Parental Socioeconomic Status and Children's Academic Achievement

Longitudinal Evidence from Ethiopia, India, Peru and Vietnam

Milagros Vandemoortele

A thesis submitted to the Department of Social Policy of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London, June 23, 2024.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I consider the work submitted to be a complete thesis fit for examination.

I declare that my thesis consists of 72,509 words (excluding the front matter, bibliography and appendices).

I authorise that, if a degree is awarded, an electronic copy of my thesis will be deposited in LSE Theses Online (in accordance with the published deposit agreement) held by the British Library of Political and Economic Science and that, except as provided for in regulation 61, it will be made available for public reference.

I authorise the School to supply a copy of the abstract of my thesis for inclusion in any published list of theses offered for higher degrees in British universities or in any supplement thereto, or for consultation in any central file of abstracts of such theses.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Abstract

For decades academics have tried to understand why children from less advantaged households perform worse than those from more advantaged ones. The evidence from high-income countries shows that the socioeconomic status (SES) of parents is a critically important predictor of children's academic achievement. There is little longitudinal evidence on the association between parental SES and children's academic achievement among low- and middle-income countries. This is partly due to the limited availability of relevant and reliable data.

I use the Young Lives dataset which, in each of four low- and middle-income countries, follows 2,000 children from 1 year old to 15 years old. These countries are Ethiopia, India, Peru and Vietnam.

With this thesis I am adding to low- and middle-income country literature what has been done to high-income country literature. First, I document SES gaps in achievement that cover the age range of 5 to 15 years old (which existing studies do not cover). Second, I examine children's trajectories in their achievement, based on their initial SES and achievement scores. Third, because early childhood education (ECE) is often promoted as an effective strategy to reduce SES gaps in achievement, I examine whether ECE is associated with gains in achievement across childhood (over and above the association with parental SES).

While I expected to obtain very different results across the countries given their distinct levels of development and diverse contexts, the findings on SES gaps in achievement and the achievement trajectories by SES, are remarkably similar, not only to each other but also to those in high-income countries. When I examine whether ECE is associated with gains in achievement across childhood my findings differ at early ages, as I expected, given distinct ECE systems across countries. However, by the age of 15 years the findings are consistent across countries.

Thank you

To my two supervisors, Stephen Jenkins and Kitty Stewart, who showed me unwavering support throughout the decade it took to write this thesis, between the birth of our two children, a Covid pandemic and the challenges of a transatlantic move. Their fulsome engagement with my work and the rigorous feedback I received time and again showed their commitment to their role as supervisors. How lucky I have been to be their supervisee.

To my husband who also patiently supported and challenged me through the PhD. Thank you for regularly taking the children on holidays to warm and wonderful corners of the world, so that I could get on with the thesis. I will now be able to join the three musketeers.

To my parents who provided tremendous moral and practical support throughout the PhD.

To my colleagues at CASE who were my professional family on this PhD journey, for your community I am grateful. I have missed you in NYC.

To my colleagues at Young Lives who hosted me for a Fellowship, and where I learnt a tremendous amount about the Young Lives data and also about the challenges of collecting longitudinal data across four low- and middle-income countries. Hats off to you.

To Roo Griffith and A. Kwan who copy edited parts of the thesis for conventions of language, spelling and grammar.

To my children for being patient while I wrote this book.

Contents

Acronyms

CDA	Cognitive Development Assessment.
CEI	Centros de Educación Inicial.
CIRCUS	Citizens' Initiative for the Rights of Children Under Six.
COHORTS	Consortium of Health-Orientated Research in Transitioning Societies.
DHS	Demographic and Health Survey.
DICE	Development of Inequalities in Child Educational Achievement: A Six Country Study.
ECE	early childhood education.
ESCALE	Estadística de Calidad Educativa.
GDP	gross domestic product.
HIV/AIDS	human immunodeficiency virus/acquired immunodeficiency syndrome.
ICDS	Integrated Child Development Services.
IEA	International Evaluation Association.
IFS	Institute for Fiscal Studies.
ISCED	International Standard Classification of Education.
LSMS	Living Standards Measurement Study.

MCS	Millenium Cohort Study.
MdE	Ministerio de Educacion.
MICS	Multiple Index Cluster Survey.
MOET	Ministry of Education and Training.
MWCD	Ministry of Women and Child Development.
NER	net enrolment rate.
OECD	Organisation for Economic Cooperation and Development.
OLS	Ordinary Least Squared.
PISA	Programme for International Student Assessment.
PPVT	Peabody Picture Vocabulary Test.
PRONOEI	Programas no Escolarizados de Educación Inicial.
SCERT	State Council of Education Research and Training.
SES	socioeconomic status.
SRoV	Socialist Republic of Vietnam.
TIMMS	Trends in International Mathematics and Science Study.
TVIP	Test de Vocabulario en Imagenes Peabody.
U5MR	under-five mortality rate.
UNESCO	United Nations Educational, Scientific and Cultural Organization.
UNU-WIDER	United Nations University World Institute for Development Economics Research.
WHO	World Health Organisation.

Chapter 1

Introduction and Literature Review

The effects of privilege (or disadvantage) on children start at birth and grow throughout childhood. Some find it begins even before, in utero, as Lu et al. (2021) show and Grantham-McGregor et al. (2007) discuss. The socioeconomic context into which a child is born shapes their academic outcomes. For example, Bradbury et al. (2015) and Feinstein (2003) use data from high-income countries to show that initially-high-achieving children diverge in their achievement trajectories as they grow up, depending on their parent's socioeconomic status (SES). From an initially-high-achieving group of children, all of whom performed well on an achievement test at a young age, those from low-SES families fell behind, on average, compared with their high-SES counterparts in subsequent achievement tests.

Arguably, this is why the initial years of a child's life are given such importance by researchers, for example Grantham-McGregor et al. (2007) and Heckman (2008). To ensure all children have a chance to achieve their potential, the argument goes, it is essential to avoid SES gaps from emerging in early years.

One policy intervention aimed at reducing SES achievement gaps at early ages is early childhood education (ECE). The Organisation for Economic Cooperation and Development (OECD) writes that ECE "can give a strong start to all children by providing equitable opportunities and experiences that support development" (2021, p. 18). Nobel laureate James Heckman (2008) has published numerous articles documenting the benefits of ECE, and in one writes that "high quality early childhood interventions foster abilities and that inequality can be attacked at its source" (Heckman, 2008, p. 290). However, recent quasi-experimental evidence of large-scale ECE does not provide definitive proof of the argument. See for example Durkin et al. (2022) and Kottelenberg and Lehrer (2019).

In this thesis I focus on the socioeconomic conditions into which children are born, their parent's SES, to examine their academic outcomes throughout childhood.

I look at academic achievement in childhood because it is a predictor of later life outcomes – including academic and financial. See Heckman et al. (2006) for evidence on the USA, and Duncan et al. (2012) for cross-country evidence.

While studies on the relationship between parental SES and children's achievement, and between ECE and achievement, on high-income countries is prevalent, those on low- and middle-income countries remain limited in coverage and in scope due to insufficient relevant and reliable data available.

With this thesis I am adding to low- and middle-income country literature what has been done to high-income country literature. I aim to contribute to the small body of research on the relationship between parental SES and children's achievement in low- and middle-income countries by: first, documenting the SES gaps in achievement at 5, 8, 12 and 15 years old (an age range existing studies do not cover); second, examining children's trajectories in their achievement, based on their initial SES as Feinstein (2003) did using British data; third, investigating the relationship between ECE and achievement by considering whether ECE is associated with an advantage in achievement, over and above the association between parental SES and achievement.

While one would expect there to be SES gaps in achievement across low- and middle-income countries, as has been widely documented for high-income countries, speculation cannot not inform policy. Documenting SES gaps in achievement serves several purposes. First, empirical evidence of SES gaps in achievement allows us to identify and measure disparities. Second, it also serves as a foundation for further investigation into the causes and consequences of the SES gaps in achievement.The empirical evidence can help policymakers formulate policy, or fund additional research, to help address SES gaps in achievement. Third, international comparisons allow policymakers to learn from successful (or unsuccessful) strategies implemented elsewhere and adapt policies accordingly to their own contexts.

1.1 Research questions

In this thesis I consider these three foundational questions for four low- and middle-income countries:

RQ1. How large is the achievement gap between children from high and low SES families at 5, 8, 12 and 15 years old? Does it change between ages? How does any change in the SES achievement gap compare across countries?

RQ2. Is progress in achievement stratified by SES? Do the trajectories of initially high achieving children differ if they are from high versus low SES families? How does this differ across countries?

RQ3. Is attendance in ECE associated with relative gains in achievement across childhood? What do estimates for each country reveal about different ECE systems?

1.2 Data and analytical strategy

One of the most informative ways to investigate the relationship between parental SES and children's achievement is by comparing data across countries. Other countries serve as reference points to compare how wide or narrow a country's SES gaps in achievement are. Comparing across countries allows me to investigate how structural (e.g. institutional, economic, cultural) differences produce similar or distinct association between parental SES and children's achievement. As Ermisch, Jäntti, et al. (2012) argue, in a cross-country analysis "genetic transmission in the out-come ... should be the same across countries, and so cross country differences should reflect different environments, policy and otherwise" (Ermisch, Jäntti, et al., 2012, p. 11). Cross-national comparison also allows us to establish how similar or different changes in magnitude of the gaps are over the course of childhood. Yet, cross-national comparative studies of low- and middle-income countries are few in number.

I use Young Lives data for my research. The Young Lives surveys are specifically and carefully designed to enable cross-country comparisons across Ethiopia, India, Peru and Vietnam. According to the Young Lives documentation, these countries were "selected to include one from each of the major regions of the developing world, along with a range of political-economic conditions and circumstances, with strong institutional capacity locally to undertake complex panel research being another crucial criterion"(Boyden & James, 2014, p. 26).

The Young Lives team, which includes research teams in Ethiopia, India, Peru, Vietnam and Oxford started collecting data from around 2,000 children per country, aged 6 to 18 months in 2001/02 and followed them up at 5, 8, 12 and 15 years old and administered the same questionnaires (for the most part). These are the children in the younger cohort. (Young Lives also surveyed approximately 1,000 older children in each country, who are referred to as the 'older cohort' by Young Lives. The older cohort is not pertinent for my analysis because real-time data on their ECE and their initial SES was not collected.)

Attrition rates are low compared to other longitudinal studies and in three of the four countries mortality (between 1 and 5 years old) is the most important cause of attrition. The attrition rate (including mortality) for the younger cohort across all countries and across five rounds averages 6.8 per cent.

1.3 Substantive findings: socioeconomic status gaps in achievement

1.3.1 High-income countries

For decades academics have documented and tried to understand why children from less advantaged households perform worse in achievement tests, than those from more advantaged ones. See Blau and Duncan (1967) for one of the earliest examples focused on the USA. The evidence from high-income countries shows that the SES of parents is a critically important predictor of children's academic achievement, as discussed in Heckman et al. (2006). After reviewing the evidence on the relationship between parental SES and children's cognitive (and non-cognitive) development in high-income countries, James Heckman writes "[t]he accident of birth is a major source of inequality" (Heckman, 2008, p. 289).

Size of the SES gaps in achievement

As I review the literature, I rely mostly, but not wholly, on findings from cross-country comparisons.

Evidence, mostly from high-income countries, shows that achievement gaps emerge prior to school entry (as reported by Blanden et al., 2012; Bradbury et al., 2012, 2015, 2019; Dräger et al., 2023; Duncan & Magnuson, 2011b; Heckman, 2008; Linberg et al., 2019; Magnuson et al., 2012, among others), and the size of the SES gap ranges from around half to one standard deviation (s.d.). Bradbury et al. (2015) report SES gaps in reading scores at age 5 of around 1.0 s.d. for the US, 0.79 for the UK, 0.47 for Australia and 0.62 for Canada (and in maths of 1.02 s.d. for the US). Bradbury et al. (2012) conduct a cross-country comparison of these same four high-income countries and find that the SES gaps in vocabulary scores at age 5, adjusted for race-ethnicity-nativity, range between 0.63 and 0.87 s.d.. Linberg et al. (2019) compare raw (and adjusted) SES gaps in achievement between Germany and the US (for children at 6/7 years old) and find the raw gap in maths scores are similar across both countries (0.96 and 0.99 s.d. in the USA and Germany respectively).

SES gaps in achievement in high-income countries also tend to persist as children grow up. See Bradbury et al. (2015), Heckman and Mosso (2014), and Magnuson et al. (2012) who report near parallel trajectories of scores through childhood, between high and low SES groups. Passaretta et al. (2022) examine SES gaps in achievement across three European countries (Germany, the Netherlands and the UK) from 5 to 11 years old, and find that gaps remained stable over primary schooling. Skopek and Passaretta (2020) report a widening of the gap in Germany (through to 16 years old).

Mechanisms

The transmission mechanisms are numerous and I group them into three categories (as discussed in more detail in the Framework Chapter 2), these include the Family Stress Model, the Investment Model, and the Neurobiological Model. The Family Stress Model sets out the psychological pathways through which low resources impact on parents' stress levels and, consequently, their mental health, affecting their parenting ability and, in turn, their children's outcomes (research using this model is reviewed in Landers-Potts et al., 2015 and Duncan et al., 2014).

The Investment Model, mostly built on work by Becker (1981), suggests that financial resources affects children's outcomes through the parents' ability to invest in better-quality or more goods and services that contribute to improved children's outcomes. In their systematic review, Cooper and Stewart (2021) find that higher household income has a positive causal effect on children's outcomes, including their academic achievement.

The Neurobiological Model is employed by neurobiologists who examine how growing up with social disadvantage shapes children's brain development, their ability to concentrate and learn, and consequently their educational achievement. Brito and Noble (2014) explain that low-income and low-parental-education affect children's brain development in different ways, as do Noble et al. (2015) and Kim et al. (2018).

1.3.2 Low- and middle-income countries

Most of the evidence on the relationship between parental SES and children's achievement is on high-income countries. Similar research on low- and middle-income countries, particularly cross-national studies, are hard to come by. The limited scope of research is due to the limited availability of relevant and reliable data, as Reynolds et al. (2017) discuss. Not only are longitudinal data expensive and labour intensive to collect, collecting achievement scores in a reliable and age appropriate manner requires them to be designed and tested on the target population (thus in their language and adapted to the local education system and culture).

There is a small body of cross-sectional studies that examine the association between parental SES and achievement. Paxson and Schady (2007) analyse cross-sectional data from Ecuador and show that the gap in vocabulary test scores between the poorest and wealthiest quartile group increased substantially between 36 and 71 months of age. They also suggest that the socioeconomic gradient in cognitive outcomes occurs at an earlier age in Ecuador than in the USA. Naudeau et al. (2011) examine datasets from Cambodian and Mozambique and find that, despite a rather homogeneous group of mostly poor children, a socioeconomic gradient in cognitive development exists even at very low levels of development.

I have identified only four published studies on SES gaps in achievement across countries, namely Das et al. (2022), Lopez-Boo (2016), Reynolds et al. (2017), and Schady et al. (2015). (I rely on published, rather than grey, literature because these are subject more rigorous review processes.)

Schady et al. (2015) is the first cross-country study that examines the relationship between parental SES and children's educational attainment across low and middle-income countries. Schady et al. (2015) report SES gaps in vocabulary scores from five Latin American countries, three of which have panel data (an one of which is Peru, for which the Young Lives data is used). For children aged between 36 and 72 months Schady et al. (2015) report SES achievement gaps ranging from 1.23 to 0.77 s.d., when an asset index is used as a measure of SES. For the three countries with panel data the SES gaps in achievement remains as they grow older.

Lopez-Boo (2016) also uses Young Lives data (specifically the younger cohort) to investigate the relationship between SES, measured with expenditure, and vocabulary scores for children aged 5 to 8 years old across all Young Lives countries (Ethiopia, India, Peru and Vietnam). (She relies on the majority language sample, so does not include children who took the test in a language other than the majority language). Between 5 and 8 years old, Lopez-Boo (2016) reports SES gaps between 0.97 s.d. and 0.33 s.d., with no statistically significant change from 5 to 8 years old. Lopez-Boo (2016) then goes on to conduct a staged regression, adding sets of control variables to the regression, to compare how adding each set of controls changes the size of the SES gap.

Reynolds et al. (2017) also use the Young Lives data and report SES gaps in vocabulary scores from 5 to 12 years old. They use an asset index (measured at 1 year old) and parental education (measured at 5 years old) to measure SES. Vocabulary scores are reported in age-adjusted percentiles. They find SES gaps in vocabulary scores at 5 through to 12 years old. Confidence intervals of the unadjusted SES gap in achievement overlap between 5 and 12 years old, when parental education is used as a measure of SES – indicating there was no statistically significant change from 5 to 12 years old.

Das et al. (2022) also use Young Lives data to examine the association between SES and college attendance, through achievement scores, across all four Young Lives countries. Authors measure SES with a composite score comprised of parental education and variables that comprise the asset index. They use maths scores to measure achievement, using data at 8, 12 and 15 years old. As SES gaps in achievement are not the aim of the paper, these gaps appear in a descriptives table and range from 0.64 to 0.99 s.d. at 8 years old, to 0.71 to 0.88 s.d. at 15 years old. I do not know whether changes from 8 to 15 years old are statistically significant, as Das et al. (2022) do not report confidence intervals.

By addressing RQ1 I add to this literature by reporting SES gaps in achievement from 5 to 15 years old, the longest age range reported. I also report gaps based on three measures of SES (maternal education, the asset index and expenditure) and use maths and vocabulary scores as measures of achievement. These allow for a critical discussion on our choice of SES and achievement measures. In considering each country's context when interpreting my results (something only Schady et al., 2015 do) I am able to provide some plausible rationales for my findings.

1.4 Substantive findings: early childhood education and achievement

1.4.1 High-income countries

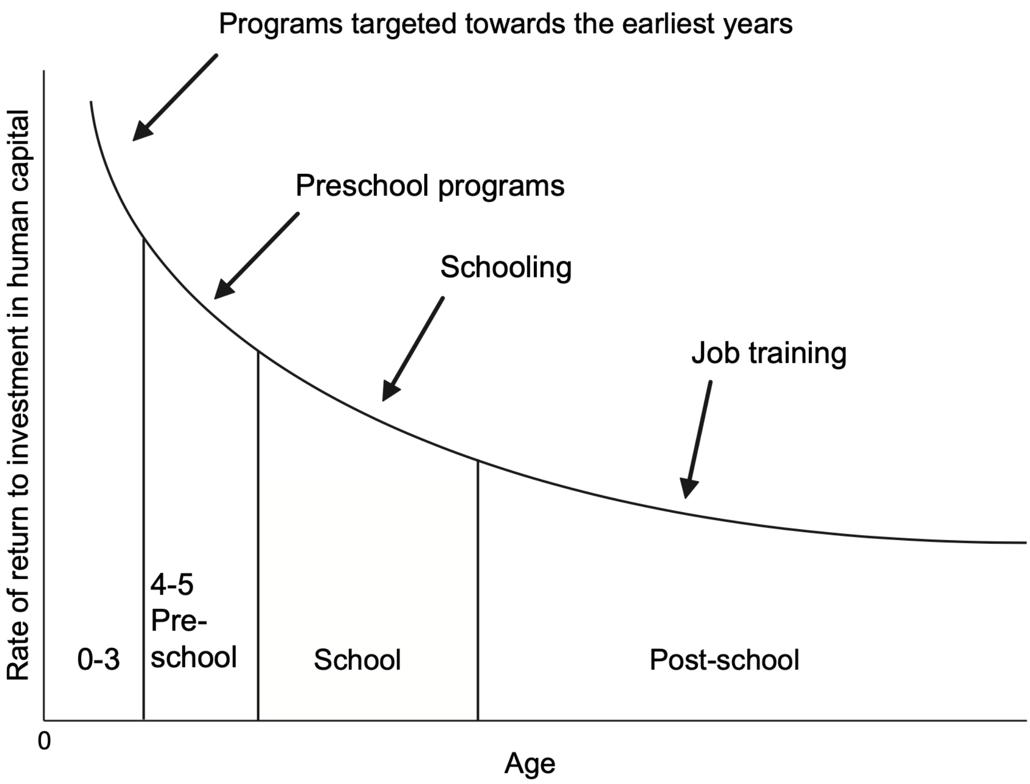
ECE's effect on children's achievement scores is crucial for policy makers as evidence shows that achievement scores are directly correlated to children's opportunities in later life. For example, the Institute for Fiscal Studies (IFS) examines UK data and find that one standard deviation increase in achievement scores at age 5 years old is associated with a 12 percentage point rise in the probability of obtaining a university degree and a £54 rise in weekly earnings at the age of 42 years (Cattan et al., 2022).

The argument in the literature that ECE pays off in the long run is best summarised in what has come to be known as the 'Heckman Curve', depicted in Figure 1.1. In his article, Heckman (2008) argues that returns to investment in human capital are greatest at early ages, and diminish as children grow older - hence the downward slope of the curve. Engle et al. (2011) and Magnuson and Duncan (2016) offer similar evidence for the USA. They write "[o]ur analysis indicates that increasing enrolments for preschoolers in the year before school entry is likely to be a worthy investment that will yield economic payoffs in the form of increased adult earnings" (Magnuson & Duncan, 2016, p. 123). Rea and Burton (2020) use evidence from the USA to refute this argument and contend there are high returns on investment throughout a child's life rather than simply in the early years.

ECE is seen as an opportunity to both address socioeconomic disparities at their source (as Heckman, 2008 writes) and to increase the earnings of individuals in one's society (i.e. increase productivity). Though, to enjoy these benefits, ECE has to be high quality. Blanden et al. (2022) find that the effect on achievement scores of an additional term of ECE in England is substantially larger for 5 year old children attending an institution with the highest quality rating (Engle et al., 2011; Ruhm and Waldfogel, 2012; Vandell et al., 2010 offer similar evidence).

The literature discusses various aspects of quality in early education. An OECD report by Edwards (2021) distinguishes between structural and process quality in early

Figure 1.1: The Heckman Curve.



Source: Heckman (2008)
Notes: Author's title: "Return to a Unit Dollar Invested at Different Ages from the Perspective of the Beginning of Life, Assuming One Dollar Initially Invested at Each Age" (Heckman, 2008, p. 311).

education. The structural dimensions of quality include child-to-adult ratios, space requirements, teacher's qualifications and an age appropriate curriculum. Process quality has to do with children's interactions and experiences in the educational institution. This can include with peers, adults, materials and other resources.

Causal evidence is somewhat mixed. Earlier studies, such as the Abecedarian Program (1970s) and the High/Scope Perry Preschool Program (1960s) demonstrate short and long term benefits to disadvantaged children (see Campbell et al., 2010; Heckman et al., 2010). Both programmes were small (n <120) and well funded, offering high-quality ECE to disadvantaged children. (The latter targeted not only low-income children but also those with lower than average IQs, which may have produced higher effect sizes.)

More recent experimental and quasi-experimental studies of large-scale ECE programmes do not provide definitive results. These cover broader populations of children, rather than just low SES children. While several studies show positive effects, for example Havnes and Mogstad (2011) in Norway, others do not. For example, the expansion of a state-funded ECE programme (referred to as a pre-kindergarten programme) in Tennessee (US) sees children who were randomly assigned to high-quality ECE obtain lower achievement tests scores in third grade (8-9 years old) and in sixth grade (11-12 years old), see Durkin et al. (2022). The introduction of universal ECE programmes in Denmark and in Quebec (Canada), also led to statistically significant declines in child outcomes, as is discussed in Kottelenberg and Lehrer (2019).

The mixed evidence regarding the effectiveness of ECE may reflect variation in the quality of the childcare in different contexts. In the case of Quebec, for example, negative effects in vocabulary test scores after the introduction of a universal child care subsidy, were explained by low-quality informal care, as discussed in Ruhm and Waldfogel (2012).

The mixed evidence may also be a product of the target population of the ECE programmes. In their review, Ruhm and Waldfogel (2012) conclude that the highest benefits from ECE tend to accrue to the most disadvantaged children. The Abecedarian and the High Scope/Perry Preschool programmes both targeted low-income children, while the large-scale ECE programmes cover the entire ECE aged population. When Kottelenberg and Lehrer (2019) re-examine the Quebec data they find that children from disadvantaged single parent households experience a boost in their outcomes (which included a vocabulary test for children aged 4-5 years old), but those from two-parent households experience a decline in parent-investment at home when they start attending ECE. This may explain their poorer outcomes. ECE, when implemented to a wider population of children, seems to have heterogeneous effects.

There is also some evidence that the achievement benefits of ECE fade over time. Blanden et al. (2022) find that the advantage of attending a high quality ECE institution fades out by 7 years old. Nores and Barnett (2010) also report fading as children move through childhood.

1.4.2 Low- and middle-income countries

Long-term evidence on the association between ECE and academic achievement in low- and middle-income countries is limited and relies heavily on Young Lives data. In their review article, Ruhm and Waldfogel (2012) identify only one article on the long-term effects of ECE in a non-high income country.

Engle et al. (2011) review the research on low- and middle-income countries, though the studies identified examine short-term associations and many are programme evaluation (i.e. grey literature). They draw similar conclusions to those drawn from high-income countries and find that benefits of receiving ECE are greatest for more disadvantaged children. Berlinski et al. (2009) also find that during an expansion of an ECE programme in Argentina the benefits of ECE on achievement tests are greater for children living in more disadvantaged municipalities where levels of poverty are higher.

The quality of education in ECE is also important. Evidence shows that being enrolled in high quality programmes is associated with better learning outcomes (as discussed in Engle et al., 2011, p. 1342). An analysis of Young Lives data from Peru finds that the attending higher quality ECE (i.e. *CEIs*) produces a higher PPVT score at 12 years old compared with not attending ECE, while the benefit of attending a lower quality form of ECE (i.e. *PRONOEI*) is not significantly different to not attending ECE (see Cueto et al., 2016). Singh and Mukherjee (2018) obtain similar results for India (also using Young Lives data) and find that children who attend private institutions scored higher in maths at 12 years old.

Harking back to the Heckman curve, Engle et al. (2011) use data from 73 countries to make the case that early childhood is the most "effective and cost-efficient time to ensure that all children develop their full potential" (Engle et al., 2011, p. 1339) and report that increasing ECE to 25 per cent enrolment in all countries would offer a benefit of $10.6 billion, and increasing enrolment to 50 per cent would offer a benefit of $33·7 billion.

Engle et al. (2011) also calculate the median effect sizes for ECE interventions which averages 0.24 s.d. (with a range of –0.14 to 1.68 s.d.). Berlinski et al. (2009) examine the impact of a large expansion of an ECE programme in Argentina, find that attending ECE increases their third-grade (8-9 year olds) achievement scores by 0.23 s.d..

Cross-national evidence is sparse in this area of research, due to limitations in data. I have identified only four cross-country comparative studies from low- and middle-income countries, namely Malmberg et al. (2011), Montie et al. (2006), Ojala and Talts (2007), and Rao et al. (2019).

Malmberg et al. (2011) is a programme evaluation where authors investigate the association between ECE and achievement for children attending the Madrasa Early Childhood Development programme in Kenya, Zanzibar and Uganda. Their sample size is relatively small, with 321 children across the three countries. Average ages for each round of data collected were 4, 6 and 7 years. Using a multi-level model they find a attendance in this for of ECE "showed a beneficial effect" Malmberg et al. (2011, p. 131) on children's achievement scores.

Montie et al. (2006) compare across ten countries (some high-income some low- and middle-income) how aspects of ECE at 4 years old are related to their achievement scores at age 7. This study also uses project's data (i.e. International Evaluation Association (IEA) Pre-primary Project data). Using a multi-level model they find that children's outcomes at 7 years old are higher if the quality of the ECE centre is high (e.g. measured by the level of education of the teacher among other aspects of ECE education).

Ojala and Talts (2007) compare the learning outcomes of two distinct ECE systems, one in a city of a high-income country, Helsinki, Finland (n=263), to one in city in a middle-income country, Tallinn, Estonia (n=163). After providing a detailed discussion of each country's ECE systems, they conclude there are no significant differences in the association between each country's ECE and children's language and mathematics skills, but learning achievements in science and the environment were higher in Estonia. These differences, authors explain, are a result of differences in cultural background for preschool education, the nature and role of preschool curriculum in society.

In their comparative study of four East Asian and Pacific countries using cross-sectional data, Rao et al. (2019) find that children (aged 36 to 71 months) who receive ECE have significantly better achievement scores than those who do not, reporting a difference of a third of a standard deviation. They also find that higher dosage (longer duration in ECE) is associated with increased achievement scores.

This study examines the long-term association between ECE and achievement, though to 15 years old, longer than the above studies. As the Young Lives data includes a broad sample of children, this study also offers a wider view on the association between ECE and achievement than one of a specific programme evaluation or that of a specific city.

1.5 Thesis plan

The relationship between initial parental SES and children's achievement has direct and indirect mechanisms. In Chapter 2 I present my analytical framework that draws on Haveman and Wolfe (1995) and Ermisch, Jäntti, et al. (2012). I use Ermisch, Jäntti, et al. (2012) to incorporate a longitudinal and cross-country approach into the Haveman and Wolfe (1995) framework. Based on this new framework I present the generic specifications I use in my analysis. I then go on to justify my selection of outcomes and predictors included in the regression. Although vocabulary and maths scores are available in the Young Lives dataset, I rely mostly on maths scores. These are a more complete measure of achievement and there are concerns with vocabulary scores that I discuss later.

The measures of SES I use are maternal education, an asset index, and expenditure per capita. As there is sparse research on the association between parental SES and achievement in low- and middle-income countries, I draw on the literature on the association between SES and children's health outcomes. In that literature, maternal education emerges as the strongest predictor of outcomes, compared to other measures of SES and father's education, see Alderman and Headey (2017) and Wamani et al. (2004). For example a systematic review by Balaj et al. (2021) shows that maternal education is a stronger predictor of reduced under-five mortality compared to paternal education.

Maternal education is also the most intuitive of the three measures to interpret, the most concrete and is a more stable (i.e. does not fluctuate annually or seasonally) measure of SES than expenditure. As academic achievement is shaped by a families long term SES, rather than variations therein, a more stable measure of SES is most practical. An intuitive and concrete measure of SES is especially useful in cross-national comparisons as they most closely reflect the reality in each country. The asset index and expenditure are both 'modelled', in that they are calculated using assumptions that complicate their interpretation and abstract them from reality.

Each of the three measures have distinct mechanisms through which they influence achievement. For example, higher mother's education may facilitate a mother's ability to navigate the health and educational institutions and ensure better educational outcomes for her child (as discussed in Mensch et al., 2019). Dräger et al. (2023) document studies where higher parental education is associated with more stimulating interactions between children and parents that involve more complex conversations, and richer vocabulary compared to those with lower education. More educated parents also provide better quality instruction and have higher expectations for their children's achievement, as discussed in Davis-Kean (2005).

Children in households with higher asset indices are more likely to live in an environment that is conducive to learning and have safer sources of water and therefore lower risk of illness, as discussed in Chowa et al. (2010). Families with higher expenditure have the resources to invest not only in children's basic needs but also in learning materials, activities and institutions (i.e. schooling) as well as live in neighbourhoods that support children's educational development, as discussed by Dräger et al. (2023).

Chapter 3 outlines my rationale and approach to conduct cross-national comparisons. Bryan and Jenkins (2016) outline various regression modelling approaches applied to multi-country datasets. In this thesis, I use the most common approach employed when investigating the association between parental SES and children's achievement – that is, a separate regression is fit to the data for each country. This means that 'country effects' (which represent the variation in the outcome variable explained by country-level variables) cannot be identified, rather they are absorbed into the intercept of each country's model, and every regression parameter is country-specific. Those who have used this approach include Blanden et al. (2012), Bradbury et al. (2012, 2015, 2019), Dräger et al. (2023), Linberg et al. (2019), Lopez-Boo (2016), Magnuson et al. (2012), and Schady et al. (2015). I then report relevant parameters in a results table or figure and use each country's contextual information to interpret estimates to gain a fuller understanding of what may explain the findings across countries.

To address my research questions I use the Young Lives dataset, a longitudinal data set that is designed for cross-country comparisons, specifically I use the data from the younger cohort. In Chapter 4 I critically evaluate the Young Lives dataset.

Young Lives is the only dataset designed for cross-country comparison among low- and middle-income countries. While there have been attempts at cross country comparisons, such as the Consortium of Health-Orientated Research in Transitioning Societies (COHORTS) and the Development of Inequalities in Child Educational Achievement: A Six Country Study (DICE) collaborations, these are cross-national data sets compiled after each country's surveys have been administered. This complicates analysis as the aims of the surveys, survey questionnaires, dates of data collection, age of the target population and follow up dates all differ across countries. For example, the COHORTS team invested an enormous amount of technical work in pooling data across the five studies which had five different aims, instruments, ages at which children were born and at which they were followed up. After that work, the scope for analysis was quite limited. In terms of attrition, Young Lives has comparatively low attrition rates compared to other cohort studies.

The Young Lives study collects a rich array of information on children's development, such as parental SES, children's achievement scores from 5 to 15 years old, whether the child attended ECE and additional controls (e.g. height-for-age, ethnicity/caste, maternal employment, household size, gender and the language of the achievement test).

Young Lives' sampling frames were designed to balance the aims of the study and costs of a cross-country longitudinal study. Following up with children over a period of 15 years is a costly enterprise, additionally the Young Lives study requires a central coordinating team to standardise instruments, which is based at the University of Oxford. For the younger cohort (whose data I analyse) they decided to purposefully sample 20 sites (or clusters) in each country and randomly sample 100 children within each of these sites. The first 20 sites were selected to capture each country's regional, geographic, ethnic and other diversity. The Peruvian team randomly selected the 20 sites from a sampling frame of all the districts, barring the wealthiest 5 per cent of them. An implication of this sampling design is that I am limited to a specific population of inference in each country. I cannot make inferences about the national population.

The Young Lives team was tasked with the challenge of implementing an age appropriate instrument that would capture the variation in maths skills across the sample populations, when children were 5, 8, 12 and 15 across the four countries and in numerous different languages. For the most part, the Young Lives team drew on already established maths assessments (i.e. Cognitive Development Assessment (CDA), Trends in International Mathematics and Science Study (TIMMS) and Programme for International Student Assessment (PISA)) and adapted them to each countries context. They piloted the tests prior to implementation and revised them accordingly. For the most part, they were able to capture the breadth of the variation in maths scores (i.e. there wasn't truncation at the top or bottom of the distribution).

With the vocabulary tests, the Young Lives team faced a similar set of challenges. The used a previously tested Test de Vocabulario en Imagenes Peabody (TVIP) (the Spanish version of Peabody Picture Vocabulary Test, PPVT) in Peru, which worked well. However, they administered a translated PPVT test, that was designed for an English speaking American population, to an Ethiopian, Indian and Vietnamese sample. This came with its challenges in terms of the accuracy of the test at 5 and 8 years old. In Rounds 4 and 5 (when the children were 12 and 15 years old) the test was amended accordingly.

In Chapter 5 I compare the the four study country's level of development, societal dividing lines, ECE systems and discuss the implication of these for analysis.

The main contextual differences across study countries are as follows. When Young Lives started collecting data, in 2001, Ethiopia, India and Vietnam were low-income countries, with Peru was the only middle-income country (as per the World Bank classification). Ethiopia and India exhibited worse human development indicators than Peru in terms of under-five-mortality (U5MR), rates of stunting and net enrolment ratios (NER) in primary school. Vietnam on the other hand, despite being substantially poorer than Peru, had better rates of U5MR and comparable NER in primary education. This difference is partly explained by political will of the Vietnamese government, after the war that ended in 1975, to improve the living conditions of all Vietnamese. Ethiopia had on average the lowest levels of education among the four study countries, and Vietnam the highest.

Important societal dividing lines to consider in Ethiopia are ethnic differences, which are correlated with access to infrastructure and public services, and there is evidence of discrimination based on gender (see UN, 2018). In India, gender and caste represent important hurdles (see IIPS & ORC Macro, 2000). In Peru differences between indigenous and non-indigenous populations are large, and these correlate with regional divides, with most indigenous populations living in the mountainous Andes region and the Amazon basin (see Barrón, 2008). Similarly, in Vietnam, divisions by ethnicity are evident, mostly between the majority Kinh and Hoa people and the ethnic minorities (see Baulch et al., 2007).

In terms of the ECE systems in each country, provision in Ethiopia catered to the wealthy urban population, as no universal public provision was in place when Young Lives children were 5 years old (UNESCO, 2008). In India, the ECE system offered poor-quality ECE across the board (except for a handful of elite ECE centres that the Young Lives children did not attend) (UNESCO, 2010). Peru and Vietnam ECE attendance rates were high and the ECE system was stratified by SES, with wealthier urban families having access to better quality education, compared with poorer families and those living in remote areas (see section 5.3 for Vietnam and Cueto et al. (2016) for Peru).

In Chapter 6 I discuss in detail how academic achievement, parental SES, ECE and control variables are measured. In Chapter 7 I address the research question RQ1. In Chapter 8 I address the research question RQ2. In Chapter 9 I address the research question RQ3.

To address RQ1 I use a regression approach to estimate the average scores for the high and low SES groups at each age in each country. I then subtract the average achievement score of the top quartile group from that of the bottom quartile group to calculate the SES gap in achievement. RQ2 I address by grouping children into initially high- and low- achieving groups at 5 years old. I then split those two groups into high and low SES groups, resulting in four groups. I then plot each group's average achievement score from 5 to 15 years old. To address RQ3 I predict the maths scores of children who did and did not attend ECE while controlling for SES, child and household characteristics and community level fixed effects.

Chapter 10 concludes with reflections on findings, implications for policy and research and suggestions for further research.

Chapter 2

Analytical Framework

2.1 Introduction

This chapter outlines the analytical framework underpinning my research. I incorporate a longitudinal and cross-country approach, drawing on Ermisch, Jäntti, et al. (2012), into the framework of Haveman and Wolfe (1995). Drawing on this proposed framework, I specify a generic regression that I in later chapters tailor to each research question and fit separately for each country.

2.2 Proposed framework

A micro-level view: determinants of children's educational achievement

This section outlines a micro-level view of the determinants of children's educational achievement, where as a macro-view would include structural determinants at the country level. Figure 2.1 sets out the main determinants of a child's educational achievement and is essentially the framework of Haveman and Wolfe (1995), modified to suit my research questions.

In line with many, including Cunha and Heckman (2007), I conceive of test scores not as pure measures of ability, but rather as affected by a number of other factors including parental socioeconomic status (SES), early childhood education (ECE), household characteristics and community and country influences (labelled in Fig 2.1 as *Contextual Characteristics*).

Child's 'ability' is a contested concept as it is not observed. Here I conceive of a child's 'ability' as scores as a result of child's genetic endowment, the environment she grew up and how these interact with each other. The genetic component, labelled *Heredity* in Figure 2.1, is inherited from the parents but is not necessarily deterministic

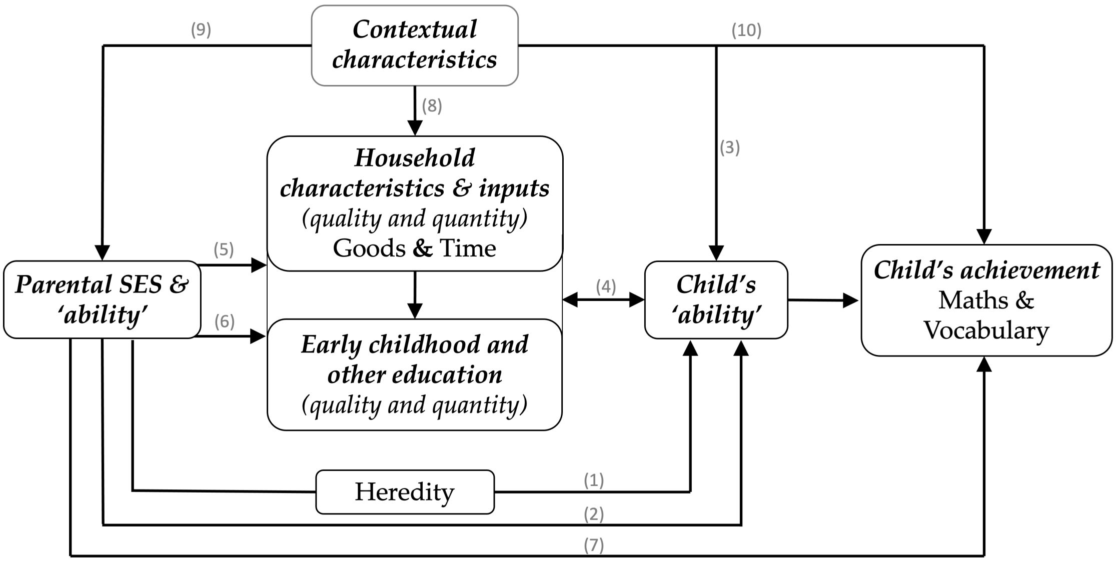


Figure 2.1: Analytical framework: The determinants of a child's achievement

Source: Adapted from Haveman and Wolfe (1995)

of one's academic abilities nor is it measurable (in my data set). Heredity, through the arrow labelled (1) shapes a child's ability.

The environmental component is comprised of parental SES, (arrow 2), which includes household characteristics, household inputs, the child's education (arrow 4) and the context in which she grows up (e.g. high pollution levels can affect brain development, arrow 3). Research on the interaction between genetics and the environment one grows up in shows that these two components interact, with certain genes manifesting themselves in particular environments, for example in a twin-study Rhemtulla and Tucker-Drob (2012) find that SES impacts on a child's cognitive development.

The bidirectional arrow, labelled (4), between the *child's ability* box and the *Household inputs* and *ECE and other education* boxes indicates that household characteristics and inputs and children's education not only determine the child's 'ability', but also that the child's ability also shapes the parents' and the child's choices in this regard. Parents with constrained resources choose to maximise their returns on investment by enrolling their more able child in ECE education or children with higher ability would be more likely to remain longer in schooling. For a more detailed discussion see Akresh et al., 2010. Other ways in which children are shaped by their own academic achievements include investing their free time in enhancing their skills, whereas a child who knows she will be married young would be much less inclined to spend her energy on academic endeavours.

Both household characteristics and inputs during childhood and education are determined by parental SES (arrows 5 and 6). A more educated mother (or parent) invests more time in her children's education and also encourages her children to attend school, as discussed in Davis-Kean (2005). More educated women also tend to have fewer children and to have them later in life, compared with women with lower education. See Alderman and Headey (2017) for evidence on low- and middle-income countries. The number of siblings in the household reduces the amount of resources available per child and also shape the domestic responsibilities that other children have. A wealthier household will be able to afford better quality goods and schooling. Children from poor households are less likely to receive sufficient or balanced nutrition, leading to stunting which impairs their ability, as discussed by Grantham-McGregor et al. (2007).

Direct links between parental SES and achievement (arrow 7) can be explained by the conditions under which the child takes the achievement test. For example, consider two children with the same ability, but from different socioeconomic backgrounds. The child from a wealthier household is more likely to be better rested if she has not had to care for other children, has not had to complete numerous household chores before taking the test, has eaten breakfast that day, has come from a more stable household, and has been better able to focus in general if her burden of responsibilities is less at home.

Parental SES shapes household inputs and education, but also plays a direct role on the child's 'ability' through softer influences (arrow 7). More educated or wealthier parents tend to have wider vocabularies than their less educated counterparts, thus moulding their children's vocabulary accordingly (see Dräger et al. (2023) and Tough (2012)). Given their education, they may also be better equipped to assist children in the learning at home, through help with homework, for example. As de Neubourg et al. (2017) suggest, they are more likely to be under less stress, freeing more cognitive resources to formulate objectives for their children's development, to set out plans to achieve this, and to execute them. They are likely to project different aspirations on their children, which shapes children's ambitions and decisions around learning, as discussed in Davis-Kean (2005).

Contextual characteristics include relevant community- and country-level institutions and investments. Haveman and Wolfe (1995) did not consider community characteristics in their schematic, hence the inclusion of the *Contextual characteristics* box. Household inputs and education are also shaped by community characteristics (arrow 8). Not all communities have their own schools, and parents may be reluctant to send their daughter to a distant school. Without a bookshop, books (goods input) and other school supplies will be harder to come by. The broader context also shapes children's achievement (arrow 10). For example, a community where returns to education are low discourages parents from investing in education, see Attanasio and

Kaufmann (2014).

The characteristics of the household's community and country are determinants of not only ability (along with the household inputs and the education system); they also shape the parent's SES and household characteristics. A largely agrarian economy limits the amount of wealth that a household can spend on education. A remote community with little access to education will influence the parent's level of education. All else being equal, policy makers who prioritise ECE will shape the quality and quantity of provision differently to those in a country with different macroeconomic circumstances.

Haveman and Wolfe (1995) did not design their framework for cross-country comparisons, and thus do not offer guidance on how to account for variation between countries. They do, nonetheless, complement their schematic with a "comprehensive economic perspective on children's achievement" (1995, p. 1838) that considers choices made by society or the government (termed *social investment in children*). Ermisch, Jäntti, et al. (2012) do offer concrete guidance on how to consider important variation between countries through an '*Investments and Institutions*' component that shapes the relationship between parental SES and children's achievement. Investments include the provision of public and private ECE programmes, that of primary school education among others. The institutional context might include processes such as how ECE is organised and regulated (if at all), how certain groups are systematically excluded from education or economic opportunities, differences between the cost of private and public education, and more.

Variation in these investments and institutions helps to explain differences or similarities in estimates (when addressing RQ1-RQ3) across countries. For instance, labour market institutions (*Institutions*) vary substantially among countries, independently shaping parents' ability to accumulate wealth. Or the quality and quantity of ECE investments (*Investments*) might also vary between countries, contributing to an explanation for why the association between ECE and achievement of five years olds differs between them. The investments and institutional context of each country will be discussed in depth in Chapter 5.

The implications for my research of not observing 'ability' or other unmeasured or immeasurable characteristics that are positively correlated with both SES and achievement is that estimates will be biased upwards in a regression that omits these biases. I acknowledge this when discussing the results in my analysis.

Figure 2.1 is a schematic for a single child, and child-specific characteristics (such as such as birth order, gender and disabilities) will also affect the relationships represented in the schematic. For example, first born children is likely to have caring or financial responsibilities for younger siblings, or a disabled child in a community without suitable school support will struggle to perform well at school.

A longitudinal view: Determinants of children's educational achievement

In Figure 2.1 the longitudinal aspects are implicit. For example, the content of the boxes will vary according to the life-course stage the framework is applied to. In the *ECE and other education* box, for example, when the framework is applied to a child in early childhood, the box will comprise only ECE, while for an 11 year old it will comprise ECE and subsequent primary school education. For my research, I make the longitudinal aspect of intergenerational status transmission explicit. To do this, I draw on Ermisch, Jäntti, et al. (2012), again adapting their schematic for the purposes of my research, as depicted in Figure 2.2. While this framework can be used to consider parental SES through out childhood, in this thesis I focus on initial parental SES.

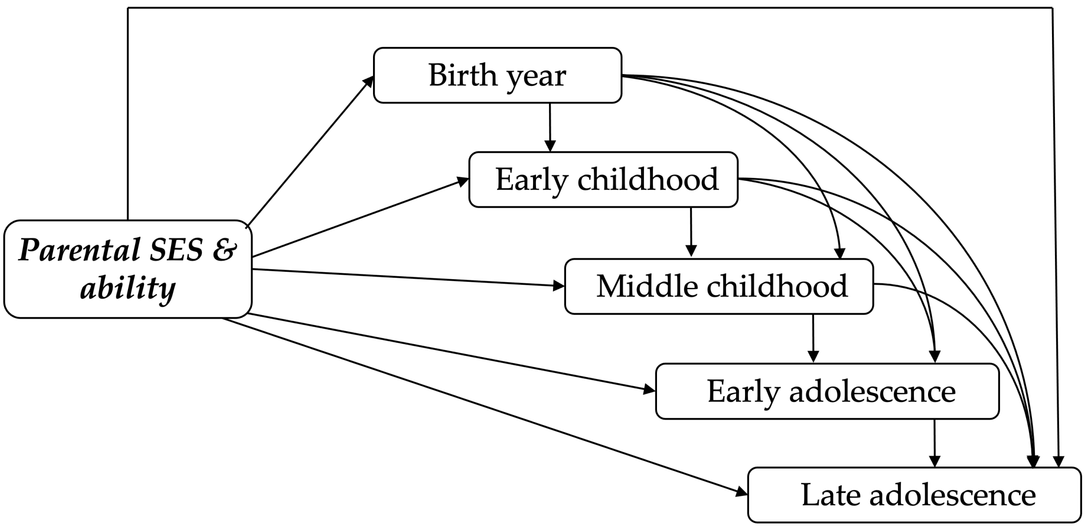


Figure 2.2: Analytical framework: A longitudinal view of the determinants of achievement

Source: Adapted from Ermisch, Jäntti, et al. (2012)

While the Ermisch, Jäntti, et al. (2012) framework includes all life stages from birth to adulthood, I examine outcomes only to adolescence. I also split adolescence into early and late adolescence, as children make important transitions in adolescence. Especially in low- and middle-income countries, children tend to marry, start a family, and begin to work or migrate at earlier ages than in high-income countries. Parental SES plays different roles on children before and after they enter the workforce or marry. I also exclude the original *Institutional and Investment* component, as I have integrated it into the Haveman and Wolfe (1995) framework.

