet Source	<1%

Research Scholar's own paper.

K. Laxminarayana

SUPERVISOR
SCHOOL OF ECONOMICS
UNIVERSITY OF HYDERABAD
Hyderabad-500 046 (INDIA)

8	Amita Chudgar, Elizabeth Quin. "Relationship between private schooling and achievement: Results from rural and urban India", Economics of Education Review, 2012 Publication	<1 %
9	brsa.org.br Internet Source	<1 %
10	drum.lib.umd.edu Internet Source	<1 %
11	Dandan Zhang, Xin Li, Jinjun Xue. "Education Inequality between Rural and Urban Areas of the People's Republic of China, Migrants' Children Education, and Some Implications", Asian Development Review, 2015 Publication	<1 %
12	capone.mtsu.edu Internet Source	<1 %
13	Irma Arteaga, Paul Glewwe. "Do community factors matter? An analysis of the achievement gap between indigenous and non-indigenous children in Peru", International Journal of Educational Development, 2017 Publication	<1 %
14	www.merit.unu.edu Internet Source	<1 %

15	people.bath.ac.uk Internet Source	<1 %
16	younglives.org.uk Internet Source	<1 %
17	Aparajita Dasgupta. "Can the Major Public Works Policy Buffer Negative Shocks in Early Childhood? Evidence from Andhra Pradesh, India", Economic Development and Cultural Change, 2017 Publication	<1 %
18	Submitted to Manchester Metropolitan University Student Paper	<1 %
19	www.younglives.org.uk Internet Source	<1 %
20	librarysearch.aut.ac.nz Internet Source	<1 %
21	www.icpublicpolicy.org Internet Source	<1 %
22	Sangeeta Goyal. "Inside the house of learning: the relative performance of public and private schools in Orissa", Education Economics, 2009 Publication	<1 %
23	repositorio.up.edu.pe Internet Source	<1 %

24	www.mathforamerica.org Internet Source	<1 %
25	socialjustice.nic.in Internet Source	<1 %
26	Duncan, Beverly. "Education and Social Background", American Journal of Sociology, 1967. Publication	<1 %
27	www.scribd.com Internet Source	<1 %
28	Alan C. Marco, Gordon C. Rausser. "The Role of Patent Rights in Mergers: Consolidation in Plant Biotechnology", American Journal of Agricultural Economics, 2008 Publication	<1 %
29	Submitted to Andrews University Student Paper	<1 %
30	airsar.asf.alaska.edu Internet Source	<1 %
31	www.unicampus.it Internet Source	<1 %
32	Bård Smedsvik, Roberto Iacono, Antonella D'Agostino. "Immigration and social assistance: Evidence from the Norwegian welfare state", Social Policy & Administration, 2021	<1 %

Publication

33	John Ermisch, Marco Francesconi. "Family Matters: Impacts of Family Background on Educational Attainments", Economica, 2003 Publication	<1 %
34	Sowmya Dhanaraj, Christy Mariya Paul, Smit Gade. "Household income dynamics and investment in children: Evidence from India", Education Economics, 2019 Publication	<1 %
35	Julian di Giovanni, Andrei A. Levchenko, Francesc Ortega. "A GLOBAL VIEW OF CROSS-BORDER MIGRATION", Journal of the European Economic Association, 2015 Publication	<1 %
36	ojs.amhinternational.com Internet Source	<1 %
37	www.asercentre.org Internet Source	<1 %
38	Submitted to Cranfield University Student Paper	<1 %
39	Submitted to National University of Singapore Student Paper	<1 %
40	mobt3ath.com Internet Source	<1 %

41	Morag Henderson. "A Test of Parenting Strategies", Sociology, 2012 Publication	<1 %
42	hdl.handle.net Internet Source	<1 %
43	eprints.utm.my Internet Source	<1 %
44	shareok.org Internet Source	<1 %
45	en.wikipedia.org Internet Source	<1 %
46	Submitted to Amity University Student Paper	<1 %
47	Geeta Kingdon. "THE QUALITY AND EFFICIENCY OF PRIVATE AND PUBLIC EDUCATION: A CASE-STUDY OF URBAN INDIA† : A CASE-STUDY OF URBAN INDIA", Oxford Bulletin of Economics and Statistics, 05/01/2009 Publication	<1 %
48	link.springer.com Internet Source	<1 %
49	openknowledge.worldbank.org Internet Source	<1 %
50	webstore.dundee.ac.uk Internet Source	

<1%

51 www.tandfonline.com
Internet Source <1%

52 Submitted to Cornell University
Student Paper <1%

53 S.K. Das, A.D. Burma, S. Amudhan, V. Mishra, P. Mahapatra, A. Ashok, M. Philip. "Do children in India grow well into adolescents? Longitudinal analysis of growth transitions from Young Lives panel survey in India", Public Health, 2022
Publication <1%

54 nces.ed.gov
Internet Source <1%

55 www.usinnovation.org
Internet Source <1%

56 Sandra E Black, Paul J Devereux, Kjell G Salvanes. "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital", American Economic Review, 2005
Publication <1%

57 dx.doi.org
Internet Source <1%

nlsinfo.org

58	Internet Source	<1%
59	philosociology.com Internet Source	<1%
60	www.migrationpolicy.org Internet Source	<1%
61	www2.druid.dk Internet Source	<1%
62	Submitted to Azim Premji University Student Paper	<1%
63	K. Deininger, D. A. Ali, T. Alemu. "Impacts of Land Certification on Tenure Security, Investment, and Land Market Participation: Evidence from Ethiopia", Land Economics, 2011 Publication	<1%
64	Submitted to University of Lancaster Student Paper	<1%
65	aura.abdn.ac.uk Internet Source	<1%
66	educationcommission.org Internet Source	<1%
67	ibs.org.pl Internet Source	<1%
	www.law.fsu.edu	

68	Internet Source	<1 %
69	www.igi-global.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		