Unlike Ermisch, Jäntti, et al. (2012), I make explicit the independent roles that previous life stages exert on subsequent stages in life (something that Ermisch, Jäntti, et al. (2012) only do with their last life stage, namely adulthood), indicated by the curved lines. An outcome at an earlier life stage, such as a financial shock, independently affects later outcomes into late adolescence. For example, a severely under-nourished baby will have impaired brain development, rendering her less able to focus in primary school and in middle childhood. In adolescence, she is more likely to be shorter and meeker, on average. In late adolescence, she is more likely to drop out of formal scooling.

Figure 2.2 depicts the different points in life that are crucial to understanding how advantage is transmitted from parents to their children. The framework starts with parental SES and 'ability'. The following boxes depict child outcomes at different stages over their life course, starting with their birth year (up to 1 year), then early childhood (ages 2 through 6), middle childhood (ages 7 to 11), early adolescence (ages twelve to fourteen) and late adolescence (ages 15 to 17). These correspond to the Young Lives data I use.

The straight arrows pointing downwards from one stage to the next reflect the cumulative nature of advantage or disadvantage. For example, a child who attends ECE in early childhood may be better prepared for primary schooling in middle childhood, or completing secondary education in early adolescence may reduce the probability of early marriage or childbearing in late adolescence.

To link these two frameworks, both frameworks include a box labelled *Parental SES & 'ability'*. These are shared across both frameworks. The rest of the framework to the right of this *Parental SES & ability* box Figure 2.1 is reduced to a single box in Figure 2.2. The content of these boxes depends on the life stage. For example, only ECE is relevant for early childhood, and primary school for middle childhood. (For an illustration of how the two frameworks are linked, see Figure A.1 in Annex A.)

2.3 Application to low- and middle-income countries

When applying these frameworks to low- and middle-income countries key components will differ to those in high-income countries.

The first is undernutrition, which is more prevalent in low- and middle-income countries. There is a large body of literature examining the links between parental wealth and childhood undernutrition in low-income countries, see Petrou and Kupek (2010) for a review, and on the impact of child undernutrition on subsequent life outcomes (see Alderman et al., 2003; Victora, Adair, et al., 2008). In a review of the research in low-income countries, Grantham-McGregor et al. (2007) find a consistent relationship between stunting in early childhood and poor child development, with moderate to large effect sizes (0.4 to 1.05 SD). Undernutrition is a vital determinant of achievement.

The contextual characteristics and education provision will also differ. Low- and middle-income countries are also poorer than high-income countries, so infrastructure available is therefore also generally poorer, with remote areas having little or no access to public services or education (see Chapter 5 for details among the Young Lives countries). The quality and quality of the education will differ, on average, given limited resources governments and the population's level of education. Though these can also offer an advantage to achievement. For example, UNESCO (2010) "found that even in programmes which would be considered to be of low to mediocre quality using western benchmarks for quality, preschool quality was positively associated with child developmental outcomes controlling for potential confounding variables" (UNESCO, 2010, pp.41-42).

2.4 A generic regression model of achievement outcomes

These frameworks illustrate the complexity of the relationship between initial parental SES and children's achievement. Isolating these links is challenging from a methodological and data quality perspective (see discussion of both aspects in Chapter 9). When I address RQ1, I am concerned with raw SES gap in achievement, one that represents the SES gap in achievement without accounting for other predictors of achievement. We can interpret this gap as including, in addition to parental SES, variables that are measured (but not included in the regression, such as attendance in ECE), unmeasured (e.g. quality of schooling) and immeasurable (e.g. heredity) predictors of achievement associated with SES that contribute to academic achievement. For example, if having high SES also means a child is more likely to live in neighbourhood that is more conducive to learning, a raw gap would not identify this link separately, as it does not take neighbourhood into account. My aim in Chapter 7 is descriptive, to document these raw SES achievement at each age across countries, to set the foundation for further investigations into the causes and consequences of these gaps. Such evidence for low- and middle-income countries is limited.

To estimate SES gaps in achievement at each age (5, 8, 12 and 15 years old), I apply this generic regression model to each country, separately. In Chapter 7 I adapt the generic equation to address methodological challenges and specific aspects of data collection. For each child at age a my generic specification is:

$$Y_a = \beta_0 + \beta_1 S + \epsilon \tag{2.1}$$

where

Y = educational achievement score for each child measured at age a,

S = initial parental SES as a categorical variable,

ϵ = error term.

I group SES into categories (e.g. quartiles) and set the bottom SES group as the reference group. Therefore the coefficient on the top SES group represents the size of the raw achievement gap (between the top and the bottom SES groups).

Chapter 8 considers whether the achievement trajectories of children who were initially high (or low) achievers (i.e. scores well or not in an achievement test) differs by SES. Rather than using a regression model, I draw on the work of Feinstein (2003), who groups children into four groups by initial high and low SES and initial high and low educational achievement and plots how their relative position changes over childhood. I use the approach suggested by Jerrim and Vignoles (2013) to address regression to the mean.

The factors and pathways depicted in Figure 2.1 can help to understand the size and persistence (or not) of these gaps. A widening or persistent SES gap in achievement is usually framed in the literature as a result of cumulative effects, as is discussed in Duncan and Magnuson (2011a), Duncan et al. (2014) and Tough (2012), for example. The central claim is that advantage in one child accumulates over time. Consequently, an individual disadvantaged early on in life will have difficulty catching up with the rest. This is demonstrated by Feinstein (2003) who, using UK data, shows that progress in education achievement is stratified by initial parental SES.

The factors and pathways depicted in Figure 2.1 can help to understand the size and persistence (or not) of these gaps. A widening or persistent socioeconomic gap in achievement is usually framed in the literature as a result of cumulative effects, as is discussed in Duncan and Magnuson (2011a), Duncan et al. (2014) and Tough (2012), for example. The central claim is that advantage in one child accumulates over time. Consequently, an individual disadvantaged early on in life will have difficulty catching up with the rest. This is demonstrated by Feinstein (2003) who, using UK data, shows that progress in education achievement is stratified by initial parental SES.

In Chapter 9, I examine whether ECE attendance is associated with gains in achievement across childhood. Within methodological and data constraints, I try to isolate the pathway from the ‘ECE education’ box to the ‘Child’s achievement’ box in Figure 2.1.

The empirical approach to this research question involves two key challenges: one is data related; the other, methodological. The available data sets that include data from several low- and middle-income countries, including real-time information on ECE attendance and achievement scores into adolescence, are sparse; I have come across only one, the Young Lives data set. Second, methodological challenges include feedback mechanisms, omitted variables, selection effects into ECE, and measurement error. Some aspects in the relationship between ECE and achievement may not be observed and therefore omitted from the analysis.

A wide range of econometric approaches are available to address these challenges. Which method to use depends on the nature of the outcome variable, the time period covered, and the research questions asked. Jenkins and Siedler (2007) discuss key data requirements and approaches to addressing these challenges in observational data for low- and middle-income countries. Nonetheless, with observational data, these challenges tend to persist despite careful specification and high-quality data, and results remain somewhat biased. I caveat the interpretation of my results taking these challenges into account.

My interpretation acknowledges both that I am unable to fully isolate the relationship between ECE and achievement, and that, given the data and methodological challenges, this is the most accurate estimate of the relationship that I can produce.

To account for factors that may confound the relationship between receiving ECE and achievement, I conduct an Ordinary Least Squared (OLS) regression which controls for such child- and household-level variables associated with achievement, while accounting for variation between communities. The regression model I use for each Young Lives child at age a is as follows:

$$Y_a = \beta_0 + \beta_1 P_5 + \beta_2 \mathbf{Z}_1 + \epsilon_a \tag{2.2}$$

where

Y = achievement score at age a

P = receiving early childhood education at 5 years old

$\mathbf{Z}$ = vector of initial conditions; child-, household- and community-level

ϵ = error term.

Because each Young Lives household includes exactly one study child of a specific age, for simplicity I omit the child (i) and household (j) subscripts.

The estimates on ECE reflect the size of the association between ECE and achievement, over and above parental SES, other individual-, household- and community-level variables.

2.5 The achievement variables and its predictors

In this section I provide a conceptual justification for the variables selected in my analyses. I do so by discussing the potential mechanisms linking the boxes in the framework discussed above.

Outcome variables: maths and vocabulary scores

Maths and vocabulary scores

The dearth of comparative research on low- and middle-income countries on the topic of socioeconomic achievement gaps is largely due to the lack of good quality data. For example, Grantham-McGregor et al. (2007) is a prominent cross-country comparison examining children's development. Achievement scores were not available to them. Instead they used stunting as a proxy for cognitive development. In contrast, Young Lives, collects data on achievement scores. See the Measures Chapter 6 for more details on how maths and vocabulary skills are measured.

Maths and vocabulary scores are the two achievement measures available in the Young Lives data for children aged 5 to 15 years old. Although the Young Lives team also collected information on reading and writing skills and executive functioning, only the maths and vocabulary scores were collected from children 5 to 15 years old . See Figure 4.6 for other tests administered only at some ages.

There is evidence that vocabulary skills are a necessary foundation for maths skills development (see Purpura et al. (2017) for references). However, there is also evidence that early maths skills are a powerful predictor of later literacy skills, more so than early literacy skills (see Duncan et al. (2007) for a meta-analysis of six longitudinal data sets from high-income countries to examine which school-entry skills are related to later reading and maths achievement). Purpura et al. (2017) explore how and why early mathematics skills predict early literacy development. These researchers argue that previous studies, as well as their own, use measures for early maths skills that may well be proxies for language skills or deeper knowledge of language in general.

Although researchers continue to explore how aspects of language skills shape maths skills, and vice versa, there is a general consensus that maths and vocabulary skills develop iteratively, not separately from each other, building on each other and other cognitive skills. See for example Cameron et al. (2019), Claessens et al. (2009), and Duncan et al. (2007). Duncan et al. (2007) conceive of maths skills as involving both conceptual and procedural competencies. They find that maths skills offer a more powerful predictor of later reading achievement than do early reading skills.

I therefore conceive of maths and vocabulary scores as different measures of educational achievement more broadly, that develop in an iterative process (maths skills building on vocabulary skills and vice versa). The Peabody Picture Vocabulary Test (PPVT) tests administered by Young Lives test receptive vocabulary and require that a child to identify an image (out of four) that represents a specific word. The maths tests not only test maths skills but also a host of other academic skills, including reading comprehension. For example, the maths questions require not only an understanding of the vocabulary in the questions; they also require the child to evaluate a mathematical expression and integrate it into the meaning of the overall question. Take, for example the following two maths questions asked of children:

- At 5 years old: *Look at the pictures of the bowls and spoons. Point to the picture that shows a spoon in every bowl.*

- At 15 years old: *In a school there were 1200 students (boys and girls). A sample of 100 students was selected at random, and 45 boys were found in the sample. Which of these is most likely to be the number of boys in the school?*

  A. 540 *B. 600* *C. 500* *D. 450*

This broader set of academic skills being tests is the logic underpinning my rationale for considering maths scores a more complete measure of academic achievement than PPVT tests. I therefore rely on maths scores for my main findings and include results with vocabulary scores in the chapter annexes.

Additionally, a challenge – especially with vocabulary tests – is that they may reflect SES differences if the test designers are not careful. Take, for example the word ‘hurdle’. It might be more familiar to children from high-SES households because they are more likely to attend a school with resourced sports facilities (i.e., they might have hurdles) or watch sports events on television (which poorer children may not do). The same may also be true for a word that is quotidian among low-SES households and rarely used among high-SES households. While this may also be the case for some mathematical concepts and the vocabulary used in the word problems, Young Lives tests basic mathematical operations (i.e. addition, subtraction, multiplication and division). These concepts are foundational to maths and also concepts used in daily life they are less susceptible to bias by SES. The main risk of bias lies in the vocabulary words used in the word problems (when they are included in the tests), thought as the maths test tests more than just vocabulary (i.e. maths skills), this bias is likely to be mitigated.

Initial SES

SES refers to a household's socioeconomic position in society. The most common measures of SES in the literature on low- and middle-income countries are those that I use in this thesis: maternal education, an asset index, and expenditure. The asset index and expenditure capture aspects of a household's material wellbeing, with the asset index reflecting the long-run material wellbeing of the family, as it takes time to accumulate assets. Expenditure is a more volatile and a short-run measure of material wellbeing as it can change from season to season and year to year, particularly in informal economies. Just as with the measures used in high-income countries, the measures of SES that I use have shared and distinct mechanisms through which they shape achievement, as Duncan and Magnuson (2012) comprehensively explain. Some specific aspects of each measure of SES that contribute to achievement are as follows: a highly educated mother will be able to impart a wide vocabulary to her children, a wealthy parent with high levels of expenditure will be able to feed her children adequately, and if she has a high asset base (e.g. good quality housing, safe water and sanitation) the child will grow up in an environment that is more conducive to learning than if the child grew up in a family with a low asset base (e.g. low quality housing, unsafe water and sanitation).

From a policy perspective, conflating the components of SES is problematic. The policy levers designed to modify each aspect of SES measured (i.e. expenditure patterns, the quality of household assets and maternal education) are manifestly distinct. If we don't know how each aspect of SES is associated with achievement then we won't be able to design effective policy. Further, Dräger et al. (2023) argue that different dimensions of SES have distinct effects on children's outcomes, using one may underestimate the level of inequality, but also combing them renders it impossible to distinguish the effects of each.

I therefore examine each component of SES separately, rather than constructing a composite index based on three indices, i.e. operationalising SES as a unitary construct (as some researchers, such as Caro et al. (2009) and Das et al. (2022) do). For example, Ermisch, Jäntti, et al., 2012, who conduct cross-country comparisons of high-income countries, use parental education as a measure of 'permanent income'. They do so "because it is a measure of permanent income and because people with different educational qualifications face different labor markets with different rewards and opportunities and make different career-path choices (artist or banker)" (2012, p. 15). Not only is conflating education and income problematic as policies aimed to change each of these are distinct (for example, tax regime vs schooling infrastructure), income can also fluctuate substantially from year to year, and Duncan and Magnuson (2012) explain that the very concept of something that is fixed over time, such as *permanent income*, is elusive, as parental incomes tend to vary across children's lifetimes, particularly in early childhood and adolescence.

To better understand the emergence and persistence of socioeconomic gaps in achievement, I now discuss the mechanisms through which maternal education, an assets index and expenditure can influence children's achievement, drawing on high-, low-, and middle-income country literature.

Maternal education

There is a long established association between parental education and children's outcomes. Of the SES variables examined in the literature, parental education is usually the strongest predictor of children's outcomes. Even when other SES variables are included as predictors in a regression, the coefficient on parent's education remains significant. In their cross-country study across six high-income countries Waldfogel et al. (2023) start their article with "Parental education is one of the strongest predictors of children's life chances" (Waldfogel et al., 2023, p. 37) (authors focus on inequalities in resources for preschool aged children, by parental education). Conducting a cross-country study using Programme for International Student Assessment (PISA) data, Zhang and Lee (2011) found that parental educational is a stronger predictor of children's maths scores than wealth. Reardon (2011) examined data on the USA, covering the period 1943–2001, and found a widening socioeconomic achievement gap (over calendar time). Throughout this period, he found that parental education remained the more powerful predictor of children's achievement compared with than family income.

Duncan and Magnuson (2012) posited that the skills obtained through education may be those that parents draw on to organise their own time and resources efficiently, and thus to carry out their parenting goals effectively. Davis-Kean (2005) found that parental education is related indirectly to children's achievement through parents' expectations and their behaviour, specifically reading and the warmth of the parent–child relationship. (In this case, parental education was defined as the education level of the most educated parent in the household.) Elliott and Bachman (2018, p. 6) cite several studies showing that more educated parents provide a more stimulating home learning environment, as do Dräger et al. (2023, pp. 9–10). Brito and Noble (2014) explain that differences in the quantity and quality of linguistic stimulation at home are associated with differences the areas of the brain that support language.

This research from high-income countries, while rich, also relies on maternal education as the measure of SES. Independent of household income, maternal education can shape children's achievement in numerous ways. Kalil et al. (2012) showed that more educated mothers not only spend more time with their children, but are also more likely to alter the composition of that time to suit their children's developmental needs, than less educated mothers. For example, in the early years, they spend more time in basic care activities, while at later ages they spend more time managing their children's time (e.g. handling extracurricular activities and social engagements).

As the research on the association between parental SES and children's achievement in low- and middle-income countries is sparse, there are few reference points I can draw on to inform my decision on which measure of SES to use or rely on. I therefore draw on the research that examines the link between parental SES and health outcomes that shows that maternal education predicts children's health outcomes more strongly than father's education. (Underpinning these results is the idea that the mother invests more time and resources on her children, than the father tends to.) As a result, a significant portion of studies examining the association between SES and health outcomes rely on maternal education. Literature reviewed in Sabates and Di Cesare (2021) on the association between maternal education and child health shows that more educated mothers are more likely to seek out antenatal care services, have a health professional at delivery and to have fewer complications at birth, than mothers with low levels of education.

A systematic review by Balaj et al. (2021) shows that maternal education is a stronger predictor of reduced under-five mortality compared to paternal education. Wamani et al. (2004) found that mother's, but not father's, education (and not household assets or land ownership) is the strongest predictor of inequalities in child health in rural Uganda. Alderman and Headey (2017) used Demographic and Health Survey (DHS) data for 56 low- and middle-income countries to examine how parental education is associated with child nutrition. They also found that maternal education was the strongest predictor of children's health, over and above the father's education and the asset index. They also found that the mother's education is more strongly associated, than father's education, with the number of children in the family (i.e. fertility), the diversity in the children's diet, antenatal and post-natal care, and the mother's ability to participate in decisions about her own health care.

When it comes to interpreting the results of maternal education, Duflo (2012) suggests some additional mechanisms through which maternal education may affect children's outcomes, particularly in low- and middle-income countries. The main focus here are health outcomes. First, a woman's education may be correlated with unobserved aspects of her ability, family or community background. For example, in a highly patriarchal society, a woman who achieves a high education must have overcome several social barriers. Having a high education may have to do with her character, and the same traits that contributed to her education will contribute also to her children's development. In those cases, a high association between maternal education and children's outcomes may also have to do with the mother's character and her education. Second, assortative mating, where more educated women marry more educated men, shapes children's outcomes through unobservable aspects of the father (e.g. the husband may be more progressive and support his wife in her education). Duflo argues that both father's and mother's education matter for children's outcomes. When I include information on the father's education in one of my predictors my findings do not meaningfully change (see Annex F).

I have incorporated Duflo's two points into my analytical framework (arrow 9). A girl's education is likely to be correlated to her own, her family's and her community's characteristics and, when the girl becomes a woman, this correlation is likely to continue. Additionally, Duflo's first point (about higher education reflecting character traits that enhance child outcomes) is relevant mostly in contexts where girls face systemic challenges in their education (which is not the case in all study countries, as I discuss in Chapter 5).

Given the limited research on achievement in low- and middle-income countries I continue to rely on literature that focuses on health outcomes, including Duflo (2012).

Alderman and Headey (2017) distinguish between urgent and less-urgent outcomes. They find that mother's education (compared with father's education) is more strongly associated with less-urgent factors that promote the child's health such as the children's dietary diversity, antenatal and post-natal care, women's ability to participate in decisions about their own health care (whether by herself or jointly) and fertility. But there is no difference in the association between mother's and father's education when it comes to children's immediate survival (e.g. vaccination and breastfeeding in the first few months). To explain the findings of Breierova and Duflo (2004) (who found that, while households have fewer children when the mother is more educated, there is no difference in infant mortality) and Fafchamps and Shilpi (2015) (who found that mother's, compared with father's, education is not a stronger predictor of child mortality) and their own findings, Alderman and Headey (2017) speculated that "at any given education level, fathers care just as much about keeping their children alive as mothers do, but devote less attention to more mundane day-to-day child care practices like feeding children an appropriately diverse diet" (2017, p. 457).

Highly educated mothers may also be better able to support learning at home. Hoff (2003) found that more educated mothers use a richer vocabulary and more complex language structures, and their children do as well. Magnuson (2007) and Magnuson et al. (2009) found that mothers who increased their education when their children were around three years old provided their children with more learning materials and were more responsive, compared with mothers who did not increase their education. Young children with mothers who obtained more education (from a low level) also improved their vocabulary comprehension and language expression. Although the authors concluded that improvements in home environment explain some of the improvements in children's language development, they do not explain all of it. (Other mechanisms suggested by the authors are changes in mother–child verbal interactions.)

Establishing causal mechanisms between maternal education and achievement is challenging. However, some experimental and quasi-experimental studies cited by Duncan and Magnuson (2012) showed that improvements in maternal education led to improvements in children's academic achievement at young ages (preschool and 7–8 years old; not 12–13 years old). One problem with these experimental and quasi-experimental studies is that they only allow us to identify how a change in a mother's level of education is associated with children's outcomes after that change in education. A more pertinent question is whether higher *levels* of maternal education shape outcomes.

Asset index

Asset indices are a mainstay in research on disparities in children's outcomes in low- and middle-income countries. Unlike for high-income countries, an asset index is not a measure of financial assets, rather it is based on observable measures of housing quality (e.g. material of the roof), access to services (e.g. access to safe water), and ownership of consumer durables (e.g. blender). These measures are combined into a single measure. Various measures are used to combine these measures, Young Lives for example use a sum score, Vandemoortele (2014) uses Item Response Theory (IRT) and Filmer and Pritchett (2001) use Principal Component Analysis (PCA). The gold standard of asset indices for low- and middle-income countries are collected in the Demographic and Health Survey (DHS) and UNICEF's Multiple Index Cluster Survey (MICS), and Young Lives' asset index is modelled on these. See the Measures Chapter 5 for more details.

An asset index offers an alternative to income and expenditure data that are often unavailable, unreliable, incomplete or incomparable. Filmer and Pritchett (1999, 2001) and Sahn and Stifel (2003) showed that asset indices capture long-term material wellbeing, and do so with less measurement error than expenditure data. Filmer and Scott (2012) demonstrated that asset indices and expenditure are only correlated in certain situations, specifically where short-term shocks to expenditure and measurement error in expenditure are small, and in contexts where the largest shares of total expenditure expenditure are individually consumed goods.

Asset indices are also used in international comparative research on high-income countries. For example, the Organisation for Economic Cooperation and Development (OECD) administers the PISA and collects information on 16 household items and combine them into a single index. These items include three country-specific items that are considered appropriate indicators of household wealth (see OECD 2017): "PISA collects data on household assets because they capture family wealth better than family income" ((Zhang & Lee, 2011, p. 468)).

Historically, most research on asset indices and children's outcomes has focused on health outcomes. There is some research on the link between asset indices and school attendance. Chowa et al. (2010) reviewed literature concerning low- and middle-income countries, and found household assets (such as having a TV) to be positively associated with schooling, but the picture is mixed with productive assets (such as livestock), as children are more likely to have to maintain them. One example they gave was a study from Ethiopia, where household assets were positively associated with school attendance but cattle ownership inversely associated with school attendance as it increased the likelihood of the child combining school with work (see Admassie, 2002). The Young Lives asset index does not include productive assets.

Research on the mechanisms linking asset indices to children's educational achievement is sparse. Indeed, after a comprehensive review, Chowa et al. (2010) concluded that conceptual frameworks linking asset indices to children's wellbeing do not include mechanism.

Chowa et al. (2013) examined the association between consumer durables (such as radios or televisions) and children's educational achievement in low- and middle-income countries. They found that children from households owning at least one of five consumer durables scored one unit higher on English than peers from households that did not own any (results for maths were not statistically significant). These authors posit that households with more consumer durables have the means through which children can access English resources outside school, such as TV and books, while children from households without these do not.

Given the dearth of literature linking asset indices to children's educational achievement, I draw on literature that examines the links between asset indices and children's health outcomes. There is an abundant body of literature cited in a review by Chowa et al. (2010) and on the links between asset indices and malnutrition, as malnutrition hampers brain development. Chowa et al. (2010) also referred to several studies (cross-sectional and longitudinal) that associate vitamin A supplementation and immunisation with asset indices. (Vitamin A deficiency can cause blindness and make a child more vulnerable to infection.) They also reported on studies showing that as asset indices increase, so does the likelihood a parent will seek out health services to treat their child. Additionally, clean water and sanitation services are positively associated with various health and schooling outcomes, as discussed in Russell and Azzopardi (2019), giving children from more privileged households a health advantage that can translate into an achievement advantage over their poorer counterparts.

Expenditure

Young Lives produces a 'total per capita expenditure' variable that includes: per capita food consumption and per capital non-food expenditure. They measure food consumption, rather than food expenditure, to ensure comparability across families, as some families are subsistence farmers who grow food on a small plot of land. When a family consumes a casava from their own plot of land this is given a monetary value and included in the 'total per capita expenditure'.

In high-income countries, income is the usual measure of SES, while in low- and middle-income countries expenditure data are used. Expenditure data in these countries tends to be more reliable than income data. First, large portions of the population work in informal circumstances, so their income is not recorded. Second, many people work in the agricultural sector, or sectors associated with it, and their income is volatile from year to year (depending on the national and international product prices), and also varies seasonally. While their income may be volatile, their consumption may be smoothed. Third, many people are subsistence farmers, living from their land, so their self-produced consumption is not an income they receive. Therefore, expenditure tends to be a better measure of SES in these countries.

Despite this, expenditure is still a noisy measure as it can fluctuate substantially from one season or year to the next. For example, in largely agricultural societies, expenditure will vary according to the weather, crop yields, transport and food prices. As food costs constitute a large proportion of poor people's expenditure, fluctuations in food prices cause high fluctuations in their expenditure across years.

A limitation of expenditure data is that savings or debt are not accounted for. If a family saves some of its income, this amount is not captured in expenditure, making the family look poorer in the data than what they are. Similarly, if a household incurs debt to finance its expenditure, this amount is included in its expenditure resulting in a family that looks wealthier in the data than what they actually are.

Mechanisms

There are three main frameworks in the literature that explain the mechanisms through which parental SES links to children's achievement. These are the Investment Model, the Family Stress Model and the Neurobiological Model. These theories have been developed for high-income countries. There is almost no research on the mechanisms underlying the association between parental SES and achievement in low- and middle-income countries.

In this discussion I draw on research that examines both parental income and expenditure as determinants of children's achievement. I use income to refer to both concepts, as most of the literature, which is based on high-income countries, focuses on income. There is a wide literature from high-income countries on what to include in the income measure (such as imputed rent, tax benefits and more) that is beyond the scope of this thesis.

The Investment Model, mostly built on work by Becker (1981), suggests that financial resources affect children's outcomes through the parents' ability to invest in better-quality or more goods and services that contribute to improved children's outcomes. See also a review by Foster (2002) which explains how, from an economic perspective, children are conceptualised as an investment, and that income forms part of a family's resources allocated to promote the positive development of its members.

The Family Stress Model sets out the psychological pathways through which low resources impact on parents' stress levels and, consequently, their mental health, affecting their parenting ability and, in turn, their children's outcomes (research using this model is reviewed in Landers-Potts et al. (2015) and Duncan et al. (2014)). Duncan et al. (2014) found that increasing family income leads to increased school achievement and achievement into early adulthood through both increased investments and reduced stress. See Cooper and Stewart (2021) for a systematic review of papers published before 2018 about OECD countries that draws on these two frameworks to examine the links between income and children's achievement, among other outcomes.

Some researchers draw on multiple frameworks and in their analysis include variables that relate to both frameworks. For example Berger et al. (2009) found that the effect of income on children's vocabulary scores is fully mediated by 'home environment'. Their definition of 'home environment' includes aspects of mother's parenting behaviours (some parenting behaviour examined include the mother being harsh or unresponsive) as well as parental investment (e.g. problems with the home interior and exterior, food insecurity, and hours of TV the child watches per week).

The Neurobiological Model is employed by neurobiologists who examine how growing up with social disadvantage shapes children's brain development, their ability to concentrate and learn, and consequently their educational achievement. This is a relatively recent and burgeoning body of literature, and one mechanism discussed is chronic exposure to stress hormones during brain development. Chronic stress is not necessarily a result of living with stressed parents, it could also come as a result of living in an unsafe neighbourhood, in poor housing or more. Noble et al. (2015) found both income and parental education to be strong predictors of brain development among a sample of US children. Kim et al. (2018) and Tough (2012) reviewed the impacts of socioeconomic disadvantage on brain development, referring to numerous studies showing that exposure to socioeconomic disadvantage is associated not only with impaired brain development, but also with brain development processes that limit cognitive development (i.e. impacting on the functioning of the amygdala, hippocampus and prefrontal cortex).

In summary, if the associations between parental SES and children's achievement in the Young Lives countries mirror findings from high-income countries, as well as those from low- and middle-income countries on children's health outcomes, I expect the strongest correlation to be with maternal education, followed by the asset index (as it is a less noisy measure of parental SES than expenditure), and then expenditure.

Early childhood education

Attendance in ECE falls under the Investment Model, where parents invest in their children's education at an early age. Heckman (2008) argues that returns to investment in human capital are greatest at early ages, and diminish as children grow older. The mechanisms through which ECE may enhance achievement scores is through age-appropriate exposure to academic concepts taught in primary school and preparing them for the routines and independence expected of them there, so they can focus on their learning.

Evidence from high-income countries shows that ECE can help to address socioeconomic disparities at their source, as Heckman (2008) writes, though these benefits accrue only when ECE is high quality and the largest gains are seen among the least privileged children, as discussed in Ruhm and Waldfogel (2012) and reported by Engle et al. (2011) and Vandell et al. (2010). Children are exposed to further influences as they grow up, and evidence shows that the advantage offered by ECE is greatest at early ages and diminishes as children grow older, as reported in Blanden et al. (2022).

Control variables

When I address RQ1 and document gaps in SES, the right-hand variables include only initial parental SES and the test language (see Chapter 9). When I address RQ2 and examine whether changes in achievement are stratified by social class, I use the initial parental SES variable. Finally, when I address RQ3 and examine the association between ECE and achievement, I use the additional control variables (see Equation 2.2). The variables included in $\mathbf{Z}$ are selected based on the amended Haveman and Wolfe (1995) framework outlined above, with additional guidance from Grantham-McGregor et al. (2007) who apply a similar framework to low- and middle-income countries. These additional child-level controls include height-for-age, gender, whether the child is firstborn, and their ethnicity/mother tongue (depending on the country). Household-level controls include household size and maternal employment. I also use community-level fixed effects by including a dummy variable for each sampling site the children lived in when they were one year old. (I tested to see if results would change if I used community level identifiers at 5 years old (Round 2) and the results were essentially identical.) Here, I provide a theoretical justification for using these controls. For details on how these were measured please see the Chapter 6.

Test languages

The meaning of test questions are likely to vary across test languages. Take, for example, the Young Lives vocabulary test that includes the word 'hurdle'. This was translated into one of the languages spoken in Ethiopia (by the Young Lives team) into a phrase that resembles 'boy jumping over a stick'. 'Hurdle' is a more challenging vocabulary word than the phrase 'boy jumping over a stick'.

It is inevitable that, through translation, some questions will get easier or more difficult. I therefore take the language of the test into account. In all of the study countries, language divisions overlap with socioeconomic divisions, so including the language of the test may narrow measured SES gaps in achievement.

Height-for-age (z-score)

Under-nutrition is reflected in a child's height-for-age. Each child's height is measured then standardised against a World Health Organisation (WHO) defined reference population. A child is classified as stunted when she lies 2 standard deviations below from median height-for-age of the reference population. Height-for-age is a reflection of physical development, and is linked inextricably to achievement. Grantham-McGregor et al. (2007) use a physical development indicator as a proxy for achievement (or cognitive development which was the terminology used by the authors) in their analysis.

Gender

Gender is an important aspect in children's development across the globe. In many societies there is a boy-preference, where boys receive better nutrition, health and educational services than girls. Research also reports a gender gap in certain disciplines. Balart and Oosterveen (2019) discuss these gender gaps (and finds that longer tests reduce the gender gaps in maths and science). Girls' and boys' brain structures also differ during puberty, as discussed in Brito and Noble (2014).

Firstborn

Two key reasons parents may invest more in the human capital of their firstborn child is they have more disposable income before subsequent children are born, and another is for cultural reasons. In some cultures firstborn boys are prioritised, benefiting their achievement score compared to their non-firstborn peers. Alternatively, older children may be expected to engage in domestic and caring duties to the detriment of the education.

Societal dividing lines

In order to account for discrimination against and preference for particular populations, I control for the characteristics that represents a main dividing line in each society, which I substantiate in the Country Context Chapter 5. In Peru and Ethiopia, it is mother tongue; in Vietnam, it is ethnicity; in India, it is caste. These tend to overlap with other dividing lines, for example ethnicity and language tend to overlap, and also to overlap with regional disparities in their respective countries.

Household size

Household size may be negatively or positively associated with achievement. Larger family size may mean more caring or work responsibilities for older children. It may also be associated with there being more income earners. In high-income countries, the number of siblings in a households is generally associated with lower achievement. Linberg et al. (2019) discuss how this may be related to resource dilution (where finite resources are shared among more and more children as the family increases). These researchers also report on studies where children in crowded households receive less language stimulation than those in less crowded households, and, in their comparison of Germany and the USA, find that the number of children in a family contributes to the SES achievement gap in maths (as reported in Linberg et al., 2019, p. 15).

Maternal employment

Parents may enrol children in ECE to facilitate maternal employment. Maternal employment may benefit the achievement scores of a child through indirect channels, for example increased income, as discussed by Ruhm and Waldfogel (2012), increased decision-making power of the mother or a wider social network. It may also disadvantage the child. Linberg et al. (2019) use maternal working status as a “measure for quantity of time resources available to invest in the child’s development” (2019, p. 7). Additionally, children with working mothers may have to take on additional responsibilities at home, or may be left to fend for themselves, having a negative impact on their achievement scores.

Community

To account for as much unobserved variation as possible in Chapter 9 (when I address RQ3) about the association ECE and achievment, I include a community-level fixed effect. My analysis exploits variations within communities. As I am interested in the association between initial conditions and outcomes across childhood, I control for community-level effects when the children were 1 year old. The benefit of accounting for unobserved time-invariant community-level effects comes with the cost of controlling for these. I am not able to identify whether variation in the provision of ECE across communities is associated with variation in maths scores, for example. Such information is important for policy making. However, as the variation in ECE provision across communities is correlated with factors that shape achievement scores across communities, I use fixed effects.

RQ3 seeks to isolate the association between ECE and achievement. I therefore prioritise reducing the bias from unobserved, time-invariant community-level effects. Nonetheless, I am still left with the problem of unobserved, time-varying community-level effects, such as changes in social norms around ECE, or the impact on social-support networks if a disease (such as HIV/AIDS) affects parental morbidity and mortality. Many of these will change slowly, thus limiting the bias they have on the estimates, but for those which do not these will bias my estimates. I take this into account in the interpretation of results.

What Young Lives does not measure

An important aspect of ECE that Young Lives does not measure is the quality of the ECE education. The estimates of my analysis therefore examine the association between attending all ECE types of centres, both high and low quality, and achievement. I am unable to examine whether attending a higher quality ECE is associated with achievement scores.

The earliest measure of achievement in the Young Lives data set is at 5 years old, it does not have an earlier measure of achievement. As Grantham-McGregor et al. (2007) show, brains develop at a tremendous speed between birth and age 5, but Young Lives' first achievement score is collected at age 5 which is quite far along in a child's brain development. Young Lives also does not provide information on the quality of the ECE institution attended by the children, nor which time of type of public options were available. This would have been useful as the quality of education differed across the options. Research has shown that the quality of the education received in ECE is positively correlated children's academic outcomes, as discussed in Ruhm and Waldfogel (2012). Young Lives did collect retrospective data on this in subsequent rounds but, given the challenges associated with recall issues the responses given at subsequent ages do not correspond to the data collected when the children were 5 years old.

2.6 Conclusion

The analytical framework presented in this chapter underpins the selection of outcomes and predictors used in the generic specification. The relationship between initial parental SES and children's achievement is complex, so this framework also offers a foundation from which to interpret my results and to construct plausible narratives for any similarities and differences that I observe across countries.

To enable a comparative study across four countries, these variables must meet several criteria which I use, in the Measures Chapter 6, to critically asses how each variable is measured.

First is the timing at which data are collected. To ensure comparability across countries, we need all children (and their households) across the four countries to receive their initial interview and all follow up interviews at the same age.

Second, that the variables used meet or are close to best practice.

Third, that the achievement tests measure the same sets of skills (in my case, receptive vocabulary and maths skills) across rounds. If the skills tested differed wildly from one test to the next, or even between countries, this would complicate my interpretation of findings.

Fourth, is the standardisation of instruments used to collect data on each variable, across countries, to facilitate comparison. This includes, but is not limited to, the question format and the order in which questions are asked, the answer options provided (though allowing for context-specific answers), and how the data are then coded into a dataset. See the Measures Chapter 6, for a detailed discussion of how the variables I use are measured.

In this chapter, I discussed my methodological approach within countries, outlining the generic regression and detailing the outcome and explanatory variables selected. In the next Chapter 3, I discuss the rationale for engaging in a cross-country comparative study.

Chapter 3

Cross-Country Comparison: Why and How?

3.1 Introduction

I examine the association between socioeconomic status (SES) and achievement and between early childhood education (ECE) and achievement, as well as the achievement trajectories of children by SES, between four countries. My study is, therefore, a small-n cross-country comparative study. To compare individual outcomes across countries, I analyse a large number of individuals within each country – approximately 2,000 children per country. These children are part of the Young Lives study, which follows about 8,000 children from 1 to 15 years old across Ethiopia, India, Peru and Vietnam. (Young Lives tracks two cohorts: the younger cohort from early childhood into adolescence (2,000 children) and the older cohort from middle childhood into their early 20s ($\sim$1,000 children). This analysis covers only the younger cohort.)

As I study thousands of children across a small number of countries, I have several issues to grapple with: How many countries would be most productive to consider? What are the advantages and limitations of a small-n cross-country comparative study? How best can I use the data to address my research questions, given the several approaches that I could choose? How best to bring together the results with the countries' contexts while also considering study limitations? Indeed, why compare at all?

Done carefully, cross-country comparisons between even a small number of countries offer a great deal of value. Cross-country comparisons provide a powerful tool to investigate the implications of similar or different social and economic policies and contexts across countries. For example, comparing how ECE correlates with children's achievement across different ECE regimes will help to identify structural differences that may mediate the relationship across different countries (see Bradbury et al., 2015; Ermisch, Jäntti, et al., 2012 for pertinent examples). Data limitations mean that there is little comparative research available on SES gaps in achievement, achievement trajectories by SES and the association between ECE and achievement in low- and middle-income countries.

How many countries should be compared? A single case study would sit in isolation, and it would be difficult to establish whether the results are country-specific or more general in nature. Too many countries, on the other hand, would prove unwieldy and superficial, with endless comparisons to be made and excessive information to be synthesised about each country's context. In my opinion using four offers adequate points for comparison and makes gaining an understanding of each country's context manageable. Further, the Young Lives project designed its data set for comparison across countries. I draw on several examples of four-country comparative studies to inform my comparative approach.

My main limitations are related to data collection (discussed in Chapter 4) and the measurement of variables across different contexts (e.g. one year of maternal education in India will be different from one year in Peru). I take these into account in the analysis and interpretation of results. Some authors of other comparative studies go a step too far, drawing causal conclusions from an associational analysis – something that I resist. Rather, I contend that there is intrinsic value in documenting how associations are similar or different between countries.

In terms of my regression approach, I run separate regression analyses for each country. Although there do exist several regression modelling approaches that can be applied to multi-country datasets, I believe that mine is the most appropriate for the research questions I consider. For example, applying a common model to a pooled dataset without modelling country effects would mean that I cannot examine variation between countries. Using a fixed-effect model, whereby country dummies are included in a regression analysis applied to a pooled dataset, would fix individual-level effects to be the same among countries, unless I interact them with country-level variables. To address my research questions, I would have to interact all individual-level variables with country-level variables and cluster my standard errors, and still would have to assume the residual variance across all countries was the same – an assumption that I have no reason to make. Using a country-level random effects model on a pooled dataset would require that I make the untenable assumption that unobserved country effects are generated by a common mechanism. I therefore run separate regression analyses for each country.

How will I bring together results and country context while considering the limitations of the analysis? I juxtapose the coefficients from my statistical results and use the contextual information to inform my interpretation of differences and similarities in the results. Based on a review of the literature on small-n cross-country comparative studies examining similar research questions to mine, I identify key analytical principles that I adhere to in my analysis and interpretation. It is important: to be explicit about the methodology of the study and its limitations; to use a staged approach to interpreting results – starting with a descriptive comparison of coefficients then incorporating each country's context; to conduct robustness checks; and to focus on the comparative nature of the study.

In the next section, I situate this comparative study in the broader context of cross-country studies of low- and middle-income countries. I then outline the advantages and challenges of conducting small-n comparative analysis in the area of child development and with Young Lives data. This is followed by a discussion of how other researchers conduct small-n comparisons, so I can situate my regression approach in the broader literature and identify the key analytical principles to which I adhere. In this chapter, I take country selection for granted, as the Data Chapter 4 discusses country selection separately.

3.2 Situating a four cross-country study in the broader context of cross-country studies

Cross-country comparisons occur across all disciplines of social science. As Hantrais (2007) states, they share a common concern to observe social phenomena in multiple countries in order to describe similarities and differences, to gain a better understanding of how social processes operate, to assess their consequences, to build or test theories, and/or to draw lessons, policy or otherwise, on best practice. Other references that critically discuss cross-country comparisons include Sigle-Rushton (2009), Moss (2010) and Jowell (1998).

My primary aim is to describe similarities and differences in SES achievement gaps, achievement trajectories by SES and associations between ECE and achievement. Secondary is the endeavour to gain a better understanding of how and why these results differ (or are similar) across countries. There is an intrinsic value to documenting differences and similarities between countries: doing so is informative and intuitive. To understand results, we need reference points. If a nine year old child scores 70 points on a test, we won't understand what 70 points means until we compare those results to the scores of a broader population. The author of a case study may reference the results of other studies to inform her results, but those other studies won't have been designed for direct comparison: measures used may differ, as may the timing of the study, the age at which the children were interviewed, and the questions asked. In this case, I have the results from studies from four different countries which were designed to be compared (as I discuss in detail in Chapter 4, the data chapter). While the country contexts differ, at least in this study we are comparing like for like in terms of the children's ages, the number of times they were followed up, the questions that were asked, and (for the most part) the tools that were used. Comparisons among countries also inform how structural differences can mediate the relationship between ECE and children's achievement.

Cross-national evidence across low- and middle-income countries is sparse, contrary to that of high-income countries. Data limitations and non-comparability are an important reason for this. Hobcraft (2007) writes that research on low- and middle-income countries is constricted and Gabel (2010) mentions also a "lack of a comparative database on child policies" (2010, p. 185). The Young Lives dataset is designed for cross-country comparisons though, as with all datasets, it comes with its own challenges.

There is a wide range of quantitative approaches to cross-country analysis. These can be grouped into two categories: those that treat countries as units of analysis (e.g. time-series cross-section (TSCS) data analysis) and those that treat individuals within countries as units of analysis. Two groupings each comprise various approaches.

In this section, I focus on individuals. Figure 3.1 offers a heuristic outline of various analytical approaches to cross-country analysis that examine individual-level differences. The horizontal axis describes the number of countries included in the comparison, and the box represents the types of analytical approaches that can be applied to this comparison.

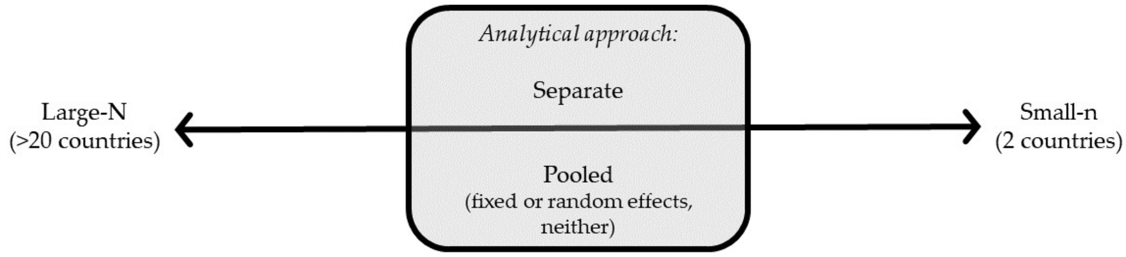


Figure 3.1: A spectrum of different types of cross-country analysis to examine individual-level differences.

At one extreme, researchers compare large number of countries ("large-N"). Large-N cross-country comparative studies are primarily good for determining a country's ranking based on an outcome measure, and tends to be used for constructing outputs such as league tables. UNICEF (2015), for example, regularly ranks countries according to their levels of inequality in various child-specific domains. Other researchers rank countries based on an outcome measure after controlling for other variables. In Figure 3.2, for example, Ho et al. (2014) rank 14 countries based on the gap in maths scores between children who attended ECE and those who did not, before and after controlling for the child's socioeconomic situation. These large-N cross-country comparisons tend not to consider country-specific context.

At the other extreme are case comparisons between two countries. Users of these are more interested in the processes that underpin differences in outcomes between countries, and, therefore, these studies tend to draw heavily on each country's context. What constitutes a "large" or "small" sample varies across the literature, and also depends on the aims of the research. For example, Bryan and Jenkins (2016) report that, in cross-country multi-level analyses, a "small number of countries ... is typically less than 30" (2016, p. 2). In literature examining research questions like mine, 30 would be considered a large number of countries. My four-country study thus represents a small-n analysis.

Whether for large-N or small-n cross-country comparisons, various analytical approaches can be employed. These can be grouped into two categories: those that pool the data set and those that analyse the data sets separately.

Cross-country pooled analysis pools individual-level data and sometimes includes country-level predictors to exploit inter-country variation, see Montie et al. (2006). Bryan and Jenkins (2016) provide a helpful classification of the most commonly used

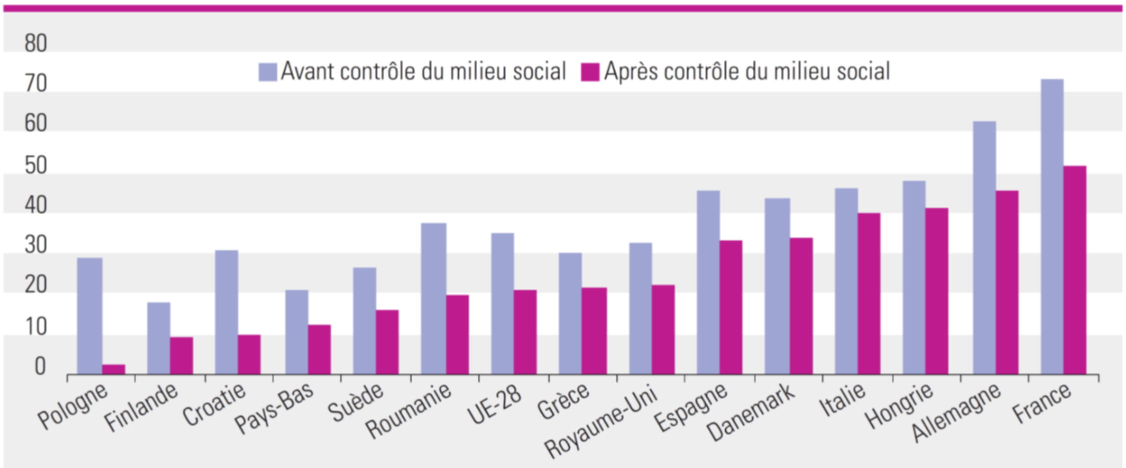


Figure 3.2: Differences in maths scores (PISA 2012) between children who attended and did not attend ECE education.

Key (translation): purple-grey columns = before socioeconomic controls, magenta columns = after socioeconomic controls.
Countries (translation): Poland, Finland, Croatia, Netherlands, Sweden, Romania, EU-28, Greece, UK, Spain, Denmark, Italy, Hungary, Germany, France. Only attendance of equal duration is considered. Controls include child-reported parents' education, parents' education and access to cultural resources and material resources.
Source: Ho et al. (2014, p. 3)

pooled models for cross-country analysis. These include those that control for but do not model country effects; those that use fixed effects to model country-level effects; and those that use random effects to model country effects, also referred to as hierarchical or multi-level models. It is conceivable to pool all individuals and use neither a fixed nor a random effects model – although I have not come across this in my review of the literature. See Table 3.1 for a more detailed discussion. I do not pool the Young Lives data for technical reasons that I will shortly discuss in Section 3.3.

Cross-country 'separate analysis', the approach I use here, entails fitting separate regression models on individual-level data for each country, as Bradbury et al. (2012), Dräger et al. (2023), Duncan et al. (2012), and Ermisch, Peter, and Spiess (2012) do. This approach is good for examining individual-level effects that differ between countries, as it is possible to estimate these reliably. It is not good for examining inter-country variation, as it does not identify country-level effects. Cross-country separate analysis is usually the approach chosen when the number of countries studied is too small to allow reliable predictions from the sample. The question of how to examine individual-level differences within countries depends on the research question(s).

3.3 Why conduct small-n cross-country comparative analysis in the area of child development?

In this section, I outline the main reasons for conducting a small-n comparative analysis and identify the main limitations of such an analysis.

The Young Lives study countries, Ethiopia, India, Peru and Vietnam are distinct, with diverse institutional systems, levels of inequality, cultures, geographies, languages and economies. They were "selected to include [a country] from each of the major regions of the developing world, along with a range of political–economic conditions and circumstances, with strong institutional capacity locally to undertake complex panel research being another crucial criterion" (Boyden & James, 2014, p. 26). Young Lives is the only survey of its kind among low- and middle-income countries. It is a longitudinal research study designed as a cross-country study. I focus on the younger cohort of approximately 2,000 children in each of these countries, whom the study followed from age 1 to 15 years old. See Data Chapter 3 for more information.

Advantages of conducting cross-country comparative analysis

A cautiously conducted and prudently interpreted cross-country, comparative study has several advantages, the main one being that it offers a powerful tool to explore the differences and similarities among countries in various domains. Examining an aspect of a country in contrast other countries reveals institutional differences that researchers may otherwise have taken for granted. As Ermisch, Jäntti, et al. (2012) explain, in a cross-country analysis, "genetic transmission in the outcome ... should be the same across countries, and so cross country differences should reflect different environments, policy and otherwise" (2012, p. 11).

For example, Blanden et al. (2012) compare the relationship between parents' SES and achievement in the UK and Australia, and provide plausible explanations for differences in this relationship – identifying differences in specific early year policies, albeit cautiously. Studying Finland and Estonia, Ojala and Talts (2007) compare children's learning achievements in ECE and establish a link between differences in the philosophy of the ECE curriculum and pedagogy and differences in learning achievement.

Here, I examine, across Young Lives study countries, SES achievement gaps, achievement trajectories by SES, and how the association between ECE and achievement varies or coincides. Young Lives countries have vastly different, and yet in some ways similar, ECE policies. With the data, I document how ECE is associated with achievement at various points during childhood. This can shed light on policy or contextual differences that mediate these relationships. For example, differences in norms across countries may have a distinct implication for children's achievement, as discussed in Young Lives (2015). In some countries, girls may be systematically disadvantaged (e.g. they perform worse in tests because of lower expectations or putting their time into housekeeping tasks); in other countries, boys may be disadvantaged (e.g. they perform worse on tests because of the need to take part in animal herding or other income-generating activities).

A second advantage is that, if results are shared across countries, it is harder to make the case that these results are exceptional – as Young Lives (2015) and Richter et al. (2012) also contend. Bradbury et al. (2015), for example, undertook a comparative study to show that "outcomes of children from different backgrounds [in Australia, Canada, and the United Kingdom] can be more equal than they are in the United States today" (2015, p. 16). I will be able to verify whether country-specific findings are not unique if they also arise for the other study countries. However, if I find a phenomenon in one country but not the others, this does not imply that this country is unique, as this phenomenon may occur also across other countries that the study does not cover.

Third, from a reader's perspective, it is much easier to understand the results from one country in reference to those of other countries. In turn, as a researcher, it is easier to communicate one country's results in reference to those of others. For example, I can state that the SES gap in Peru, let's say, is larger or narrower than that in other countries, and that Peru differs in a potentially relevant manner, thus grounding the Peru result in reference to that of another country. If I were interpreting Peru's results in isolation, I would struggle to communicate my finding meaningfully – as there is little comparable research on the topic in low- and middle-income countries – thereby leaving the reader questioning what "good associations" are as there is no absolute scale for these (leaving only comparative scales to offer meaning). Bradbury et al. (2015), for example, show that a comparison between the UK and the US is informative as "[t]he fact that the SES gap in children's achievement in the United Kingdom is not as large as it is in the United States may be due to more universal supports for low-SES families, such as universal health insurance, universal preschool, child benefits, and so on" (2015, p. 17).

Advantages: Young Lives designed as a cross-country study

The Young Lives surveys are specifically and carefully designed to enable cross-country comparisons. Before considering that in more detail, let us compare the Young Lives dataset to the only other cross-country, longitudinal database with individual-level data on early childhood from low- and middle-income countries that I am aware of: the Consortium of Health-Orientated Research in Transitioning

Societies (COHORTS) database. As I shall discuss in Chapter 4, COHORTS is a collaboration between Brazil's Pelotas Birth Cohort, the Institute of Nutrition of Central America and Panama (INCAP) study in Guatemala, the New Delhi Birth Cohort, the Philippines' Cebu Longitudinal Health and Nutrition Survey (CLHNS) and South Africa's Birth to Twenty (BT20) cohort. There are three challenges of using COHORTS data that do not apply to Young Lives (and hence give Young Lives an advantage); a further challenge is shared but is substantially less pronounced when using the Young Lives data.

First are differences in variable definitions and measurement across sites. Each of the COHORTS studies is designed and implemented separately by a different set of researchers with its own aims. The Young Lives team, by contrast, uses the same questionnaire and the same instruments, and strictly coordinates its implementation strategy for measurement across all countries.

In contrast, at COHORTS, a "major effort has gone into producing a common data set" (Richter et al., 2012, p. 625), which has limited the number of indicators studied. This has meant that variable selection often amounts to choosing the lowest common denominator. For example, not all surveys have data on household assets in early life or on individual incomes, as Richter et al. (2012) describe. To deal with these challenges, some analyses use "different outcome variables ... (e.g. pre-hypertension for adolescents and hypertension for adults) because of the different ages of the individuals across cohorts" (Richter et al., 2012, p. 625). In the Young Lives data, the main questionnaires are the same across countries, allowing researchers to run the same regression analyses across the four countries. In their four-country comparative study using different datasets, Bradbury et al. (2012) also "adopt a lowest common denominator approach to minimise the risk that differential measurement error affects [their] results" (2012, p. 97). (A similar cross country study using national survey data from six high-income countries, Development of Inequalities in Child Educational Achievement: A Six Country Study (DICE), shares also that "considerable processing was required to harmonize the data" (Waldfogel et al., 2023, p. 37).)

Second, as Figure 3.3 illustrates, COHORTS surveys collect data in different years, at different ages, and at different intervals between each survey: The Pelotas 1982 Birth Cohort study in Brazil collects data from children at age 2, 4, 23 and 30 years, whereas the INCAP survey in Guatemala collects data for children in the age range 0–7 years first, then in the age range 11–26 years, then 26–41 years and then 29–44 years. BT20 in South Africa collects data from individuals much more regularly: 16 times between birth and 20 years of age. This complicates comparisons, particularly when data points are not consistent across infancy and childhood because the effect of an early-years intervention differs according to age and the time since

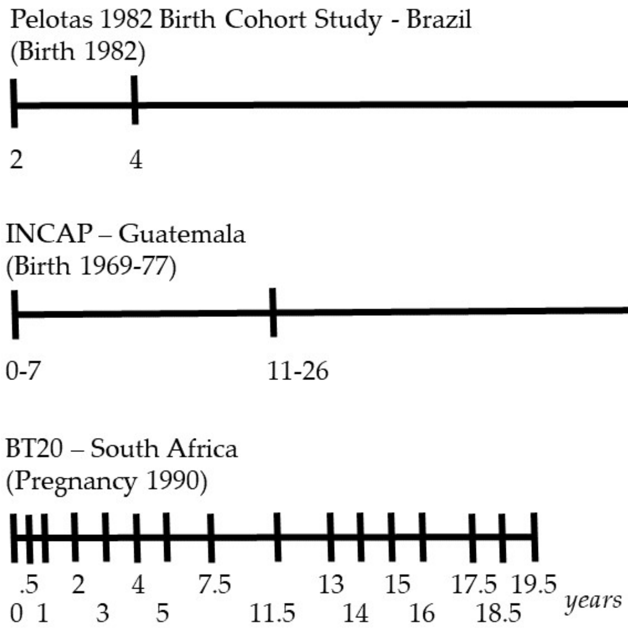


Figure 3.3: Participants' ages at data collection illustrate the challenges of comparing across the three COHORTS surveys.

intervention. With Young Lives, data are collected from all children at the same age in the same years with the same time period between rounds.

Third, COHORTS surveys include children of particular ages regardless of when they were born, resulting in comparisons between children born in the 1960s and in the three decades following. The context in which children grew up in the 1960s is obviously distinct from that of the 1990s, which further complicates comparisons across studies. All Young Lives children are born around the same time – in 2001/02.

Challenges of cross-country analysis with Young Lives data

The findings from cross-country studies pose several limitations. As it is not possible to overcome all of the challenges of cross-country research using existing data and methodologies, it is essential to consider their implications for the interpretation of results.

It is important to note that the challenges of using national surveys also apply when using cross-country surveys. However, the extent of heterogeneity between countries is far greater. As Jowell (1998) highlights, researchers using surveys rely on a principle of equivalence whereby, for example, questions have a broadly equivalent meaning to all respondents. A key challenge for cross-country researchers lies in reconciling context differences – such as in cultural norms, language, service provision (e.g. education) and methodological approaches – with the practical aspects of data collection, measurement, analysis and interpretation of results. The data chapter

(Chapter 4) discusses challenges of Young Lives data collection and, in the section below, I discuss various approaches to cross-country analysis and their advantages and limitations, focusing on measurement and interpretation.

Measurement: Different meanings

Researchers have not yet validated, and achieved international recognition for, instruments to measure achievement in most low- and middle-income settings. While such instruments do exist for high-income countries, translating them to low- and middle-income contexts, where children have different levels of literacy and different ages and are growing up in diverse social and cultural contexts, poses an important methodological challenge for Young Lives, as Boyden and James (2014) show.

There are challenges with measuring achievement and its validity across countries (and also within countries, particularly those with multiple languages), and this is the case with all the variables that I use. Even the most careful translation of a question can change its meaning, thereby compromising the validity of any finding. The same words can mean different things across countries; some words may not even exist. For example validation for the instrument to test receptive vocabulary (the Peabody Picture Vocabulary Test, PPVT) is available for the Spanish context but not for the Peruvian context. I discussed the word “hurdle” in Chapter 2, where the absence of a local-language term resulted in a decision to translate it as “child jumping over a stick”. Such a large change results in testing a different level of vocabulary in different languages. Although the Young Lives team validated their instruments prior to full implementation at each round and drew on technical expertise to compile them, such instruments can never be perfect. They are, nonetheless, the best available measures to gauge children’s achievement over time across the four low- and middle-income countries.

Also, the provision of services differs across the four countries. In education, we see this manifest in relation to equity, curricula and quality of education. One year of education in Ethiopia, for example, is not equivalent to one year in Peru. Nor is one year of education necessarily the same for children within each country, although the differences between countries will be starker. UNESCO offers a tool to render education levels more comparable across countries (the International Standard Classification of Education, ISCED).

Interpretation

Here, I consider five questions that reflect key challenges to interpreting analytical results. While I cannot fully overcome them, I try to attenuate them as much as possible.

First, how would one interpret SES gaps in achievement, achievement trajectories by SES, and the association between ECE and achievement across countries? In my first two empirical chapters 7 and 8, I describe similarities and differences of the SES gaps in achievement and achievement trajectories by SES across countries, while taking each country's context into account. In my third empirical Chapter 9, where I report the association between ECE and achievement over and above parental SES and other factors, I interpret cautiously. There are aspects of these relationship that are neither measured or accounted for in the regression I use, and the influences of these factors may differ across countries.

Second, how would one draw policy conclusions? My primary aim is to document SES gaps in achievement, achievement trajectories by SES, and the adjusted association between ECE and achievement to enhance the evidence base for social policy, as evidence can prompt and support an evaluation of policy options. Other researchers or policy-makers can use my work as a foundation to examine which policies are most conducive to closing an achievement gap. Nonetheless, should consistent similarities (or differences) become evident in the results across countries, I will identify these and make educated guesses as to the policy reasons for these similarities (or differences).

Third, how would one focus on the substantive differences and similarities between countries without stating superficial generalities or getting lost in fine-grained particularities? I address this rhetorical challenge by considering the story that my findings tell, and I draw in the relevant information in order to narrate it. I discussed a framework in Chapter 2, and will draw on it for guidance.

Fourth, how would one compare between countries in which the participants have been sampled according to different processes? The sampling design, and therefore the population of inference, differs between countries. Although the children will be the same ages across countries, I acknowledge that comparisons are not like-for-like. Indeed, Young Lives carefully selected the sample with the diversity of regions and circumstances in mind. A comparison between the results in Ethiopia and those in Peru, for example, represents a comparison between Ethiopian children born in only the 20 sampling sites with relatively good roading and the Peruvian children from 95 per cent of the districts of Peru. That is why I am explicit about the population of inference in each country.

3.4 How do researchers go about conducting small-n cross-country analyses?

Common regression modelling approaches to cross-country comparisons

Bryan and Jenkins (2016) outline four modelling approaches commonly applied to multi-country datasets. Table 3.1 summarises these. Three approaches involve pooling the individual-level data from all countries; one involves running a separate regression analysis for each country. Researchers may choose to pool to achieve a larger sample size and, therefore, more precise estimates. However, as highlighted in the table and the discussion, standard error calculations must be carefully considered to avoid artificially precise estimates.

Table 3.1: Common regression modelling approaches applied to multi-country datasets.

	Approach	**Remarks about specification**
1	Common model for all countries, pooled data, country-specific clustered SES.	Country effects controlled for, not modelled.
2	Separate model fitted to the data for each country.	Country effects not separately identified (absorbed into the intercept of each country's model). Every model parameter is country-specific.
3	Common model applied to pooled data (as in approach 1), except that model has country fixed effects.	All country-level factors are absorbed into the country fixed effect; estimates refer to specific sample of countries.
4	Common model applied to pooled data (as in approach 1), except that model has country random effects (multi-level model).	Country effects can be specified in terms of a country error variance and fixed effects of country-level predictors; 'exchangeable' estimates.

Note: In their supplemental material, Bryan and Jenkins (2016) also discuss a fifth approach, the two-step approach that entails one regression at the individual level, then another at the country level. However, they provide three reasons why using this approach with a small number of countries will result in unreliable or imprecise estimates (see Bryan & Jenkins, 2016, sup. mat. p. 10). Since I am fitting a regression with four country observations, I exclude this approach from the table and discussion below.
Source: Bryan and Jenkins (2016, p. 3)

For clarity, I refer to the first approach as the "common model", the second as the "separate model", the third as the "fixed effects model", and the fourth as the "random effects model".

I use the second of these approaches – the separate model – and run separate regressions for each country, as Bradbury et al. (2012), Duncan et al. (2012), and Ermisch, Peter, and Spiess (2012) all do using high-income country data. Underlying my rationale to run separate regression analyses for all countries, rather than using one of the pooled-data approaches, is that separate analyses constitute the most general approach. The reason is that separate analysis allows coefficients on variables and error terms (i.e. parameters) to differ across countries. As soon as I start pooling, I will have to start assuming that things are similar (e.g. the same intercept, slope coefficient, or error variance). As I explain, I see no reason to start making these assumptions yet.

To justify this modelling approach relative to the other approaches, we need a regression model as a base for discussion. Here, I refer to the basic linear model that Bryan and Jenkins (2016) propose. (For consistency with the equations in Chapter 2, I place the variable's parameter in front of the vector.) It is as follows:

$$y_{ic} = \beta \mathbf{X}_{ic} + \gamma \mathbf{Z}_c + u_c + \epsilon_{ic} \tag{3.1}$$

where

$i = 1, \ldots, N_c;$

$c = 1, \ldots, C$

where y_{ic} is the outcome for each person i in country c and depends on the following factors: $\mathbf{X}_{ic}$ is a vector of individual-level characteristics (and β their associated parameters); $\mathbf{Z}_c$ is a vector of country-level variables (and γ their associated parameters); and the unobserved country-level effects (u_c) and individual-level effects (ϵ_{ic}). Both unobserved country- and individual-level effects are assumed to independent and identically distributed (iid) according to a normal distribution with mean 0 and variance σ^2, i.e. $\epsilon_{ic} \sim_{\text{iid}} \mathcal{N}(0, \sigma^2_\epsilon)$ and $u_c \sim_{\text{iid}} \mathcal{N}(0, \sigma^2_u)$. This means that the error terms do not correlate with $\mathbf{X}_{ic}$ and $\mathbf{Z}_c$, and are independent of each other.

Fundamentally, I am comparing the differences in achievement associated with a unit change in an individual-level characteristic (i.e. SES or ECE) across countries, when I fix all other explanatory variables to some specific value (within each country). For example, in the case of ECE, to do this I need to estimate β, or the 'average marginal effect' (AME) of ECE attendance (versus non-attendance) on achievement. Given the small sample of four countries, I am unable to robustly estimate country-level effects, i.e. γ.

I select the second approach (fitting separate models) over the first approach (a common model) because the former allows the AMEs (expressed in β) for SES and ECE to vary across countries. In the first, or common, model controls for country-level effects (γ) through country cluster robust standard errors, although without modelling them. Constraints hold the parameters on each variable to be the same across countries. The AME reported in the first approach would essentially be the average of the four countries' AMEs of ECE education, for example. In other words, the process would implicitly average the change in achievement associated with ECE attendance (rather than non-attendance) for children in India, Ethiopia, Peru and Vietnam and treat them as equal. This assumption is not tenable in my research, nor is the average of interest to me. Rather, it is differences in AMEs across countries that I am interested in. The first approach does not, therefore, match my research.

The assumptions of the fourth approach (the random effects model with random intercepts) are also untenable in my research. It assumes that countries are a random sample of a broader population of countries that have a distribution with a mean of zero; the model estimates the variance. Equation 3.1 describes the random effects model, as it includes unobserved country-level effects (u_c) in the model. In multi-level modelling, we would use the variance of (u_c) to estimate the intra-class correlation $\rho = \sigma_u^2/(\sigma_z^2 + \sigma_u^2)$ (where σ_z^2 and σ_u^2 are the individual and country random effects, respectively), which is the proportion of the total variance in the outcome that data structure (specifically clustering) accounts accounted for. Bryan and Jenkins (2016) explain that the random effects model treats unobserved country effects (u_c) as generated by a "common mechanism" which makes them, therefore, interchangeable between countries. However, it is unreasonable to assume that the unobserved country effects in Ethiopia, India, Peru and Vietnam are generated by a common mechanism, and that generalisations can be made from such a small sample of countries. My regression modelling approach does not make such ambitious assumptions; rather, running separate regression analyses by country absorbs unobserved country-level effects (u_c) into the model intercept.

I choose to run separate regressions over the fixed effects model (the third approach) for the main reasons that doing so would necessitate assuming that the residual variances are the same across countries. The fixed effects model does not allow AMEs for individual-level variables to vary across countries unless all individual-level variables are interacted with country-level variables. (As I am interested in how AMEs for SES vary across countries, I would have to interact these with the country-level variables $\mathbf{Z}_c$.) Doing so would fix the residual variance across all countries. Bryan and Jenkins (2016) explain that estimating a model with a full set of interactions between individual-level explanatory variables and country dummies is not the same as running separate regression analyses for each country. Rather than estimating four separate residual variance parameters (i.e. $\sigma^2_{\epsilon_{ic}}$) for each country, the fixed effects model entails estimating only one residual variance parameter for the pooled sample. The unobserved country-level effect u_c is not estimated separately in the fixed effects approach; it is only modelled separately in the random effects model – the fourth approach.

Fixing the residual variance across countries is problematic for my research. As a regression model with a continuous outcome variable is identified without the residual variance, the coefficient β can be interpreted as a straightforward AME (i.e. a unit change in X corresponds, on average and ceteris paribus, to Y changing by β). When the residual variance is fixed across groups, this is "less serious ... because there is no bias in the coefficient estimates" (Allison, 1999, p. 199). Gould (2024) also explains this well. The situation is perhaps considered "less" serious because, although the coefficient estimate is not biased, fixing the unobserved variance to be the same across countries may result in an incorrect estimation of the error variance for each country – particularly if they are wildly different. A key implication is that, if the estimated standard errors are incorrect, so are the test statistics and confidence intervals, rendering statistical inferences invalid. To address this, researchers tend to cluster standard errors because they assume that people within groups are more similar than they are between groups.

The fixed effects model can be extended to allow the residual variance to vary across countries. Gould (2024) illustrates that it is possible. This is, however, not a typical approach to pooling, where residual variances are usually fixed, as discussed above. This approach falls between the second and third approach (running separate regressions and the fixing the effects) – let us call it approach 2.5. Gould (2024) explains that the advantage of doing this is that we can assess the differences in the magnitude of the residual variance across groups and determine whether it is appropriate to fix residual variances across countries. Another advantage of this approach is a researcher can "test the equality of coefficients between the [fitted] equations" (Gould, 2024, §3). In other words, I would be able to test whether the association between attending ECE and achievement is statistically different between Vietnam and India.

In my research, there is no reason to believe that residual variation will be the same across countries, but there is also no way to know without examining the differences. In the case of ECE attendance, for example, Ethiopian children may have a more heterogeneous learning experience at ECE ages, given that most children do not attend ECE and are left to their own, or their caregiver's, devices. In contrast, most children in Vietnam attend ECE and the quality of provision is among the least heterogeneous of the four study countries. The residual variance in Ethiopia is hence likely to be larger than that of Vietnam.

A reason not to use approach 2.5 (i.e. pooling all data, including a full set of interactions and allowing the residual variances to vary across countries) is practical is that it requires a lot of interactions, and cell sizes can become quite small or empty (rendering the results challenging to interpret). Another reason not to use approach 2.5 on the data from the four study countries is that, when I do test for equality, I am arguably not comparing like-for-like. The sampling designs differ between all four countries (three are non-probability samples) and so does the population of inference. Interpreting the results of a statistical test across countries would be problematic.

To summarise, the second approach, running separate regressions for each country, provides the most robust parameter estimates of the available options, given the small sample of countries. Other approaches will either not address my research question (first approach), make unreasonable assumptions (fourth approach), provide only unreliable country-level estimates with confidence intervals that are likely too narrow (second approach), or conflict with the different sampling designs across countries (approach 2.5).

However, with the second approach, I am not able to identify differences in country-level effects across countries from a statistical perspective (i.e. to determine whether the difference in β between countries is statistically significant). Given that the above approaches do not allow me to assess statistical differences in SES achievement gaps or the association between ECE and achievement across the four countries in a reliable manner without unreasonable assumptions, I choose not to attempt such comparisons. My statistical analysis is limited by the small sample size of the countries. This small sample size, however, lends itself to a rich descriptive comparison of results, taking country contexts into account.

Common approaches to comparing regression coefficients across countries in small-n comparisons

Having established which analytic approach to take, I now focus on how to interpret the results – specifically, how I compare the fitted coefficients between the countries, how I take the countries' contexts into account, and what analytical principles I adhere to. To do this, I draw on other examples of small-n cross-country comparative studies that investigate SES gaps in achievement, achievement trajectories by SES and the association between ECE and achievement, and that use the same analytical approach as mine. I take 'small' values of n to mean 10 countries or fewer.

The research on the association between ECE and achievement is dominated by case studies, with only a few comparative studies, and fewer small-n cross-country studies. You will notice, therefore, that cross-country literature examining the link between SES and achievement dominates this section's discussion (for a brief review of international comparative research between ECE and achievement, see West, 2016, 9–10).

Most of the papers employ the same regression approach as I do and run separate regression analyses for each country. Of the 18 studies identified, only 2 do not run a separate regression analysis for each country. Rather, Montie et al. (2006) and Malmberg et al. (2011) analyse a pool of 10 and 3 countries, respectively, and use a multi-level-model. I do not include these in the discussion below as they offer little guidance on how to interpret my results.

Based on a review of the literature, I identify four broad approaches to interpreting coefficients, incorporating each country's context and considering study limitations in a small-n country comparison. Authors may:

1. consider neither country context nor the limitations of a comparative study;

2. consider country context but not the limitations;

3. not consider country context but do consider the limitations;

4. consider both country context and the limitations.

I consider both grey literature (including working papers) and published papers here while, in the empirical Chapters 7–9, I restrict my attention to published papers as they have undergone more rigorous review processes. Table 3.2 summarises all the sources (grey and published literature) corresponding to each category. For further information on these sources, see Annex B.1 for the literature review that underpins this section.

I employ the fourth approach and juxtapose the regression coefficients, drawing on each country's context to interpret the results while also considering the limitations of a small-n cross-country comparative study.

Table 3.2: Summary of small-n cross-country comparative studies that examine the association between SES or ECE and achievement

Reference	No. of countries	SES or ECE	Context	Limitations
Consider neither country context nor limitations of a comparative study				
Duncan et al. (2012)	4	SES	No	No
Jerrim and Micklewright (2012)	9	SES	No	No
Jerrim and Micklewright (2014)	10	SES	No	No
Georgiadis (2017)	4	SES	No	No
Lopez-Boo (2016)	4	Mostly SES, ECE as mediator	No (except for a few descriptive statistics)	No
Das et al. (2022)	5	SES	No	No
Consider country context but not limitations				
Ojala and Talts (2007)	2	ECE	Yes	No
Singh (2014a)	4	SES	Yes. Half a paragraph, not integrated into interpretation of results	No
Jaramillo and Tietjen (2001)	2	ECE	Yes	No
Lockheed et al. (1989)	2	SES	A little bit	No
Does not consider country context but does consider limitations				

Continued on next page

Table 3.2 – *Continued from previous page*

Reference	No. of countries	ECE or SES	Context	Limitations
Fernald et al. (2012)	4	SES	No (except for a few descriptive statistics)	Yes
Consider both country context and limitations				
Blanden et al. (2012)	2	SES	Brief context on policy differences in early childhood, incorporated into interpretation of results	Yes
Magnuson et al. (2012)	2	SES	Brief context and light use of context in interpretation of results	Yes
Schady et al. (2015)	5	SES	Context provided and incorporated into interpretation of results	Yes
Crookston et al. (2014)	4	SES	Yes. Half a paragraph, not integrated into interpretation of results	Yes
Bradbury et al. (2015)	4	SES	No context provided prior to results section, integrated into interpretation of results	Yes
Bradbury et al. (2012)	4	SES	Context provided and integrated in to the discussion of results and conclusions	Yes
Bradbury et al. (2019)	4	SES	Yes, neatly integrated into the article's discussion	Yes
Linberg et al. (2019)	2	SES and ECE	Yes, integrated into the whole article	Yes
Other studies that do not run separate regressions for each country*				

Continued on next page

Table 3.2 – *Continued from previous page*

Reference	**No. of countries**	**ECE or SES**	**Context**	**Limitations**
Montie et al. (2006)	10	ECE	No	Yes
Malmberg et al. (2011)	3	ECE	Yes. Half a paragraph, not integrated into interpretation of results	No

Notes: * I include these for completeness, to include all small-*n* cross-country studies I identified.

Comparing parameter coefficients

When researchers run separate regression analyses for each country, they try to ensure that the scales of both dependent and independent variables are comparable. Most standardise their measures using national z-scores or Item Response Theory (IRT), except Georgiadis (2017) who uses the raw scores. Few authors discuss the implications of that choice when comparing results across countries.

When comparisons are made, they tend to be descriptive, focusing on differences in the gaps between the better- and less-well-off children or between those attending ECE and those not attending, a widening or narrowing of the gap over time, and whether socioeconomic gradients in achievement are similar or different in each country. For example, Duncan et al. (2012) observe that "[t]here is little indication that Swedish gradients are flatter than gradients in the United States or United Kingdom" (2012, p. 223); Lopez-Boo (2016) writes that, "although differences in early language development by SES are present in all countries, they arise more starkly in Peru" (2016, p. 506).

When the links between SES and achievement are so complex and bidirectional, it seems all the more important to interpret results cautiously and to test whether one's results are robust to changes in measurement or model specification. However, not all authors interpret their results carefully – nor do they all test the robustness of their results.

Some papers make a blanket comparison of coefficients without considering the implications of how they measure their outcome variable on cross-country comparisons (e.g. Georgiadis, 2017; Jaramillo & Tietjen, 2001). Others are more cautious in their interpretation. Magnuson et al. (2012) write that "[a]lthough results also differ by country with regards to gradients in socioemotional development, here we hesitate to draw strong conclusions given the differences in measurement across the two countries" (p. 257). Blanden et al. (2012) are equally circumspect in their interpretations, using cautious terminology: "It is possible that . . . " (2012, p. 159); "an alternative explanation is one of . . . "(2012, p. 159); "appear to be . . . " (2012, p. 143); "a great deal of research remains to be done in this area . . . " (2012, p. 160); "[a]lthough these findings are consistent with the overall body of research . . . we can only speculate here on the reason for these differences" (2012, p. 160).

Some authors also employ robustness tests to check whether substantive results remain the same. Jerrim and Micklewright (2012, 2014) use various ways to measure SES and find that the results across countries are not stable. Lopez-Boo (2016) tests whether results change if the wealth measure is calculated separately for urban and rural areas (they do not). Other authors use different measures of the variable of interest to check if results across countries persist (e.g. Schady et al., 2015), or to test the robustness of their age groupings (e.g. Fernald et al., 2012).

Incorporating country context

There is instrumental value in incorporating country context into analysis. First, understanding a country's context can inform relevant variable selection and regression specifications. If a researcher knows that being firstborn may bestow advantage on a child in a particular country, she would include this variable in the regression equation. Second, understanding the country context assists with the types of robustness checks that should be conducted. For example, if a researcher did not know that, in Ethiopia, ECE-aged children sometimes enrol early in primary school rather than attending ECE, she would not test whether including children who had enrolled early in primary school affects results (which I did, finding no difference in results; see Vandemoortele, 2018). Third, having a good understanding of the country context enables a more informed and critical interpretation of results. Without a critical discussion of results, it is difficult to ascertain whether the results are plausible, or merely an artefact of the data. Fourth, relevant background information grounds the study in something concrete, investing the reader in the discussion and enabling a better understanding of the data and methodology than would otherwise have been the case. Fifth, not only does incorporating this context into the discussion of results render them more intuitive and accessible to the reader, but also the results become more policy-relevant, as they are situated in a concrete understanding of the country context. This can also stimulate a discussion and further research on the topic. It is, however, difficult and time-consuming to incorporate each country's context into a piece of research, as gaining familiarity with this requires a substantial amount of reading and investigation, and if possible travel.

Most papers examining the links between SES and achievement provide descriptive comparisons across countries that place no or minimal emphasis on the country's context. Some papers provide some context in the introduction but do not incorporate this into the interpretation (e.g. Crookston et al., 2014; Fernald et al., 2012; Jaramillo & Tietjen, 2001). When the specific context is taken into consideration, it is often done only briefly (e.g. a half a paragraph in Singh, 2014a and in Schady et al., 2015 or a few words in Lopez-Boo, 2016). This may be intentional in the context of word limits that journals impose, but may also be a disciplinary preference towards describing quantitative results rather than interpreting them in the context from whence they came.

Other papers introduce information on the country's context only when interpreting results (e.g. Lockheed et al., 1989; Lopez-Boo, 2016). In this case, contextual information is not grounded in a broader discussion of the country, so can be perceived as conveniently selected contextual information to support results. Lopez-Boo

(2016), for example, provides no context but, when discussing results on ECE, offers some context in parentheses: "in Ethiopia (where ECE attendance is lowest) and Vietnam (where ECE attendance is highest but probably heterogeneous in quality)" (2016, p. 507). Lockheed et al. (1989) do the same in their study on Malawi: where they are unable to explain a result, they provide some contextual information (see discussion on Table 9 on page 251 of their article).

Country context is taken into more account in books and book chapters (e.g. Blanden et al., 2012; Bradbury et al., 2012, 2015; Jaramillo & Tietjen, 2001) – although Jaramillo and Tietjen (2001) do not consider context in their interpretation.

The number of countries compared and the comparisons made may influence how much context is taken into account. In examining the links between ECE and achievement, Ojala and Talts (2007) and Jaramillo and Tietjen (2001) take each country's context into account in great detail (though Jaramillo and Tietjen, 2001 do not when interpreting results). Over two pages, Ojala and Talts (2007) critically discuss the core curriculum of ECE in each country, then identify similarities and differences in these curricula. Jaramillo and Tietjen (2001) provide a short paragraph on the following main topics: ECE access and distribution, ECE provision, ECE personnel, and ECE students, then descriptive tables on the characteristics of ECEs in each country. They may have elected to give this much detail because of the small number of countries – only two – made it practicable, as well as the nature of the comparison. The parameters are much more defined when discussing mechanisms that link ECE and achievement, than those linking SES to achievement. Perhaps because of this, authors linking SES to achievement avoid providing contextual information and, when they do incorporate it, they do so only briefly, focusing instead on differences in each country's gross national income, level of inequality and other key variables of interest in their analyses (e.g. Crookston et al., 2014).

Most of the literature that compares a larger number of countries seldom includes the country context in the discussion (e.g. Chmielewski, 2019; Jerrim and Micklewright, 2012, 2014), whereas that comparing two countries does so more frequently (e.g. Blanden et al., 2012; Ojala and Talts, 2007). For the literature comparing between 2 and 10 countries, the picture is more mixed. Some literature does take the country context into account (e.g. Bradbury et al., 2012, 2015; Crookston et al., 2014); some does not (e.g. Duncan et al., 2012).

Authors who consider context when interpreting results do so in steps (e.g. Bradbury et al., 2012). They start with descriptive comparisons, then incorporate the context with the aim of enabling a better understanding of the differences and similarities in their results. Some take an additional step and use results to inform policy. Blanden et al. (2012), for example, aim to identify policies that "appear to be most effective at improving social mobility" (2012, p. 143). Others focus extensively on descriptive comparisons and provide numerous regression results, diluting the value of a comparative analysis, such as Georgiadis (2017).

Bradbury et al. (2019) start by providing one piece of contextual information prior to discussion results, a piece of information essential to the motivation of their research question: that the US income distribution is wider than that of the other three countries studied (UK, Australia and Canada). They then report on important determinants of SES gaps in achievement and how these are distributed across each country (using absolute measures of SES), consequently providing the reader with the relevant country context. Among the determinants of SES gaps in achievement, they discuss maternal education, family structure, maternal age, paternal nativity, maternal employment patterns and ECE attendance.

One successful attempt at integrating country context is Bradbury et al. (2012). I use this article to inform my cross-country comparisons. Authors document the emergence of inequality in the early years (specifically five years old) across four countries: Australia, Canada, the UK and the US. Bradbury et al. (2012) succinctly provide country context at the beginning of their chapter, then integrate it into the interpretation of results and the conclusion. To facilitate a discussion comparing countries across key areas of interest, the authors provide a table summarising these, which include: inequality, child poverty, per capita social expenditure on a specific population, and public expenditure as a share of total health expenditure. This initiates a discussion on the performance of these countries according to these areas, which leads on to a detailed discussion on the differing policy approaches to early childhood care and education. Each country receives approximately a paragraph of attention. Before delving into the analysis, the authors provide some descriptive statistics for each country, illustrating key demographic differences between the countries that cannot be overlooked when comparing across them, and discuss these in detail. Finally, when discussing the results and concluding their analysis, the authors again incorporate country contexts.

Limitations

The literature I reviewed takes the limitations of cross-country comparative studies into account to varying degrees. There is a clear association between those who consider the limitations of their study and those who err on the side of caution when interpreting results. The latter prefer to draw cautious conclusions rather than "blockbuster" or superficial ones. Crookston et al. (2014), for example, write that "collecting data in different contexts also introduces complications. Education systems vary and so it is not possible to use identical measures of parental schooling in each country" (2014, p. 7). Magnuson et al. (2012) write that "[school] factors differ across countries, we can not attribute all cross-country difference to the influence of schools" (2012, p. 256).

Fernald et al. (2012) acknowledge that their study is severely limited by not incorporating "economic, cultural, political, or historical information about any of the countries [leaving them] unable to comment on any country-related issues that may explain" their findings (2012, p. 17276). Linberg et al. (2019) do similarly, acknowledging that "finding comparative measures for educational attainment is not easy, as educational systems very considerably between industrialised countries…" (2019, p. 5).

Georgiadis (2017), on the other hand, states, "my results suggest that policies that seek to improve the material circumstances of the household and mother's education and socioemotional competencies may be effective in promoting child cognitive and socioemotional development in low- and middle-income countries" (2017, p. 28). Bradbury et al. (2019) acknowledges that, due to data limitations across countries, they are not able to adjust incomes for taxes paid and conduct a robustness check (using alternative data to estimate taxes paid). Jaramillo and Tietjen (2001) conclude that "ECE improve[s] all children's cognitive development scores" (2001, p. 29), pointing to causality where, given their methods and data, a causal link cannot be demonstrated. Lopez-Boo (2016) concludes, rather than starts with "not all mediators [e.g. urban residence, ECE, height-for-age, or parental education] operate equally in all countries" (2016, p. 507). Linberg et al. (2019) come to a similar conclusion.

3.5 Conclusion

The aims of this comparative study are to situate one country's results in its own country context, relative to that of the others, and to use policy and contextual differences and similarities across countries to inform the interpretation of results.

In this chapter, I have justified the analytical approach to cross-country comparisons that I employ in the rest of the thesis. I have reviewed the literature on small-n country comparisons examining the associations between SES or ECE with achievement. After considering the options, I have chosen the approach best suited to my research questions, the approach that, will produce the most robust statistical results (separate regression analyses for each country) and that offers the most informative and policy-relevant discussion of results. When applying this approach, I adhere to the following analytical principles. I aim to:

- be explicit about the limitations of the comparative study, particularly when interpreting results;

- provide relevant information on the countries' context to both ground the study in something concrete and strengthen my analytical approach;
- begin with a descriptive comparison of the results, then draw on the pertinent context of the country to interpret results and finally discuss plausible reasons for these similarities and differences;
- conduct robustness checks of results;
- focus on the comparative nature of the study, rather than on a specific country.

Reflections on conducting a cross-country comparison

Conducting a cross-country study is not an easy task. It involves considerable data cleaning, Stata codes to write, various country contexts to take into account and several methodological challenges to consider. However, I found it a useful and rewarding exercise, as I exploited the Young Lives data set for its intended purpose. Scores of researchers have worked tirelessly over the decades to collect this data, to design and coordinate measures that are as consistent, reliable and comparable, collected at the same time (i.e. when children are the same age), all necessary to produce a data set to support cross-country analyses.

This cross-country comparison offers a broader perspective on three main research questions, which single case-studies would not have allowed to answer. Indeed, with a single case-study it would have been challenging to ascertain whether one country's findings were exceptional or not, for no other reference points would have been available to evaluate how similar or different they are. A case-study would not have kindled the same intellectual curiosity to try to understand why, despite quite diverse country contests, there is such consistency across countries in terms of SES achievement gaps across childhood, as well as in achievement trajectories by SES. I would not have found that India appears to be an exception among the four countries when it comes to the relationship between ECE and achievement.

There is also something about the nature of the human mind where it needs a reference point from which to understand something new. I posit that a cross-country comparison is perhaps more intuitively comprehensible than a single case-study as it provides various reference points. This is especially important in a research field with limited studies to use as comparisons.

Chapter 4

Young Lives Data

4.1 Introduction

A thorough understanding of the strengths and limitations of one's data is key to informing analytical choices, as well as to assess the quality of the research that uses the data.

This chapter offers a detailed discussion of the Young Lives data, specifically its aims, design features, implementation of the design, data collection methods, data quality, and implications for analysis. First, however, I situate Young Lives in a broader context of similar studies, particularly those in low- and middle-income countries. Two aspects of Young Lives stand out from the comparison: its cross-country design and its comparatively low attrition rates.

The chapter is structured as follows: Section 4.2 offers a brief background to Young Lives, Section 4.3 situates Young Lives in the broader context of other child cohort studies, Section 4.4 outlines how the Young Lives project implemented its study design, Section 4.5 discusses the panel's quality, and Section 4.6 concludes.

4.2 Young Lives: a brief background

Context and aims

At the dawn of the new millennium, there was great optimism when world leaders came together and agreed to achieving the Millennium Development Goals (MDGs) by 2015. Many of the MDGs relate to child development, such as those on achieving universal primary education and reducing child mortality, hunger and extreme poverty.

In order to facilitate achievement of the MDGs, the UK Department for International Development (DFID) sought to better understand the causes and consequences of child poverty in low- and middle-income countries with the aim of shaping policy debate and programme design (see Young Lives, 2015).

DFID therefore commissioned the Young Lives study, a panel study tasked with tracking children's lives over the 15-year life span of the MDGs across four low- and middle-income countries, and has since received funding to extend the study. Study countries include Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam. (At the time of selection, Andhra Pradesh and Telangana were one state, Andhra Pradesh; they were split in 2014.) The former Young Lives Director, Professor Jo Boyden, explained that countries were "selected to include one from each of the major regions of the developing world, along with a range of political-economic conditions and circumstances, with strong institutional capacity locally to undertake complex panel research being another crucial criterion" (Boyden & James, 2014, p. 26).

At the start of the project, in 2002, according to World Bank classifications, Ethiopia, India's state of Andhra Pradesh (which in this thesis I refer to as India, in line with Young Lives documentation) and Vietnam were low-income countries and Peru a lower-middle-income country.

Young Lives' initial aims were as follows:

- "To produce good-quality long-term panel data about the changing nature of the lives of children growing up in poverty.
- To trace linkages between key policy changes and child welfare.
- To inform and respond to the needs of policy makers, planners and other stakeholders" (quotes from both Attawell, 2003, p. 2 and Wilson et al., 2006, p. 352).

Sampling design features

Young Lives is a panel study, specifically a birth cohort study (because the children were born at a similar time; Young Lives started surveying the younger cohort when they were aged six to eighteen months old). Panel studies follow the same individuals across time. Their advantage over repeated cross-sectional studies is they allow us to track change for individuals over time.

As a birth cohort study, Young Lives' observation units are children, not households. It is not a household panel. Birth cohorts offer three key advantages over household panel studies, the first being wider samples of the target population (children) and hence more precision. Bradbury et al. (2012) note that panel surveys are usually small in size, given the costs of panel maintenance and tracking respondents. Bradbury and Jenkins (2001) write that, in high-income countries, household survey "[s]ample sizes of around 2000 to 5000 households are common" (Bradbury & Jenkins, 2001, p. 48) but that these include childless households. As sampling errors may be sizable, Bradbury and Jenkins (2001) stress it may be impossible to establish, with precision, changes for particular groups of interest, such as children. This problem can be overcome by studying a similarly sized sample of children rather than households.

The second advantage of birth cohort studies is the level of detail in the data collected. Given differences in the population of interest, child-level data from household panels are usually less detailed than those from a birth cohort study. For example, measuring children's educational achievement through household panels would require as many age-appropriate instruments as there are children's age ranges.

For research questions investigating phenomena in an age group, a third advantage of birth cohort studies is that the children are of roughly the same age. In household panels, children's ages vary substantially, and the timeframe of the panel may not include the life stages of interest. For example, a household panel would limit a researcher interested in the role of early childhood in subsequent development to a smaller sample of children aged three to six years old in the early stages of the survey, whereas a birth cohort study can include all children in such an analysis. For research questions that require variation in age, this is not an advantage.

Young Lives collects some retrospective data (e.g. a history of school attendance). Such data are simple and cheap to collect, but this comes at the cost of measurement error (see Rose, 2000). Recollection of past events is not easy for most people, and the longer into the past they are asked to recall, the higher the risk of measurement error. In addition, people's current situation shapes how they remember the past. Retrospective surveys are better suited to measuring prominent events in people's lives – events that are easily remembered, such as births, mortality and marriage. They are not well suited to examining processes or regular events such as school attendance.

The Young Lives samples, as Table 4.1 shows for the younger cohort, include three non-probability samples and one probability sample at the country level. The first stage of sample selection saw 20 sites per country chosen to capture each country's regional, geographic, ethnic and other diversity (see Boyden et al., 2012, p. 476). In Ethiopia, India and Vietnam, this stage involved the judgement sampling of 20 sites, whereas in Peru it involved multistage cluster random sampling.

Table 4.1: Young Lives sample design for the younger cohort (born in 2001/2)

Young Lives (younger cohort)	**Cohort size**	**Sampling**	**Attrition (Round 1-5)**	**Refusal rate in Round 1**[1]	**References**
Ethiopia	1,999	Non-probability (judgement and convenience) of 20 sites, random selection of 1 village within the site, random sampling of households with a singleton child born from April 2001 to June 2002 in the village and surrounding area until reached quota of 100	9.4%	0%	(Outes-Leon & Sanchez, 2008; Woldehanna et al., 2008; Young Lives, 2014a)
India	2,011	Non-probability (judgement) of 20 sites, random selection of 4 villages within 19 sites, judgement sampling of 3 slum areas in Hyderabad; selection of households with a singleton child born in 2001-2002 in village and surrounding area until reached quota of 100	5.5%	< 1% (n=14)	(Galab et al., 2008; Kumra, 2008; Young Lives, 2014b)
Peru	2,052	Probability sample (multistage cluster random) of 20 sites; census of eligible households in site and surrounding area until reached quota of 100	9.4%	<5 % (n=130)[2]	(Escobal & Flores, 2008; Escobal et al., 2008; Young Lives, 2014c)
Vietnam	2,000	Non-probability (judgement) sample of 20 sites; simple random sample of 100 eligible households in site.	3.1%	< 2% (n=36)	(Nguyen, 2008; Young Lives, 2014d)

Notes: [1]Nonresponse rate is for both the younger and the older cohort in each country. [2]Personal correspondence with the Peru principal investigator shows 130 children refused. Escobal et al. (2008) reported 88 children, but this was subsequently revised.

4.3 Young Lives in context

There are various ways to collect information on a birth cohort. When the target population is small, a census may be practical. Census data are representative of the target population because censuses collect information from all eligible participants. However, when the population is large, a sample offers two main advantages: (i) it can provide reliable information on the population at much lower cost and, with a probability sample, it is possible to quantify the sampling error; (ii) data collection is quicker so that data are available sooner (see Lohr, 2010, p. 18).

According to Lohr (2010), it is possible to classify survey designs into the following groups:

- probability samples that represent a target population, including simple random samples, stratified random samples, cluster samples, systematic samples and multistage probability samples that involve a mixture of the aforementioned designs; and
- non-probability samples that are not representative of the target population and include samples of convenience and judgement samples. Judgement sampling is also referred to as *deliberate* or *purposive* sampling.

Some studies are randomised controlled trials (RCTs), whose sampling design we can group according to the classification by Lohr (2010). However, RCTs take an additional step and randomly assign units to two (or more) groups: an experimental group and a control group. The marginal difference between the groups is the outcome of interest. In a cluster RCT, the unit of randomisation is not the individual or the household but rather a group of subjects exposed (or not) to an intervention (e.g. villages).

To achieve representativeness in a longitudinal survey, researchers take a probability sample from the target population. However, representativeness has to do not only with initial sample selection but also with the following rule employed. The following rule specifies which persons to interview in subsequent rounds. Rosenzweig (2003) showed how the survey's following rule shapes how representative the initial sample remains. Using Living Standards Measurement Study (LSMS) panel data, he showed that the following rule of excluding split households in follow-up surveys resulted in a biased sample when analysing reasons for moving. Following rules may also include only tracking children who remain within a particular area and excluding those who migrate; this would bias the sample if the characteristics of those migrating are different from those remaining in the area. A following rule that will least affect representativeness would track migrating children and collect information on their new location.

The Young Lives team also collects rich longitudinal qualitative data, though it's use is beyond the scope of this thesis.

Why non-probability sampling designs are chosen over probability sampling designs

When the target population is national, a non-probability sample is usually preferred over a probability sample for several reasons. The sampling frames are often inadequate; accessing areas that are almost inaccessible for geographic or conflict-related reasons would increase study costs to unreasonable levels; or governments restrict access to some populations or areas (see Boyden & James, 2014, p. 15). Sampling frame inadequacy in low-income countries is a key constraint to taking probabilistic samples. Unlike the UK's Millenium Cohort Study (MCS), which made use of administrative data from UK electoral wards (existing on 1 April 1998) to randomly select clusters, not all Young Lives countries had such information. Only Peru, the wealthiest country of the four, had the information available to achieve a probability sample at the site level. For the remaining countries, the study used non-probability sampling to select the 20 sites.

In addition to sampling frame limitations, time spent travelling to respondents represents a substantial cost element. A probability sample of 20 sites in a country with poor road infrastructure may select sites located in areas so remote that the cost of reaching them is exorbitant. In Ethiopia, for example, which arguably had the worst road infrastructure of the four countries, the team chose to purposively sample sites with better road access.

Why cluster sampling may be chosen

In most settings, but particularly low-income settings, achieving a simple random sample of children aged one year old with a limited budget is challenging, given sample frame inadequacies and geographic and logistical limitations. Cluster sampling is one way to construct sampling frames at a more reasonable cost than through constructing a sampling frame for the whole country (see Groves et al., 2009, p. 106). Cluster sampling also addresses some of the geographic and logistical constraints of low-income countries, where road infrastructure is poor or non-existent and some areas may be off-limits. In countries with a rainy season, poor road infrastructure may mean that roads are non-navigable at times. Thus, restricting the sampling frame to specific areas, rather than across the whole country (as simple random sampling would require), is logistically and financially more efficient, particularly given the trade-off between spending person-hours on travelling to and from sites versus actually interviewing respondents. Cluster sampling is often the sampling design of choice for low-income countries. For example, many Demographic and Health Surveys (DHSs) and Multiple Indicator Cluster Surveys (MICSs) use a multistage-clustered sampling approach. Cluster sampling is also the preferred sampling strategy in high-income settings, for the same reasons.

Attrition is a major concern in longitudinal research, and keeping track of respondents can be cumbersome and expensive. Cluster sampling is particularly helpful in a study which involves young people, who are more likely than older people to migrate (whether for education, marriage or work). In the Philippines' Cebu Longitudinal Health and Nutrition Survey (CLHNS) and the Jamaican birth cohort study, for example, over the first 11 years there was 30 and 83 per cent attrition, respectively. With high levels of attrition, there is no reassurance that the sample will become biased.

Through community-level cluster sampling, researchers can establish relationships with communities, their members and leaders. Relationships facilitate the tracking of children who have moved, as neighbours and leaders can then supply information on their movement (as discussed in Wilson et al., 2006, p. 355). Tracking individuals is especially important for longitudinal research in all settings, and especially in settings where means of communication (e.g. telephone or e-mail) and administrative data are limited.

Challenges of sampling children in low-income settings

While sampling eligible children, Young Lives country teams also faced sample frame inadequacies. Unlike the MCS, which used the UK Child Benefit Register to sample children, up-to-date and accurate databases are rare in low-income settings, in particular for target populations of households with young children. Indeed, for an up-to-date sampling frame of households with children aged one year old, researchers need a sampling frame with information dating back only one year prior to the start of the research. Additionally, given the high infant mortality rates in low-income settings, such a sampling frame may already be out of date a year later. Each Young Lives country therefore developed its own approach to randomly selecting eligible children within each site. I discuss these below.

The higher infant mortality rates in low- and middle-income countries also contribute to higher rates of attrition. This is why Young Lives sampled children over six months of age in the younger cohort, so as to avoid high attrition due to infant mortality.

Young Lives in the context of birth cohort studies in low- and middle-income countries

It is important to situate Young Lives in a broader context of similar studies, specifically birth cohort studies in low- and middle-income countries. I identified eighteen studies of low- and middle-income countries by searching for “cohort profiles” in the *International Journal of Epidemiology*, and from papers related to the UN Children’s Fund (UNICEF) International Symposium on Cohort and Longitudinal Studies in Low and Middle Income Countries, Banati and Zharkevich (2014), and those that Dworsky (2014), Harpham et al. (2003), and Hill (2004) reference. Of the eighteen studies identified, eight are censuses of a target population, five are probability samples, and five are non-probability samples.

In Annex C, Table C.1 outlines the studies, giving the study name, location, cohort size, rounds completed at the time of writing, sampling design for a target population, following rule, attrition rate, notable features and the information sources. While some studies include as observation-units the mothers of index children, their offspring or other members of the household, or subsequently add other children to the sample, this table focuses on the initial birth cohort the studies sampled.

The reported attrition rate reflects all mechanisms of sample loss over the whole study, so includes death, nonresponse, refusal, non-contact and inability to respond. To compare with the Young Lives sample that spans 15 years, I report attrition rates across studies at 11 years (or thereabouts) and also at the most recent data point available. (Table 4.8 provides this information for the Young Lives country samples.)

Two high-income country studies

The two birth cohorts from high-income countries are the UK’s MCS and the US’ Fragile Families and Child Wellbeing Survey (henceforth Fragile Families). The MCS provides an almost gold standard of birth cohort sampling designs. Using available sampling frames, the MCS designed a multistage stratified random sampling design that sampled 18,827 children born at the turn of the century, around the same time as the Young Lives younger cohort. In both the MCS and Young Lives, the sampling unit is the household but the unit of observation is the child. The MCS did not restrict itself to singleton births as Young Lives did. It also collected data on children at nine months and at three, five, seven and 11 years. Young Lives covered the same age ranges but did not collect data at three years old. At 11 years, the MCS attrition rate was 31 per cent, much larger than that of Young Lives (5.5 per cent) at the time.

Fragile Families provides an example of a sampling design from a high-income country that does not aim to achieve a nationally representative sample, but rather to represent a particular section of society: non-marital hospital births between 1998 and 2000 in US cities with populations of 200,000 or more (see Reichman et al., 2001). Like Young Lives, Fragile Families combines probability and non-probability sampling in the first stage of sampling. Using a multistage random sampling design, they first selected the 16 cities included in the study, then hospitals, then births. The survey also used non-probability (judgement) sampling to select an additional four cities. In total, 4,898 children make up the initial sample. Fragile Families did not sample births in smaller cities and in rural areas; Young Lives was not interested in the wealthiest children in its respective countries, rather in poorer children, or the 95 per cent, as is specifically the case in the Peruvian sample. Fragile Families collected information on children at birth and at one, three, five and nine years old; by nine years it had experienced 31 per cent attrition.

Differences between the MCS and Young Lives lie in sampling frame availability, study resources, and country infrastructure. Without limitations in these areas, Young Lives would have aimed for a sampling design similar to that of the MCS. Given their research focus on a specific population, Young Lives' and Fragile Families' sampling designs are somewhat similar. Both use probability and non-probability sampling and sample specific sites based on specific criteria. One key difference is that Fragile Families selected three-quarters of its cities through probability sampling, whereas only one-quarter of Young Lives sites (i.e. the sites in Peru) were selected this way.

Eight census studies in low- and middle-income countries

Six of the eight census studies concern city-level censuses: Pelotas, Brazil; Soweto, South Africa; West African cities; and a defined area of New Delhi, India. One study was a census of births at a hospital (CSI Holdsworth Memorial Hospital, Mysore, India). All of these followed only the children who remained in or near the city. In South Africa, the Birth to Twenty study (BT20) initially followed only those in the vicinity, but then gradually expanded the area to include children who migrated within Gauteng province. As mentioned earlier, the exclusion of children who migrate out of the sample may bias the results.

Sampling births in an urban area (or a hospital) is more convenient and cost effective than doing so in rural areas, where travel costs can be substantially higher. The same is true when following children who remain in the city's vicinity. Obviously, a census of births in a city provides an accurate perspective of the situation in that city but does not reveal how rural children or children in other urban centres fare.

The sixth census study is a national-level study: the Jamaican 1986 Birth Cohort Study. From an initial sample of all births in Jamaica in 1986, the study experienced the highest attrition rate of all birth cohort studies identified – 91 per cent at 18 years. This high loss rate is due mostly to the study's following rule, applied for financial reasons, stipulating that only children who remained in or moved into two south-east parishes (Kingston and St. Andrews) be followed. Representativeness obtained from the initial census of births was lost because of this following rule.

For Young Lives to achieve its aims, the sampling design of Young Lives Peru is preferable to any of the above designs as it captures the experiences of children across the country (rather than in just one city). In addition, its following rule does not undermine its representativeness.

Five probability sampling designs in low- and middle-income countries

Three of the five probability samples are country-level samples. Two of these are programme evaluations that spanned two years: the Transfer Modality Research Initiative (TRMI) evaluation in Bangladesh and the El Proyecto Integral de Desarollo Infantil (PIDI) evaluation in Bolivia. Both studies had two waves. Concerns related to attrition, loss to follow-up and transport costs are less important in a two-year two-wave study than in a multiple-wave study that spans 15 years, as Young Lives does. Resources not spent on tracking children over 15 years and travelling to respondents for multiple waves can otherwise be spent on obtaining a country-level probability sample. That said, the PIDI evaluation lost 51 per cent of its sample after just two years.

The other country-level probability sample is Chile's Encuesta Longitudinal de la Primera Infancia (ELPI). If we were to rank the sampling designs of all birth cohort studies identified in low- and middle-income countries (see Table C.1), the ELPI's would top the chart. It is comparable with that of the MCS. It covers 15,175 children sampled from a multistage stratified sample and follows them if they remain in Chile. It has completed two rounds, in 2010 and 2012, with 15 per cent attrition between these rounds.

Two of the five probability samples are city-level samples: the CLHNS in Cebu (Philippines) and the Limache Birth Cohort Study in Limache (Chile). Their following rules also stipulate the following of children remaining in the city or nearby.

On the one hand, for the purposes of the Young Lives study, the probability sampling designs in the programme evaluations are not appropriate as their span is too short. Young Lives is concerned with the changing nature of children's lives over a period of 15 years and beyond. Nor are the city-level probability samples appropriate – for the same reasons that city-level censuses are not appropriate.

While the ELPI sampling design suits Young Lives' aims quite well and, had the resources and infrastructure been available, the Young Lives sampling design may have looked quite similar. However, there are two important considerations to bear in mind when comparing the ELPI with Young Lives. First, the most important challenge for Young Lives in terms of sample frame was to select a design that would be feasible (financially and logistically) given that the study covers four countries. Unlike any of other cohort studies in low- and middle-income countries, Young Lives needed to administer a relatively standardised set of instruments across all study countries and maintaining a coordinating mechanism (i.e. the Young Lives team at the University of Oxford) over 15 years is quite costly.

Second, the Young Lives countries were low- or lower-middle-income countries at the outset, whereas in 2010 Chile was an upper-middle-income country and, since 2010, a member state of the Organisation for Economic Co-operation and Development (OECD). The resources, administrative data and infrastructure available to achieve a nationally representative probability sample of births and children were, hence, superior to those in most Young Lives countries.

Five non-probability sampling designs in low- and middle-income countries

Two of the non-probability samples are also cluster RCTs: the Institute of Nutrition of Central America and Panama (INCAP) Nutrition Trial Cohort Study in Guatemala and the Andhra Pradesh Children and Parents Study (APCAPS) in Andhra Pradesh, India. Both examine the effect of a nutritional intervention and chose clusters, or villages, based on researchers' judgements. They then followed all eligible children in each village. The INCAP study followed all children who remained in Guatemala, whereas the APCAPS followed only those children who remained in the site.

Two studies used judgement sampling to select sites that captured the diversity of the population of interest. The Prospective Cohort Study of Thai Children (PCTC) is interested in children born in Thailand and the Anhui Birth Cohort Study in children born in Anhui Province in China. Similar to the Young Lives sampling design, the PCTC used judgement sampling to select five study sites in different parts of the country. In its second stage, it invited all pregnant women intending to remain in the area for at least five years to participate. It then followed children born to these women if they remained in the study site's vicinity. The Anhui study purposively sampled six cities from all regions of the province and invited all women in their first or second trimester to participate. It followed children born in the hospital where the mother had had her antenatal appointments.

The Mauritius Longitudinal Child Health Study – the fifth non-probability study – focused on two cities, selected by convenience and judgement sampling. Reasons for selecting these cities were "their central locality to the research laboratory [...] and because they contained a racial mix very similar to the racial distribution of the island as a whole" (Raine et al., 2010, p. 1443). Using administrative records, it sampled all children born in 1969, who were three years old at the time.

For the aims of the Young Lives study, the sampling designs of the cluster RTCs are not relevant as Young Lives is not examining the effect of an intervention. The Mauritius Longitudinal Child Health Study and the Anhui Birth Cohort Study are limited in scope compared with Young Lives, as they focus on cities.

The first stage of the PCTC sampling design is similar to that of Young Lives, but the second stage generates additional biases in the sample, whereas Young Lives takes a random sample of eligible children within the sites. The PCTC invites only those children whose mothers plan to remain in the area, excluding those who may migrate. Its following rule further excludes children who move or who have mothers who gave birth at home or somewhere other than the hospital where they received antenatal care, thus biasing results further, as those who migrate may be different in an important aspect, to those who do not migrate.

Among the birth cohort studies with non-probability sampling designs, Young Lives is the most appropriate for its own objectives as it samples children from both urban and rural areas and randomly samples children within sites, reducing any further bias in the sample.

Cross country panel studies

Young Lives is the only cross-country study of children in low- and middle-income countries. In 2006, a collaboration began between five of the birth cohort studies listed in Table C.1. According to Richter et al. (2012), this collaboration "had at least 15 years of follow up and an initial sample size of 2000 or more newborns" (2012, p. 621). The collaboration took COHORTS as its acronym, which stands for the Consortium of Health-Orientated Research in Transitioning Societies. It includes the 1982 Pelotas Birth Cohort, the INCAP study, the New Delhi Birth Cohort, the CLHNS and the BT20 cohort. Using these five cohort studies, COHORTS addresses a series of research questions, its most recent one being: What are "the effects of rapid weight gain vs rapid linear growth during early life on adult human capital and health, as well as consequences of rapid subsequent weight gain in relation to stunting at 2 years"? (Richter et al., 2012, p. 622).

The advantage of designing a cross-country panel study from its inception is that it does not face the challenges that cross-country studies like COHORTS does, which include:

> (i) differences in variable definitions, or in measurement techniques across sites (this affects exposures, outcomes and confounding variables), which means that major effort has gone into producing a common data set; (ii) the different ages of individuals across the five cohorts and the different time periods they reflect; (iii) the different ages for which data are available throughout infancy and childhood; (iv) heterogeneity in the

> results for some of the analyses, for example, those on body composition (Richter et al., 2012, p. 625).

As a result, the COHORTS team invested an enormous amount of technical work pooling their datasets across the five studies. (Chapter 3 discusses the advantages and limitations of a cross-country study with different sampling designs.)

Unique aspects of Young Lives

Two aspects of Young Lives stand out from Table C.1: none of the birth cohort studies provides a cross-country comparison and Young Lives has the lowest attrition rate among studies that span at least 11 years. Boyden and James (2014) attribute the low attrition in Young Lives to the team carefully tracking children between data rounds and maintaining good relationships with the children, their families, communities and local officials (see page 21). Section 4 of this chapter, on panel quality, considers attrition in detail.

The dual cohort and the mixed-methods design are also unique to the Young Lives study, but I do not exploit these in this study. With an older cohort, results can show how outcomes are changing for the cohorts across a 7-year period (for example, comparing the two cohorts at the same age shows a decline in learning outcomes in India and no improvement in Ethiopia, as reported in Young Lives (2017a, 2017b). The longitudinal qualitative research provides granular information that can help us to better understand some of the trends from the survey data.

Summary

For its stated aims, the ideal sampling design for Young Lives would have been a design akin to that of the ELPI. However, to work within several limitations, Young Lives' sampling design had to be adapted. Its actual design is more appropriate than those of any of the existing census studies as it collects data on both rural and urban areas and has low attrition rates. Compared with non-ELPI probability sampling designs, the Young Lives sample has a broader scope than a purely urban focus and a longer timeframe than the two-year, two-wave programme evaluations. Compared with the non-probability sampling designs that are also not RCTs, Young Lives' scope is wider than a city focus, and its sampling design within sites and its following rule contribute to maintaining the original sample. Comparison with RCT sampling designs is not appropriate for Young Lives as it is not an RCT. Two aspects of Young Lives stand out from the comparison: its cross-country design and its comparatively low attrition rates.

4.4 Young Lives sample design and data collection

All Young Lives country teams were asked to sample 20 sites and then randomly sample 100 children aged 6–18 months within each site. Sample selection varied between countries; the country-specific reports referred to below describe the relevant details. As will become clear from these descriptions, each country took great care to make the Young Lives sample as illustrative as possible of the country's diversity (in terms of location type (rural/urban areas), ethnicity, caste, language, religion, and more).

While the study was primarily interested in children growing up in poverty, researchers appreciated that children move into as well as out of poverty. Young Lives sampling designs therefore did not limit themselves to sampling "poor" children, but rather over-sampled them, and also included children who are not poor.

Only singletons of the appropriate age were eligible for the study; it excluded twins and triplets. In households with more than one eligible child, the study randomly chose one child (as discussed in Wilson et al., 2006).

Targeting the population of households with children aged 6–18 months took into account higher mortality risk in early infancy and an increase in maternal mobility around the time of birth (e.g. travelling to her parents' village) (see Wilson et al., 2006, p. 353). Selecting this age group, therefore, reduced attrition rates across the four countries.

The Young Lives team selected an additional 50 households within each site, with children aged 7.5–8.5 years old, to compare with the younger cohort as well as to pilot data collection on older children (see Attawell, 2003, p. 2). (Owing to resource constraints, in Peru only 714 older cohort children were enrolled. Although Escobal and Flores (2008) state 716 children, the Young Lives internal documentation on attrition and the Young Lives dataset show 714 children.) However, as there is no information on the initial conditions of these children nor on current attendance in early childhood education (ECE), these cross-cohort data are of limited use for my research aims. Retrospective data were collected in Round 3 and/or 4 depending on the country. This implies a recall period for the older cohort of up to 16 years in Ethiopia, to when they were three years old. Given the measurement error associated with such long recall periods, I do not use these data. Figure 4.1 summarises the younger cohort's dataset.

As its following rule, Young Lives tracks children to their new communities, creates a new community ID, implements new community questionnaires and collects information on the index children's new household, if they form part of a new household.

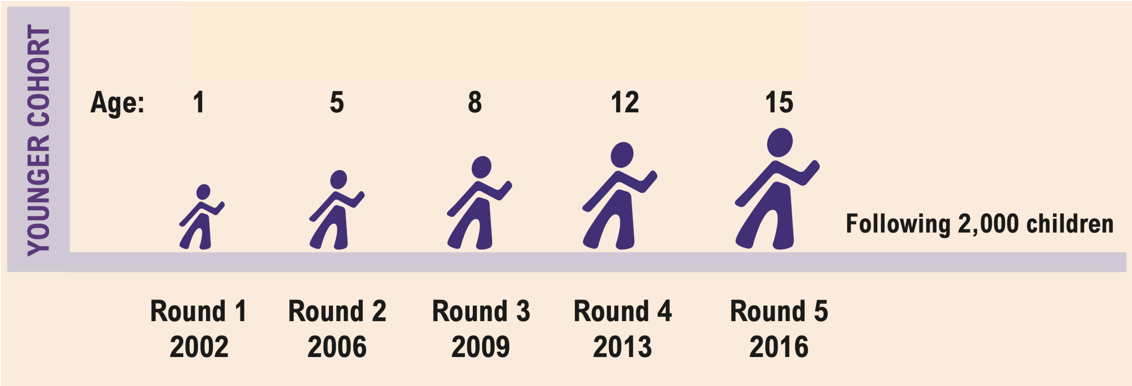


Figure 4.1: Young Lives birth cohort

Sampling terminology

Young Lives documentation refers to the sampling strategy as adapted from “sentinel site surveillance” methods, which are commonly used in epidemiology and public health. (A basic search of the term “sentinel surveillance” in the *International Journal of Epidemiology* (ranked among the top journals in its category) produced 59 articles, mostly involving human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) research but also other diseases, including measles in Zaïre and respiratory diseases in Switzerland (search on 30 July 2015).) According to World Health Organization (WHO), data collected from a well-designed sentinel surveillance study “can be used to signal trends, identify outbreaks and monitor the burden of disease in a community, providing a rapid, economical alternative to other surveillance methods” (2018). This is echoed in Wetterhall and Noji (1997), who write that sentinel surveillance is “useful in monitoring the incidence of disease and injury” (Wetterhall & Noji, 1997, p. 53) in certain situations, such as when “time and resource constraints prohibit collecting information through population-based surveys” (Wetterhall & Noji, 1997, p. 53). As its name indicates, sentinel site surveillance monitors sites rather than individuals. Therefore, it does not follow individuals who move away from the site but does include new entrants into the site.

The team leading Young Lives at the time of sampling was made up largely of public health academics, which perhaps indicates why the term “sentinel site surveillance” was chosen to describe the sampling strategy. However, Young Lives sampling differs from a sentinel site surveillance approach in two key aspects. First, Young Lives’ aim is not to monitor the burden of disease or to identify outbreaks. It is not an epidemiological study and, if it collects data on public health issues, that is because these relate to the aims of the study. Second, Young Lives does not limit itself to the study of selected sites; it follows children when they leave their communities. Indeed, Wilson et al. (2006) recognise that “[t]he analogy is not perfect” (Wilson et al., 2006, p. 358). They note that site selection in Young Lives is akin to what Shadish et al. (2002) term "purposive sampling of heterogeneous instances" where "the key methodological implication is to deliberately sample for heterogeneity on as many population characteristics as are known and practical" (Shadish et al., 2002, p. 356).

Early Young Lives documentation does not use the term "cluster sampling" to describe its sampling approach. Harpham et al. (2003) explains that a probabilistic cluster sampling approach would not be appropriate for a "small" sample of 20 sites, as it would not account for the wealth of information available about the country (see Harpham et al., 2003). These authors argue that a research team may know that one area of the country is especially prone to famine and malnutrition and wish to include it, but simple random sampling would not ensure its selection. That said, the Peruvian sampling design achieved a probabilistic sampling design that selected a diverse set of sites. (Sánchez and Escobal (2020), however, rephrase "sentinel sites" as clusters.)

As the next section will make evident, a more accurate description of the Young Lives sample design, for all countries but Peru, is "judgement sampling of 20 sites (or clusters), followed by probability sampling within sites (or clusters)". In Peru, it is "multistage clustered random sampling".

Implementation of the Young Lives sampling design

Ethiopia

For Ethiopia, Young Lives used a non-probability (judgement and convenience) sample of sites, then a probability sample of households within each site (see Outes-Leon & Sanchez, 2008, pp. 2–6). The team purposefully selected sites to over-sample areas with food deficiency to capture Ethiopia's diversity across regions and ethnicities in both urban and rural areas, and to keep sampling costs manageable, which entailed selecting accessible (rather than remote) sites (see Outes-Leon & Sanchez, 2008).

The sampling frames included a list of nine regions, districts (*woredas*) in selected regions, communities (referred to as peasant associations in rural areas or *kebeles* in urban areas) within each district, and villages in each peasant association or *kebele* selected.

Sampling involved the following steps:

1. **Regions**: Judgement sampling of five of the nine regions in the country, in order to maximise national coverage (the five regions account for 96 per cent of the national population).

2. ***Woredas***: Judgement sampling of three to five districts in each region (to gain a balance of rural and urban, poor and non-poor). (There were no national

statistics at the district level, so the team carried out classification and selection in consultation with local officials in each district (see Outes-Leon & Sanchez, 2008, p. 6).).

3. ***Kebeles***: Judgement and convenience sampling of at least one peasant association (mostly rural) or *kebele* (mostly urban) in each district, considering road access. These are the sentinel sites (see Table 4.2 for a description and Figure 4.2 for their geographic location). If there were not enough households to fulfil the sampling criteria of 100 households with a one-year-old child and 50 households with an eight-year-old, the peasant association or *kebele* was considered the centre point around which the sentinel site was established. This happened in five cases.

4. **Households**: Interviews of all the households on the periphery of the village until it was possible to enrol 150 eligible households (100 for the younger cohort and 50 for the older cohort) (see Outes-Leon & Sanchez, 2008, p. 6). Woldehanna et al. (2008) say the approach was simple random sampling, which appears to contradict Outes-Leon and Sanchez (2008), unless the team employed a random route approach that started at the periphery of the village (see European Social Survey (ESS) sampling documentation that uses this approach – ESS, 2012). However, Young Lives documentation on the matter is not clear.

Eligibility criteria for the younger cohort included the child being a singleton birth between April 2001 and June 2002, still alive at the time of sampling. For the older cohort, children had to be born between April 1994 and June 1995 (see Woldehanna et al., 2008). If a selected family had children eligible for both the younger and the older cohort, the younger child was included in the sample because that cohort required a larger number needed to be enrolled (see Young Lives, 2014a).

Nonresponse was zero; no caregiver refused to respond to the survey. According to Woldehanna et al. (2008), this is not unusual for Ethiopian surveys in their first rounds.

Outes-Leon and Sanchez (2008) compared the 2002 Young Lives data with the 2000 DHS and showed the Young Lives households in all rural and urban areas, barring Addis Ababa, to be wealthier and with better access to services. In Addis Ababa, Young Lives households were poorer and had less access to services.

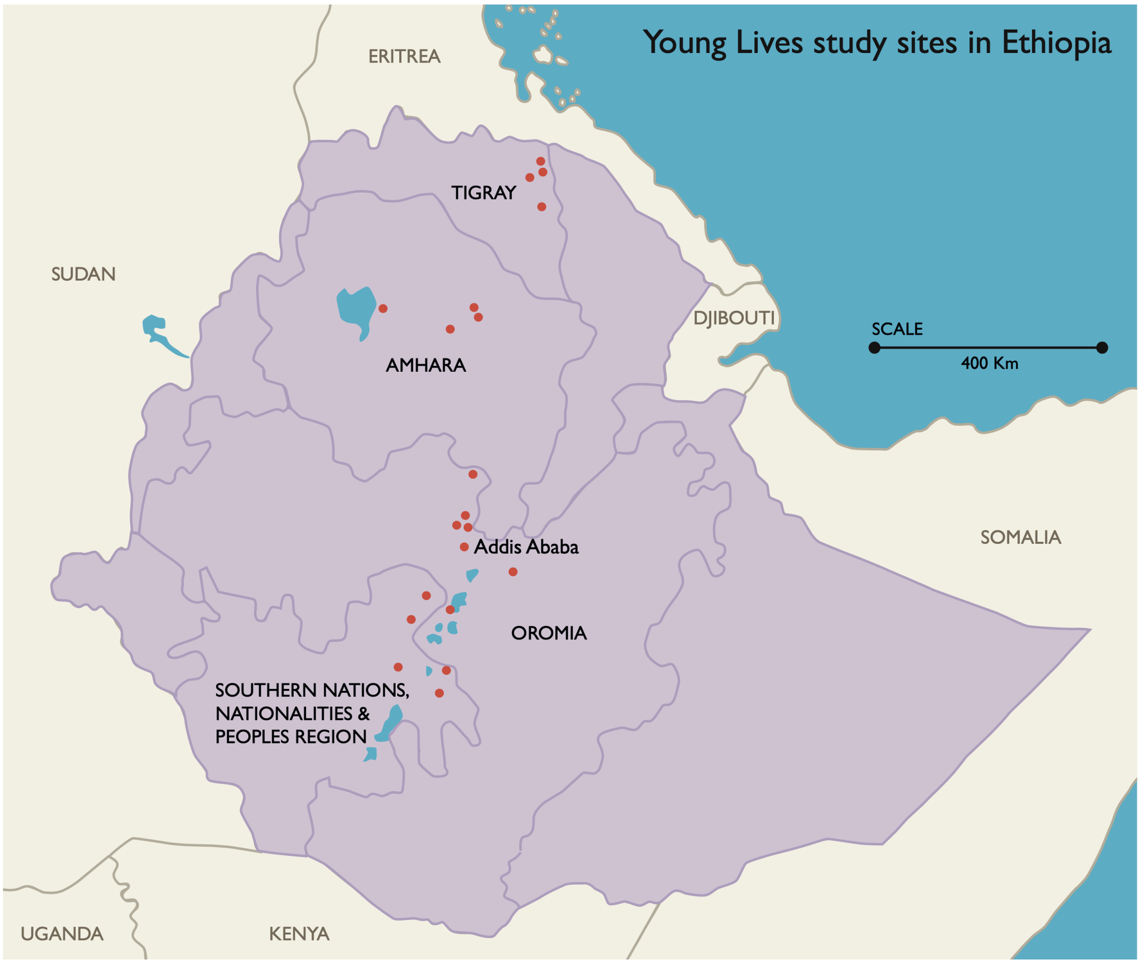
Outes-Leon and Sanchez (2008) also compared Young Lives data with the 2000 Welfare Monitoring Survey (WMS) and found that Young Lives households were poorer based on several assets such as land, home and livestock ownership. Contrary to the DHS wealth measure, which includes access to services, the WMS data do not cover access to services. In terms of access to services, the 2000 WMS data were

Table 4.2: Young Lives sites in Ethiopia by region

No.	Zone	Description
Capital city		
1	Addis Ababa	Overcrowded area in the centre of the city
2	Addis Ababa	Industrial area in the southern part of the city
3	Addis Ababa	Slum area in the city
Amhara		
4	North Wello	Tourist town with some extremely poor neighbourhoods
5	North Wello	Poor rural community
6	South Gondar	Rural area near Lake Tana
7	South Gondar	Rural food-insecure area
Oromia		
8	East Shewa	Rural area near Lake Ziway
9	Arsi	Drought-prone rural area
10	North Shewa	Fast-growing town
11	East Shewa	Relatively rich rural area on the outskirts of Debrezeit town
Southern Nations, Nationalities and Peoples Region		
12	Gurage	Densely populated rural area growing enset ("false banana")
13	Wolayita	Densely populated town
14	Hawassa City Administration Zone	Fast-growing business and tourist town
15	Sidama	Coffee-growing rural area
16	Hadiya	Poor and densely populated rural community
Tigray		
17	Southern Tigray	Drought-prone rural area highly dependent on government support
18	Eastern Tigray	Extremely poor rural area dependent on the Productive Safety Net Programme and other government support
19	Eastern Tigray	Small, very poor urban town
20	Eastern Tigray	Model rural area known for its success in soil and water conservation

Source: Adapted from Alemu et al. (2003)

Figure 4.2: Young Lives study sites in Ethiopia



Source: Young Lives (2018a)

similar to the 2000 DHS data. These findings are in line with the sampling design that prefers poorer areas with better road access.

The population of inference would thus be children born at the beginning of the millennium, from the selected villages and urban neighbourhoods (covering three–five districts within each of five selected regions of Ethiopia) that have higher levels of food deficiency and relatively good road access.

India (Andhra Pradesh and Telangana)

The Young Lives India team took a non-probability (judgement) sample of sites, then a probability sample of households within each site (Kumra, 2008, pp. 2–6). The team selected sites to ensure a broad spread of sites across three regions of the then-state Andhra Pradesh and at least one poor and one non-poor site in each region.

At the time of sampling, Telangana was part of Andhra Pradesh. In 2014, it became a separate state (with a population of over 35 million). The sampling frames used by the Andhra Pradesh team included a list of regions (Coastal Andhra, Rayalseema and Telangana); 23 administrative districts in the state; and administrative *mandals* (akin to counties) in each district.

Sampling involved the following steps:

1. **Region**: Stratification of districts by region and poor/non-poor status.

2. **District**: Judgement sampling of six out of 23 districts: one poor and one non-poor per region.

3. ***Mandals***: Judgement sampling of 20 *mandals* based on a classification of relative development (that included economic, human development and infrastructure indicators). Since there were few urban *mandals* (where more than two-thirds of the population lived in an urban area), the team selected the state capital, Hyderabad, as one site. The remaining 19 sites came from the six districts previously selected. These are the sentinel sites (see Table 4.3 for a description and Figure 4.3 for their locations).

4. **Villages**: Division of each *mandal* into four contiguous geographical areas and random selection of one village from each area. Where these villages did not give rise to sufficient eligible households, the team included additional villages (see Kumra, 2008). For Hyderabad, three slum areas were selected. When there were insufficient eligible children in the selected sample villages, the team included additional villages.

5. **Households**: Compilation, prior to data collection, of a door-to-door listing to identify eligible children (see Kumra, 2008). However, Kumra (2008) does not state what this list was used for: whether to take a simple random sample, to ensure sufficient children in the study area to sample, or for another reason.

Nonresponse in Round 1 was less than 1 per cent for both the younger and older cohort of Young Lives children: 14 households refused to participate in the study (see Galab et al., 2003, p. 20). When this occurred, the team identified and enrolled replacement households. A total of 3,033 eligible households were invited to participate in the study.

Kumra (2008) compared 2002 Young Lives data and the 1998/99 DHS survey and found Young Lives households to be slightly wealthier, with more assets and better access to public services. This may be due, at least in part, to the overall improvement of living standards in Andhra Pradesh between 1998 and 2002.

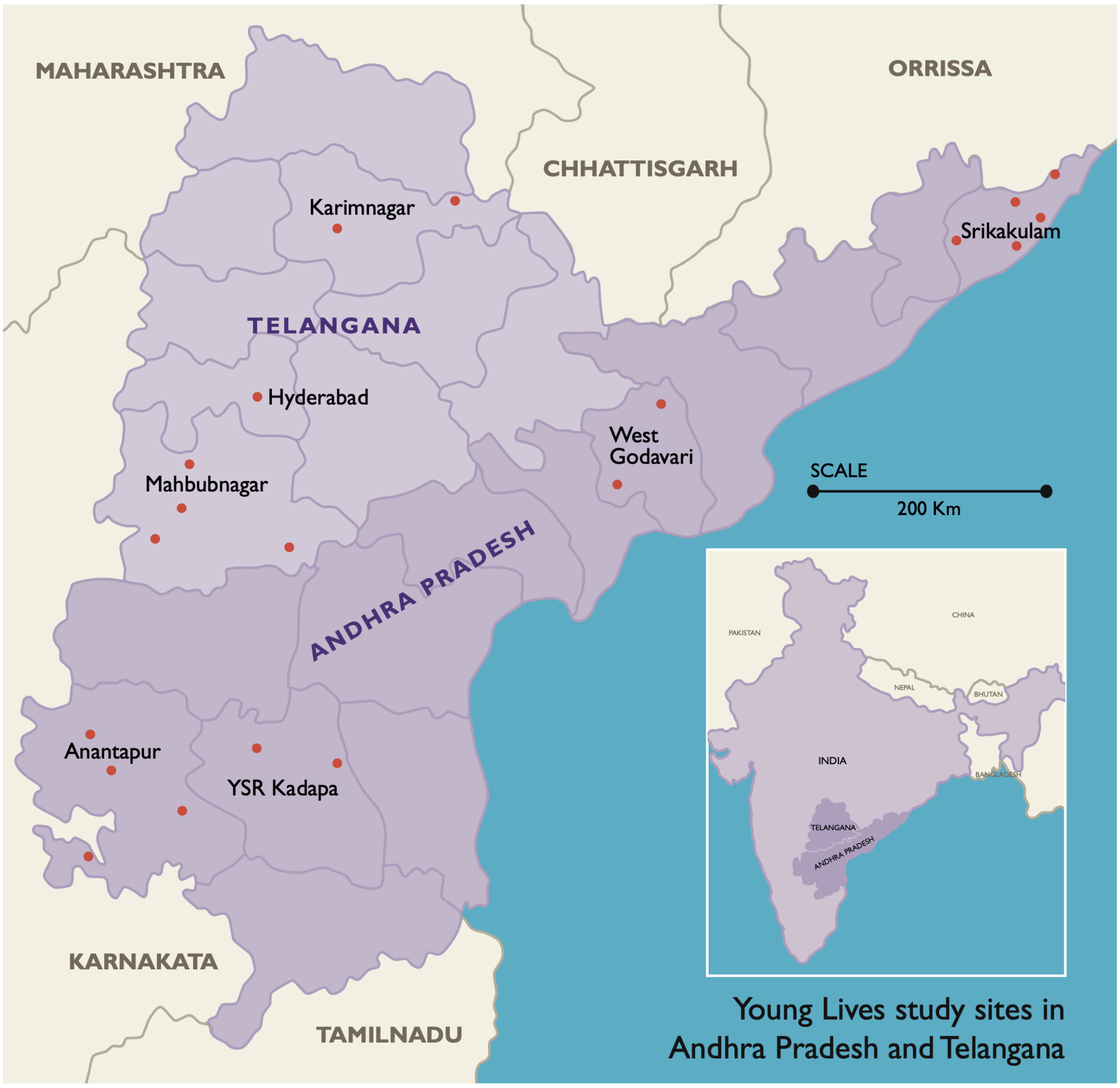
On some individual indicators, the Young Lives sample appeared to be worse off. Households were less likely to own a house; mothers were less likely to breastfeed,

Table 4.3: Young Lives sites in India by region

No.	Zone	Description
Coastal Andhra (Andhra Pradesh)		
1	West Godavari	Urban area in a well-developed coastal region
2	West Godavari	Tribal *mandal* in a well-developed coastal district
3	Srikakulam	Town in the north
4	Srikakulam	Tribal *mandal* in the north
5	Srikakulam	Rural *mandal* in the north
6	Srikakulam	Rural *mandal* in the north
7	Srikakulam	Rural *mandal* with a mix of tribes and non-tribes in the north
Rayalaseema (Andhra Pradesh)		
8	Kadapa	Rural *mandal* in the heart of the region where agriculture is the main occupation
9	Kadapa	Remote rural *mandal* in a forested part of the region
10	Anantapur	Urban site, which is a district headquarters
11	Anantapur	Poor rural *mandal* affected by Naxalite movements
12	Anantapur	Poor rural area spread across hilly areas and affected by Naxalite movements
13	Anantapur	Rural *mandal* bordering the neighbouring state
Telangana		
14	Karimnagar	Medium-sized town in the north with people of mixed religion
15	Karimnagar	Rural area in the north affected by Naxalite movements
16	Mababubnagar	Rural tribal *mandal* in the forest areas of the south
17	Mababubnagar	Rural *mandal* in the south with people moving in seasonal migration
18	Mababubnagar	Rural *mandal* in the south with high incidence of child labour and seasonal migration
19	Mababubnagar	Very poor *mandal* in the south
State capital		
20	Hyderabad	Densely crowded area in the state capital

Source: Adapted from Kumra (2008) and Young Lives (2014b)

Figure 4.3: Young Lives study sites in Andhra Pradesh and Telangana



Source: Young Lives (2017d)

receive antenatal visits or be vaccinated against tetanus; and caregivers had lower levels of education (Kumra, 2008).

The population of inference would, therefore, be children born at the beginning of the millennium from the 20 selected *mandals* from six regions of the states of Andhra Pradesh and Telangana.

Peru

The team in Peru used a probability (multistage cluster sample) sampling design, see Escobal and Flores (2008, pp. iii–4) and Young Lives (2014c). Peru is hence distinct from all the other study countries, which used non-probability sampling to select sites.

The sampling frames included a poverty map of the 1,818 districts of Peru from

2000 (the Poverty Index included infant mortality rates, housing, schooling and access to services), a list of census tracts within districts (a census tract is a small geographical area that one census worker can cover in a short time), and a list of *manzanas* (blocks of houses) in each census tract.

Sampling involved the following steps:

1. **Districts**: Systematic sampling of 95 per cent of districts that had been divided into equal population groups and ranked according to their poverty levels. The team excluded the wealthiest 5 per cent of districts (all in the capital city, Lima). It then conducted 10 computer-generated systematic selection runs and chose the list that best satisfied the project requirements (see Escobal & Flores, 2008).

2. **Census tract**: Random selection of one census tract in each district. These are the sentinel sites (see Table 4.4 for a description and Figure 4.4 for their geographic location);

3. ***Manzanas***: Random selection of one *manzana* per district.

4. **Households**: Within each block, a census of eligible households until it was possible to select 100 eligible households. If necessary, the team visited a neighbouring block to reach the stipulated number of eligible households. A total of 36,375 households were visited (see Escobal et al., 2008, p. 12). (Boyden and James (2014) state that 36,153 were interviewed, but, based on personal correspondence with the principal investigator of the Young Lives Peru team, the correct number is 36,375.)

Nonresponse in Round 1 was less than five per cent for both the younger and the older cohort of children: 130 households refused to participate in the study, 77 from the younger cohort and 53 from the older cohort. (Although Escobal et al. (2008) say that 88 households refused to participate, 53 from the younger cohort and 35 from the older cohort, personal correspondence with the principal investigator of the Young Lives Peru team revealed that the true figures were closer to 130 households refusing to participate – 77 from the younger cohort and 53 from the older cohort.) When this occurred, the team identified and enrolled replacement households. A total of 2,896 eligible households were invited to participate in the study.

Annex C, Section C.2 provides the pertinent data and weights. However, in my analysis, I do not use weights as I do not aim to generalise my results to 95 per cent of the districts in Peru. Most of the analysis on the Young Lives data does not use weights.

Based on comparisons by Escobal and Flores (2008), Young Lives households are slightly wealthier than DHS 2000 households. They also have better access to public services, are better educated, and have high levels of vaccinations and antenatal care. Compared with the 2005 census data, Young Lives households also have better access to electricity and drinking water. However, when applying post-stratification weights to both Young Lives and DHS 2000 data, using the 2005 census as the base, the differences between the two narrowed considerably – with the exception of differences in access to health services and antenatal care. When Escobal and Flores (2008) compared Young Lives with the Peruvian LSMS 2001, poverty rates of the Young Lives 2002 sample were similar in both urban and rural settings.

The population of inference is, therefore, children born at the turn of the millennium, excluding those from the wealthiest 5 per cent of districts in 2000.

Table 4.4: Young Lives sites in Peru by region

No.	Zone	Description
1	Tumbes	Small city on the northern coast
2	Piura	Poor coastal rural area
3	Piura	Very poor rural area in the northern Andean highlands
4	Amazonas	Very poor rural area in the north of the country
5	San Martin	Poor rural area
6	San Martin	Medium-sized city
7	Cajamarca	Medium-sized city in the northern Andean highlands
8	La Libertad	Shanty town on the outskirts of a medium-sized city on the northern coast
9	Ancash	Poor rural area in the central Andean highlands
10	Ancash	Medium-sized city in the central Andean highlands
11	Huánuco	Very rural area in the centre of the Andean highlands
12	Lima	Large urban district located in the north of the city
13	Lima	Large urban district located in the eastern part of the city
14	Lima	Large urban district located in the south of the capital city
15	Junín	Poor rural area in the Amazon
16	Ayacucho	Very rural poor community in the southern-centre of the Andean highlands
17	Ayacucho	Poor rural area in the southern-centre of the Andean highlands
18	Apurimac	Poor rural area in the southern Andean highlands
19	Arequipa	Small city on the southern coast
20	Puno	Medium-sized city in the southern Andean highland

Source: Adapted from Young Lives (2014c)

Figure 4.4: Young Lives study sites in Peru

Source: Young Lives (2018b)

Vietnam

The team in Vietnam took a non-probability (judgement) sample of sites, then a probability (simple random) sample of households within each site (see Nguyen, 2008, pp. 2–5). Sites were selected based on poverty levels, representation of common provincial and regional features, commitment from local government, feasibility for research logistics, and population size, though Young Lives faced some political complications as the government refused to allow Young Lives to study in certain areas.

The sampling frames included a list of the eight socioeconomic regions of Vietnam (North-West, North-East, Red River Delta, North Central Coast, South Central Coast, South-East, Central Highlands and Mekong River Delta) and a "Cities" region containing all of the major urban centres (Hanoi, Ho Chi Minh City, Da Nang, Hai Phong and Ba Ria-Vung Tau) (see Nguyen, 2008, pp. 2–5), provinces by region, communes by province, and eligible children per sentinel site. A commune is an administrative area with a local government, primary school, health centre, post office and market.

Sampling involved the following steps:

1. **Regions**: Judgement sampling of five out of nine regions in Vietnam (eight regions and one that included all major urban centres). The team purposively selected regions to include the North, Central and South regions, urban, rural and mountainous areas and the over-poor, and to reflect some unique aspects of the country including the consequences of natural disaster and war (see Nguyen, 2008, p. 5).

2. **Provinces**: Judgement sampling of one province per selected region,

3. **Communes**: Judgement sampling of four commune sites in each province. (This selection was made in consultation with provincial governments and based on ranking communes by poverty level (poor, average, better off and rich) on the following criteria: infrastructure, percentage of poor households and child malnutrition status (see Nguyen, 2008, p. 5).) If the selected commune had a population of less than 6,000 persons, the team visited a commune with similar socioeconomic conditions to ensure that it would be possible to identify 100 eligible children in the sentinel site (see Tuan et al., 2003, p. 14). Therefore, the 20 clusters are made up of 31 communes. See Table 4.5 for a description and Figure 4.5 for their geographic location). Due to administrative boundary changes since the initial selection in 2002, there were 33 communes in 2008.

4. **Households**: Within each site, the team made a list of all eligible children. Criteria for eligibility are broadly similar, but not clear in Young Lives documentation. Young Lives (2014d) states that the sample included children born between January 2001 and May 2002; Tuan et al. (2003) state that children eligible for selection included children born between 1 January 2000 and 31

> December 2001; the Young Lives data include children with birthdates between May 2000 and September 2002. A simple random sample with replacement was taken from this list of eligible children.

Nonresponse was below 2 per cent for both the younger and the older cohort of Young Lives children: 36 households refused to participate in the study (see Tuan et al., 2003). When this occurred, the team identified and enrolled replacement households. A total of 3,036 eligible households were invited to participate in the study.

Consistent with the sampling methodology applied in Vietnam, Nguyen (2008) found that Young Lives households were generally poorer than the average Vietnamese household, in terms of both assets and access to services. Nguyen's comparison was with the DHS 2002 and the Vietnam Household Living Standards Surveys (VHLSS 2002).

The population of inference is, therefore, children born at the beginning of the millennium from 20 selected communes: from five (of nine) regions, one district was chosen, and within the district, four communes.

Summary

Young Lives researchers in all countries used a cluster sampling approach, but the site selection method varied across countries. In all countries, these sites were purposively sampled, with the exception of Peru, where they were randomly sampled (from 95 per cent of the possible districts, the wealthiest 5 per cent having been excluded). When compared with their respective nationally representative samples, the Ethiopian Young Lives sample was found to have better access to services but to be slightly poorer in terms of assets. This was in line with the sampling design that preferred poorer areas with better road access. In India, the Young Lives sample was slightly wealthier than nationally representative sample households and had better access to services. A reduction in poverty between the time of the nationally representative surveys and the Young Lives survey may explain this difference. In Peru, Young Lives households were slightly better off than nationally representative samples in terms of assets and access to services. In Vietnam, Young Lives households were significantly poorer and had less access to services than did national samples.

Table 4.6 outlines the baseline characteristics of the children sampled. Young Lives sampled approximately 2,000 children per country, their average age just below one, with around half in each country being male, and the proportion living in urban areas ranging from 20 per cent in Vietnam to 69 per cent in Peru. Caregivers in Peru had the highest average level of education, at 7.7 years.

Table 4.5: Young Lives sites in Vietnam by region

No.	Zone	Description
South Central Coast		
1	Phu Yen	Inland flood-prone rural community with high rate of poverty in 2002 but has improved since and is now not so poor
2	Phu Yen	Coastal community with average rate of poverty
3	Phu Yen	Very poor mountainous community with mostly ethnic minority groups
4	Phu Yen	Relatively prosperous coastal community, with shrimp farming
Mekong River Delta		
5	Ben Tre	Poor flood-prone coastal area with difficult transport links
6	Ben Tre	Inland area with a slightly above-average poverty rate
7	Ben Tre	Inland flood-prone area, with difficult transportation but a relatively low poverty rate
8	Ben Tre	Relatively prosperous inland area with good transport links
North-East		
9	Lao Cai	Among the poorest mountainous communities in Lao Cai province, with mostly ethnic minority groups, very difficult transportation and little infrastructure
10	Lao Cai	Very poor mountainous area, with mostly ethnic minority groups and underdeveloped infrastructure
11	Lao Cai	Poor mountainous area with mixed ethnic groups
12	Lao Cai	Very poor mountainous area, with mixed ethnic groups and underdeveloped infrastructure
Red River Delta		
13	Hung Yen	Prosperous rural area, with high population density and good infrastructure
14	Hung Yen	Poor rural area, near a major city and with good infrastructure
15	Hung Yen	Rural rice-producing community, with good infrastructure
16	Hung Yen	Poor rural area, with a high population density and good transport infrastructure
Cities		
17	Da Nang	Urban neighbourhood with mostly blue-collar labour and average infrastructure
18	Da Nang	Mostly prosperous urban area with very good access to services
19	Da Nang	Relatively poor suburb, with quite poor environmental conditions and transportation
20	Da Nang	Newly developed urban and fishing community, with average infrastructure and poor environmental conditions

Source: Adapted from Nguyen (2008) and Young Lives (2014d)

Figure 4.5: Young Lives study sites in Vietnam

Source: Young Lives (2018c)

Data collection

Who collects the data?

Data collection occurred in partnership with various national research organisations. In Ethiopia, the partner was the Ethiopian Development Research Institute (EDRI)

Table 4.6: Baseline characteristics of children born in 2001-2002

	Ethiopia	**India**	**Peru**	**Vietnam**
Number of children	1999	2011	2052	2000
Age in months, mean (SD)	11.7 (3.6)	11.8 (3.5)	11.5 (3.5)	11.6 (3.2)
Male, n (%)	1056 (52.8)	1081 (53.8)	1027 (50.0)	1030 (51.5)
Urban, n (%)	700 (35.0)	508 (25.3)	1406 (68.5)	400 (20.0)
Caregiver: highest grade completed, mean (SD)	2.4 (3.6)	3.3 (4.5)	7.7 (4.4)	6.4 (3.5)

Source: Author's calculations.
Notes: Values in parentheses are as indicated in the first column.

across all five rounds. In the first round, the Department of Economics at the University of Addis Ababa was also involved. In India, data collection was in partnership with the Centre for Economic and Social Studies (CESS) across all five rounds. In Peru, data collection was done by the Instituto de Investigación Nutricional (IIN), with the Grupo de Analisís para el Desarrollo (GRADE) providing some supervision and technical assistance across all five rounds. In Vietnam, the data was collected by the General Statistical Office (GSO) across all five rounds, in collaboration with the Research and Training Centre for Community Development (RTCCD) in the first round and the Centre for Analysis and Forecast at the Vietnamese Academy of Social Sciences (CAF-VASS) in subsequent rounds. Partnership with the GSO entails the surveys being administered by government census enumerators (see Morrow, 2009, p. 10).

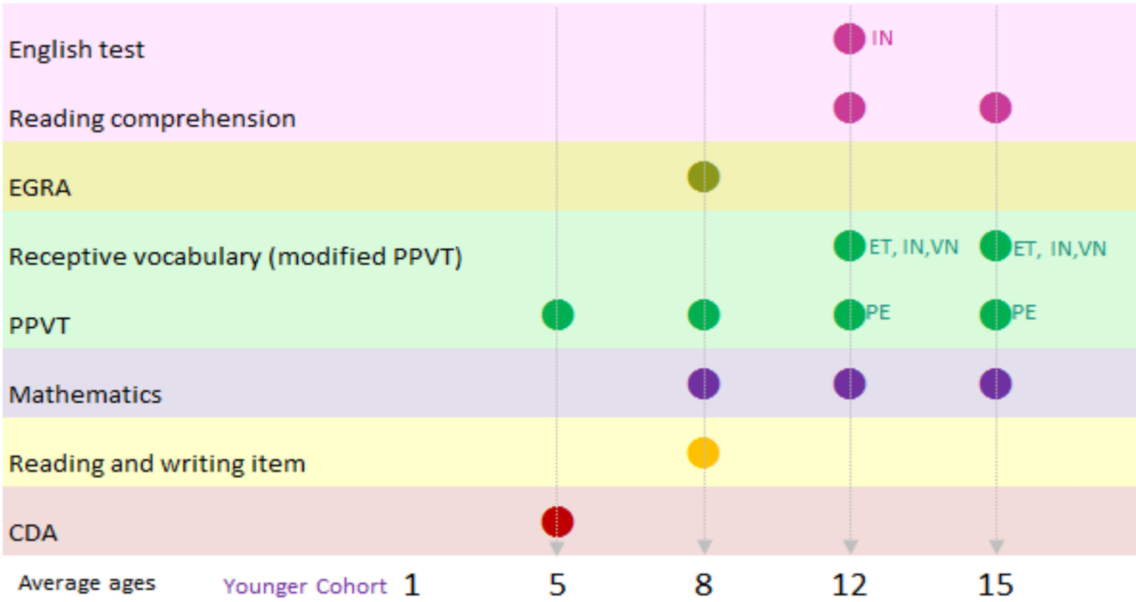
What instruments are used?

Young Lives used three different questionnaires: a household questionnaire to collect basic information on all household members, detailed information about the child's caregiver, as well as information on the household (expenditure, housing quality, etc.); a child questionnaire to collect information on child-related topics of interest; and a community questionnaire to collect community-level information. Young Lives piloted all survey instruments in all countries prior to their implementation (see Young Lives, 2011b).

To enable comparisons across countries, Young Lives used a common set of instruments and research methods. Further, the design of the survey questionnaires are intended to produce data that are comparable across countries and time (see Boyden & James, 2014, p. 16).

Young Lives applied several instrument to test academic achievement, covering receptive vocabulary, reading, and maths (see Figure 4.6). (To allow for sibling-level fixed effects, in Rounds 3 and 5 the Peabody Picture Vocabulary Test (PPVT) was administered to the siblings of the younger cohort children in Ethiopia, Peru and Vietnam. In Rounds 4 and 5, it was administered to a younger sibling. In India in Round 4, a maths test was administered to the younger siblings of the younger cohort children.)

Figure 4.6: Summary of achievement tests administered to the younger cohort



Notes: Cognitive Development Assessment (CDA) is the maths tests administered to 5 year olds

Source: Revollo and Scott (2022, Figure 1)

Who are the respondents?

The respondent changed from caregiver to child as the children grew up; as the children matured, the questionnaires were reformulated and adapted to the respondent's age (see Boyden & James, 2014). Young Lives administered child questionnaires to children when they were 7 or 8 years old or older (see Young Lives, 2011c). The younger cohort children, therefore, had the child survey administered in Round 3 onwards. Prior to this, the household questionnaire collected all child-level information, from an adult in the household, often the caregiver.

The teams gave self-administered questionnaires (SAQs) with sensitive questions on issues including smoking, alcohol and violent and sexual behaviours to the older cohort in Rounds 3 and 4 and to the younger cohort in Round 5 (see Young Lives, 2017c).

Community informants, who may include local mayors, municipal leaders, government officials, village headmen, health, education or agricultural authorities, leaders of women's groups, religious leaders or representatives of grassroots organisations, received community questionnaires (see Young Lives, 2002a, 2002b). Where secondary data, such as census data, were available, Young Lives used these, too.

What data are collected?

Table 4.7 provides an overview of the type of data the Young Lives survey collected. When administered, the child questionnaire collects data on the child's school and work activities, their education and schooling environment, their health, feelings, attitudes and perceptions, their social networks, social skills and social support, how they use their time and household issues. The SAQ collects information on sensitive issues. Children also take age-appropriate tests to assess their receptive vocabulary, literacy and numeracy development (see Table 4.6).

The household questionnaire contains a roster that collects basic information on all household members and on household education, parental background, livelihoods and assets, food and non-food expenditure, social capital, shocks, wealth, child activities, child health, caregivers' perceptions, attitudes, aspirations for the child and the family, and their mental health. Round 1 collected information on pregnancy, delivery and breastfeeding from the younger cohort.

All rounds collected anthropometric information on the index child and, in particular rounds, on the younger cohort's mothers, fathers and younger siblings.

Young Lives collects some retrospective data on children's school histories, listing, for every year since the child's birth, the type of school and grade attended and the school's location. In India and Peru, this is collected in Round 3, and a top-up is collected in Round 4. In Ethiopia and Vietnam, all the information is collected in Round 4, although, for Vietnam, information is only collected back to 2005. For the younger cohort in Ethiopia, information on ECE requires respondents to recall up to nine years; the older cohort needs to remember up to 16 years.

Arguably, Young Lives school history data provide only an imprecise indication of children's schooling trajectories rather than a true record of their schooling history, year by year. A comparison of "current" and "recent" preschool attendance (collected in 2006/07) in Ethiopia and "past" preschool attendance (2013/14) show notable discrepancies in both directions: caregivers claiming the child did and did not attend preschool when in the earlier survey they said otherwise. The total discrepancy is 248 children. As a reference point, 476 children's caregivers reported that the child had attended or was attending preschool in 2007. Such discrepancies are not uncommon, when the recall periods are so long. I therefore do not use the retrospective data, rather the data collected when children were 5 years old.

The community questionnaire collects data on the community's socioeconomic, demographic and environmental context. Specific variables include population, ethnicity, religion, language, economic activity, employment, infrastructure and

Table 4.7: Summary of data types collected in the Young Lives surveys

Child-level data
General characteristics (e.g. sex, first language, ethnic group, religion, age in months)
Anthropometric measures
Birth (e.g. birth weight, birth attended by skilled health personnel)
Immunisation
Illness, injury and disability
Smoking and drinking habits
Reproductive health knowledge
Subjective health and wellbeing
Time-use (including time spent commuting to school or work, in leisure, studying, at school, working, on domestic tasks, sleeping)
Educational services (including ECE, enrolment, type of school attended, highest grade completed)
Achievement scores (Vocabulary, maths and reading*)
Caregiver characteristics (age, sex, level of education, subjective well-being)
Biological parent characteristics (mother and father's age, location, level of education)
Household-level data
Household head characteristics (Age, sex, education, relation to child)
Household size and composition (by gender and age groups)
Livestock ownership
Land and house ownership
Credit and food security
Take up of public programmes
Household shocks (to land, crops, land, livestock, consumer goods, death/disablement or through crime, natural disasters, changes to the family structure or employment, public policy or more)
Household asset index (housing quality, access to services, consumer durables)
Community-level data
General community characteristics
Social environment
Access to services
Economy
Local prices
Educational services
Child day care services
Health services
Child protection

Source: Briones (2018)

political representation. The questionnaire also provides information on a range of health, education and child protection services available to the community (see Young Lives, 2011c).

See Annex H for a detailed discussion on communities. When conducting community-level fixed effects, I use the initial sampling sites (i.e. clusters), and these overlap substantially with the initial 'communities'.

Compensation

It is important to compensate respondents in settings where they are poor and where time spent responding to surveys can disrupt their livelihoods (see Boyden & James, 2014). Each Young Lives country team deals with compensation for time spent responding the survey. Care is taken to ensure that people do not see compensation as a form of coercion or as an incentive (see Morrow, 2009). This is complicated in a context of poverty, and also in cultural settings where consent to government requests is expected and where respondents perceive Young Lives interviewers as government officials. Additionally, as Boyden and James (2014) highlight, there are risks of data "contamination", especially in longitudinal research, where the compensation may affect the household's circumstances.

Morrow (2009) documents the compensations, in Peru the team gave small gifts to the respondent, including a token of 'thank-you' along with some school supplies. In Ethiopia, the team provides monetary compensation and encouraged children to buy materials for school. In India, the team provided resources to local schools but insights led to a decision to compensate respondents for their time in subsequent rounds. I did not identify documentation on compensation in Vietnam. (Personal correspondence with a Young Lives researcher conveyed that Young Lives compensates households who participate in pilot rounds, but not in the data rounds.)

Young Lives also shares their research results with their respondents through 'research reciprocity activities' which provide wider feedback.

4.5 Panel quality and implications for analysis

Lohr (2010) offers a useful framework to assess survey quality:

> Total survey error is the sum of five components: coverage error, non-response error, measurement error, processing error and sampling error. The main concern with undercoverage and nonresponse is bias. Sampling error produces variability in the estimated. Measurement and processing error have both bias and variance aspects (2010, p. 545).

Coverage error

The main concern of coverage error is undercoverage – where a sampling frame does not include part of the target population. Estimating the bias that undercoverage causes is difficult, as the survey provides no information on the characteristics of those not covered.

Undercoverage error in probability samples may bias results if the means for the excluded population differ from the included population means or if undercoverage is substantial. Being explicit about undercoverage is, therefore, important. In the case of Peru, the sampling frame intentionally excluded the wealthiest five per cent of the population. For researchers interested in the whole population, this may constitute coverage error.

Sampling error

While coverage error has to do with units excluded from the sampling frame, sampling error has to do with the sampling procedures.

Shadish et al. (2002) highlight that judgement sampling designs (or "purposive sampling of heterogeneous instances") are not underpinned by a statistical theory that allows for formal generalisation. Inferences made from probabilistic samples assume that all sample members have a known probability of selection, making it therefore possible to make adjustments for population-level inference. With judgement sampling, in contrast, researchers do not know the probability of selecting individuals. Using weights is, therefore, not possible. In addition, formal generalisation requires that selection not be endogenous, in other words that the selection of units not be correlated with the measure of interest (the outcome measure), hence the value of probability sampling, where researchers know the probability of selection, and where weights can be calculated. Where sampling is endogenous, sampling weights can help to correct for the endogeneity (see Solon et al., 2015).

Peru's is the only country team that used probability sampling to select sites. As described above, the team excluded the wealthiest five per cent of sites. The sampling is, therefore, arguably endogenous if the outcome variable is correlated with the poverty level of the site. In such cases, site-level weights are available. As a second stage, across all sites, the team used random sampling techniques to select eligible households, and then eligible children. Inference to the site level is thus possible.

Ethiopia, India and Vietnam used judgement sampling to select study sites. This has implications for representativeness. Estimates derived from Young Lives data should be explicit about which population they are referring to. Country teams have carefully described procedures for site selection, so it is therefore possible to describe the population of inference of the sample. I have done this above, following the discussion on the implementation of each country's sampling design.

Measurement error

Many of the instruments commonly used to measure achievement in high-income countries have found limited application in low- and middle-income settings (see Boyden & James, 2014). Young Lives has pioneered the adaptation of some of these instruments to the local languages and cultures of some low- and middle-income countries. This has come with some challenges. Measurement error also depends on from whom, or from what, the information is collected. Chapter 5 discusses measurement error, outlining all measures used in analyses, item nonresponse and relevant measurement error issues.

There are also challenges in translating key study concepts into child-friendly formats and into the local languages and cultural context of the countries while ensuring comparability of concepts across countries (Chapter 3 on cross-country comparisons discussed this in more detail).

Nonresponse error

There are two types of nonresponse: unit nonresponse and item nonresponse. Unit nonresponse occurs when information for an entire observation unit is missing. Item nonresponse occurs when information for a unit exists but at least one item is missing (see Lohr, 2010, p. 329). I discussed item nonresponse earlier in this chapter.

In Young Lives, unit nonresponse in Round 1 entailed refusal to participate. This averaged 2 per cent across the four countries (see Table 4.1). Survey analysts refer to nonresponse between rounds as attrition, and define it in various ways. Some definitions examine attrition between specific rounds; they may also exclude from their calculations sampling units that drop out for particular reasons, for example that the child died, that they were not contacted (e.g. owing to migration), that they refused to respond (nonresponse), or that they were not capable of responding (e.g. because of illness or disability). As my analysis does not distinguish between types of attrition, I include all reasons for attrition in my calculations of attrition. I also use all five rounds of Young Lives data to examine the determinants of learning trajectories. Thus, the most useful definition of attrition here reflects the portion of the sample lost over the whole study – total attrition.

Total attrition rates range from 3.1 per cent in Vietnam to 9.4 per cent in Peru (see Table 4.8). The highest attrition rate occurred between the first and the second rounds, across all countries (mostly due to mortality). Yet, overall attrition remains extremely low, even compared with other birth cohort studies in low-, middle- and high-income countries. Attrition in the UK MCS over a similar period (from 2001–2002 to 2012) was notably larger, at 31 per cent (see Connelly & Platt, 2014), most of which, across the five waves, was due to refusal (see Mostafa, 2012).

Table 4.8: Attrition for the younger cohort: round-on-round and total attrition

	Ethiopia	**India**	**Peru**	**Vietnam**
Round 1 (6–18 months)				
No. of children	1999	2011	2052	2000
Round 2 (4–5 years)				
No. of children	1912	1950	1963	1970
Attrition* (%)	4.4	3	4.3	1.5
Round 3 (7–8 years)				
No. of children	1885	1931	1943	1961
Attrition* (%)	1.4	1	1	0.5
Round 4 (11–12 years)				
No. of children	1875	1915	1902	1932
Attrition* (%)	0.5	0.8	2.1	1.5
Round 5 (14–15 years)				
No. of children	1812	1900	1860	1938
Attrition* (%)	3.4	0.8	2.1	-0.3
Total attrition (%)	9.4	5.5	9.4	3.1

Notes: * Attrition since previous round.
Source: Young Lives dataset (Constructed Rounds 1–5).

See Table 4.9 for a breakdown of reasons. Between rounds 1 and 5, in Ethiopia, India and Vietnam mortality was the main cause for attrition (with 85, 47 and 14 children deceased by Round 5, respectively). In Peru, almost half of attrition is due to refusal. Note that attrition is general in the Young Lives sample; children may have reentered the sample after leaving (or being lost).

Young Lives had major costs in Peru where the initial sample was more dispersed (across districts) and where migration is very common. In the cases of Vietnam and Ethiopia, the governments discourage migration by requiring that households obtain permission in order to do so; households that migrate without permission would lose their access rights to services such as health and education.

When children are not found in a round, teams continue to track them regardless and, when they find them, they re-enrol them in the sample. Across the four rounds, this happened in 7 cases in Ethiopia, 4 in India, 46 in Peru and 17 in Vietnam.

Sánchez and Escobal (2020) examined bias in attrition across rounds 1 to 5 and found that attrition varies by parental SES, area of residence and ethnicity (or caste

Table 4.9: Reasons for attrition for Rounds 1–4, younger cohort

	Ethiopia	**India**	**Peru**	**Vietnam**
Mortality, n	85	47	25	14
Refused, n	10	3	97	1
Untraceable, n	14	7	38	4*
Living abroad, n	6	0	32	10

Source: Young Lives (2017d, 2018a, 2018b, 2018c).
Notes: * One child in Vietnam was inaccessible as they were in the army or in prison, as reported in (Young Lives, 2018c).

in India). Interestingly, those lost in Peru were relatively poorer, while those in the other three countries were wealthier. Consequently, to address this bias, authors suggest controlling for SES when conducting analyses with Young Lives data.

Processing error

The data processing that underpins the Young Lives databases is impressive. In each data round, Young Lives processes around 12,000 questionnaires of more than 30 pages, administered in 80 sites across four countries (see Young Lives, 2011a). The teams administered paper child and household questionnaires in all countries in Rounds 1 and 2. In Rounds 3 and 4, they used Computer Assisted Personal Interviewing (CAPI). Personal digital assistants (PDAs) were used in Peru and Vietnam in Round 3 to collect 50 and 70 per cent of the data, respectively (see Young Lives, 2011a). In Round 4, all four teams used PDAs. The Young Lives team administered the achievement tests on paper in all rounds. They also inputted community questionnaire responses into a PDA.

Round 2 adopted double entry of all data to reduce processing errors in relation to data from paper questionnaires. Two different people entered the data from each questionnaire and a software programme checked for entry errors (see Young Lives, 2011a). Each country's data manager then checked for discrepancies against the hard copy. Subsequent data cleaning focused on identifying contradictions and checking those against hard copies. Then, as researchers began analysing the data, other inconsistencies emerged and were checked and cleaned. Data cleaning takes several months. Processing is never without errors, yet Young Lives appears to perform quite well at minimising it.

4.6 Conclusion

Using Young Lives data limits my analysis to specific populations of inference in each country: I cannot make inferences about the national population. While there are nationally representative datasets for many low- and middle-income countries, most are not longitudinal and none are designed for cross-country comparisons. To address my research questions, I need a longitudinal data set that is designed for cross-country comparisons.

Another implication is that a non-probability sampling design complicates the derivation and interpretation of standard errors. In a non-probability sample, some characteristics may be over represented or underrepresented in a systematic, but unknown, manner. I therefore consider the standard errors in my analysis as measures of precision in the estimates as they relate to the population of inference. In other words, the standard error of a sample mean, for example, is the standard deviation of the sampling distribution of the mean.

A third implication for analysis is that, although attrition in the younger cohort of the Young Lives study is small (compared to those of other longitudinal studies), it is biased, and varies according to SES, area of residence, and ethnicity/caste across countries. To address this biased attrition, Sánchez and Escobal (2020) highlight the importance of controlling for parental SES, which I do in my analyses.

Young Lives data offer a unique opportunity to examine the factors shaping child development in low- and middle-income countries. Here, I specifically focus on factors shaping children's academic achievement or learning trajectories. The dataset offers a wealth of information on early childhood education between the ages of 3 and 5 years old and on SES at 6–18 months, and at each subsequent life stage observed. In addition, Young Lives data contain a wealth of information to control for several factors that may confound the relationship between early childhood education and parental resources, and achievement. This information is available across the four low- and middle-income countries. Unfortunately, Young Lives does not collect data at three years old, a key stage in achievement and a useful baseline for examining how SES gaps in achievement evolve. I therefore rely on achievement data collected at 5 years old.

Compared with other birth cohort studies, the longitudinal data of Young Lives is unique in its cross-county, cross-cohort and multi-sectoral comparison, as well as in its low attrition rates. For the purposes of examining the changing nature of children growing up in poverty, Young Lives offers insights into children's lives across each country, covering various and key regions and both rural and urban areas. Most birth cohort studies in low- and middle-income countries focus on urban areas, usually a single city.

The Young Lives sampling design shares broad principles across study countries: teams selected 20 sites or clusters and, within each site, randomly sampled eligible children. Ethiopia, India and Vietnam used non-probability (judgement) sampling designs to select the 20 sites, whereas Peru used a probability (multilevel cluster) sampling design. For Young Lives' aims, Peru's sampling design is one of the best among birth cohort studies in low- and middle-income countries. The following rule across all countries is to track children into new communities when they move, limits the introduction of additional bias in the sample and contributing to low attrition rates.

Although the sampling design in three Young Lives countries is not nationally representative, their populations of inference are nonetheless relevant for policy. The surveys I analyse follow approximately 8,000 children born at the turn of the millennium, between ages 6–18 months through to 15 years old. The longitudinality of the data offers unique insights into the factors shaping children's lives across these four countries.

The Young Lives dataset is not without its limitations: there are several challenges to conducting longitudinal data collection in low- and middle-income countries and Young Lives sampling designs reflect these. In some countries, sampling frames were inadequate or simply non-existent, costs of accessing some areas of the country would have been exorbitant given their remoteness (lack of transport infrastructure) or ongoing conflict and government restrictions precluding researcher access. Within these limitations, Young Lives teams across all these countries did, and continue to do, their best to minimise coverage, sampling, measurement, nonresponse and processing errors.

Chapter 5

Contextual Information

Ethiopia, India, Peru and Vietnam differ in many ways and given the regression approach I use, I cannot control statistically for their differences. Therefore, I use contextual information to inform my comparisons across countries, as well as to inform variable selection. In this way, this chapter bridges between the analytical framework and the empirical work that follows it.

Academic achievement is the outcome variable in my regression analysis of the Young Lives data so, in this chapter, I focus on how contextual information might relate to gaps in achievement and achievement trajectories by soceioeconomic status (SES), and also how they may shape the relationship between early childhood education (ECE) and later achievement.

First, to provide an idea of the country context in which the Young Lives children grew up, I discuss each country's levels of social and economic development. Second, I discuss the main dividing lines in society and discuss how these may further embed achievement disparities. Third, I discuss the ECE context and how attendance might contribute (or not) to advantages in achievement and for whom. Finally, I conclude with a discussion on the implications of this contextual information on my analysis.

5.1 The four countries' levels of development

I first compare countries' general level of development when the children were born against that at the most recent round of data collection (at the time of writing) when the children were around 15 years old, in 2016.

There are various indicators of development. Two of these – under-five mortality rate (U5MR) and the prevalence of stunting – are also important predictors of academic achievement. One indicator, the net enrolment rate (NER) in primary school, captures educational participation. The next two indicators relate to economic development: the gross domestic product (GDP) per capita, and how evenly economic resources are distributed in each country (the Gini index). Finally, where data are available, I present Program for International Student Assessment (PISA) maths test scores.

According to UNICEF (2012), U5MR constitutes a good summary measure of development because it reflects a host of social, economic and service-related factors including: nutritional health and education of mothers; level of immunisation and oral rehydration therapy use; availability of maternal and health services; family income; food availability in the family; availability of safe water and safe sanitation; and overall safety of a child's environment. These factors relate directly or indirectly to a child's brain development and ability to engage in learning. U5MR is measured at the individual level and is calculated from the total number of live births occurring during the five years preceding the survey, and is reported as the number of deaths per 1,000 live births.

According to UNICEF, "[s]tunting is considered the most reliable measure of undernutrition, as it tends to reflect recurrent episodes or prolonged periods of inadequate food intake, calorie and/or protein deficiency or persistent or recurrent ill health" (UNICEF, 2012, p. 20). Stunting is measured by height-for-age and reports on the percentage of the child population under five years old whose height-for-age falls more than 2 standard deviations below the WHO Child Growth Standards median. Undernutrition is an important determinant of brain development and thus achievement (see Grantham-McGregor et al., 2007).

According to the UN (2023), NER in primary education measures education coverage. It reflects a country's ability to provide a primary education system to its primary school aged children. Although the parents of the Young Lives children may have been in primary school several years before 2000, the earliest good quality data available for NER in primary school is for 2000. This coincides with the start of the Millennium Development Goals (MDGs). Prior to 2000, data are sporadic. NER in primary school is the proportion of children enrolled at the appropriate level for their age. It is expressed as percentage; 100 per cent means that all children of primary school age are enrolled in primary school.

Levels of inequality provide an indication of social inequality in a society. United Nations University World Institute for Development Economics Research (UNU-WIDER) compile Gini coefficients from countries across the globe. The Gini coefficient (or index), reported in Table 5.1, is based on consumption data for all countries except for Peru where it is based on income data. The Gini index ranges from 0 to 100, with higher numbers denoting greater inequality.

From 2000 to 2015, Ethiopia was the poorest and least developed country of the four under consideration. Ethiopia had the highest U5MR and stunting rates. In 2000, its U5MR was 140 and stunting stood at 57 per cent, though falling: in 2015, its stunting rates were on par with India (at 38 per cent). Ethiopia also had the lowest levels of primary school participation with 40 per cent of primary school aged children attending primary school in 2000 (increasing to 85 per cent by 2015.) It was also the poorest country of the four, with a GDP per capita of 123 current US$ in 2000 that rose to 630 current US$ in 2015.

India was next to last (of the four study countries) in terms of its level of development. India had the next highest rates of U5MR and stunting in both 2000 and 2015 (in 2000, U5MR was at 92 and stunting at 55 per cent), and the next lowest rates of primary school participation (at 80 per cent in 2000). India was also slightly wealthier than Vietnam at the time, with a GDP per capita of 442 current US$, above Vietnam's 395 current US$.

Peru was by far the wealthiest, though arguably not the most developed, nation of the four, with a GDP per capita of 1,941 current US$ that rose to 6,180 current US$. Despite this, its U5MR was worse than that of Vietnam in 2000 (38 versus 30 in Vietnam) though, by 2015, it was less than Vietnam's (at 16 compared with 22). Stunting rates in Peru were better than in Vietnam in 2000 (at 31 per cent compared to 43 per cent in Vietnam and remained so in 2015.

While Peru's under-five children died at a higher rate than Vietnam's, Peru's exhibited a lower percentage of stunting. The rates of primary school participation in Peru from 2000 to 2015 remained high, around 97 per cent, on par with Vietnam.

Vietnam's development indicators stand out prominently in Table 5.1. It's GDP per capita was relatively close to that of India and less than a fifth of Peru's, yet its U5MR rates were a third of India's (92 in India and 30 in Ethiopia). Compared with Peru, Vietnam's U5MR was better and primary school participation rates were remarkably close at around 98 per cent. How Vietnam organised its society seems to reflect that, with the limited resources it had, it was able to do more than India to support children and families. As I discuss with a co-author in Vandemoortele and Bird (2011), following the end of the war in 1975, Vietnam's living standards and macroeconomic performance were poor and the government's legitimacy was precarious. The government focused on improving the lives of Vietnamese people. With a strong centralised government pursuing a long-term development vision and a long-rooted tradition of consensus-based decision-making that avoided creating political losers, Vietnam achieved high-levels of development on par with wealthier countries (such as Peru). Whether these gains were shared equally is debatable, as discussed in Baulch et al. (2007)

The Programme for International Student Assessment (PISA) maths scores for Vietnam are also notable, with a 2015 score of 495, significantly more than Peru's 387, and slightly above the Organisation for Economic Co-operation and Development (OECD) average of 490 (2016). Glewwe et al. (2021) investigate this exceptional performance in 2015 relative to other Young Lives countries:

> Vietnamese 15-year-olds have several advantages over 15-year-olds in the other three Young Lives countries that could lead to higher learning outcomes, such as better nutritional status, fewer siblings, greater wealth, and (except for Peru) better educated parents. They also spend more hours studying at home, and their parents spend much more on private tutoring. In most respects, the primary and secondary schools that they attend appear to be better, including primary school maths teachers with better pedagogical skills. (Glewwe et al., 2021, p. 4).

It should be noted that these observations were made when the children were 15 years old, around 2016.

All of the development measures discussed so far are country-wide summary figures that do not reflect how evenly development is distributed across each country's population. I therefore look also at measures of inequality. Of the four countries, Peru had the highest level of inequality in 2000 and Ethiopia had the lowest, with Gini indices of 49 and 30, respectively. When inequality is calculated on income, as it is for Peru, it is usually higher than when it is calculated on consumption data, as it is for the other three countries. This difference in the measure used may explain some of the higher inequality in Peru, but it is unlikely to explain the difference entirely. Latin American countries have historically had higher levels of inequality than those in other regions. Vietnam had the second highest levels of inequality at 35, followed by India at 32 and then Ethiopia. Between 2000 and 2015, inequality rose in Ethiopia and India by 5 points each, stayed the same in Vietnam at 35, and fell in Peru from 49 to 45. In subsequent chapters, I consider whether levels of income/consumption inequality among adults in each of these countries are reflected in gaps in SES achievement scores and across their achievement trajectories based on SES.

These are a diverse set of countries at different levels of development in 2001, with the starkest differences between Ethiopia and Peru where I would expect my findings to differ the most.

Table 5.1: Countries' levels of development when children were around 1 and 15 years old

	Ethiopia	**India**	**Peru**	**Vietnam**
U5MR[a]	140 (2000)	92 (2000)	38 (2000)	30 (2000)
	62 (2015)	44 (2015)	16 (2015)	22 (2015)
Stunting[b]	57 (2000)	55 (1999)	31 (2000)	43 (2000)
	38 (2016)	38 (2015)	15 (2015)	25 (2015)
NER	40 (2000)	80 (2000)	97 (2000)	98 (2000)
in primary[c]	85 (2015)	92 (2013)	96 (2015)	98 (2013)
Inequality[d]	30 (2000)	32 (1999)	49 (2000)	35 (1998)
	35 (2016)	37 (2012)	45 (2015)	35 (2016)
GDP per capita	123 (2000)	442 (2000)	1,941 (2000)	395 (2000)
(current US$)	630 (2015)	1,590 (2015)	6,180 (2015)	2,595 (2015)
PISA (*Maths*)			387 (2015)	495 (2015)

Notes: [a] Under Five Mortality Rate (per 1000 live births); [b] Stunting: Children aged < 5 years stunted (height-for-age < -2 SD); [c] Total net enrolment ratio in primary education; [d] Income inequality for Peru; consumption inequality for all other countries. *Sources*: U5MR from United Nations Inter-agency Group for Child Mortality Estimation (2023); Stunting from UNICEF et al. (2023); NER in primary from UNESCO (2023); Gini coefficient from United Nations University World Institute for Development Economics Research (2022); GDP per Capita from World Bank (2023); PISA scores from OECD (2016).

5.2 Societal dividing lines

The SES achievement gaps across countries and also achievement trajectories by SES that I report are situated in a broader country context. In this section, I examine what may underpin these.

I discuss the societal dividing lines in each country. When analysing data and interpreting results some of these are important to consider. In the analysis in Chapter 9, I employ community-level fixed-effects, in order to control for time-invariant variation outside communities (e.g. geographical variations), and within-community variation is also important to consider (e.g. gender or caste differences). Though, as not all my analysis uses community fixed effects, I discuss societal dividing lines that would exist within and between communities.

Ethiopia

In Ethiopia, ethnic and gender differences are important. However, empirical evidence on disparities along ethnic lines is lacking; see Mohammed et al. (2016) who also finds this. This may be due to its political history, where, according to Abbink:

> In 1991 the TPLF [Tigray Peoples' Liberation Front] presided over the deconstruction of Ethiopia as a nation-state seen as dominated by one ethnic group [the Amhara] and reconfigured it on the basis of a model of an alleged 'voluntary federation' of the 75 or so ethnic groups in the country. This was a new model dictated partly by ... an ideological programme aimed at reversing 'ethnic' hierarchies, ousting the perceived elites in place, and impose a new political dispensation (Abbink, 2011, p. 597).

So, it may be that, in a political system where one's politics are inextricably linked to one's ethnicity, it would be politically volatile to report that one ethnic group is doing substantially better/worse than another.

Ethnicity and language in Ethiopia tend to overlap. In 1994, Ethiopia introduced a mother-tongue education policy, marking a shift away from Amharic-only instruction to the use of multiple local languages in primary schooling. James (2018) shows that there are between-language disparities in learning, but that these reflect differences in school quality across communities (including, among other things, differences in terms of literacy levels, linguistic development, and standardisation).

Ethnic lines also overlap with regional lines. Disparities across regions vary substantially. In 1999/2000, among the five regions sampled by the Young Lives team, the capital city, Addis Ababa, had the lowest poverty rates (at 36 per cent), followed by Oromia (at 40 per cent), Southern Nations, Nationalities and People's

(SNNP) region (at 51 per cent), Amhara (at 54 per cent) and Tigray (at 61 per cent), see MoFED (2012) report. A more recent study Tiruneh et al. (2021) reports the gains in numeracy over the 2018-19 academic year across regions. The highest progress was seen in the capital city, Addis Ababa (followed by schools in Tigray, Oromia, and Amhara regions). Authors found that schools in SNNP did not contribute to progress in numeracy.

Gender disparities in Ethiopia also exist. Hoot et al. (2004) explain that, at early ages, traditional female gender roles (e.g. housework and caring) are important barriers to early education, and to additional human capital investment, and that this is more common among families with the lowest socioeconomic status. This is evident in divergent literacy rates where, in 2004, for every 100 literate males aged 15 to 24 years, only 60 women of the same age were literate (see UN, 2018).

India

Gender and caste represent important divisions in India. According to UNESCO, gender discrimination "cuts across all strata of society" (UNESCO, 2010, p. 13). Based on the 2001 census, the female-to-male ratio, for children aged up to 6 years, was 927 to 1,000 (see UNESCO, 2010) – a difference of 73 girl children. The preference for boys manifests throughout the child's life through lower levels of nutrition for girls, or exclusion from schooling. For example, in 2001, for every 100 literate males aged 15 to 24 year, only 80 women of the same age were literate (see UN, 2018).

Caste in India also reflect divisions in society. The most excluded castes are the scheduled castes (SC) and scheduled tribes (ST). While I do not have information on poverty level by caste in Andhra Pradesh, U5MR in the state speaks volumes: in 1998/9, scheduled caste and scheduled tribe U5MR were nearly double that of "non-backward classes" at 122 and 116 for scheduled castes and scheduled tribes, and 64.7 for non-backward classes (see IIPS & ORC Macro, 2000).

Peru

An important divide in Peru is between the indigenous and non-indigenous population, often measured by the language spoken by the individual (e.g. Spanish, Quechua or Aymara), which practice dates to the Spanish invasion. Estimates of the indigenous population in Peru reported in Barrón (2008) range from around 50 to 60 per cent of the population. He also shows that the indigenous population is substantially less educated than the non-indigenous population. Forty percent of the indigenous population had no education, in contrast to 10 per cent of the non-indigenous population in 2003. Across all education levels, the average income of indigenous people is 11 to 43 per cent lower. Barrón (2008) also reports the yearly income, by age, of both indigenous and non-indigenous population. In every age group, the non-indigenous have higher incomes, with the difference increasing until around 62. According to the UN (2018), the women-to-men parity index for literacy rates (for 15–24 year olds) was near parity in 2004.

There are also important geographical divides in Peru. Peru comprises three areas: the western coastal plain, the mountainous Andes in the centre, and the Amazon basin. The coastal plain, where Lima, the capital city, is located, is the most developed and the wealthiest. The mountains and the Amazon basin, where most of the indigenous population live, are less developed and notably poorer.

Vietnam

In Vietnam, there is an important distinction between the majority Kinh and Hoa people and the ethnic minorities, with a strong overlap of ethnicity and regional disparities. Work by Baulch et al. (2007) shows that the living standards of the majority Kinh and Hoa households are considerably higher than that of minority households (from Vietnam's 52 other ethnic groups). Baulch et al. (2007) show that differences in education constitute a major driver in the differences between living standards between ethnic groups. Gender does not appear to be a dividing line in Vietnam. On average, Baulch et al. (2007) show that primary school enrolment is relatively balanced between boys and girls, with the exception of a few ethnic minority groups. Also, according to UN (2018), the women-to-men parity index for literacy rates for 15–24 year olds was near parity in 2000.

Stark regional variation in Vietnam has historical roots (dating back to the end of the Vietnam War in 1975), particularly between the northern and southern regions. Poverty rates are highest in the north and lowest in urban centres, as is reported in Minot et al. (2003). Provincial authorities have to mobilise their own resources to fund most services and activities, which has created stratification. As I report in Vandemoortele and Bird (2011), infrastructure has usually been built by local institutions, resulting in infrastructure disparities across regions, though a cross-regional transfer system is in place now to address these disparities.

5.3 Early childhood education systems

To better answer my third research question on whether attending ECE is associated with gains in achievement across childhood and what this reveals about different ECE systems, it is important to understand the contexts in which these children attended ECE.

In this section, I first discuss ECE options in each study country at the time the

Young Lives children's households were administered the Young Lives questionnaire (i.e. Round 2), around 2005 when the children were about 5 years old. I then discuss how these differed in terms of quality and unpack how access, provision and quality of ECE differed across societal dividing lines.

In India, Peru and Vietnam, primary school started at six years although, in Andhra Pradesh, five year old children were eligible for primary school, according to Woodhead et al. (2009). In Ethiopia, primary school entry age was one year later, at age seven.

The Young Lives data does not include reliable data on ECE quality. While the researchers asked parents to rate the quality of their child's ECE, these reports are subjective assessments by people who are not experts in the matter and who also have a vested interest in perceiving it as good, especially if they are paying for the education. Unsurprisingly, most parents rate their children's ECE as good quality.

Rather, I draw on available research documenting the quality of ECE at the time. ECE quality is notoriously difficult to measure as it is not objectively or directly observable. Here, I draw on various means of assessing the quality: from qualitative observations of ECE systems to indicators used as a process measure for quality, such as teacher:student ratios.

Table 5.2 outlines the main providers of ECE in each study country, along with the ECE centres' funding sources, the SES of families these ECE centres served and the perceived quality of the ECE centres.

Ethiopia

Unlike in other study countries, no government-supported ECE system was in place when the Young Lives children were 3 to 6 years old. A report by the Ministry of Education (MoE, 2008) recounts that, in 2006, primary school education was still being consolidated and there was no budget allocation for ECE. In 2006/07, the gross enrolment ratio (GER) for ECE was 3.1 per cent. ECE provision in Ethiopia was dominated by private and fee-charging schools with some hybrid schools (funded by government stipends and parent fees). Other options included religious, community and non-governmental organisations' (NGOs') ECE centres, but these represented a miniscule portion of ECE provision. According to a report by UNESCO (2008), in 2006, 95 per cent of children aged 4–6 years attending ECE were enrolled in private institutions, with most of them being in urban areas, especially the capital, Addis Ababa.

Table 5.2: Main types of ECE centres in Ethiopia, India, Peru and Vietnam in the mid-2000s

Type	Funding	Children served	Perceived Quality
Ethiopia			**(3% – 2005/6)***
Private	Parent fees	Upper SES	Very high
Public	Government stipend plus parent fees	Middle- and upper-SES	High
Government	Government stipend plus parent fees	General population/Lower SES	Basic
India			**(40% – 2006)**
Public: *Anganwadi* Centres	Government funded, free	Low SES; disadvantaged communities, particularly those in rural areas and urban poor	Mixed, but mostly very poor
Public (linked to primary schools)	Government funded, free	Low-and middle-SES	Mixed, mostly poor
Private	Parent fees	Middle, low and high SES	Mixed. Low in poor areas, better in wealthier areas
Peru			**(59% – 2005)**
Private: CEI	Parent fees	Upper SES	High
Public: CEI	Government	Middle- and upper-class	High
Public: PRONOEI	Government	Lower SES. Mostly rural and urban-slum population	Basic
Vietnam			**(57% – 2006)**
Public: State-owned	Government; parent and community contributions	Lower and middle SES	Mixed, mostly high
Community: Non-state-owned	Parent fees; community contributions; little government funding	Lower SES. Mostly rural and remote population	Basic
Private: Non-state-owned	Parent fees	Upper SES	Mixed, mostly very high

Source: for Ethiopia, Hoot et al. (2004, p. 5); other countries from various sources referenced in the text.
*Gross Enrolment Rates of early childhood education (3–6 years old)

How do these options vary in terms of quality?

The various options in Table 5.2 varied greatly in terms of their quality but there was little variety in terms of the options available, as Hoot et al. (2004) discuss. Most children attended private ECE centres which, Hoot et al. write, offered a “very high quality of education” (Hoot et al., 2004, p. 5). For the remaining children who did not attend private ECE centres, the ECE quality varied from a basic education in government schools to high quality in public ECE centres. A qualitative review of ECE found that “the majority of the existing personnel had irrelevant or only slightly relevant qualifications” (Tigistu, 2013, p. 154).

How ECE differs across societal dividing lines?

We already know that ECE provision was concentrated in urban and wealthy areas, mostly around the capital city. Gender gaps were smallest in Addis Ababa, Amhara, and Gambella (1 percentage point), and largest in the SNNP Region (5 percentage points) (MoE, 2008, p. 20).

India: states of Andhra Pradesh and Telengana

In 2005, when Young Lives children were of ECE age, Andhra Pradesh and Telangana formed one single state – Andhra Pradesh. Henceforth I will refer to the two as Andhra Pradesh. Where I have not identified state-level information, I will present national-level data, with the necessary clarification.

According to UNESCO (2009), India’s GER for the academic year ending in 2006 was 40 per cent. ECE provision and access varied across the Indian states. While there was no universal take-up of ECE in the mid-2000s, several ECE options were available. According to Kaul and Sankar (2009), there were four main types: public, private, NGO and religious. Public and private schools made up the large majority of options, while NGO and religious schools constituted only a small minority.

Anganwadi centres accounted for the main form of public provision in India. (*Anganwadi* is from the Hindi, ‘courtyard shelter’.) In 1975, the Indian national policy on childhood care converged around the Integrated Child Development Services (ICDS) programme. As part of the ICDS, these *Anganwadi* centres were set up and modelled after the US HeadStart programme, though ICDS was universal, not targeted as HeadStart was. The ICDS’ aim was to provide a wide range of services including pre-natal, child health and education services. Children who enrolled in *Anganwadi* centres should, therefore, have received a package of services (educational and health-related), all provided by a single *Anganwadi* worker who sometimes had an assistant. As we will see, this was not always the case in practice. Accessing these services was free. Based on a 2007 government report by the (Ministry of Women and Child Development (MWCD)), in Andhra Pradesh, approximately 25 per cent of children aged 3–6 years were attending an *Anganwadi* MWCD (2007). According to Kaul and Sankar (2009), *Anganwadi* centres represented, in 2006, approximately 75 per cent of ECE provision in India. There were also *Sarva Shiksha Abhiyan* (SSA) centres, although their numbers were too small to merit an in depth discussion.

In 2006, there were also ECE provisions attached to public primary schools; these were also free. Mehta (2007) report that in Andhra Pradesh in 2005/06, only 8.9 per cent of primary schools had these pre-primary classes.

According to Kaul and Sankar (2009), approximately 6 per cent of ECE centres were private in 2006.

While, in 2006, private ECE provision represented a small portion of total provision, it was expanding rapidly. Kaul and Sankar (2009) posits that this increase was driven by the growth of India's consumer class, enabling more families to afford ECE. In Andhra Pradesh specifically, Woodhead et al. (2009) finds that key attractions for parents were both the English language of instruction, and the focus on academic skills learning also reported in Singh (2014b).

While there were also NGO and religious ECE centres, their numbers were small compared with those of the government and private schools, as discussed in Kaul and Sankar (2009). NGO schools primarily catered to specific excluded communities such as tribal people, migrant labourers and rural children in specific circumstances.

How do these options vary in terms of quality?

In 2006, there was no national ECE curriculum. Nor was there a system of control or accreditation for ECE centres. In addition, there was no formal qualification system for ECE teachers. Available training courses identified by Kaul and Sankar (2009) ranged in duration from a few days to two years. Therefore, the quality of ECE varied greatly across the different types of providers.

In the public sector, Kaul and Sankar (2009) find that the approach to ECE was 'minimalist'. The quality of education provided in *Anganwadi* centres depended on two key components: funding and the qualifications of both the workers and the assistants (who were responsible for managing the centre). Funding, which came from the central government and should have been topped-up by the state government, was commonly delayed, as reported in Woodhead et al. (2009).

Anganwadi centre workers and helpers' qualifications were generally low, and ECE was not a focus. A Citizens' Initiative for the Rights of Children Under Six (CIRCUS) (2006) report shows that approximately 40 per cent of the assistants had no education. Among the *Anganwadi* workers, some 28 per cent had completed a maximum of 10 years of formal education (equivalent to UK GCSEs or year 11). ICDS workers were also often overworked and, as a consequence, a UNESCO report states that the "education components of the ICDS do not receive adequate attention" (UNESCO, 2010, p. 46). This is reinforced by a government report that states that the ECE component of *Anganwadi* centres was "very weak" (Planning Commission, 2008, p. 11).

Between 2005 and 2007, data were collected for ICDS-commissioned research on the quality of *Anganwadi* centres in Andhra Pradesh. The following account demonstrates the poor quality of these *Anganwadi* centres:

> In most villages, ICDS centres were either missing or not functioning properly. Ainole village in Nalgonda district has a population of 500, but no *Anganwadi*. In Gatevalesa village in Arakku Valley there are just 28 households. So there was no AWC [*Anganwadi* Centre]. But people complained that their children were not allowed to register at the ICDS in Gangudy village (1 km away). In Kota Bhallaguda (Arakku Valley), the ICDS worker was missing. At 10 am there were less than 10 children present at the *Anganwadi* in Bospeda, a village with a population of 650. The centre is supposed to start at 9 am. It was so cold that we were clad in shawls and sweaters, but the children were sitting inside in a cold, dark room. They were in rags and some were crying. There was no toilet. The weighing scale needed replacement and the baby bag was extremely filthy. The ICDS centre in Kalvakurti was no better. It was situated next to the primary school. 25–30 children and 6 mothers were present. We learnt that food is provided twice a day for children between 3 and 6 years of age. But pregnant and lactating woman [sic] and young children got no take home rations; just soyabean [sic] powder. There were mosquitoes all around the ICDS centre" (ICDS, 2007, p. 4).

A government report states a similarly dismal situation for the ICDS across the country, though without specific examples from Andhra Pradesh (see Planning Commission, 2008, p. 204). The evidence indicates that children attending *Anganwadi* centres were likely to receive a poor-quality education. While there may have been some exceptions, these appear to be few and far between.

While it is not clear from the literature that private ECE provided better-quality education, it does appear that the quality of the ECE centres tended to depend on the community they catered to. Woodhead et al. (2009) found that private ECE centres in poor areas were generally of poor quality, and of better quality in wealthier communities. UNESCO reports that, while a few prestigious private ECE centres offer very high quality education, approximately 95 per cent of ECE centres use "age-inappropriate methods, focus on academic objectives and are downward extensions of primary education" (UNESCO, 2010, p. 56). Indeed, Kaul and Sankar

(2009) find that private ECE centres tended to adopt the primary school curriculum, designed for much older children, which they argue to be rather inappropriate and potentially counterproductive to learning. This has resulted in a dual-track provision, as explained in by the authors, where *Anganwadi* centres aim to provide a holistic approach (but lack the staffing and resources), while the private ECE centres focus on teaching reading, writing and arithmetic.

How ECE differs across societal dividing lines

There is some evidence of ECE provision differing by caste. A report by CIRCUS refers to a study that included Andhra Pradesh which found that "in none of the surveyed mixed-caste villages was the ICDS centre located in the dalit [lowest caste] hamlet'. ... [and there was] extensive evidence of 'everyday caste discrimination' at the Anganwadi" (CIRCUS, 2006, p. 49). Qualitative research by the ICDS programme reports that "[t]he most vulnerable groups ... are excluded from this system either because there are no ICDC centres in their areas or because they are not allowed to visit existing centres due to caste politics" (ICDS, 2007, p. 3).

Enrolment in private ECEs differs by gender also. Kaul and Sankar (2009) report that more boys attend ECE and explain that "parents tend to prefer enrolling their sons in private institutions, as an investment due to better perceived quality" (Kaul & Sankar, 2009, p. 28), though the quality of provision in these centres remained low.

Peru

Peru's history of ECE dates back to the 1970s. ECE became mandatory in 2004 and, by 2005, 59 per cent of children were attending ECE, according to Estadística de Calidad Educativa (ESCALE) (2018). There were two types of ECE available to children aged three to five: formal and non-formal. Centros de Educación Inicial (CEIs) provided formal education, were public or private, and catered mostly to the urban populations. In the 1970s, to achieve wider coverage and to integrate mostly rural children into the education system, non-formal Programas no Escolarizados de Educación Inicial (PRONOEIs) were rolled out. They offer a lower-cost and more flexible approach to ECE, so were usually found in rural communities and poor urban areas. PRONOEIs were community-based entities that received support from the state. Beltrán and Seinfeld (2013) state that 99 per cent of PRONOEIs in 2010 were public.

A Ministry of Education (Ministerio de Educacion (MdE)) report stated that, in 2007, public ECE provision was eight times larger than private provision (MdE, 2007). That said, at the time, private provision was burgeoning: the private supply of ECE had increased by 70 per cent in the decade preceding 2007 (MdE, 2007).

How do these options vary in terms of quality?

Private CEIs were mostly privately owned and managed. They were either for-profit or not-for-profit (e.g. managed by a charity or a church), and do not necessarily adhere to the national curriculum. For the most part, they provide age-appropriate education. Both private and public CEIs were mandated to have a qualified ECE teacher and to provide five days of education per week. Public CEIs were managed and funded by the Ministry of Education, and followed the Ministry's curriculum.

PRONOEIs were not required to have qualified teachers, and the state provided them with minimal training. PRONOEI teachers, or 'animators', were volunteers from the community, usually mothers, who received a nominal stipend. They were monitored by a coordinator who reported back to the Ministry of Education. Beltrán and Seinfeld (2013) explain that the official policy was for each coordinator to monitor 8 to 10 PRONOEIs but, in reality, they could monitor as many as 40, limiting their ability to support the PRONOEI educators and centres. According to Beltrán and Seinfeld (2013), the government did not treat PRONOEIs as permanent establishments by the government, and therefore provided only limited public resources and little educational material. The community usually provided the building and furniture. PRONOEIs were mandated to provide four days of schooling.

How ECE differs across societal dividing lines

Beltrán and Seinfeld (2013) show that, in 2010, the coast had a ten-percentage point higher ECE attendance rate (at 92 per cent) than the mountainous region and the Amazon basin (81 and 79 per cent, respectively). They also showed that students who attended ECE in the mountainous regions and the Amazon basin performed worse in language and maths tests than coastal ECE attendees.

Vietnam

Vietnam has a long-established public ECE system in place, dating back to before the 1980s (London, 2011). Early childhood education for three- to five-year-olds was organised into two age groups: for children three to four years old, and ECE for five year olds, according to a Socialist Republic of Vietnam (SRoV) report (2003) and London (2011). These were complemented with parental education programmes to promote ECE's importance. Henceforth, my mentions of "ECE" refer to education for children aged three to five years old. National statistics report that, in 2006, 57 per cent of children aged three to five years old were enrolled in ECE (GSO & UNICEF, 2007).

The main ECE options included state-owned, community-owned and privately owned institutions (United Nations Educational Scientific and Cultural Organization, 2006). For brevity, I refer to these as public, community and private ECE centres, respectively. The private ECE centres were often referred to as *people-founded* schools, as reported in Nguyen and Nguyen (2008).

Public ECE centres were distributed widely across the country, and less so in some rural and remote areas. They were not necessarily free; they followed the national curriculum. In a government report (SROV, 2003), the government states that ECE (which they refer to as “ECCE”, short for Early Care and Child Education) received limited public funding, and therefore, the cost of provision had to be shared between the government and the community. Specifically, the government report states that “a notable proportion of ECCE costs are covered directly by parents and communities” (SROV, 2003, p. 95). London (2011) also reports that most public ECE centres charged fees.

Community ECE centres were usually found in hard-to-reach or poor rural areas, and therefore accounted for only a small portion of ECE provision. The communities established the schools, provided the infrastructure and footed operating costs, according to UNESCO (2006). Most of these ECE centres were financed by parental contributions, with only a small minority receiving government and commune subsidies. These ECE centres were monitored by the district’s Education Departments (UNESCO, 2006).

Private ECE centres were often found in urban and advantaged areas (UNESCO, 2006). As in the community ECE centres, school founders supplied the infrastructure and covered the operating costs, mostly through parent fees. They were regulated by the government. Given their location and cost, they catered mostly to better-off urban families. The education provided was mostly of high quality. In 2005/06 private ECE centres accounted for approximately 60 per cent of total enrolment in Vietnam (Ministry of Education and Training (MOET), 2015).

Figures on which sector dominated ECE are somewhat inconsistent. UNESCO (2006) states that private schools accounted for approximately 8 per cent of ECE provision, but does not state which year this pertains to. Nguyen and Nguyen (2008), on the other hand, report a substantial rise in the proportion of non-public ECE centres between 1994 and 2004, increasing from 30 to nearly 60 per cent (p. 144). London (2011) reports that, in 2008, 75 per cent of ECE groups were people-founded and, therefore, private. Boyd and Phuong (2017) report that, in 2005/06, 55 per cent of children attended non-public ECE centres and 45 per cent attended public ECE centres. These latter figures are broadly consistent with the 58.3 per cent figure for 2005/06 reported by the MOET (2015). Both the MOET (2015) and Manji et al. (2015) report that private sector provision has declined since 2000, in response to a government promotion of public ECE centres.

How do these options vary in terms of quality?

When the Young Lives children were attending ECE, there was an age-appropriate play-based kindergarten curriculum in place (UNESCO, 2006). However, its implementation in public and community ECE centres was limited. Most public ECE centres were staffed by "underpaid young women on short-term contracts" and lacked the appropriate materials (London, 2011, p. 33), a situation documented also by the government: "...the status and employment position of ECCE teaching personnel is unclear and their remuneration often uncertain. Many teachers in both towns and rural areas are employed on a contract basis, providing low job security and limited professional training and career development opportunities" (SROV, 2003, p. 96). Nonetheless, the MOET (2015) reported that, in 2006/07, 87 per cent of ECE teachers were trained to national standards.

The availability of appropriate learning and play material in these public ECE centres was very limited, according to SROV (2003) which reports that, in 2001, only 6 per cent of all crèches (for children aged 0–3 years) and ECE centres had any appropriate play materials (disaggregated figures are not available for only the three- to five-year-old group).

I have not identified research assessing the quality of private ECE in Vietnam but, based on aforementioned documentation, it appears, given the resources and financial-security available to private ECE centres through parent fees, and the age-appropriate curriculum required of them by the government, the quality is higher than public ECEs.

Community ECE centres were likely to provide poor-quality education given their limited infrastructure, materials, staff and funding – all provided by the community. By extension, community ECE centres in poor areas would have had lower quality infrastructure, limited materials, less-trained staff and little funding.

How ECE differs across societal dividing lines

In 2006, official figures show that the highest attendance rates were in the Red River Delta provinces (80 per cent) and the lowest rates were in the Mekong River Delta (40 per cent) (GSO & UNICEF, 2007). Moreover, Giang et al. (2016) show that, compared with children in the Red River Delta, children from all other five regions were less likely to attend ECE, even when controlling for other variables.

The majority Kinh were the dominant ethnicity in the Red River Delta (Baulch et al., 2007), and children from non-Kinh households were less likely to attend ECE (Giang et al., 2016).

In relation to ECE provision, "several rural areas, mostly in the south and central areas, as well as in the remote mountain villages of the north" were excluded from provision, and community ECEs – which provided poorer quality education – being more prevalent in the poorer and more remote northern provinces than in the rest of the country (see UNESCO, 2006, p. 5).

5.4 Conclusion: Implications for analysis

RQ1. SES gaps in achievement

My discussion on country context has shown that Ethiopia, India, Peru and Vietnam are diverse countries with diverse contexts. They have developed at different rates, with Vietnam outperforming the rest, in terms of economic development. Similarly sized SES gaps in achievement and changes in them are, therefore, not what I would expect.

All countries exhibited a moderate to high level of economic inequality. I am interested to see how Vietnam compares to the other countries in terms of SES gaps in achievement, as its absolute level of development and maths skills seem to be comparatively high. While Vietnam's economic inequality remains unchanged from 2000 to 2015, I wonder whether the notable progress in absolute levels of development is associated to reduction in the SES gap in achievement across childhood. After all, achievement scores have a natural limit: top scorers near 100 per cent don't have a lot of space for improvement, while those further down do. If those further down the achievement scale are there because they were born into challenging circumstances, if they have experienced improvements in their learning environment (both at home and at school), we might see them catching up to the high-scorers.

RQ 2. Achievement trajectories based on SES

In a subsequent empirical Chapter 8, I plot achievement trajectories based on SES. I group high-achieving children into low- and high-SES groups and do the same with initially low achieving children children. The present discussion of how different groups are subject to discrimination underlies the make-up of these attainment-SES groupings, and will thus help to inform my interpretation of their trajectories.

RQ 3. Does ECE offer an advantage in terms of achievement?

While the ECE system excluded the poorest and lowest-caste children in India, it also offered poor-quality ECE across the board (except for a handful of elite ECE centres that the Young Lives children did not attend). Because of poor quality, I would not be surprised to find little difference in achievement between children who did attend ECE and those who did not, when controlling for caste and gender. In India, nutrition may also have been an important determinant of achievement. Although India was three and a half times wealthier than Ethiopia in terms of GDP per capita in 2000, its stunting rate was equivalent and remained so until 2015 – pointing to possible malnutrition among India's children.

Peru and Vietnam's ECE system seems to be stratified by SES though the Young Lives data does not disaggregate within the public and private category in either country (for example between PRONOEIs and CEIs in Peru and between community and other public ECE centres in Vietnam). When interpreting my results, it will be important to consider the heterogeneity in these groups.

When the Young Lives children were of ECE age, there were no large-scale supplementation or stimulation programmes that targeted children of ECE age. Although the ICDS had a feeding component, "[t]he quality of food, the mode of preparation and the frequency of distribution varied [across centres]" (ICDS, 2007, p. 2). I am, therefore, not able to control for it directly. However, using children's height-for-age allows me to account for it to some extent, as supplementary nutrition that impacts upon physical growth is likely also to impact upon cognitive development and educational achievement.

A selection effect is likely to explain some of my results in all countries, as, in all countries, the quality of ECE a child received depends on one's SES. For this reason, I analyse for community-level fixed effects to address some of this selection effect. While these communities are quite homogeneous, I nonetheless control for additional factors that may vary within communities such as gender, ethnicity, caste or mother tongue and SES.

Next, before delving into the empirical analysis, I discuss the measures I use in the next chapter.

Chapter 6

Measures

In this chapter, I detail how researchers measure the variables that I use in my analysis. In Chapter 2, I provided the conceptual justification for the variables I selected, and in the empirical chapters 7, 8 and 9, I will explain how I use them.

At the end of Chapter 2, I listed the main criteria that the variables need to meet in order to enable cross-country comparisons. The first criterion concerns the coordination of participants' ages and the timing of data collection across all countries. This criterion is met by the Young Lives data, and I shall discuss it in detail in Chapter 5. The second criterion is that the achievement tests meet or are close to best practice and that they measure the same sets of skills across rounds (which, in this case, would be maths and vocabulary skills). I discuss this in the next section, below. The third criterion is that the measures are measured using a standardised set of instruments across rounds. In Chapter 5, I mentioned that the Oxford team worked hard on standardising the survey instruments; here, I discuss the work I put into cleaning and/or standardising most of the variables for analysis across countries.

I start with the outcome measures, educational achievement, followed by the three measures that I use for parental socioeconomic status (SES), then early childhood education (ECE) variables. Finally, I will discuss how other covariates are measured.

Note that I do not use all of the variables in all analyses. My first research question (RQ1) requires the educational achievement variables, SES variables and the language of the test. To address my second research question (RQ2), I use educational achievement variables and SES variables. For my third research question (RQ3), I use educational achievement variables, an ECE variable, parental SES, and a set of child-, household-, and community-level control variables (specifically height-for-age, gender, whether the child is firstborn, ethnicity/caste, language of the test, household size, maternal employment and the cluster site).

6.1 Educational achievement

The Young Lives achievement scores are intended to be interpreted relative to their distributions. As Revollo and Scott (2022) explain, the Young Lives study does not administer the achievement tests in schools so cannot compare how a child has performed relative to an 'average' student. It is, therefore, important that each test capture as wide a distribution in maths and vocabulary skills as possible.

Young Lives children took a maths and receptive vocabulary test at 5, 8, 12 and 15 years old, which I discuss below. The children took the tests at their homes, in conjunction with the rest of the questionnaire. Prior to the tests, all children also received instructions on what to do, with examples, to check that they understood the tasks. According to Cueto and León (2012), children who did not understand the instructions did not take the test. In later rounds, illiterate children could not take the tests. The study used tests in the main local language and, in earlier rounds, in multiple translations.

6.1.1 Maths

Round 2: Five years old

The Young Lives teams worked both to assess children's maths skills at 5 years old, and to pilot the International Evaluation Association (IEA)'s Cognitive Development Assessment (CDA) for four-year-olds. The original instrument included three subsets of questions about spatial relation, quantity and time; the Young Lives team decided to use only the component evaluating quantity. Revollo and Scott (2022) explain that the component for measuring spatial relations was too time-consuming for the children and that the time component did not provide reliable measures.

The CDA–Q (Q for quantity) comprised 15 items. Children were asked to choose an image from a selection of three or four that best represented a concept communicated by the examiner (e.g. few, most, nothing, etc.) For example, the examiner directed, "Point to the plate that has a few cupcakes". Each child scored one point for a correct answer and nil for an incorrect or no answer.

Round 3: Eight years old

When the children were 8 years old, another age-appropriate instrument was necessary for testing maths skills. According to Cueto and León (2012), to design this (and the tests for Rounds 4 and 5), the Young Lives team drew from items of existing national and international testing programmes that were freely available, as well as developing a few new items based on these (total 29 items). All Round 3 maths tests were the same across countries.

The test was divided into two sections. The first section, comprising nine items, sought to measure basic quantitative and number notions, and included items on counting, number knowledge, number discrimination and basic operations (e.g. "Please count how many balls there are here", and "Please tell me the answer of this calculation: Two times four?"). According to Young Lives (2009), an examiner read the non-numerical questions out loud to the child, to ensure that maths skills were being tested rather than reading skills.

The second section sought to measure ability to perform basic mathematics operations with numbers. It comprised 20 items involving addition, subtraction, multiplication and division, with whole-number answers. Results from pilot testing informed the ordering of the items by increasing difficulty. Examiners encouraged each child take the test at his or her own pace, according to Cueto and León (2012), they stopped this portion of the test after 8 minutes. Round 3 results showed that maths skills varied across countries.

Round 4: 12 years old

In Round 4, the Young Lives team adapted the tests to each country's context. This was needed partly to allow for differences in maths skills between countries, as the children in Vietnam exhibited the most advanced maths skills. Administering an identical test in all countries would not have captured the variation in maths skills within each country's sample.

Vietnam's test was the most difficult and Ethiopia's the easiest. According to Young Lives (2013b), Young Lives drew from a pool of items, including some from the Trends in International Mathematics and Science Study (TIMSS) test. According to Young Lives (2013a), tests ran for up to 40 minutes. There was some nominal overlap of questions between the Round 3 and Round 4 tests in each country. The Ethiopian test had 28 items, the Indian and Peruvian tests 29, and the Vietnamese test, 34 items.

Rounds 5: 15 years old

Round 5 tests were also tailored to each country's maths skills. The Young Lives team again drew from a pool of items, including the TIMSS test, plus new items from the Programme for International Student Assessment (PISA), for which they procured translations and other adaptations to each country's context (as discussed in Revollo & Scott, 2022, p. 16). The Ethiopian fieldworker's manual allowed each child up to 50 minutes to complete the test (Young Lives, 2016; see also Revollo & Scott, 2022). There is some overlap of questions between the Round 3, 4 and 5 tests in each country. The Ethiopian test had 30 items; the other countries' tests, 31 items.

For the most part, the tests were able to capture the variation in their maths scores. See distributions reported in the chapter Annex Section E.4. Though, in Ethiopia, the maths test was not able to capture as much variation at the bottom of the distribution (at 8 and 12 years old).

6.1.2 Vocabulary

Rounds 2 & 3: five & eight years old

In Rounds 2 and 3, the Young Lives team used the Peabody Picture Vocabulary Test (PPVT) to test receptive vocabulary in all countries. As Cueto and León (2012) describe, the PPVT tests for vocabulary recognition. It was developed in the US in the late 1950s and is now in widespread use as a general measure of educational achievement. Revollo and Scott (2022) explain that the PPVT is designed to test children from the age of 2.5 years old through to adults. It is essentially a long list of words (grouped into sets) in order of difficulty. Beginning with the starting set – determined by the respondent's age – the examiner reads out a stimulus word and the respondent selects, from four pictures, the one that best represents its meaning.

In Ethiopia, India and Vietnam, teams administered the PPVT-III in Rounds 2 and 3. It consisted of 204 words (in local translation), divided into 17 sets of 12 items, with each set becoming progressively more difficult.

Dunn et al. (1986) developed a Spanish version of the PPVT, the Test de Vocabulario en Imagenes Peabody (TVIP). The Peru team used it in all rounds. The test comprises 125 words that also become progressively more difficult. Quechua-speaking respondents received a version in Quechua translation (Cueto & León, 2012).

Not all of the PPVT items have to be administered: the test starts, usually at the set corresponding to the child's age, and stops when a predetermined number of errors are made within a set. Calculating the raw PPVT score involves establishing a baseline and a ceiling. For the PPVT-III test in particular, the baseline is the first number in the first set in which the child makes one or no errors; the ceiling is the last item in the set where the child makes eight or more errors (Cueto & León, 2012).

León et al. (2015) explain that, with the TVIP, the highest eight consecutive correct responses define the baseline; administrators select the first item of the eight to code as the baseline. The first eight consecutive responses containing six errors defines the ceiling; administrators specifically select the last of them for the ceiling coding. If there is no identifiable baseline set, then the first item answered correctly defines the baseline, according according to León et al. (2015).

As León et al. (2015) do, I code all non-administered items below the baseline as correct, and those above the ceiling as incorrect. Those below the baseline are easier, and the tests' assumption is that the child would have answered these correctly. Those above the ceiling are more difficult so the test assumes that the child will not be able to identify them. The raw score is the sum of all correct items.

Although the test provides age-standardised scores, León et al. (2015) explain that Young Lives did not use them because the reference population differed too much from the populations studied in the study countries. Specifically, the standardised reference scores are the American population for the PPVT-III, it is the American population; for the TVIP, it is the Mexican and Puerto Rican populations.

The Round 3 Young Lives data provide 'corrected scores' which, according to Cueto and León (2012), exclude items which appear biased by gender or language or had poor psychometric characteristics. Young Lives does this only for Round 3, so I do not incorporate these 'corrected scores' into my analysis.

Rounds 4 & 5: 12 & 15 years old

In Ethiopia, India and Vietnam, the format of the vocabulary test changed for Rounds 4 and 5. According to Leon and Singh (2017), this was because it became evident that different items performed differently in different languages (i.e. were much easier in one language than in an other). Recall the hurdle example discussed in Chapters 2 and 3, where Ethiopia's translation of 'hurdle', 'boy jumping over stick', was easier than the original. Revollo and Scott (2022) also share that, when the PPVT test was administered to the older cohort at 8 years old, some children were able to reach the most difficult items. This indicated that the tests were not well designed in their translated form, as this should not have been the case as the original PPVT was designed to test receptive vocabulary through to adulthood.

The Round 4 and 5 tests involved a subsample of the words that varied between countries, but were the same for both rounds. The selection for Ethiopia contained 55 words; for India, 57 words; for Vietnam, 76 words. Leon and Singh (2017) describes the selection criteria: "(i) adequate item fit using data from Round 2 and 3; (ii) items without DIF [Differential Item Functioning (i.e. different performance)] by round and cohort using data from Rounds 2 and 3; and (iii) items across the different range of item difficulty" (2017, p. 6). The Peruvian team continued to administer the full 125-item TVIP test.

Assessing the children with the exact same test at 12 and 15 years old (Round 4 and 5) seems counter-intuitive to me, as the measurement error, which may be distributed unevenly across the sample (e.g. some words are more common in particular regions or SES groups) would be quite similar in both rounds. It also perhaps complicates the interpretation of the results as, when administering the test at 15 years old, we may have been testing not only receptive vocabulary, but also memory or initiative. For example, had exposure to a new word in the test at 12 years old motivated a child to learn its meaning, then, in the next round, we would also be measuring the child's ability to remember the word and their initiative to learn its meaning.

The distributions (Annex, Section E.4) are quite flat at 5 years old in India, 8 years old in Ethiopia, India and Vietnam. This shows, as Young Lives identified, that the PPVT test translations were not optimal for discriminating between achievement levels, hence their revision for ages 12 and 15 years.

6.1.3 Standardising maths and vocabulary scores

To compare SES gaps in achievement across both childhood and countries, I need to standardise scores so they have a common metric. Here I discuss my choice for such a common metric for educational achievement.

Comparing raw scores between countries and between rounds is analogous to comparing apples with oranges. Tests between countries vary to accommodate different educational achievement levels (e.g. the maths test for 12-year-olds in Vietnam is more difficult than that for 12-year-olds in Ethiopia). Only the Peruvian vocabulary test remained constant in every round (although, by 15 years old, the test exhibits ceiling effects). The tests also vary in their length. The Maths test in Round 2 had 15 items, so a score of 15 is excellent. However, in Round 3, a score of 15 means the child answered approximately half of the questions correctly. At each stage of childhood, tests differ in their content, having been adapted to be age-appropriate.

It is important that this common metric also take children's ages into account. Within each round, children are interviewed over several months, so their ages vary within each round. Brito and Noble (2014) explain that, when examining socioeconomic disparities in brain development, "[f]irst and foremost, the age of the participant must be taken into account, as brain structural volumes change significantly across childhood and adolescence" (2014, p. 9). Indeed, a child's age may explain variation in achievement scores, as an older child has a more developed brain, and has also had more time to learn maths and vocabulary skills.

So, the ideal standardisation to enable comparisons across childhood and across countries should include a summary measure that meets the following 5 criteria: it should

i. have an established meaning: the scores have a meaning that the reader readily understands, one that may already exist in the educational testing literature;

ii. is a common metric: the unit that represents the distance between two groups (high vs low SES) is the same across childhood and across countries;

iii. describes distance from the mean score: I can ascertain from the score how far from the sample average the group's score lies;

iv. describes dispersion from the mean: the unit representing the distance between the two groups reflects the dispersion of the scores (so that a gap where scores range from 60 to 70 can be compared with one where the scores range from 600 to 700);

v. translates across the child's age: takes the child's age (in months) into account (this last criterion is specific to within-country comparisons across childhood, not cross-country comparisons).

There are various options I could use to standardise maths and vocabulary scores including a rank score, a percentile score, percentage correct (out of 100), a z-score and an item response theory (IRT) score. I use a country- and age-specific z-score as Schady et al. (2015) do. In Annex E, I critically discuss alternative options and why I do not use each respective approach.

Country- and age-specific z-score

Age-specific z-scores are calculated by dividing the sample into the smallest possible age groups, in this case months, and computing z-scores within each age group. Therefore, z-scores have a mean of 0 and a standard deviation of 1.

Researchers who conduct cross-country analyses on educational achievement tend to use this measure. Engle et al. (2011) use z-scores with Young Lives data, and other seminal texts do similarly with other data, for example several chapters from Ermisch, Jäntti, et al. (2012) (specifically Chapters 5, 6, 10, 11 and 19), Bradbury et al. (2015), and Schady et al. (2015) use country specific, age specific z-scores.

As a z-score's unit is a standard deviation, when interpreting results, we discuss gaps in standard deviations. This has meaning in the field of education, being useful to enable comparison of results across studies. The distance between groups of 0.40 standard deviations has the same meaning regardless of the sample size. However, a standard deviations can be sensitive to outliers and to ceiling or floor effects. If, among the 105 60-month old children in India, there is a exceptionally bright child, she may pull the group average up. This would result in a child who would otherwise have an 'average score', for a 60-month-old child, having a lower than average score because the outlier pulled the average up.

One way to deal with this is to ensure that the number of children in each group is large. In the tails of each age group, the number of children in each month is small, sometimes as little as 1. So, at the tails of each distribution, I have grouped children into a group of at least 20 children. For example, in India, I grouped a

190-months-old, four 189-months-old and eight 188-months old children into the 187-months-old group that had 29 children, creating a group of 42 children. While this should attenuate the effect of an outlier in the month-group, it does simplify the actual age range down into a single month. Given that the number of children affected by this represent less than two per cent of the sample in each country, this should not affect my results.

Also, depending on whether the test is *too easy* or *too difficult*, there may be ceiling or floor effects, where the test was not able to capture the variation at the top or bottom of the achievement distribution, respectively. Under these circumstances, one standard deviation above the mean is not equivalent to one standard deviation below the mean. I am comparing between group averages, which should attenuate the ceiling and floor effects within months-groups.

The z-score indicates how far from the month-group average an individual child's score lies. In my results, I group children into high- and low-SES groups. So, an average score of 0.40 means that the group's score lies 0.40 standard deviations above the mean, and -0.40 means it lies 0.40 standard deviations below the sample mean.

The z-scores also take the dispersion of the distribution into account by expressing distance from the mean in terms of the standard deviation. It is likely that month-groups have different dispersions.

If the dispersion among 60-months-old children is wider than at 65 months old, one standard deviation from the mean at 60 months old is, in absolute and percentage terms, larger than that at 65 months old. A reason for calculating a z-score is precisely to account for differences in dispersion and enable comparisons. However, doing so conceals differences in raw scores.

Some authors control for age in their regression models. However, as I am pooling all rounds, if I control for age, I assume a linear relationship between age and test scores. For example, I associate a one-month age difference with the same test score increase regardless of whether it happens at 5 years or 15 years old. I could model a logarithmic relationship between the two, but, either way, putting age into the regression equation means that I need to make assumptions about the relationship between test scores and age, which I avoid doing when I calculate an age-specific z-score.

6.1.4 Comparing gaps in achievement across childhood and countries

In this section, I focus only on comparisons of gaps in achievement across childhood and between countries. (See Chapter 3 for a critical discussion of cross-country analysis, and the approach that I employ.)

There are significant cultural, economic and historic differences between (and within) countries, making it impossible to construct strictly comparable tests for each country. The exact same test, in translation, will change: there are some words that are common in some countries, but not in others. Even if translation were not necessary, concepts vary in meaning between countries.

Comparing test scores over childhood poses challenges, as most tests take on customised versions for each age group and, therefore, differ at each stage in childhood. Within countries, the meanings of concepts and words will differ by groups: there may be differences between girls and boys, between ethnic groups, and between children in one part of the country versus another. A test that, in one round, prefers the words and concepts used by one gender, ethnic group or region and, in another round, those of another will produce different scores for the same children. This is something that test designers try to take into account, but cannot do perfectly.

All educational tests that compare across childhood and between countries face these challenges. While this reality complicates comparisons between and even within countries, it is still useful to document achievement gaps between countries and over time, a case that I made in Chapter 3. But it is worth repeating that I compare cautiously: I highlight the limitations of comparisons across childhood and between countries, and avoid making grand claims.

6.1.5 Cleaning test scores

The Young Lives data available from the UK Data Services required a substantial amount of cleaning. The Young Lives data set provides summary scores for the maths and vocabulary scores. I recalculated all of them from the raw data, having found numerous errors in the Young Lives data.

Some of the maths tests had been scored using an incorrect answer key. In other cases, correct answers were not marked as correct, missing codes were used inconsistently, or marking was inconsistent between countries. (There is also the possibility that the Young Lives teams made some of these amendments intentionally but did not document them in the principal documentation.) Table 6.1 describes the difference between the raw score Young Lives provides in the data, and the final score I calculate after cleaning the data. While the (pairwise and rank) correlations between both sets of scores remains high, thousands of children's scores were affected by the cleaning. (This is especially relevant when an IRT approach has been used to standardise the scores (as done by Das et al., 2022). As this approach weights each item according to the probability of answering the question correctly, it is important that correct items are coded correctly.) In Annex E, I provide details of the cleaning I conducted.

Table 6.1: After cleaning maths scores: comparing my scores with the Young Lives scores

Country	**Round**	**No. of children's scores affected**	**Correlation***
Ethiopia	2	454	0.99
	3	20	0.99
	4	10	1.00
	5	4	1.00
India	2	1,439	0.99
	3	26	0.99
	4	3	1.00
	5	6	0.99
Peru	2	1,846	0.96
	3	72	0.99
	4	2	1.00
	5	1,348	0.99
Vietnam	2	491	0.99
	3	56	0.99
	4	1,523	0.99
	5	1	1.00

* Correlation between Young Lives' score and my score, after cleaning and recalculating.

The teams calculating the raw scores incorrectly identified ceilings and baselines and incorrectly recorded raw scores. For example, sometimes a raw score of 25 should have been 125; the digit '1' was missing. Table 6.2 documents the differences between the raw score Young Lives provides in the data, and the final score I calculate after cleaning the data.

Table 6.2: After cleaning vocabulary scores: comparing my scores with the Young Lives scores

Country	**Round**	**No. of children's scores affected**	**Correlation***
Ethiopia	2	1861	0.98
	3	537	0.98/0.96
	4	1576	1.00
	5	1675	1.00
India	2	1847	0.99
	3	590	0.99/0.96
	4	10	1.00
	5	0	1.00
Peru	2	1280	0.98/0.99
	3	397	0.99
	4	308	0.97/0.99
	5	405	0.95/0.98
Vietnam	2	1747	0.99
	3	99	0.99
	4	0	1.00
	5	1933	1.00

* Correlation between Young Lives' score and my score,
after cleaning and recalculating (pairwise/rank).

6.1.6 Missing observations

The results that I present in the empirical chapters are from a balanced panel that includes all the children for whom I have observations of the variables needed for the analysis. In other words, within each country, I compare the same group of children from 5 to 15 years old. When examining SES gaps in achievement, the size of the balanced panel changes according to the measure of SES I use, as the number of children who have data on each measure of SES varies.

As I am following the analytical style of Feinstein (2003) in Chapter 8, his methodology necessitates that I use a balanced panel. A balanced panel will be needed also in the other empirical chapters as I feel that comparing the same group of children over time makes more intuitive sense. While I could conduct a case-wise analysis, when I conduct a sensitivity analysis, this does not produce substantially different findings.

For those children with missing test scores, I assume they were either unable or unfit to take the test (i.e. did not understand instructions). Therefore, for sensitivity analysis, I test the worst case scenarios and assume all children with missing test scores received a poor score of minus one of a standard deviation (-1 s.d.).

I have a choice between imputing missing test scores and other missing data or using the raw data. Both options have their own strengths and limitations. An argument for imputing is that children without test scores may be low-performing and/or less privileged, biasing the results. However, imputing itself may bias in a different way, depending on the imputation choices made. I therefore prefer to analyse the raw data, rather than to impute.

Table 6.3: Average socioeconomic status for children with and without achievement scores

	Ethiopia		**India**		**Peru**		**Vietnam**	
	With	Without	With	Without	With	Without	With	Without
Maths scores								
Maternal education	1.61	1.42	1.71	1.42	2.55	2.48	2.34	2.04
Asset index	0.37	0.24	0.83	0.78	0.66	0.60	0.86	0.77
Expenditure	4.60	4.36	6.55	6.45	5.01	4.89	5.78	5.75
n	1489	510	1793	218	1751	301	1799	201
Vocabulary scores								
Maternal education	1.56	1.55	1.69	1.64	2.55	2.49	2.33	2.26
Asset index	0.36	0.26	0.83	0.84	0.66	0.62	0.85	0.86
Expenditure	4.56	4.46	6.54	6.56	5.02	4.92	5.75	5.87
n	1537	462	1777	234	1645	407	1570	430

6.2 Parental socioeconomic status (SES)

I use maternal education, an asset index and parental education as measures of parental SES. Each represents a distinct aspect of SES; see Chapter 2 for an in-depth discussion.

6.2.1 Maternal education

How maternal education is measured by Young Lives meets best practice, as it reports both the number of years in education and the mother's country-specific level of education. (Without the latter, the earlier is irrelevant, as in some education systems, students may repeat years.)

Each country has a distinct education system. The International Standard Classification of Education (ISCED, 2011) facilitates comparisons of education systems between countries by offering a comprehensive framework to convert national education programmes and qualifications onto uniform and internationally agreed classifications. These ISCED mappings are available from the United Nations Educational, Scientific and Cultural Organisation (UNESCO) website (see UNESCO, 2011). While the conversions were mostly self-explanatory, I make a few adjustments. For Ethiopia, the Young Lives team coded 'masters and doctorate' to a single category, whereas the ISCED mapping codes them separately. I recode them to the ISCED 'masters' code (7). There is no adult literacy or religious education option in the ISCED mapping, so I code these as 'no education' (0).

For India, the ISCED mapping offers an 'adult education' programme that I apply to the 'adult literacy' programme in the Young Lives data. As in the Ethiopian data, India's Young Lives team coded 'masters and doctorate' into a single category, whereas the ISCED mapping codes them separately. I recode them to the ISCED mapping 'masters' code (7). 'Religious education', I coded as 'no education' (0), as it not appear in the mapping.

For Peru, what Young Lives team coded as 'complete CETPRO', I consider as equivalent as to 'educación superior no universitaria (técnica)' (post-secondary non-university technical education) which the ISCED mapping situates under 'short-cycle tertiary education' (5). I recode 'incomplete CETPRO' as the ISCED mapping equivalent of 'post-secondary non-tertiary education' (4). 'University complete' I recode as 'bachelor's degree'. I code an 'incomplete university education' at one level lower than complete university education (5). While there is no 'adult literacy' programme per se in the ISCED mapping, there are various sets of alternative education programmes for older children and adults, so I code these as the lowest level of alternative adult education (1), equivalent to 'primary education'.

What Young Lives codes as 'post-secondary, vocational' in Vietnam, I code as 'Collegiate vocational' in the ISCED mapping. As there is no adult literacy programme, or an equivalent, in the Vietnam ISCED mapping, I recode it as 'no education'.

I then grouped these into a single variable with three categories: 'none' (i.e. no education), 'primary and lower secondary education', and 'upper secondary education and above'.

6.2.2 Asset index

The best-reputed assets data from low- and middle-income countries are the Demographic and Health Surveys (DHS) and UNICEF's Multiple Index Cluster Survey (MICS). The Young Lives index builds on the concepts underpinning these indices. For example, Young lives follows DHS and MICS in collecting information based on housing quality, access to services, and ownership of consumer durables. The data collected for the asset index, therefore, also meets best practice.

As Briones (2017) explains, the Young Lives asset index is based on three components: housing quality, access to services, and ownership of consumer durables. Housing quality refers to the main materials of the walls, roof and floor, as well as a scaled household density (number of rooms per household member). Access to services includes access to safe drinking water, a sanitation facility, adequate cooking fuel, and electricity. For each country, Young Lives selected a set of country-specific consumer durables. Young Lives constructs an asset index as discussed in Briones (2017), which is a sum score of the three components.

As I did in Vandemoortele (2014), I use an IRT model to calculate the asset index for each round in each country. The main reason for doing so is that, while the sum-scores weight each asset equally, IRT models allocate weights based on the item's ability to discriminate between households' wealth.

Each aspect pertaining to housing quality and service access can be described by a binary variable. For example, I identify households with good-quality walls, roofs, flooring, with safe drinking water sources, a safely managed sanitation service, adequate fuel for cooking. (For the definition of 'good quality', see Briones (2017), summarised in Table E.7.) Note that the only difference between the Young Lives team criteria and mine is the crowding definition. Starting with the number of rooms per household member, Young Lives scales the ratio to range from 0 to 1.5. I define those in the bottom third as living in a crowded condition.

The IRT model I use is analogous to the one in Figure E.1. The more items, the better for IRT (preferably at least 30). The latent variable is 'asset wealth and access to services', and each box represents a variable used to calculate the asset index (e.g. quality of walls, floors, roof, access to electricity, owns a radio, etc.). The loadings (i.e. b_i) in this model represent the probability of having good-quality housing, access to safe services or owning a consumer durable. The correlations between the Young Lives sum score and the IRT score are 0.96 for Ethiopia, 0.76 for India, 0.84 for Peru and 0.75 for Vietnam.

6.2.3 Household expenditure

Household expenditure data are always challenging to collect. It demands quite a lot of recall from the respondent and then a demanding job from the research team to convert the responses into a monetary value that is comparable across households (as prices can vary within countries and across seasons) (see Deaton & Grosh, 2000). These challenges are not specific to the Young Lives survey, and are common across surveys that collect expenditure data in low- and middle-income countries. The instrument that Young Lives uses is in line with those used in the World Bank's Living Standards Measurement Study (LSMS) surveys.

Data on household expenditure was collected from households in Rounds 2 to 5. According to Marion (2018), data collection covered two categories: food and non-food items. The recall period for food consumption was 15 days, unless an unusual or special event occurred in this period (e.g. festival, wedding, feasting or fasting period). In those cases, interviewers asked about consumption in the 15 days prior to the event. The Young Lives team also collected data on the value of the consumed items. (Table 1 in Marion (2018) offers an informative list of food items included in the consumption aggregates by country.) To ensure a consistent measure within countries, the Young Lives team collected data on the same items in each round. Items varied across countries, as the list was context-specific.

To identify frequent non-food purchases, researchers asked respondents to identify which items from a list they had consumed in the last 30 days. Infrequent non-food purchases were identified from another list, over a 12 month recall period. These recall periods are long, and hence prone to measurement error. Marion (2018) assigns non-food expenditure into four categories: education, health, clothing and footwear and other non-food items.

Researchers converted all data with different reference periods into monthly values and summed them, applying regional consumer price index (CPI) to adjust for pro-rata inflation in Ethiopia, Peru and Vietnam. There was no CPI available for Andhra Pradesh and Telengana, so information came solely from the Young Lives community questionnaire. Marion (2018) provides a detailed discussion on how this was done. As the data for Round 4 in Peru were the only set not already inflation-adjusted, I did those calculations using Lima's CPI. All results are adjusted by their current household size, and expressed per capita. Ethiopia's expenditure data are reported in 'per adult' terms (an equivalence scale that takes gender and age into consideration is used). However, for consistency across countries, I recalculate results to obtain per capita expenditure levels.

Among these three measures of SES, maternal education is the least modelled and most intuitive, and is thus my preferred measure. The asset index offers little scope for measurement error as its components are observable (e.g. material on the roof). Though, depending on how this index is constructed, measurement error may arise. I try to avoid that by using IRT. Expenditure, on the other hand, has more scope for measurement error, due to recall issues and variation in prices within countries and across seasons. It was also collected when the children were 5, not 1, years old, so is not as timely a measure of 'initial SES' as the other two measures.

6.3 Early childhood education (ECE)

6.3.1 ECE attendance (yes/no)

The Young Lives team collects data on ECE participation. The Young Lives survey of the younger cohort (unlike the older cohort) collects these data at the age when children are attending ECE, so no recall issues arise.

In the Round 2 questionnaire, respondents were asked: "Since the age of 36 months, has [NAME] regularly attended a [LOCAL NAME FOR FORMAL AND INFORMAL PRESCHOOL] i.e. for a whole morning, afternoon, evening or night almost every week?" (Young Lives, 2006, p. 60). I use the responses to this question for the ECE variable in Equation 2.2. If the answer is yes, the interviewer asks a set of follow-up questions related to the age when the child started attending this school, how long they attended, what type of institution it was, whether payment was required, and the numbers of days and hours per day the child attended. I report these in Chapter 9.

In subsequent rounds and for the older cohort, Young Lives collects retrospective data on children's attendance in ECE, but there is a stark discrepancy between real-time data (collected at 5 years old) and that collected at older ages (see my discussion in Chapter 4, section 4.4).

6.3.2 ECE type

While I do not use the type of ECE centre the child attended in my analysis, I do use it in my descriptive statistics (Chapter 9). Interviewers asked: "Who runs this [LOCAL NAME FOR PRESCHOOL]?" (Young Lives, 2006, p. 60) and presented the set of options listed in Table 6.4. As discussed in Chapter 5, the type of ECE institution a child attends usually reflects the quality of education they receive. I therefore group the ECE institution type into the following categories: public, private and other. See Table 6.4 for the details of this conversion.

In some cases, children attended more than one ECE centre. In Ethiopia, there are 12 children who moved from an ECE centre once. In India, 123 moved once, and four moved twice. In Peru, 110 moved once, and 6 moved twice. In Vietnam,

Table 6.4: Recoding of ECE type

Country	Original code	Original label	New code	New label
Ethiopia	1	Private	1	Private
	2	Public (part student, part gov. fees)	1	Private
	3	Community	3	Other
	4	Government-funded	2	Public
	5	Other	3	Other
India	1	Private	1	Private
	2	NGO/charity/church	3	Other
	3	Public	2	Public
Peru	1	Private	1	Private
	2	NGO/charity/church (not for profit)	3	Other
	3	Public, local municipality	2	Public
	4	Public of the National Government	2	Public
	6	Informal	3	Other
	7	Half Public/Half Private (i.e. paying)	1	Private
Vietnam	1	Private	1	Private
	2	NGO/charity/church	3	Other
	3	Public	2	Public
	4	Informal community	3	Other
	5	Other	3	Other

102 moved once, and nine moved twice. In my analysis, I use the ECE centre they attended for the longest duration or, when durations are the same, the most recent ECE centre.

6.4 Controls

The Young Lives team combined 5 rounds of data for each country into a single data set. This data set includes a selection of variables, but does not include educational achievement scores, ECE information or SES variables. Most of the controls below, however, are included in this combined data set. Briones (2018) defines and discusses the variables included in the data set.

6.4.1 Child level

Age in months

The child's age in months, for each round, is included in the Rounds 1–5 combined data set provided by Young Lives. It is calculated by subtracting the interview date from the child's reported birthdate and rounding the difference to the nearest month.

In Ethiopia, the children surveyed were born between December 2000 and September 2002. In India, they were born between April 2001 and April 2002, and in Peru, between December 2000 and June 2002. In Vietnam, most were born between January 2001 and September 2002 (with three who were born in 2000).

Maths test: Languages of administration

Test difficulty can differ with the language of administration. The Ethiopian team administered, the Round 2 and 3 tests in 9 languages (Afarigna, Amharic, Guraghigna, Hadiya, Oromifa, Sidamigna, Siltigna, Tigrigna, Wolayta), and the Round 4 and 5 tests in 7 languages (Amharic, English, Hadiya, Oromifa, Sidamigna, Tigrigna, Wolayta). The language variable I created for all rounds includes the seven languages of Rounds 4 and 5, plus an 'Other' category to capture the additional languages in earlier rounds.

The team in India administered the Round 2 test in 5 languages (Hindi, Kannada, Oria, Telugu, Urdu); and 'Other' local languages. The team administered the Round 3 the same 5 languages without the 'other' category. In Rounds 4 and 5, they administered the test only in Telugu. The language variable I created for all rounds includes Telugu and an 'Other' category to capture all of the additional languages from earlier rounds. 'Other' represents around 200 children in Round 2 and 40 children in Round 3, of a balanced panel where the following information is available across all rounds: test score, asset index and test language.

In Peru, the team administered the Round 2 test in 5 languages (Spanish, Quechua, native of the Amazon, Spanish and Nomatsiguenga). They administered the Round 3 and 4 tests in 4 languages (Spanish, Quechua, Aimara, Spanish-and-Quechua), and the Round 5 test in 3 languages (Spanish, Quechua and Spanish). The language variable I created for all rounds includes the 4 languages included in Rounds 3 and 4, along with an 'Other' category to capture the additional languages in Round 2.

The Vietnam team administered the Round 2 test in 5 languages (Vietnamese, H'Mong, Nung, Dao, Glay), plus an 'Other' category, and in Vietnamese only in all subsequent rounds. I created a language variable for all rounds that includes Vietnamese and an 'Other' category for the additional languages of earlier rounds. 'Other' represents around 100 children in Round 2 of the balanced panel.

Vocabulary test: Languages of administration

The interviewers administered the vocabulary tests in a similar, though not identical set of languages, as the maths tests. Based on the Young Lives data, in Ethiopia the Round 2 languages were Amharic, Oromifa, Tigrigna, and an 'Other' category (4). In Round 3, the test languages were Amharic, Hadiya, Oromifa, Sidamigna, Siltigna, Tigrigna, Wolayta (7). The Round 4 test languages were Afarigna, Amharic, Agewigna, Guraghigna, Hadiya, Kembategna, Oromifa, Sidamigna, Somaligna, Tigrigna and Wolayta (11). The Round 5 tests were administered in only three languages – Amharic, Oromifa and Tigrigna. The language variable I created for all rounds includes the six most common languages (Amharic, Hadiya, Oromifa, Wolayta, Sidamigna, Tigrigna), with an 'Other' category to capture the additional languages in Rounds 3 and 4. The 'Other' category represents 139 children in Round 2, 49 in Round 3 and 11 in Round 4, of the balanced panel.

The India team administered the Round 2 test in 2 languages (Kannada and Telugu) and included an 'Other' category. In Round 3, they administered the test in 5 languages (English, Kannada, Oria, Telugu, Urdu). In Rounds 4 and 5, the team administered the test in only Telugu. The language variable I created for all rounds includes English, Kannada, Telugu and an 'Other' category to capture the additional languages administered. 'Other' represents around 118 children in Round 2 and 37 children in Round 3, of a balanced panel where the following information is available across all rounds: test score, asset index and test language.

In Peru, the researchers administered the Round 2 test in 2 languages (Spanish and Quechua), and included an 'Other' variable. In Rounds 3 to 5, they administered it in 3 languages (Spanish, Quechua and Spanish-and-Quechua). The language variable I created for all rounds includes the three language categories included in

Round 3 to 5, along with an ‘Other’ category to capture the additional languages in Round 2. The ‘Other’ category represents 31 children in Round 2, of the balanced panel.

In Vietnam, the team administered the Round 2 test in 2 languages (Vietnamese and H’Mong), including an ‘Other’ category. In Round 3, they used two languages also (Vietnamese and H’Roi). In all subsequent rounds, 4 and 5, the team administered the test only in Vietnamese. The language variable I created for all rounds includes Vietnamese and an ‘Other’ category, to capture the additional languages in earlier rounds. ‘Other’ represents 108 children in Round 2 and one child in Round 3 of the balanced panel.

There is quite some variation in the number of languages the vocabulary test was administered, raising concerns as to whether they all were equally able to capture variation in vocabulary skills in the sample populations. Reassuringly, the number of children in ‘other’ categories is relatively small.

Height-for-age

A child’s height-for-age reflects under-nutrition. Each child’s height is measured and then standardised against a World Health Organization (WHO) defined reference population. A child is classified as stunted when she lies at least two standard deviations below the median height-for-age of the reference population. The Young Lives combined data set for Rounds 1–5 includes the child’s height-for-age in each round.

Gender

Interviewers asked caregivers for the child’s gender in Round 1, offering binary options (male or female). The Young Lives combined data set for Rounds 1–5 includes gender.

Firstborn

The Young Lives team asks about the child’s birth order. Using data for Round 1, I created a dummy representing whether a child is first-born or not.

Ethnicity/caste

In order to account for the possibility that particular populations may be discriminated against or, conversely, may receive an advantage, I control for a variable that represents a main dividing line in each society. In Peru and Ethiopia, the variable is mother tongue; in Vietnam, it is ethnicity and, in India, it is caste. These selections are based on my discussions with the Young Lives team members who shared with me the most useful indicator of dividing lines, as recommended by country teams.

These tend to overlap with other dividing lines, for example ethnicity and language overlap, and also overlap with regional disparities in each respective country.

6.4.2 Household level

Household size

The Rounds 1–5 combined data set includes household size for each round. It is simply the number of people that respondents report to be residing in the household, including children, in each round.

Maternal employment

I use the maternal employment data collected in Round 2 (when children were 5 years old, the age when they would attend ECE centres) and group the categories into: no employment, agricultural employment and non-agricultural employment. 'No employment' includes: unemployed, household chores, household dependent, begging, and other unpaid activity. Agricultural employment includes: self-employed (food crops), self-employed (non-food agriculture), self-employed (aquaculture), self-employed (livestock), wage employment (agriculture), annual farm servant, and other agriculture. Non-agricultural employment includes: self-employed (manufacturing), self-employed (services), self-employed (business), self-employed (business), wage employment (non-agriculture), regular salaried employment, and house maid.

6.4.3 Community level ID

To control for time-invariant community level effects, I include a dummy for the community where the child resided when they were1year old. I explain my logic for doing so in the Analytical Framework Chapter 2. The children in Round 1, belonged to 20 communities. Community ID is available in the Rounds 1–5 combined data set.

To clarify, the variables I was able to use without cleaning were most of the control variables (height-for-age, gender and the community ID). Those that I spent time cleaning, calculating and, in some cases standardising (across countries), were whether the child was firstborn, ethnicity/caste, participation in ECE, type of ECE attended, and the language of test administration. Those that I constructed myself, using cleaned-up Young Lives data, were maternal education which I classified according to the ISCED categories, the asset index (which I calculated using IRT), and the maths and vocabulary scores (where I calculated a country and age specific z-score).

6.5 Conclusion

For the most part, the variables I use meet the standards of best practice. The instruments Young Lives use are remarkably similar to those used in DHS and LSMS surveys. Young Lives seems to have made an effort to standardise the Young Lives survey using well-established survey instruments (e.g. DHS, LSMS, MICS). The SES measures, similar to standard population survey measures, come with their own inherent challenges that I consider in my analysis.

The Young Lives team faced the challenge of implementing an age-appropriate instrument that would capture the variation in maths and vocabulary skills across the sample population when children were aged 5, 8, 12 and 15 across four different countries, in numerous languages. This is no easy feat. For the most part, the Young Lives team drew on already established maths assessments and adapted them to each country's context (i.e. CDA, TIMMS and PISA). They piloted the tests prior to implementation and revised them accordingly.

Drawing on established instruments did not work out for the PPVT, however, whose difficulty levels did not seem to translate well in all countries, as indicated by the challenges encountered in Rounds 2 and 3 (when the children were 5 and 8 years old) and the amendments made for Rounds 4 and 5 (when the children were 12 and 15 years old). While I make the theoretical case that maths would be a better measure of achievement than vocabulary, here, I make the case that, for reasons of data collection, I also prefer to use the maths scores as my main measure of achievement.

Chapter 7

Socioeconomic Status Gaps in Achievement

7.1 Introduction

The idea that all children should succeed regardless of their circumstances at birth is widely accepted. Yet most countries continue to witness an incongruous gap in academic achievement between children from high and low socioeconomic status (SES) families.

In this chapter I apply a set of research questions, oft considered for high-income countries, to data for low- and middle-income countries around the magnitude and the progression of SES gaps in achievement. For Young Lives countries, my first research question (RQ1) asks: How large are achievement gaps between children from high and low SES families at 5, 8, 12 and 15 years old, across countries? Additional questions I consider include: What happens to the gaps as children grow up? Do findings differ depending on the measure of SES used? What similarities and differences are there in the magnitude and pattern of SES gaps across the four countries?

Most evidence, particularly cross country evidence, on SES gaps in achievement is based on high-income country data (see Blanden et al., 2012; Bradbury et al., 2012, 2015, 2019; Dräger et al., 2023; Duncan & Magnuson, 2011b; Heckman, 2008; Linberg et al., 2019; Magnuson et al., 2012, among many others), with little evidence from low- and middle-income countries. I have identified four cross-country studies that examine SES gaps in achievement, namely Das et al. (2022), Lopez-Boo (2016), Reynolds et al. (2017), and Schady et al. (2015). This literature shows that first, SES gaps in achievement emerge before children commence formal schooling. Second, that the SES gap in achievement remains relatively stable though out childhood, a phenomenon termed ‘near-parallelism’. Third, that the magnitude of the raw

SES gaps in achievement seem to vary from around 0.5 to around 1.2 s.d.. See the Introduction Chapter 1 for a detailed review of the literature.

As I discussed in the Introduction, this thesis contributes to the research on low- and middle-income countries in three ways: first, I report SES gaps in achievement from 5 to 15 years old, the longest age range examined. Second, I report gaps based on three measures of SES (maternal education, the asset index and expenditure) and use maths and vocabulary scores as measures of achievement. These allow for a critical discussion on our choice of SES and achievement measures. Third, I consider country contexts in the interpretation of findings.

7.2 Differentials in parental socioeconomic status?

In the Country Context Chapter 5 I discuss in detail how these country differ in terms of their development and socioeconomic status. Education levels were highest in Peru and Vietnam, while those in Ethiopia were the lowest. Rutstein and Stavetei (2014) develop a comparable asset index, calculated from DHS, using a set of common assets, indicators of housing quality and access to services across countries. Based on this data, Peru has the highest asset index, Ethiopia the lowest, and India and Vietnam approximately the same.

Based on gross domestic product (GDP) per capita, in purchasing power parity (PPP) values based on constant 2011 international $s, a rough measure of expenditure potential in each country, Peru was by far the wealthiest nation (with ~$6,500), followed by India and Vietnam (at ~$2,600) and finally Ethiopia (at ~$600) (see World Bank, 2018). These countries also exhibited wide-ranging levels of inequality. Based on the Gini coefficient reported in United Nations University World Institute for Development Economics Research (UNU-WIDER) (2022), Peru was the most unequal country in the group (0.52 in 2005); it was followed by India and Vietnam (0.37 in 2004) and Ethiopia (0.30 in 2004). These differences may inform any differences we see across countries.

7.3 Data

To examine SES gaps in achievement, I draw on the multi-country longitudinal birth cohort data set of Young Lives. See Data Chapter 4 for a comprehensive discussion on the Young Lives data. The advantages of the Young Lives data are that it covers four diverse countries; the cohorts in each country were born at the same time; the follow-up rounds of data collection are conducted simultaneously; and the surveys share similar measures of achievement from age 5 to 15 years of age. Data has been collected at ages 1, 5, 8, 12 and 15 years (these are the average ages of the children in each round). Attrition rates are relatively low for a longitudinal study (between Round 1 (when the children were 1 year old) and Round 5 (when the children were 15 years old) attrition was 4.5 per cent in Ethiopia, 3.0 per cent in India, 8.2 per cent in Peru, and 2.3 per cent in Vietnam).

The data set is not nationally representative. Ethiopia, India and Vietnam have used non-probability samples, and in Peru a probability sample. As a reminder for the following three empirical chapters (chapters 7-9), the population of inference for each of the four countries consists of children born at the beginning of the millennium (2001/2) who:

- in Ethiopia come from the 20 selected villages and *kebeles* (urban neighbourhoods) and have higher levels of food deficiency and relatively good road access
- in India come from 20 selected *mandals* (akin to a county in the UK) from 6 (out of 23) districts of the states of Andhra Pradesh and Telangana
- in Peru are from a probability sample, excluding those from the wealthiest 5 per cent of districts in 2000
- in Vietnam come from 20 selected communes (akin to a council in the UK) selected from a province in 5 (out of 9) regions.

7.4 Measures

In Measures Chapter 6 I discuss in detail measures I use in this chapter, and in the Framework Chapter 2 I provide justification for the use of each of these variables. This section serves as a reminder of the essential information needed to interpret the estimates.

Achievement scores

The outcome variables I use to measure educational achievement are age-specific z-scores calculated from maths and vocabulary tests children took at age 5, 8, 12 and 15 years old. As I explain in Chapter 2 (the Framework Chapter), I conceive of maths and vocabulary scores as different measures of educational achievement, and maths scores as a more complete measure of academic achievement (compared with receptive vocabulary test scores). Maths tests not only demand an understanding of the vocabulary in the questions, but also require the child to identify and complete a mathematical expression and integrate it into the meaning of the overall question.

Measures of socioeconomic status

In their respective cross-country studies (of low- and middle-income countries) Reynolds et al. (2017) and Schady et al. (2015) use an asset index as a measure of SES, Das et al. (2022) use a composite measure of parental education and an asset index and Lopez-Boo (2016) uses expenditure data. I report the gap in achievement for each SES measure: maternal education, an asset index and household expenditure. I do so because each of these measures capture a different aspect of SES.

Maternal education

When the children were 1 year old, the Young Lives team collected information on maternal education and assets. Maternal education is reflective of a long-term measure of SES as it remains relatively constant through out the child's life. Maternal education not only captures social aspects of SES such as a mother's ability to navigate health care and education systems for herself and on behalf of her children but also reflects the mother's earning potential and other economic aspects.

I group maternal education into the International Standard Classification of Education (ISCED) developed by the United Nations Educational, Scientific and Cultural Organization (UNESCO). I define mothers with no education as having low SES and those with an upper secondary education or above as having high SES. According to UNESCO's (2020) online glossary entry for Upper Secondary Education an upper secondary education is "typically designed to complete secondary education in preparation for tertiary education, or to provide skills relevant to employment, or both." An upper secondary education or above provides mothers with the necessary literacy and analytical skills to enable them to navigate health and educational systems and guidelines and to support their children's education.

Some studies use both parents education level, so I conduct some sensitivity analysis, whereby I use the mother and father's average education level as a measure of socioeconomic status. The findings remain essentially the same (see figures in Annex F.3).

Asset index

Asset indices are a mainstay in research on disparities in children's outcomes in low- and middle-income countries. An asset index offers an alternative to income and expenditure data, which are often unavailable, unreliable, incomplete or incomparable. Filmer and Pritchett (1999, 2001) and Sahn and Stifel (2003) show that asset indices capture long-term SES, and do so with less measurement error than comes with expenditure data. I calculate the asset index using Item Response Theory (IRT) within each country, and scale it to have a mean of zero and a standard deviation of one.

For the asset index I estimate quartile groups from each country sample. Those in the bottom and top quartile group I define as having low and high SES, respectively.

Expenditure per capita

I conceive of expenditure data as a short term measure of SES, as it can fluctuate substantially across and within years, depending on seasonal variation (e.g. droughts, floods, locusts and more). I log household expenditure per capita, and henceforth, rather than referring to "logged per capita real household expenditure" I simply use "expenditure".

As with the asset index, for the expenditure data I estimate quartile groups from each country sample, and those in the bottom and top quartile group I define as having low and high SES, respectively.

For two reasons, I rely more heavily on SES gaps measured with maternal education and the asset index. First, I am interested in the initial circumstances into which children were born. It was only when the children were 5 years old that Young Lives started collecting household expenditure data. Second, maternal education and an asset index are more stable measures of parental SES.

Language of test

I include the language of the test in the regression to account for any variation that differences in the difficulty of the test cause. A question phrased in one language may be easier or more difficult to understand than it is in another language. This is especially the case when it comes to the vocabulary test. At younger ages, there is more variation in the language of the tests; at older ages, in some countries there is no variation in the language of the test. Including a control for the language of the test also allows me to include all the children in my analysis, whereas Reynolds et al. (2017), for example, exclude children who took a vocabulary test in a language that was not deemed a "major language" (at ages 5 and 12) and Lopez-Boo (2016) includes only children who are majority language speakers (see the Measures Chapter 6 for more details).

The language of the test may also be associated with SES. In Peru, for example, it is more likely that it is children from lower SES families who take the test in Quechua. An unintended consequence of including the language variable, therefore, is that the coefficient on language may absorb some of the variation in SES, therefore reducing the size of the coefficient on SES. This means that, where there was variation in the language of the test, the reported SES gaps may be smaller than if the variable had been excluded.

7.5 Methods

SES gaps in achievement: emergence and change

The SES achievement gap is measured as the difference between the average test scores of children from high SES families and the average test scores of children from low SES families. The Young Lives dataset has information on children's achievement and SES at age 5, 8, 12 and 15 years. To ascertain whether the gap changes with age, I need to know whether the SES gap at age 5 is significantly different from that at subsequent ages. This section discusses how I go about doing so.

First I use a regression-based approach to calculate the mean score for the high and the low SES groups at every age (5, 8, 12 and 15 years old). I do this for each measure of SES separately (i.e. maternal education, asset index and expenditure). I then predict the difference in the average scores between the high and the low SES groups. I do this for each country.

In Chapter 2, I provide the a generic equation 2.1. The specific equation I apply separately to each country is as follows:

$$y_a = \beta_0 + \beta_1 S_1 + \beta_2 L_a + \beta_3 R_a + \beta_4 (S_1 \cdot R_a) + \beta_5 (L_a \cdot R_a) + \epsilon \tag{7.1}$$

where

y = educational achievement score for each child measured at age a,

S = initial parental SES as a categorical variable,

R = round at which data was collected at age a,

L = language of the test at age a,

ϵ = error term.

S, R and L are categorical variables, and y is continuous. S is either maternal education, an asset index or expenditure. When I run the regression with maternal education as S, it has three categories: one group comprises the children with mothers with no education, the next includes children with mothers with a primary or lower secondary education and the third covers children with mothers with an upper secondary education or above at 1 year old. When I run the regression with the asset index as S, these are quartile groups, and the data were collected when the children were 1 year old. When I run the regression with expenditure as S, the children are again grouped into quartile groups using the data collected when the children were 5 years old. I cluster standard errors at the community level at 1 year old.

I set the bottom SES group as the reference group. Therefore the coefficient on the top SES group represents the size of the achievement gap (between the top and the bottom SES groups). This is the raw SES gap in achievement (accounting only for the language of the test). We can interpret this gap as including unmeasured predictors associated with SES that contribute to academic achievement.

My aim in this chapter is descriptive, to document these raw SES achievement at each age across countries, to set the foundation for further investigations into the causes and consequences of these gaps. Such evidence for low- and middle-income countries is limited.

Samples in each round are not independent because the same sample is interviewed repeatedly at different stages of childhood (see the Data Chapter 4). To account for the dependence between observations, I pool test scores from all rounds within each country. The relationship between initial SES (S) and test scores is likely to vary across rounds, as is the relationship between the language of the test and test scores. Therefore, to allow the coefficient on S and on L to vary across rounds, I include an interaction between initial SES round in which the test scores were collected ($S \cdot R$) and I do the same with the language of the test ($L \cdot R$).

Doing so fixes the error variance across rounds. Constraining the error variance to be equal across rounds is problematic at a theoretical and statistical level. In each round the variance of the unobserved factors contributing to achievement is more likely to differ than to be the same. For example, the error term might include the quality of schooling, and its variance is likely to be different at 5 and at 15 years old. If the error variances differ, the standard errors obtained from the pooled regression will be incorrect. To allow the error variances to differ across rounds, I follow the approach suggested by Gould (2024) (and also make the finite-sample adjustment discussed therein).

To allow the variance of the error term to vary across rounds, I run equation 7.1 with analytic weights. I first run the variance-constrained regression, equation 7.1, and predict the variance of the residual and collect the number of observations. With this information I derive the weight w for each round with the following equation:

$$w = v \cdot (n - 1)/(d) \tag{7.2}$$

where v is the variance of the predicted residual, n represents the number of observations, and d represents the degrees of freedom (in a Round r regression).

I then run equation 7.1 again, this time using the analytic weights derived from equation 7.2 (Young Lives data does not contain any additional weights). The standard errors of my estimates now take into account dependence between rounds and variable error variances across rounds. (Without these weights, equation 7.1 is a variance-constrained equation.) Henceforth, for brevity, when referring to equation 7.1, it will be the version that allows a round specific variance.

There are 4 countries, 2 measures of achievement and 3 measures of socioeconomic status, so in total I run 24 (weighted) regressions in the analysis.

To facilitate interpretation of estimates, I produce post-estimation summaries. (I use the *margins* command in Stata.) These post-estimations report the predicted gap in the average test score between the top and the bottom quartile SES group in each round with the language variable fixed at its average value. The estimates also include a 95 per cent confidence interval that I use to see whether the change is statistically significant from one round to the next.

I apply equation 7.1 (with weights) separately to each country as I find it is theoretically problematic to pool all country data into one regression. The exact definition of maternal education in one country, for example, will differ from that in another, and therefore its relationship with the outcome variable, in this case maths scores. The meaning of other measures will also differ across countries and their relationship with the outcome variable. Many cross-country studies for high-income countries also apply separate equations to each country's data set (see among others Blanden et al., 2012; Bradbury et al., 2012; Magnuson et al., 2012). In Chapter 3, I discuss in detail my rationale and approach to comparing across countries. The approach of applying separate regressions to each country signifies that statistical comparisons are appropriate within countries (i.e. in the section when I look at patterns across age). When I compare across countries, the comparisons are descriptive.

Sample

The younger cohorts of children in the Young Lives survey comprise 1,999 children in Ethiopia, 2,011 in India, 2,052 in Peru, and exactly 2,000 in Vietnam. The attrition rates are relatively low for a longitudinal study, but are not random. Sánchez and Escobal (2020) found that attrition varies by parental SES, area of residence and ethnicity (or caste in India). Those lost in Peru were relatively poorer, while those in the other three countries were wealthier.

I conduct a robustness check to see how my findings change by taking missing test scores into account. I test the worst case scenario by assuming that all children with missing test scores received a poor score of minus one of a standard deviation (-1 s.d.). When I conduct this robustness check, the story line of my analysis remains unchanged (see Figures in Annex F.3).

Whether I use a balanced or an unbalanced panel the findings remain substantively the same (see Figures in Annex F.3). In this chapter, I use a balanced panel including children with all pertinent data across the 5 rounds. This makes for a more intuitive interpretation of estimates as it uses the same set of children from 5 to 15 years old. This brings my sample to 1,435 in Ethiopia, 1,783 in India, 1,687 in Peru and 1,746 in Vietnam for maths scores. For vocabulary scores the equivalent figures are 1,490, 1,770, 1,603 and 1,555, respectively.

7.6 Descriptive statistics

In this section I describe the distribution of the SES measured and the impact it has on how to interpret the estimates. Table 7.1 provides the number of children in each low and high SES group for the analytic sample.

Table 7.1: SES group size (maths score sample)

	Ethiopia		**India**		**Peru**		**Vietnam**	
	Low	High	Low	High	Low	High	Low	High
Group size (n)								
Maternal education	831	82	1079	338	187	649	445	221
Asset index	395	321	455	406	559	343	509	325
Expenditure	359	358	446	445	422	421	437	436

Notes: Maternal education: low – none, high – upper secondary and above. Asset index and expenditure: low – bottom quartile group, high – top quartile groups

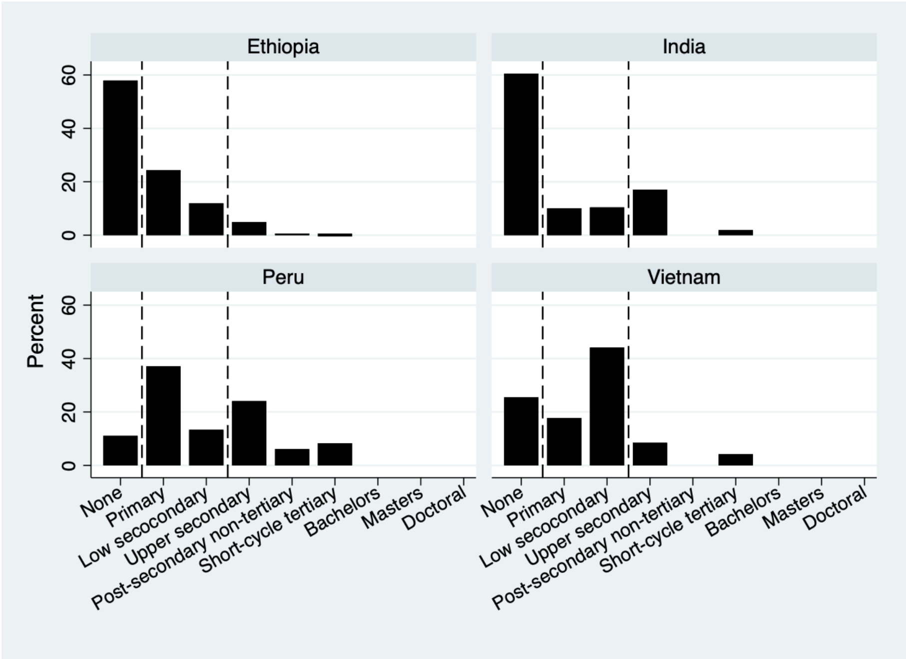
Maternal education

Figure 7.1 reports the percentage of children whose mother falls in each ISCED category (measured at one year old). Vertical dashed lines in Figure 7.1 indicate cutoff lines for low and high SES households. (The line to the left is the cutoff line for the low SES group and the line to the right is the cutoff line for the high SES group.)

Because maternal education is an absolute measure of SES, group size can vary substantially. In Ethiopia, 82 children had a mother who has an upper secondary education or above at 1 year old is 82, while 831 had a mother with no education (see Table 7.1).

In Ethiopia and India, the majority of mothers have no formal education, whereas in Peru and Vietnam, most have an education. Further, Peru has a larger proportion of mothers with a post-secondary education. The high SES group in Peru may

Figure 7.1: Distribution of maternal education (maths score sample)



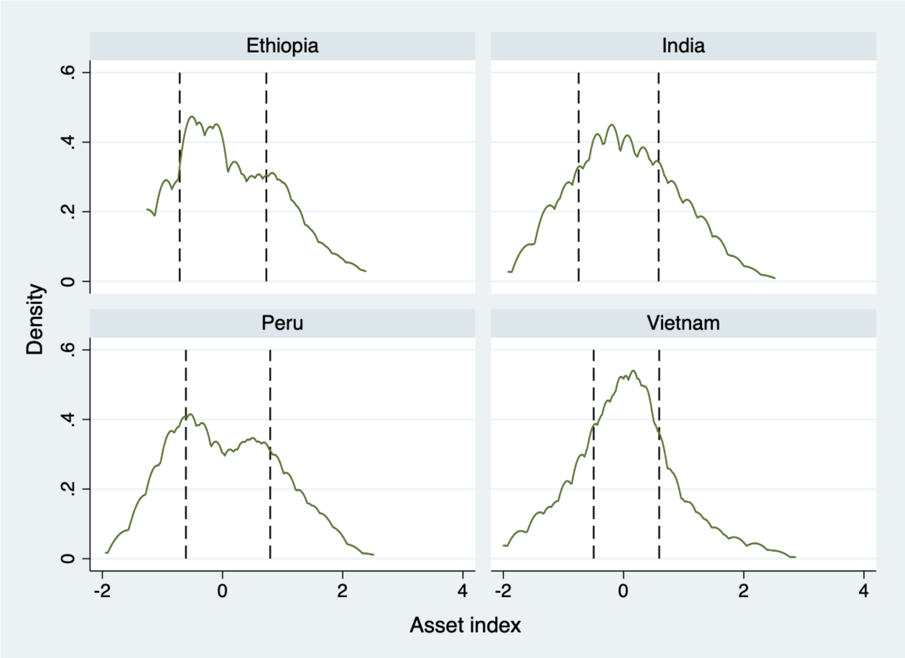
Notes: Vertical dashed lines on the left and right indicate cutoff lines for low and high SES households, respectively.

therefore be quite a heterogeneous group, with literacy and analytical skills on average more advanced than in the other countries.

Asset index

Figure 7.2 provides the distribution of the asset index at 1 year old. Here I again include the cutoff lines for the low and high SES households. Most notable is that, in Ethiopia, the asset index is truncated on the left. (I checked against the Young Lives asset index, which uses an additive rather than an IRT approach to calculate the index, and find that Ethiopia's asset index is also truncated on the left, indicating that the IRT methodology did not cause this). In 2001, Ethiopia was the poorest country in the study, and it appears that the information collected to compile the asset index did not capture variation among the poorest children in sample. The poorest children are therefore compressed at the bottom of the distribution. The truncation should not affect my grouping, as the 25th percentile cutoff line (the dashed line on the left) lies after the truncation. The distribution for the other countries does not show any truncation. India and Vietnam display a normal distribution whereas Peru follows more a lognormal distribution, indicating higher inequality.

Figure 7.2: Distribution of asset index (maths score sample)



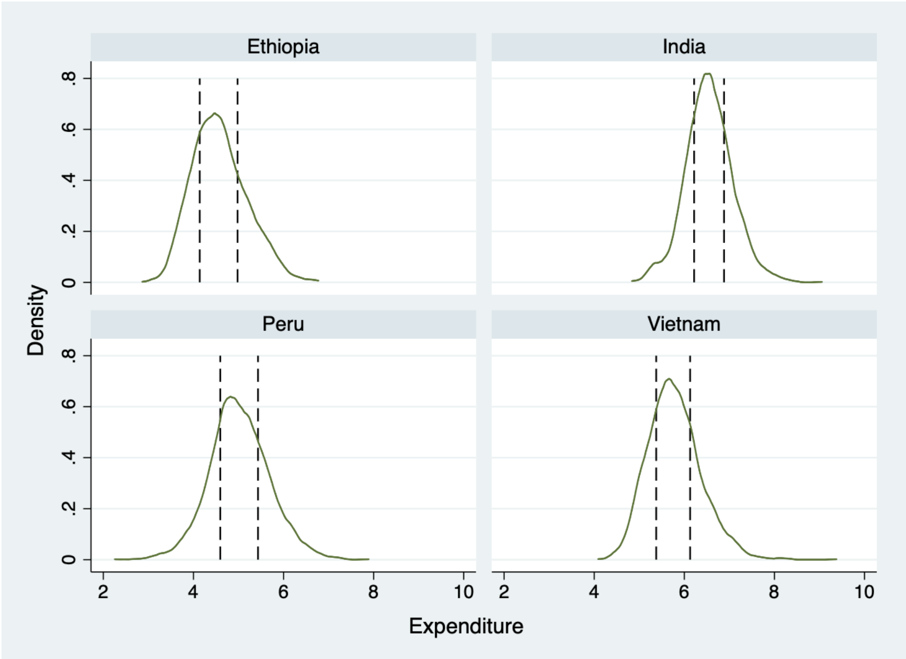
Notes: Vertical dashed lines on the left and right indicate cutoff lines for low and high SES households, respectively.

Expenditure

Figure 7.3 reports on the distribution of logged expenditure per capita, measured at 5 years old. The narrowest range is in India, Vietnam appears to have a long tail on the right side, and Ethiopia and Peru appear to have similarly shaped distributions.

The equivalent estimates for vocabulary scores are in Annex F.2 and are remarkably similar to those for maths scores. Annex F.2 also explains why the quartile groups vary in size.

Figure 7.3: Distribution of expenditure (maths score sample)



Notes: Vertical dashed lines on the left and right indicate cutoff lines for low and high SES households, respectively.

7.7 Estimates and their interpretation

Annex F.1 reports the estimates of equation 7.1 applied to Young Lives data for each country separately (specifically Tables F.1 to F.3). The equation includes interactions which complicates the interpretation of estimates. In order to facilitate interpretation, in the figures and table below I report the predicted SES gap in achievement (calculated by subtracting the mean predicted score of the low SES group from that of the high SES group). For brevity, henceforth I use the term 'SES gap', rather than 'predicted SES gap'.

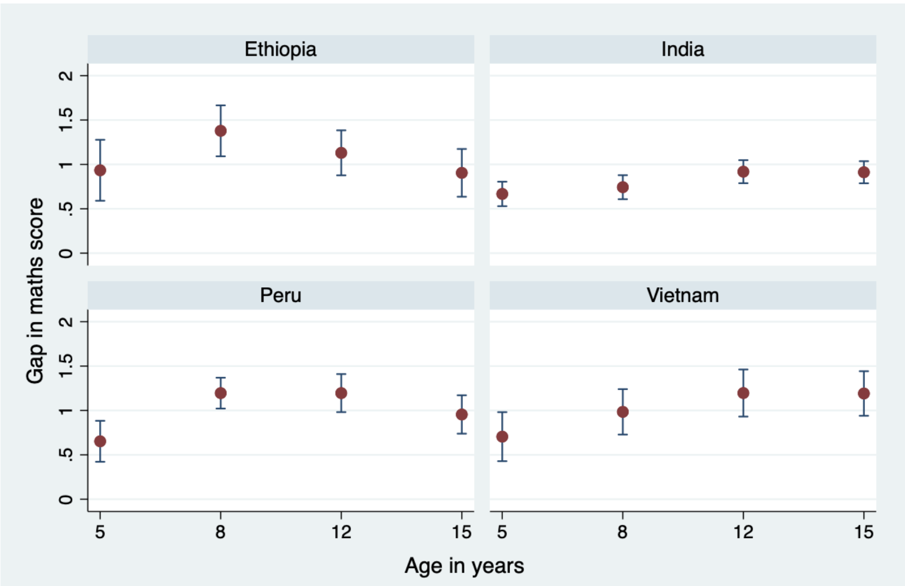
The SES gap in achievement when using maternal education is the difference between the average score of children whose mother obtained (at least) upper secondary education (based on ISCED categories) and that of those whose mother had no formal education. The SES gap in achievement when using the asset index and expenditure as measures of SES is the difference between the average score of the top and the bottom quartile groups.

Figure 7.4 shows the SES gap in average maths scores at 5, 8, 12 and 15 years old, for each country, when SES is measured using maternal education. Figures 7.5 and 7.6 show the equivalent gaps when SES is measured using the asset index and expenditure, respectively. These figures also includes the 95 per cent confidence interval for each estimate. On the y-axis are the age-specific z-scores whose units are standard deviations. A point at nil would signify no difference in average test scores between the high and the low SES group (i.e. there is no gap in achievement scores between the high and the low SES groups). The higher the point lies, the larger the SES gap in achievement.

If, within countries, the confidence intervals at two different ages do not overlap, we can say there is a statistically significant difference (at the 95 per cent level) in the size of the gap between each age. When they do overlap, as they do in most cases, then we cannot conclude there is a difference in the SES gap when children are one age compared with when they are another age.

I answer research questions 1 (how large are SES achievement gaps at 5 years old) and 2 (how to these gaps change as children grow up) in two separate sections below. The answers to research questions 3 and 4, on how these estimates compare with those of other studies and how the estimates differ or are similar across the four study countries, I integrate into all the sections below.

Figure 7.4: SES gaps in maths score (SES is maternal education at one year old)

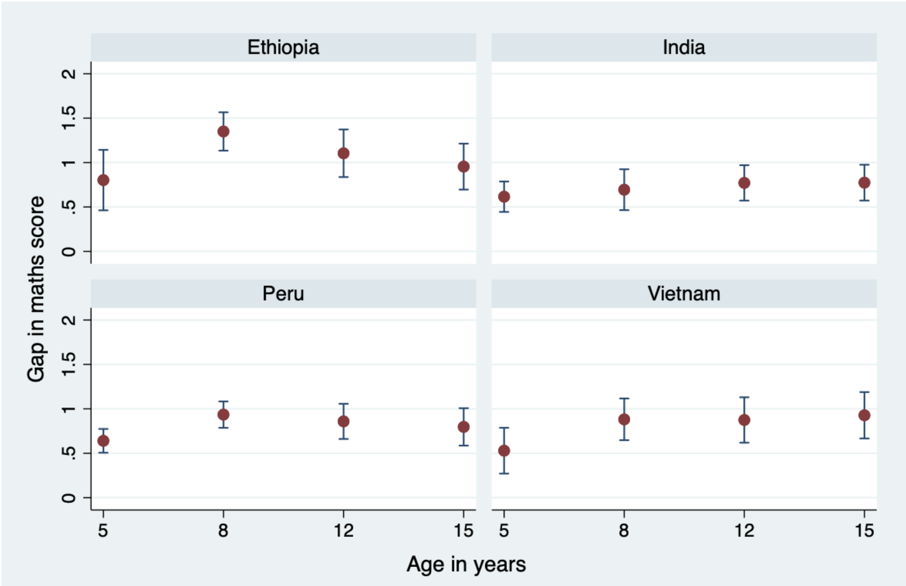


Notes: Gap = average score for children with mother with an upper secondary education and above minus average score for those with mother with no formal education, with 95 per cent confidence interval.

Gaps at 5 years old across the study countries

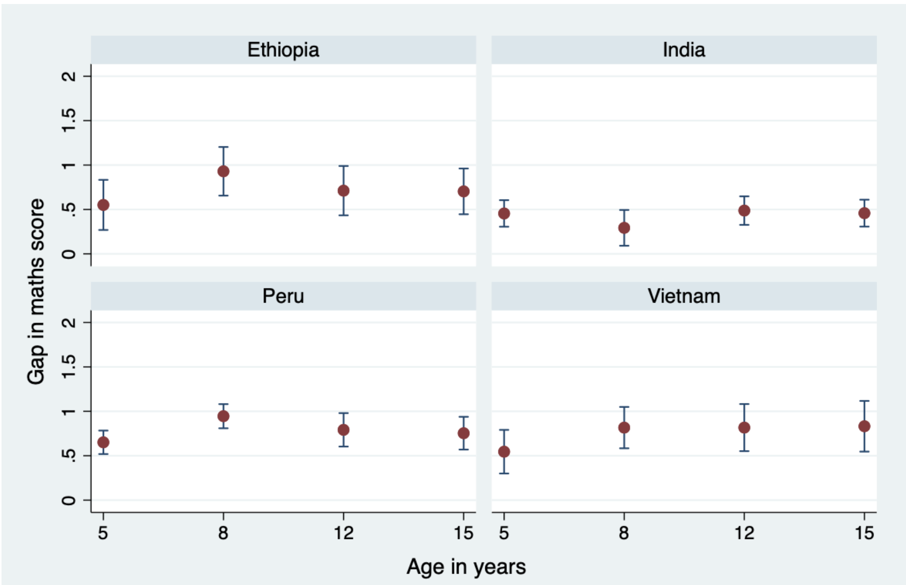
SES achievement already exist, across all countries, prior to school entry. The estimates show that SES gaps in maths scores already exist at 5 years old across all study countries. When I use maternal education as the measure of SES, the gaps at 5 years old are 0.93 s.d. in Ethiopia, 0.67 s.d. in India, 0.65 s.d. in Peru and 0.70 s.d. in Vietnam. The magnitude of the gaps is similar to that estimated for high-income countries and for other middle- and low-income nations, as reported in Bradbury

Figure 7.5: SES gaps in maths score (SES is asset index at one year old)



Notes: Gap = average score for children in the top quartile group minus the average score for those in the bottom quartile group of the asset index, with 95 per cent confidence interval. Outcome variable is maths scores (age-specific z-scores) and predictors are, in turn, maternal education, the asset index and expenditure. Additional controls include the language of the test.

Figure 7.6: SES gaps in maths score (SES is logged real expenditure per capita at 5 years old)



Notes: Gap = average score for children in the top quartile group minus the average score for those in the bottom quartile expenditure group, with 95 per cent confidence interval.

Table 7.2: SES gaps in maths scores (age-specific z-scores, s.d.), with 95% confidence interval

	Ethiopia	**India**	**Peru**	**Vietnam**
Mother's education				
5 years old	0.93	0.67	0.65	0.70
	(0.59-1.28)	(0.53-0.81)	(0.42-0.88)	(0.43-0.98)
8 years old	1.38	0.74	1.20	0.98
	(1.09-1.67)	(0.61-0.88)	(1.02-1.37)	(0.73-1.24)
12 years old	1.13	0.92	1.20	1.20
	(0.88-1.38)	(0.79-1.05)	(0.98-1.41)	(0.93-1.46)
15 years old	0.90	0.91	0.95	1.19
	(0.64-1.17)	(0.79-1.04)	(0.74-1.17)	(0.94-1.44)
Asset index				
5 years old	0.80	0.62	0.51	0.79
	(0.46-1.14)	(0.44-0.79)	(0.51-0.77)	(0.27-0.79)
8 years old	1.35	0.69	0.79	1.12
	(1.13-1.57)	(0.46-0.92)	(0.79-1.08)	(0.65-1.12)
12 years old	1.10	0.77	0.66	1.13
	(0.84-1.37)	(0.57-0.97)	(0.66-1.06)	(0.62-1.13)
15 years old	0.95	0.77	0.59	1.19
	(0.7-1.21)	(0.57-0.98)	(0.59-1.01)	(0.67-1.19)
Expenditure				
5 years old	0.55	0.46	0.65	0.55
	(0.27-0.83)	(0.31-0.6)	(0.52-0.78)	(0.3-0.79)
8 years old	0.93	0.29	0.94	0.82
	(0.66-1.2)	(0.09-0.49)	(0.81-1.08)	(0.58-1.05)
12 years old	0.71	0.49	0.79	0.82
	(0.43-0.99)	(0.33-0.65)	(0.6-0.98)	(0.55-1.08)
15 years old	0.70	0.46	0.75	0.83
	(0.45-0.96)	(0.31-0.61)	(0.57-0.94)	(0.55-1.12)

Notes: The SES gap when using maternal education as a measure for SES is the average score for children with a mother with an upper secondary education and above minus the average score for those with a mother with no formal education. The SES gap using the asset index and expenditure as a measure for SES is the average score for children in the top quartile group minus the average score for those in the bottom quartile group. Controlling for language of test.

et al. (2012, 2015, 2019) and Schady et al. (2015).

How SES gaps in achievement change with age across the four study countries

From 5 to 8 years old

Irrespective of the SES measure, the gaps in achievement persist with age in the four countries. All SES gaps are statistically significantly different from zero at the 95 per cent level. In other words, there is statistical evidence that there is a socioeconomic gap in achievement.

There is a pattern. In Ethiopia, Peru and Vietnam, there is an apparent increase in the gap between 5 and 8 years old, for all SES measures. In India, the gap increases only slightly between 5 and 8 years old when using maternal education and the asset index as a measure of SES, and not when using expenditure. Peru is the only country where the increase between 5 and 8 years old is statistically significant, and it is so for all measures of SES. When using maternal education, the SES gap in maths scores for Peru increases from 0.65 s.d. to 1.20 s.d. between age 5 and age 8 years.

An important transition happens for children between 5 and 8 years old, as they enter primary school. This transition may explains the apparent widening of the SES gap in maths scores. Perhaps it is that the children from high SES families are better prepared for primary school as they are more likely to have attended ECE of good quality. The increase in India is smaller because, unlike in other countries, ECE was generally deemed of poor quality when the Young Lives children were attending ECE (see ICDS, 2007 and SCERT, 2011), except for the children for the richest families (whom the Young Lives survey did not sample).

Another explanation could be that the ECE schooling system in Ethiopia, Peru and Vietnam is stratified by SES so that the children from high SES families receive better quality education, explaining the jump in their achievement score between age 5 to 8 years. In addition, children from higher SES families live in an environment that is more conducive to learning, where they my have fewer caring or work responsibilities and better nourishment, and may be better able to learn. For example, research from high-income countries shows that living in poverty affects children's brain development (as discussed in Kim et al., 2018). Further research is necessary to establish why differences emerge.

After 8 years old

Beyond 8 years old the SES gaps in achievement scores appear to narrow in Ethiopia and Peru but widen in India and Vietnam, when using maternal education as the SES measure. Yet the changes are not statistically significant. Their confidence intervals all overlap. This is consistent with the findings of Reynolds et al. (2017) who report (in thier Figure 2) that the confidence intervals of the unadjusted SES gaps in vocabulary at 5 and 12 years old overlap (when using both an asset index and parental education as measures of SES). Schady et al. (2015) find a similar, near-parallelism across the three countries with panel data (For example, in the country with the longest duration of panel data, Ecuador, authors find that the SES gaps at 12-13 years old is very similar to that those at 5-6 years old). While I cannot say whether the SES achievement gap narrows or widens with age, it certainly persists with age (see Annex F.2).

By 15 years old, the SES gaps in achievement are not statistically significantly different to those when the children were 5 years old (see Figures 7.4 to 7.6). This is the case for both maths and vocabulary scores. My findings confirm the near-parallelism found that Bradbury et al. (2015), Heckman and Mosso (2014), Magnuson et al. (2012), and Schady et al. (2015) find in both high- and low- and middle-income countries.

At 12 years old the SES gap (when using maternal education) in maths scores ranges from 0.90 s.d. in Ethiopia to 1.19 s.d. in Vietnam. Das et al. (2022) find that a standard deviation increase in scores at 12 years old is associated with an increase of about one to two years of education by the time children are 22 years old, when controlling for the asset index and parental education.

Difference in measures across study countries

Are there differences in my findings when I use different measures of SES? The largest gaps are evident when mother's education is used, except in Vietnam at 12 and 15 when they are quite similar. Indeed, in three of the four countries, SES gaps using maternal education and the asset index were larger than those using expenditure data. This is perhaps not surprising given that both are measures of long-term SES (i.e. they do not tend to fluctuate as much across seasons and years) and both were collected when the children were one year old. Achievement scores reflect an accumulation of investments in children's health and learning, so it is perhaps not surprising that the largest gaps are evident when using long-term measures of SES. It was not the case in Peru, where SES gaps based on expenditure were larger than those based on an asset index.

In Vietnam, SES gaps using the asset index and maternal education were of a similar size (when using maternal education as a measure of SES the gap is slightly smaller at 5 and 8 years old but slightly larger at age 12 and the same at 15 years old, when compared with the SES achievement gap using the asset index).

SES gaps are smallest when using expenditure, except in Peru. At the time of data collection, Peru was the wealthiest of the four countries studied. Therefore, expenditure may not have been as unstable as a measure of SES. Seasonal effects may have a lesser impact on Peru's more formalised economy. Further investigation into the reasons for these differences is merited.

Across all SES measures, the pattern of the gaps is similar: gaps already exist at age 5 and they persist throughout childhood. They are not statistically significantly different at 15 years old from what they were at age 5.

SES context and differences across countries

Kim et al. (2019), after reviewing the research on the relationship between SES and academic outcomes in low- and middle-income countries, hypothesise that the relationship will be weaker in poorer countries and will progressively strengthen with rising wealth. Reynolds et al. (2017) write that "[l]arger gaps arise from stronger associations between SES and child outcomes" (p. 770). This would imply that the gaps in Ethiopia, the poorest of the four countries, ought to have the smallest SES gaps in achievement, and those in Peru, the wealthiest country of the sample, the widest. My findings do not bear this out, however. At age 5, and using maternal education as the measure of SES, my findings contradict the hypothesis of Kim et al. (2019): the SES gap in Ethiopia is 0.93 s.d. and that of Peru 0.65 s.d.. At 5 years old, and using the asset index and expenditure, the hypothesis also does not hold. Based on the expenditure measure, Peru has a larger SES gap than Ethiopia (0.65 s.d. and 0.55 s.d., respectively), though India has a smaller gap than Ethiopia (at 0.46 s.d). When using the asset index, Peru also does not have the largest SES gaps; rather Ethiopia's are larger (contrary to the hypothesis).

The inconsistent relationship between a country's SES context and the SES achievement gaps signifies a broader set of influences on achievement gaps in each country. Other aspects, such as historical, demographic, social and economic context factors, are likely to influence the size and pattern of SES achievement gaps. For example, levels of maternal education are quite varied across the study countries, with Peru having the largest proportion of mothers with tertiary education, yet the SES gaps based on maternal education are largest in Vietnam at 15 years old (though, this is not the case with vocabulary scores).

If indeed genetic transmission in the outcome is the same across countries and cross-country differences reflect different environments, policy and otherwise, as

Ermisch, Jäntti, et al. (2012) argue, these findings indicate there is perhaps something in the DNA of each country's environment and policies that is shared. The pattern of these SES gaps across countries are remarkably similar.

Sensitivity analysis of vocabulary scores

Annex F.2, Figures F.4 to F.6, give the equivalent figures for vocabulary test scores. These findings are essentially the same: socioeconomic gaps in vocabulary scores exist at 5 years old and they persist at about the same level through to 15 years old. The pattern of an apparent increase in scores between 5 and 8 years old is not as pronounced with vocabulary scores. In Peru, the increase is not statistically significant for vocabulary scores, whereas it is for maths scores.

7.8 Conclusion

This chapter contributes to the research on the relationship between parental SES and children's achievement in low- and middle-income countries, by adding to the spare evidence of these gaps. This study documents gaps from 5 to 15 years old across four different countries, and shows that they persist through to adolescence. It uses two different measures of achievement and three different measures of SES.

Despite the diverse levels of development across the Young Lives countries, there is a remarkable similarity in the size of the SES gaps in achievement and their pattern across age, as well as with those reported for high-income countries. One can interpret the near-parallelism of the SES gaps from ages 5 to 15 as proof of an effective social policy programme in all countries to keep advantage (or disadvantage) from intensifying over childhood. A less optimistic view is to conclude that there are structural problems in society that limit the potential of low SES children, while reinforcing that of high SES children.

Before drawing any strong conclusion, it is necessary to look behind the average scores of these groups. In the next chapter I examine whether the trajectories of children who were initially high achievers are related to their family's SES. If the relationship proves to be positive, then the environment to which children are exposed during their schooling years will matter a great deal for their academic development.

This near-parallelism from 5 and 15 years old invites the question whether the gap at age fifteen would narrow if the SES gaps in achievement at age 5 were narrower. Between conception and age 5, a child's brain develops quickly (as discussed in Grantham-McGregor et al., 2007). A plethora of influences shape this development. Kim et al. (2018), for example, show that for children in high-income countries, poverty has negative impacts on brain development. Understanding when and why

SES gaps in achievement emerge is essential to inform policy that aims to narrow them.

A widely used policy instrument aimed at reducing SES gaps in achievement early in life is ECE. In Chapter 9, I explore the effectiveness of this policy option by examining the association between preschool education and achievement scores in the four study countries.

My analysis uses three different measures of SES: maternal education, an asset index and expenditure. While the findings I obtained with these measures lead to similar conclusions, they yield differences in the size of the SES gaps. I argue this is mostly down to which aspects of SES each measure captures and the nature of how each is measured. Given the limited research into SES gaps in achievement in low- and middle-income countries, it would be useful for researchers to continue to use various measures of SES in their analysis, and compare and contrast their estimates. A larger body of evidence will be better able to understand why and how these measures behave differently in different contexts.

Chapter 8

Socioeconomic Status Differences in Children's Achievement Trajectories

8.1 Introduction

Numerous studies report that the socioeconomic status (SES) gap in achievement has an almost parallel pattern from early ages well into childhood (see Bradbury et al. (2015), Heckman and Mosso (2014), Magnuson et al. (2012), and Schady et al. (2015)). A SES gap in achievement is the difference in the average achievement scores between children from high and low SES families. I also report this near-parallelism in the SES gap in achievement from five to 15 years old in Ethiopia, India, Peru and Vietnam (see Chapter 7). In all four countries, children from low SES families lag behind their wealthier counterparts, in terms of achievement scores. This persistent SES gap is of concern as it means that a significant portion of the population is unable to reach their potential and are therefore unable to contribute to society to their full potential, not because of anything they do, but because of the family they are born into.

Looking at these persistent SES gaps in achievement, it may seem that almost nothing happens to the SES gap after school entry, that the SES gaps that emerged by five years old are the reason the gap has persisted through to 15 years old, and that the simple solution to these persistent SES achievement gaps is to focus solely on narrowing the SES gap prior to school entry. However, the story is not as simple as it may appear.

What complicates matters is that over time there is considerable movement of children within the achievement distribution. As Feinstein (2003) shows using British data, Bradbury et al. (2015) using US data and Blanden et al. (2012) using UK and

Australian data, is that a child's position in the achievement distribution is not fixed at school entry. We can see from Table 8.1 how much movement there is across the four study countries, in terms of maths scores. Table 8.1 shows what percentage of children from a particular quartile group at five years old score in each quartile group at 15 years old. The diagonal (in bold) is the percentage of children who remain in the same achievement group at five and 15 years old.

The first observation to note in Table 8.1 is that only about a third of the children (32% in Ethiopia and Peru, 31% in India and 35% in Vietnam) are in the same achievement group at five and 15 years old. The other two thirds of children moved into another achievement group. The second observation to note is the proportion of children from the initially highest or lowest achievement groups end up in the opposite (lowest and highest, respectively) achievement groups. For example, in Ethiopia, 14 per cent of the children who were in the highest achievement group at five years old are in the lowest achievement group at 15 years old (the equivalent figures for India, Peru and Vietnam are 15, 17 and 16 per cent, respectively). This is more movement than is reported in some high-income countries. (The percentage of children who end up in opposite quartile groups between five and seven years old in the UK is about 5 per cent, in Australia it is about 8 per cent (both reported in Blanden et al. (2012)). In the US the percentage of children who end up in opposite quartile groups between four/five and 13/14 years old is between 4 and 7 per cent (as reported in Bradbury et al. (2015)).) While some of this movement may reflect some imprecision of the maths test administered to five years old, this does not explain all of the movement.

Research using high-income country data shows that this movement is linked to the SES of the children's families (see Blanden et al. (2012), Bradbury et al. (2015), Feinstein (2003), Jerrim and Vignoles (2013), and Washbrook and Lee (2015)). These studies show that the trajectories of initially high, or low achieving children diverge depending on their family's SES. If this is the case in Ethiopia, India, Peru and Vietnam then, there is indeed 'something going on' in the school years that may be contributing to the persistent SES gap in achievement.

Plotting the achievement trajectories of initially high and low achieving children by their parent's SES may seem to be an academic exercise, but it has caused quite some political and academic controversy. When Feinstein (2003) plot the trajectories of high and low achieving British children by their SES, he showed how initially low achieving children from high SES families overtook their initially high achieving counterparts from low SES families (in their average rank, based on test scores) (see Figure 8.1). His analysis had a significant influence on UK policy, with a leading researcher writes that "[f]ew pieces of recent longitudinal research have had as much influence in UK policy circles as Leon Feintein's..." (Bynner, 2015,

Table 8.1: Transition matrices for maths scores: the percentage of children scoring in a particular quartile group at five years old scoring in each quartile group at 15 years old, by country (%)

Quartile group of maths score at 5 years old	**Quartile group of maths score at 15 years old**				
	1 (Lowest 25%)	2	3	4 (Highest 25%)	Total
Ethiopia					
1 (Lowest 25%)	**31**	33	20	17	100
2	31	**24**	27	17	100
3	24	26	**27**	23	100
4 (Highest 25%)	14	20	22	**43**	100
India					
1 (Lowest 25%)	**38**	28	21	14	100
2	30	**25**	27	18	100
3	18	27	**24**	30	100
4 (Highest 25%)	15	19	28	**37**	100
Peru					
1 (Lowest 25%)	**39**	28	20	13	100
2	23	**30**	25	23	100
3	22	20	**26**	32	100
4 (Highest 25%)	17	23	29	**31**	100
Vietnam					
1 (Lowest 25%)	**40**	27	21	12	100
2	26	**30**	26	19	100
3	18	23	**29**	29	100
4 (Highest 25%)	16	20	24	**40**	100

Notes: The numbers in bold on the diagonal are the percentage of children who remain in the same quartile group at five and 15 years old. N = 1,457 for Ethiopia, 1,714 for India, 1,675 for Peru and 1,601 for Vietnam.

p. 331). It was instrumental in bringing issues of socioeconomic disadvantage to the fore and bolstered the case for major early childhood interventions under the then Labour government, such as the Sure Start programme, as Bynner (2015) claims. As forcefully as it was used to support policy, so did its methodological scrutiny (by Jerrim and Vignoles (2013) and others) contribute to the eventual reversals of these policies by the then Conservative (and Coalition) governments. This has raised a vibrant debate on: first, the methodological rigour and reliability of this type of analysis (i.e. how much Feinstein's findings are affected by regression to the mean, the age one classifies children into initial ability and the arbitrariness of the cut-off line is for the top and bottom initial achievement level); second, the meaning one can ascribe to the findings; and third, the reliance of this type of analysis to shape policy.

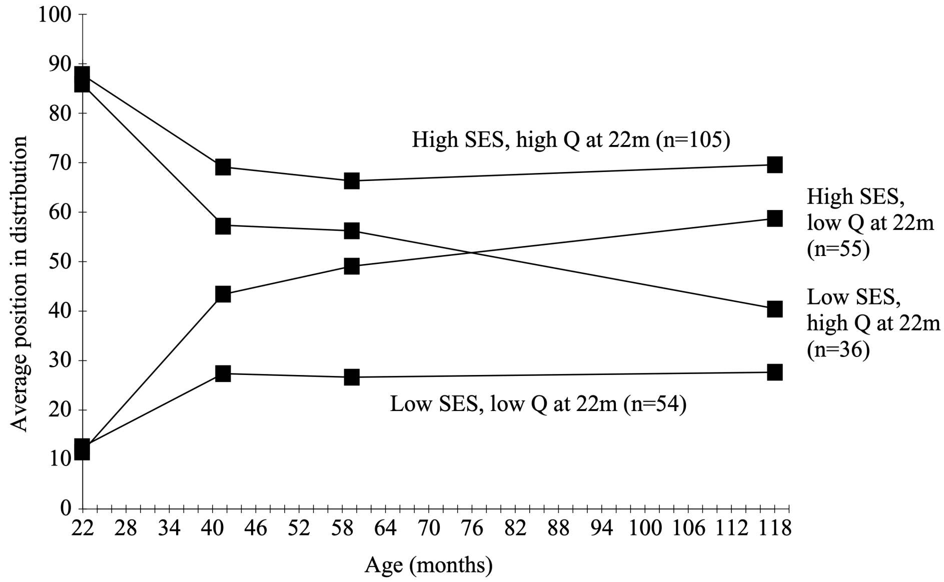


Figure 8.1: Feinstein's Famous Figure: Average rank of test scores at 22, 24, 60 and 120 months, by SES of parents and early rank position (using the British 1970 cohort data)

Source: Figure 2, Feinstein (2003)

In this chapter I focus mostly on the first area of debate: the methodological issues raised by the analysis in Feinstein (2003). I replicate this analysis, with some adjustments to address regression to the mean, using data from four low- and middle-income countries. Attempts at examining children's trajectories based on initial achievement level and SES are sparse. In his conclusion Feinstein (2003) writes "[i]t would be very interesting to know how much of these associations are reproduced elsewhere, in countries with perhaps less, or more, social inequality" (p. 89). There are only a few other studies that do so, namely Jerrim and Vignoles (2013) using UK data, Bradbury et al. (2015) and Washbrook and Lee (2015) using US data (who use the adjustments suggested by Jerrim and Vignoles (2013), which I discuss further on in the chapter) and Blanden et al. (2012) using UK and Australian data (who apply the approach used by Feinstein (2003)).

Plotting achievement trajectories by parental SES has not yet been done with data from low- and middle-income countries. This is most likely due to stringent data requirements for such an analysis. The Young Lives data set offers a unique opportunity to apply this type of analysis across multiple countries. It measures outcomes in a comparable measure for approximately 2,000 children per country on four occasions between five and 15 years old. Young Lives has the detailed data required to examine individual children's trajectories.

To illustrate the challenges of this type of analysis across multiple countries, Bradbury et al. (2015) attempt this analysis across four high-income countries, but face challenges as different country data sets have children born in different years (sometimes decades), with different ages at which they are followed up on and receiving different achievement tests. They end up dropping one country due to data limitations and focus their analysis on one country, the US.

First I discuss the data and measures I use in my analysis. Second, I discuss the main methodological critiques of the Feinstein (2003) approach and discuss alternative approaches to address these (suggested by Jerrim and Vignoles (2013)). I also clarify the methodological choices I make in my analysis. Third, I test both approaches on the Young Lives data and discuss the similarities and differences in the trajectories depicted. I also discuss what these findings reveal about the achievement trajectories in children in Ethiopia, India, Peru and Vietnam. In the conclusion I discuss the implications of these findings on policy recommendations related to the persistent SES gaps in achievement I reported in Chapter 7.

8.2 Data and measures

Implications of using Young Lives data

The Young Lives team intended for all country samples to be 'pro-poor', with a focus on lower SES populations. Hence why in Peru the wealthiest five per cent of districts were excluded from the sample. The samples used by Feinstein (2003) or Jerrim and Vignoles (2013) include families from across the SES distribution. Therefore, the trajectories I report in this chapter are likely more conservative (i.e. narrower in their distance from each other), than those I would report had the Young Lives samples included children from the highest SES families in society.

The first achievement test in the Young Lives data is at five years old, perhaps quite far along for a child's educational journey to be affected by parents SES.

Feinstein (2003) uses data for children as young as 22 months, and Jerrim and Vignoles (2013) from 36 months, for example. I consider this in the interpretation of the data.

Educational achievement

The Young Lives data offer two sets of achievement measures derived from tests administered from 5 to 15 years old: maths and vocabulary tests.

In Peru children are tested with the same Peabody Picture Vocabulary Test (PPVT) (adapted to Spanish with 125 words) in all rounds. In the other countries, Ethiopia, India and Vietnam, at 5 and 8 years old they are administered the same PPVT test (comprising 204 words), then at 12 and 15 years old they are administered a subset comprising of around 30 of these words. So at 12 and 15 years old children are administered the exact same set of words (within each country). This poses some methodological complications that I discuss below.

During the maths test a child faces a mixture of simple maths operations, multiple choice questions and word problems. While the maths tests from 8 to 15 years old share a few items, they are different tests across rounds.

I calculate age-specific z-scores by dividing the sample into the smallest possible age groups, in this case months, and computing z-scores within each age group. These scores have a mean of zero and a standard deviation (s.d.) of one.

Socioeconomic status (SES)

Three measures of parental SES are available in the Young Lives data. They are maternal education, an asset index and household expenditure. I use all three in this analysis, although my main findings are based on maternal education.

To identify children in high and low SES groups with the asset index and expenditure, I select the top and bottom quartile groups to represent each group, respectively. The asset index and expenditure are continuous variables, whereas maternal education is a categorical variable.

How I group maternal education is based on how maternal education is distributed in the sample population. For Ethiopia, India and Vietnam, the low SES group (based on maternal education) comprises of children with mothers with no education. The high SES group includes children with mothers with lower secondary education or above. Peru has the highest level of maternal education among the four study countries, so to have a grouping that takes this into account. In Peru the low SES group comprises of children with mothers with none or a primary education and the high SES group includes those with mothers who have an upper secondary education and above. (Note this is a different grouping for high and low SES in Chapter 7, where I apply the same cut-off lines to all countries.) The group sizes for high and low SES I report in Table 8.2 For further details on the measures used in this chapter please see the Measures Chapter 6.

Table 8.2: Socioeconomic status group size

	Ethiopia		**India**		**Peru**		**Vietnam**	
	Low	High	Low	High	Low	High	Low	High
Group size (n)								
Maternal education	845	262	1,034	507	806	648	414	905
Asset index	619	347	442	383	605	421	524	386
Expenditure	370	369	431	430	429	428	403	403

Notes Maternal education - Ethiopia, India and Vietnam: low = none, high = lower secondary and above. Peru: low = none or primary, high = upper secondary and above. Asset index and household expenditure - low = bottom quartile group, high = top quartile group. Grouping based on complete case analysis. This table differs to Table 7.1 as in Chapter 7 I apply the same cut-off lines to all countries, here Peru's cut-off lines for low and high SES are different to that of the other countries. I also have fewer variables to include for complete case analysis, in Chapter 7 I include the language of the test and the community ID, which reduces the size of the sample.

8.3 Methods

In this chapter I plot the trajectories of initially high and low achieving children by SES. To address some of the methodological issues raised around the work of Feinstein (2003), in this section I explain the following:

1. how I address regression to the mean;

2. which variable I choose to classify children into their initial achievement groupings, and which variable I use to track progress;

3. which variable I choose to classify children into SES groups;

4. what cut-off line I use to group children into high and low achievement and SES groups;

5. how I deal with attrition; and

6. discuss potential implication of classifying children at five years old.

Regression to the mean

The first and principal area of dispute in the analysis of children's achievement trajectories by SES, is how to address regression to the mean. Regression to the mean is a statistical phenomenon identified by Francis Galton (1886), and as Tu and Law (2010) state, it "never loses its power to surprise" (2010, p. 1251).

Take for example the following equation for a child:

$$Y_m = \beta_m M_m + \epsilon_m \tag{8.1}$$

where, at each time point (m) the observed maths score (Y) is a combination of the child's true maths ability (M) and classical measurement error (ϵ). If a child's observed maths score happens to be particularly high (in an outlier sense) at an initial time point, given the error process it is unlikely the child will have the same high score again. Indeed, she is likely to have a smaller error, hence taking the observed score back closer to the (longitudinal) mean for the child, hence the term regression to the mean. We see this in sports competitions as well where in one season a team is particularly lucky and wins the league, but in subsequent seasons reverts to their usual (i.e. longitudinal) position in the table (think Leicester City FC in the Premier League, or Emma Raducanu in the tennis competitions).

Jerrim and Vignoles (2013), who scrutinise the methodology of Feinstein (2003), explain that a reliance on tests administered to young children, poses a high risk of misclassifying children into initial achievement groups. They argue that tests administered at young ages tend to be noisy and have relatively large error variances, this means more young children are more likely to be misclassified (compared with more precise tests administered to older children). A group of children who, in an initial test, obtain a relatively high (or low) result owing to random error (i.e. good or bad luck) will produce a fewer extreme results in a subsequent test.

As SES gaps in achievement already exist at young ages, Jerrim and Vignoles (2013) contend that misclassification is most likely to happen in the middle groups (i.e. initially high achievement-low SES and initially low achievement-high SES). Consequently, it is in these groups that we would expect to observe a more acute regression to the mean. They contend the convergence of the trajectories of these two middle groups is largely due to regression to the mean. Jerrim and Vignoles (2013) draw on work by Davis (1976) and Ederer (1972) to suggest using two different tests to account for regression to the mean: one to divide children into initial ability groups and a second to set the baseline score from which change is measured.

When they propose this alternative approach, Jerrim and Vignoles (2013) assume the error structures of both measures are not correlated. That is, any random error (i.e. good or bad luck) experienced by a child on the classifying test will not be part of their baseline score. The authors acknowledge, however, that the tests they use are administered on the same day (as is the case with the Young Lives tests I use) and their errors may be correlated. For example a child may be ill on the day of the test. This means the approach suggested by Jerrim and Vignoles (2013) would not completely eradicate the effects misclassification. Rather, as they state, this approach would calculate "an upper bound" (Jerrim & Vignoles, 2013, p. 8) of the amount of change over time (rather than a point estimate).

Jerrim and Vignoles (2013) also caution that if the luck a child experiences on the second test is correlated with that of the first test, then the second test scores will still be partly conditional on the error in the first test. This means the measurement error that results from the effect of luck (i.e. error) will continue to affect the findings. This partially motivates which measure of achievement I use to track children's progress, and which one I use to classify children into initial achievement groups.

The debate on whether or how to address regression to the mean remains contentious. Goldstein and French (2015) consider the method proposed by Jerrim and Vignoles (2013) as a poor instrumental variable (IV) approach. They also contend that by dint of how children are initially selected, we expect the initially low achievement group of children from low SES families to, on average, have lower average achievement over time compared with their initially low achievement-high SES counterparts (and the same would apply within the high achievement group, with the high SES children expected to, on average, have higher average achievement over time compared with their initially high achieving-low SES counterparts). While this may be the case, this rationale cannot fully explain the convergence observed in the studies they critique.

Goldstein and French (2015) consider the approach of Jerrim and Vignoles (2013) as "inherently flawed and is not to be recommended" (Goldstein & French, 2015, p. 353). Crawford et al. (2017) endorse the Jerrim and Vignoles (2013) approach, which they implement in their own analysis. Bradbury et al. (2015) and Washbrook and Lee (2015) apply a similar method, only rather than use two different tests administered on the same day, they use two different tests administered six months apart (this reduces the chance that the error structures of the tests are correlated), and are explicit about their approach being akin to an IV approach. Blanden et al. (2012) are not concerned with the issues posed by regression to the mean and apply Feinstein's original methodology.

There is no consensus on how to most effectively address the methodological challenges posed by regression to the mean. At the moment, most of the analysis is focused on UK data (with the exception of Washbrook and Lee (2015) using US data, Blanden et al. (2012) comparing the UK with Australia and Bradbury et al. (2015) comparing the US with Australia and the UK). Most of the recent literature adopts the Jerrim and Vignoles (2013) approach (for example Bradbury et al. (2015) and Crawford et al. (2017)) and it seems to be the predominant approach in addressing regression to the mean, with no established alternative suggested to date. I therefore use the Jerrim and Vignoles (2013) approach in my analysis.

I conduct sensitivity analyses to examine whether findings change if I do not address regression to the mean. To do this, I apply the Feinstein (2003) approach to the same dataset, and use the same test to classify children and set the baseline from which to measure progress.

Variable choice for initial achievement groupings and tracking progress

Here I provide my rationale for the variables I select to classify children into initial achievement groups, as well as the measure of progress. I consider maths scores a more complete measure of educational achievement compared with receptive vocabulary test scores. Maths tests not only demand an understanding of the vocabulary in the questions, but they also require the child to identify and complete a mathematical expression and integrate it into the meaning of the overall question. This is evident in this question administered to fifteen-year-old children in Vietnam:

> Tickets for a train cost either 10 [dollars], 15 [dollars], or 30 [dollars]. Of the 900 tickets sold, 1/5 cost 30 [dollars] each and 2/3 cost 15 [dollars] each. What fraction of the train tickets was sold for 10 [dollars] each?

Therefore, I chose maths scores to track progress, and vocabulary scores to classify children into initial achievement groups.

There are also statistical reasons not to use vocabulary scores to track progress. An assumption underlying the Jerrim and Vignoles (2013) approach is that error structure of the test scores are not correlated across time. However, in the vocabulary tests, there is substantial repetition in the items used in these tests across rounds. Recall, in Ethiopia, India and Vietnam children are administered the exact same test, comprising of about 30 words, at 12 and 15 years old. This can lead to a correlation in error terms across rounds.

Test scores and 'true ability'

Jerrim and Vignoles (2013) assume that two different tests measure the same underlying 'true ability', and each test will rank children on the same scale of this 'true ability'. Feinstein et al. (2015) critique this definition as a statistical definition (not one that accords with a popular understanding of true ability) and one that can be (and has been) interpreted by those who do not understand this statistical concept, as being fixed and innate, eventhough it is not what the authors intend to communicate.

Test scores, however, are a reflection of a child's skills in answering those items correctly and random factors (such as the testing environment and how familiar the child is with the specific items in the test, that can be summarised as measurement error), as Goldstein and French (2015) discuss. The maths and vocabulary tests contain different items in their tests and also test different academic skills. Test items may also be more or less familiar to children from different strata in society (see Dorling (2010) who discusses how structures of privilege are reflected in intelligence tests). I therefore use the term educational achievement (rather than 'ability'), as I feel it better characterises what test scores represent. A test score reflects the number of correct items a child has achieved.

When I classify children using the vocabulary test, I am ranking children based on their vocabulary skills related to the items included in the test, and random luck. When I plot children's trajectory over time using their maths scores, those scores reflect children's maths skills related to the items included in those tests and random luck. We can therefore expect a difference in average test scores, within each initial achievement group.

I also conduct sensitivity analyses and use maths scores to classify children into achievement groups, and vocabulary tests from which to set the baseline score from which to measure progress. From a substantive perspective, these findings are different, as distinct skills are measured with different tests. They are also problematic from a statistical perspective, given the risk of correlation in the error structures across rounds. Given these caveats, the story they tell is relatively consistent with the findings reported here.

Variable choice for SES grouping

Few authors provide an explicit rationale for their choice of SES variable or dwell on the implications of their choice. The choice of SES variable used in analysis is as important to consider, as is the choice of variable to measure achievement.

When studying achievement trajectories of children by SES, we want a societal measure of long-term SES that reflects the way people are stratified in the society they belong. This can vary from country to country, context to context. We also want a measure that is able to identify both low and high socioeconomic status households effectively. Feinstein (2003) uses parental occupation when analysing UK data, and findings are similar when SES is based on parental education.

The Young Lives dataset offers three measures of initial SES shared across countries: mother's education, an asset index and expenditure. Different measures of SES may be more or less appropriate in distinct contexts.

My variable of choice for this analysis is mother's education. It is the most concrete and intuitive measure of the three, and involves no abstraction from reality or additional modelling (as do the other two measures of SES). Mother's education also provides a better measure of *initial* SES as it is measured at one year old, rather than five years old when expenditure data are first collected. Mother's education is also more stable than expenditure, particularly in the context of low- and middle-income countries. (Expenditure can fluctuate substantially from one season or year to the next. For example, in largely agricultural societies expenditure will vary according to the weather, crop yields, and transport and food prices. As food costs comprise a large proportion of poor people's expenditure, fluctuations in food prices will result in high fluctuations in expenditure within and across years.) Mother's education is also not subject to reverse effects, as would be the asset index or expenditure data. For example, were a child with special needs to be born into a household, families may to draw down on household assets or reduce expenditure, while mother's education will not be affected.

Mother's education has its own limitations. Compared with the asset index, it is more unevenly distributed across the populations of the four study countries. Low rates of maternal education in Ethiopia and India, mean about 60 per cent of mothers have no education. On the other hand, high rates of maternal education in Peru mean just 11 per cent of mothers have no education.

My aim, when using a measure of SES in this analysis, is not to imply anything deterministic about SES. Rather the SES measure serves as a societal measure of how people are stratified. While this does not have an implication on the findings, it may affect the meaning one ascribes to them.

To test whether findings substantively change with different measures of SES I use all three available measures of SES in my analysis, and compare and contrast the findings.

Choice of cutoff line to use to group children into high and low achievement and SES groups

As we are examining extremes of test score and SES distributions, the choice of where we draw the cutoff line to define what the high and the low groups are important to consider. The number of children and the children's average scores in each group may vary substantially depending on the cutoff lines, particularly if there are outliers in the group.

While no article I have come across has justified their selection of cutoff lines, Feinstein (2015) acknowledges that questions on the 'arbitrariness' of these cutoff lines have been raised. Indeed, cutoff lines are arbitrary, and acknowledging this is important.

The general convention seems to separate children into quartile groups, as Blanden et al. (2012), Feinstein (2003), Goldstein and French (2015), Jerrim and Vignoles (2013), and Tu and Law (2010) do (at least with initial achievement, and with

SES variables where they are continuous). I therefore follow this convention in my approach, and test whether separating children into tercile or quintile groups affects findings. (I apply these to both the achievement grouping and the continuous SES measures, simultaneously.) I select the top and bottom quartile group to correspond to the high and low initial achievement or SES group.

Mother's education is a categorical variable. It is not possible therefore to identify the top and bottom quartile groups. Here I apply country specific cutoffs, as my aim is to identify low and high SES households in the sample. Ethiopia, India and Vietnam have the same cutoff lines, and Peru, the country with the highest levels of maternal education has a slightly different cutoff line, that I discussed earlier.

Attrition

Common in longitudinal surveys, is the fact that children with lower SES have higher attrition rates, and also tend to have low performance on achievement tests , as raised by both Schady et al. (2015) and Jerrim and Vignoles (2013). Different researchers use different approaches to address attrition. For UK data, Bradbury et al. (2015) impute test scores using interval regression. Jerrim and Vignoles (2013) conduct complete case analysis, that is they examine only children with test results across all rounds.

Attrition rates in the Young Lives dataset are relatively low for a longitudinal study (see Chapter 5, Data chapter for more details). Between Round 1, when the children were one year old, and Round 5, when the children were 15 years old, attrition in Ethiopia was 4.5 per cent, in India 3.0 per cent, in Peru 8.2 per cent and in Vietnam 2.3 per cent. But attrition may not be random. Table G.1 it breaks down by asset index quartile groups for Round 5 (when the children were 15 years old). It shows a relatively even distribution of attrition from the top and bottom quartile groups. I therefore conduct complete case analysis. The sample I use therefore includes children with vocabulary scores at five years old, and maths scores from five to 15 years old, as well as data on maternal education at one year old. For Ethiopia this results in an analytical sample of 1,457 children, India 1,714 children, Peru 1,675 children and Vietnam 1,601 children. As the Young Lives data is not nationally representative across countries, and I do not aim to generalise findings to the country's populations.

Classifying children at five years old

Classifying children at five years old is perhaps quite far along for a child's educational journey to be affected by parents' SES. Grantham-McGregor et al. (2007) explain that the first few years of life are particularly important because vital brain development occurs. These authors document studies from low- and middle-income countries that report an association between early exposure to poverty and malnutrition and lower achievement from six months to five years old, and older (most are cross-sectional studies).

By five years old, children may have already passed some critical stages of development. This is why Feinstein (2003) uses data collected at 22 months, to measure initial achievement. Doing so, however, comes at a cost of measurement error, as at 22 months the instruments used to measure achievement are less precise than those used with older children. Goldstein and French (2015) and Jerrim and Vignoles (2013) use data that allows them to plot trajectories in achievement from three years old. Blanden et al. (2012) compare the UK and Australia and plot trajectories from three years old in the UK, and from four to five years old for Australia.

However, it is not unusual that trajectories start around five years old, as Blanden et al. (2012) did in Australia. For example, Bradbury et al. (2015) and Washbrook and Lee (2015) use US data and track children from six to 14 years old. Bradbury et al. (2015) in their analysis of the UK data start at five years old (using data from three year old children to classify children into achievement groups), and in their analysis of Australian data start tracking children at seven years old (using the data from five year old children to group children into achievement groups).

In my interpretation of findings I acknowledge that parental SES may have shaped children's achievement at five years old. Even so, it is of value to look forward from five years old to 15 years old. Doing so allows us to observe what happens to children's trajectories during their school years, and examine whether achievement gaps at five years old are fixed over that period, and if they are or not, to try an better understand why this is the case, with the aim of informing policy aimed at reducing SES gaps in achievement. (In a subsequent study design, I might suggest taking an earlier measurement of achievement.)

To summarise, when I plot the achievement trajectories of children from five to 15 years old by SES I:

- use the Jerrim and Vignoles (2013) approach to address regression to the mean;
- use vocabulary scores to classify children into initially high and low achievement groups, and the maths scores to set the baseline score from which to measure progress;
- use the maternal education to measure SES;
- to classify high and low achieving children, I group children into quartile groups, and to group children into high and low SES I use a country specific cutoff line

for mother's education; and

- I conduct complete case analysis.

I also conduct sensitivity analysis to examine how the findings change with different choices.

8.4 Findings and discussion

Figure 8.2 produces the trajectories of initially high and low achieving children, by SES group across the our study countries. Solid lines represent initially high achieving children's trajectories, and dashed lines those initially low achieving children. Circle markers indicate high SES and triangles low SES. The trajectories start at five years old, when the children were first tested and finish at the most recent data point, when children are 15 years old. I use vocabulary tests administered at five years old, to group children into high and low achieving groups, and maths scores to track progress over time.

Those who report trajectories by SES, as I do in Figure 8.2, do not report confidence intervals in their analysis (see Blanden et al. (2012), Crawford et al. (2017), Feinstein (2003), and Jerrim and Vignoles (2013) as examples). These authors do not mention why they do not. It is only Bradbury et al. (2015) who do report confidence intervals, but unlike the aforementioned authors they use predicted scores (from a regression they apply to their data) in their trajectory, and plot the confidence intervals of these predicted scores. (An option for those who might want to produce confidence intervals is to use cluster bootstrapping.) We therefore cannot rely heavily on point estimates, rather we ought to exercise caution in our interpretation when drawing conclusions from the findings.

Here I conceive of children's achievement level neither as stable, innate or fixed, rather as areas that can evolve with changes in a child's circumstances, environment and physical development. The findings below show what happens to average test scores across childhood, for different groups of children in an interaction between an indicator of family SES and an average measure of achievement, starting at five years old.

The trajectories in Figure 8.2 look remarkably similar across countries, and similar to that of high-income countries (for example Bradbury et al. (2015), Crawford et al. (2017), Feinstein (2003), and Jerrim and Vignoles (2013)), where we observe a divergence of trajectories within the initially high (and low) achieving groups, by parental SES.

For ease of discussion, I refer to the initially high achieving-high SES group as the top group, as their trajectory remains on top from five to 15 years old for all four countries. The initially low achieving-low SES group I refer to as the bottom group, as that group also maintains its position with the lowest average scores from five to 15 years old in all four countries. The two groups in between I refer to as the middle groups, and when I refer to them individually I refer to them by their full description (e.g. the initially low achieving-high SES group). In Table 8.3 I report the size of each of there groups, by country.

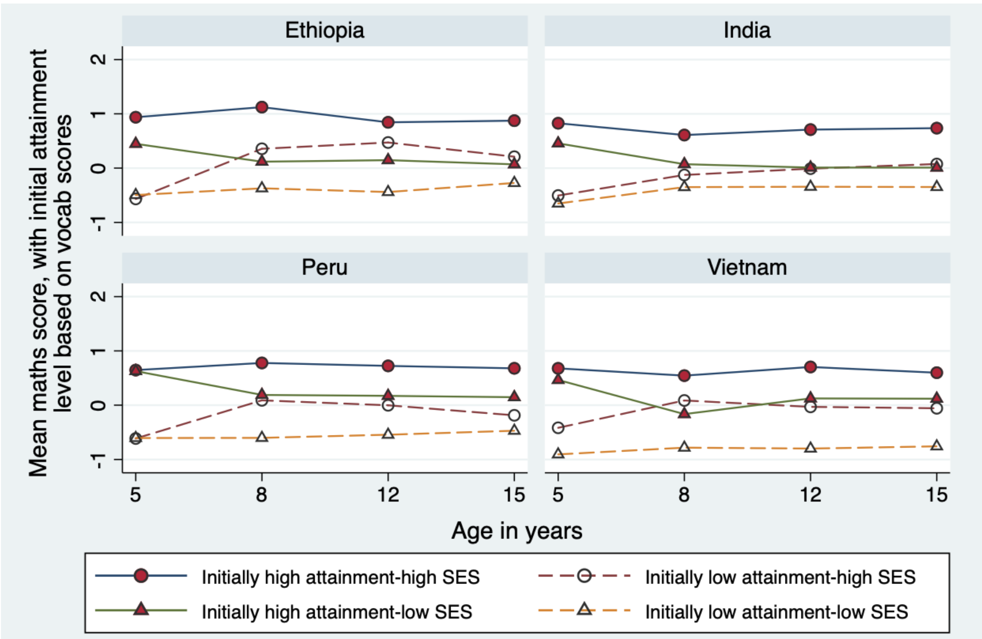


Figure 8.2: Trajectories in maths scores from five to 15 years old by initial achievement (defined using vocabulary scores at 5 years old) for the high and low SES children, based on mother's education

Table 8.3: Group sizes for Figure 8.2, using maternal education for SES

	Ethiopia	**India**	**Peru**	**Vietnam**
Initially high achievement - high SES	145	214	286	297
Initially low achievement - high SES	18	58	53	158
Initially high achievement - low SES	113	160	84	44
Initially low achievement - low SES	290	340	323	176

Notes: Initial achievement based on the top and bottom quartile groups vocabulary scores. SES based on maternal education. Ethiopia, India and Vietnam: low SES = none, high SES = lower secondary and above. Peru: low SES = none or primary, high SES = upper secondary and above.

Movement from five to eight years old

While the top and bottom groups maintain their position in the group ranking, there is a substantial amount of movement in the middle groups between five and eight years old. We see a convergence in scores for the initially low achieving-high SES children and the initially high achieving-low SES children. This convergence is most likely due to a combination of two factors specific to the five to eight year old age range.

First, from five to eight years old the Young Lives children are passing through an important transition: they were moving from ECE (if they were enrolled in such an institution) to primary school. The quality of the education they received at preschool and in their first year(s) of primary school may have had an influence on their scores at eight years old.

ECE in Ethiopia, Peru and Vietnam was stratified by socioeconomic status. The convergence we observe between five and eight years old in those countries may reflect the benefits of good-quality ECE on test scores at eight years old. High SES children, both initially low and high achieving, may have seen their achievement scores boosted from their exposure to good quality ECE, prior to entering primary school. Whereas the initially high achieving-low SES group did not have access to ECE (in Ethiopia) or poor quality preschool (in Peru and Vietnam) may have experienced a decline in their scores compared with the initially low achieving-high SES group. In India ECE at the time was generally of poor quality (see Context Chapter) giving us an indication as to why initially low achieving-high SES children may no have seen as large a boost in their scores, as we see in the other countries.

This initial advantage, combined with better quality primary schooling that high SES parents could purchase through private schooling or living in a neighbourhood with better schools, would offer a cumulative advantage in test scores to the high SES children. They would not only enter primary school better prepared, but in their first exposure to formal schooling would have received better tuition. All this in addition to the mechanisms described in teh Framework Chapter 2 summarised as the Investment Model the Family Stress Model and the Neurobiological Model.

Second, some of the convergence may be due to regression to the mean. Although I account for regression to the mean, by using the approach suggested by Jerrim and Vignoles (2013), there may be correlation in the error structures of the maths and vocabulary tests administered on the same day, and this may contribute to some regression to the mean between five and eight years old. As we do not know how much the error structures are correlated between tests, it is not possible to establish with certainty how much of the convergence is due to the effects of measurement error. We can only assume that the correlation is low, and that most of the regression to the mean has been accounted for.

From eight to fifteen years old

At eight years old the trajectories of the middle groups stabilise and remain relatively flat through to 15 years old. While there may be some regression to the mean (due to miss-classification of children into their initial achievement groups) between five and eight years old, by eight years old it should have washed out almost entirely. Regression to the mean would only persist beyond eight years old if there were a correlation in the error structures of the tests across rounds. This may be the case as there are a few items shared between maths tests for 8, 12 and 15 year olds. There is, however, no evidence of regression to the mean in these trajectories.

By the time the children are in primary school, their trajectories appear relatively fixed, particularly the top and bottom groups. Examining the composition of these groups may help to illuminate these findings. If, for example, the initially high achieving-high SES group (the top group) is comprised of children who are from advantaged groups of society, and the initially low achieving-low SES group is comprised predominantly of children from disadvantaged groups in society, we might surmise that broader structures in society may be supporting these divergent trajectories (both inside and outside of school).

One point to note is that initially low achieving-high SES children do not catch up with their initially high achieving counterpart. This is likely to reflect natural variation in achievement levels in our population, where some children are simply better are something than other children.

In Tables 8.4 to 8.7 I report the composition of the four initial achievement-SES groups by various social divisions. For example in India I report on the caste breakdown of each of these groups. I also report the composition of all the children in these four groups in the column titled ‘Total’. This column allows us to ascertain whether initial achievement-SES groups have a disproportionate representation of advantaged (or disadvantaged) groups.

Ethiopia is the only country where gender divisions are evident in the composition of the initial achievement-SES groups. The top group (initially high achieving-high SES) is comprised 63 per cent of boys and 37 per cent of girls. This indicates that disproportionately more boys make up this group.

For the initially low achieving-high SES group, 40 per cent are boys and 60 per cent are girls, indicating that has a disproportionately large number of girls in it. (Note, however, that this group comprises of 18 children.) We could say that the initially high SES group is split into the initially high achieving group, made up disproportionately of boys, and the initially low achieving group ,made up disproportionately of girls. Growing up as a boy in Ethiopia does offer several advantages over girls, and might explain why the top group has maintained higher scores, on average, than the other three groups (for more details see Context chapter).

Based on caste in India, region in Ethiopia and Peru, language in Peru and ethnicity in Vietnam, the top groups are consistently comprised, disproportionately of the advantaged groups, and the bottom groups, disproportionately of the disadvantaged groups.

In Ethiopia, children from the top group are disproportionately from the capital city Addis Ababa. While there are some children from Addis Ababa in the middle groups, there are disproportionately less of them, and only one per cent of the children in the lowest group lives in the capital city. In India, the high SES groups are disproportionately made up of children who are not in the most disadvantaged castes, while the bottom group is mostly comprised of children belonging to the disadvantaged castes. In Peru the high SES groups are made up entirely of Spanish speaking children, while Spanish speakers are under-represented in the low SES groups, and include native-language speakers. In Vietnam, the top and the middle groups are almost entirely made up of the majority ethnicity (Kinh), while the bottom group is disproportionately made up of minority groups who tend to live in remote areas that are under developed.

Table 8.4: Composition of achievement-SES groups in Ethiopia, for Figure 8.2 (main findings), column percentages

	Initially high achieving - high SES	**Initially low achieving - high SES**	**Initially high achieving - low SES**	**Initially low achieving - low SES**	**Total**
Gender					
Male	63	40	51	55	55
Female	37	60	49	45	45
Urban/rural					
Urban	92	63	31	13	38
Rural	8	37	69	87	62
Region					
Addis Ababa	61	10	11	1	17
Amhara	3	37	17	37	25
Oromiya	7	20	20	13	14
SNNP	26	20	18	15	18
Tigray	2	13	35	34	26

In Ethiopia, India and Peru, the high SES groups have a disproportionately higher composition of urban children. Urban areas in these countries have better infrastructure and provision to education (e.g. better public schools and also private

Table 8.5: Composition of achievement-SES groups in India, for Figure 8.2 (main findings), column percentages

	Initially high achieving - high SES	**Initially low achieving - high SES**	**Initially high achieving - low SES**	**Initially low achieving - low SES**	**Total**
Gender					
Male	54	58	57	53	55
Female	46	42	43	47	45
Caste					
SC	6	28	19	25	19
ST	11	3	17	20	16
BC	41	33	48	47	45
Other	41	36	16	7	20
Urban/rural					
Urban	57	42	19	9	26
Rural	43	58	81	91	74

**Notes*: SC-Scheduled caste, ST - Scheduled tribe, BC - Backwards caste, Other - none of the above

Table 8.6: Composition of achievement-SES groups in Peru, for Figure 8.2 (main findings), column percentages

	Initially high achieving - high SES	**Initially low achieving - high SES**	**Initially high achieving - low SES**	**Initially low achieving - low SES**	**Total**
Gender					
Male	48	50	45	54	50
Female	52	50	55	46	50
Language					
Spanish	100	100	62	74	84
Quechua	0	0	33	21	14
Spanish & Quechua	0	0	4	2	1
Nomatsiguenga (native)	0	0	1	3	1
Urban/rural					
Urban	95	88	53	47	69
Rural	5	12	47	53	31
Region					
Costa (Coast)	57	62	20	18	37
Selva (Tropical)	10	10	13	18	13
Sierra (Mountainous)	33	29	67	63	50

Table 8.7: Composition of achievement-SES groups in Vietnam, for Figure 8.2 (main findings), column percentages

	Initially high achieving-high SES	Initially low achieving-high SES	Initially high achieving-low SES	Initially low achieving-low SES	Total
Gender					
Male	49	47	49	55	50
Female	51	53	51	45	50
Ethnicity					
Kinh	98	96	93	38	80
H'Mong	0	0	2	30	9
Ede	0	0	0	1	<1
Bana	0	0	0	3	1
Nung	0	1	0	2	1
Tay	1	1	2	1	1
Dao	0	0	2	10	3
Other	1	3	2	16	6
Urban/rural					
Urban	40	8	13	3	19
Rural	60	92	87	97	81

schools). This might explain the improvement in scores of the initially low achieving-high SES children, and the maintenance at the top of the initially high achieving-high SES children. The low SES children, in all four countries, are predominantly located in rural areas, with worse infrastructure, poorer services and poorer education. (In Vietnam the only group that has an over representation of urban children is the initially high achieving-high SES group.) In all the other groups children from rural areas are over represented.

While the composition of the groups may help us to understand why children's trajectories are essentially flat from eight to 15 years old, they can also help us understand why the gaps emerge at five years old, as well as explain the movement we observe between five and eight years old. After all, if being in a high SES family in Peru for example also means living in a city (which tends to be better resourced) in a better resourced region (the coast) and being a Spanish speaker - then it is understandable that these children will have access to better services, ECE, primary education and more. Conversely, if having low SES means living in the countryside in a poorly resourced area (Selva and Sierra) and speaking an indigenous language as your mother tongue, then access to resources and obtaining support for one's education will be a challenge.

Does accounting for regression to the mean change the findings?

Here I consider whether these findings would differ had I not taken regression to the mean into account, and instead used maths scores both to classify children into their initial achievement group and track their progress. This is the approach Feinstein (2003) used in his paper. The findings from applying this approach are reported in Figure 8.3 and Table 8.8.

There are several points to note when comparing between the findings that take regression to the mean into account (see Figure 8.2 and for those that do not, see Figure 8.3). First, the initial difference between the high and low achieving group is wider when maths scores are used to group the children at five years old. This makes sense, as when I use vocabulary scores to group children into high and low achievement groups at five years old and calculate their average score using a different test, then some of variation from measurement error washes out. When I do not, the variation due to measurement error contribute to the wider difference in maths scores at five years old. As there is more measurement error in the initial achievement groupings there is more scope for regression to the mean to occur, and so it does.

Second, we observe more drastic convergence of trajectories from five to eight years old across all groups, particularly among the middle groups (as Jerrim and

Vignoles (2013) discuss would occur). However (this is the third point to note), trajectories from eight to fifteen years old remain mostly similar to those when regression to the mean is not taken into account, apart from perhaps Peru. In Peru, between five and eight years old, the initially low achieving-high SES group in this case over takes (rather than catches up to) the initially high achieving-low SES group. After eight years old the initially low achieving-high SES children's trajectory remains slightly above (rather than slightly below) that of the initially high achieving-low SES children. However, given I do not have confidence intervals here, I cannot say whether there is a statistically significant difference in the trajectories of these two different groups in Peru.

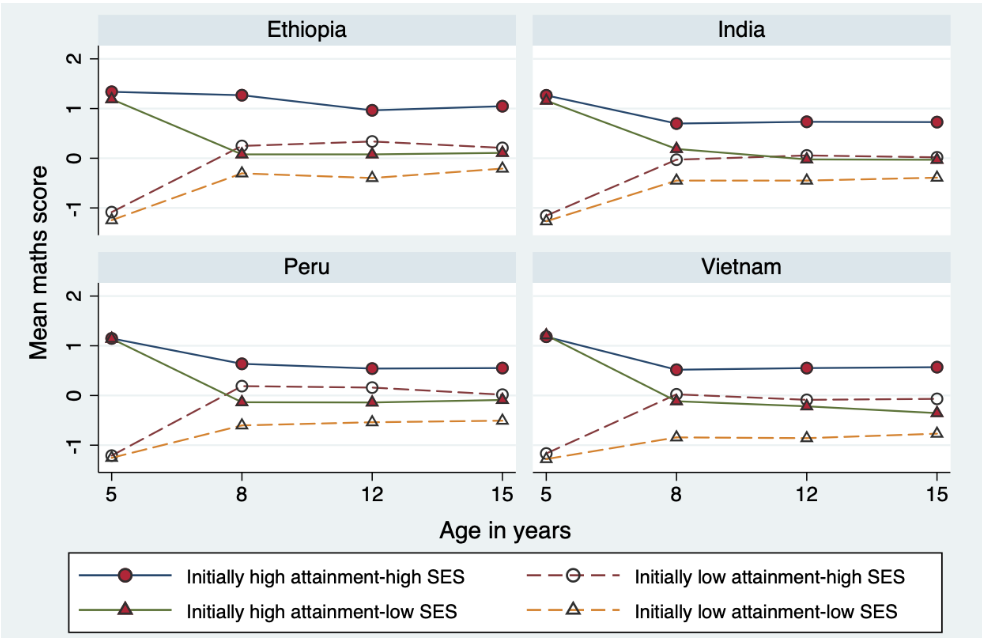


Figure 8.3: Feinstein (2003) approach: Trajectories in maths scores from five to 15 years old by initial achievement defined using maths scores at five years old
Notes: Initial achievement based on the top and bottom quartile groups in maths scores. SES based on maternal education. Ethiopia, India and Vietnam: low SES = none, high SES = lower secondary and above. Peru: low SES = none or primary, high SES = upper secondary and above.

When Jerrim and Vignoles (2013) scrutinise Feinstein (2003) using a different UK data set to that Feinstein (2003) used, the difference in their findings between those that take regression to the mean into account and those that do not are much starker (see their Figure 5). This may be because the UK context is distinct to that of the Young Lives countries, but the age range for the children in Jerrim and Vignoles' study is much narrower than that from Feinstein's and this study (the age range of children in Jerrim and Vignoles (2013) is from three to around around six and a half years old, while that in Feinstein (2003) is from 22 months to 10 years old, and here from five to 15 years old). I wonder how different findings would be

Table 8.8: Group sizes for Figure 8.3

	Ethiopia	**India**	**Peru**	**Vietnam**
Initially high achievement - high SES	124	186	224	252
Initially low achievement - high SES	30	69	104	152
Initially high achievement - low SES	132	180	135	55
Initially low achievement - low SES	269	333	265	183

Notes: Initial achievement grouping based on the top and bottom quartile groups in maths scores. SES based on mother's education. Ethiopia, India and Vietnam: low SES = none, high SES = lower secondary and above. Peru: low SES = none or primary, high SES = upper secondary and above.

with data from the children as they have grown older.

8.5 Sensitivity analysis

In Annex G.3 I report on the sensitivity analyses I conduct. Therein I consider several methodological choices one can make when examining trajectories of initially high (and low) achieving children by parental SES. These are: the choice of achievement variables used; the choice of SES variable used; and the cutoff line to group children into high and low initial achievement groups. I do this across the four countries. These sensitivity analyses do not produce any meaningfully different findings, although some are subject to regression to the mean to different extents.

When I switch the variable I use to classify children into initial achievement groups, and the one I use to plot their trajectories the findings are slightly different. In Figure G.3 I use maths scores at five years old to group children into achievement scores, and vocabulary scores to track children's progress. As I discuss, using vocabulary test scores to plot children's trajectories is problematic given the likelihood the error structures of the tests are correlated across rounds, as the tests are the same across rounds. The findings for India and Ethiopia, for example, may raise questions of regression to the mean across rounds.

Furthermore, the vocabulary and maths test scores, are not only different types of tests (the vocabulary test requires you to choose one picture of four that corresponds to a word said, and the maths test include complex word problems that are open ended), they also capture different aspects of achievement. I would therefore expect somewhat different findings. The main findings however, that the top and bottom groups maintains their positions from five to fifteen years old hold. We also see convergence of scores of the middle groups across all four countries. However, I would interpret these findings with caution given the likelihood of correlation in their error structures with age.

When I use the asset index and expenditure as an alternative measure of SES, the findings I produce are remarkably similar (see Figure G.4), albeit with slightly larger gaps. Recall that the categories for maternal education are not evenly distributed, for example about 60 per cent of Ethiopian and Indian mother's have no education (which corresponds to the low SES group). So the asset index and expenditure data capture a narrower and poorer band of children in the low SES grouping, resulting in wider gaps in most findings.

When I group children into quintile or tercile initial achievement groups, rather than quartile groups, the findings also remain similar (see Figure G.5).

8.6 Conclusions

To discern who is falling behind and when, it is useful to conceptualise SES gaps in achievement as having two components, those that emerge prior to school entry and those that emerge though schooling. Concerning gaps that emerge prior to school entry, a low SES child who enters school well prepared would not be disadvantaged in her subsequent educational career, relative to an equally prepared high SES child. If SES gaps emerge only through children's educational career, independent of school-entry achievement levels, then any advantage gained by an initially high achieving-low SES child starting school would erode over time relative to her initially high SES counterpart, and SES gaps in achievement would widen as children progressed through schooling.

If gaps in achievement emerged only prior to school entry, then eliminating SES gaps at school entry would suffice to ensure SES gap would not emerge into adolescence. Solely focusing on early years would be justified. If SES achievement gaps emerged only as children progressed through their educational career, then eliminating SES gaps prior to school entry would not address the problem. Low SES children would continue to systematically fall behind their high SES counterparts. Of course, in reality, SES gaps among adolescents reflect a combination of both gaps in initial achievement and subsequent disadvantage (or advantage).

In this chapter I have focused on the second component of these SES gaps: those that emerge through the schooling years. (In the previous chapter we identified SES gaps at five years old, establishing that yes, the first component of SES gaps does exist across all study countries).

While the persistent near-parallel gaps identified in Chapter 7 and in other studies (see Bradbury et al. (2015), Heckman and Mosso (2014), Magnuson et al. (2012), and Schady et al. (2015)) might lead some to conclude there is 'nothing happening' to SES gaps in achievement during the school years, evidence in this chapter has shown otherwise. In all study countries, low SES children have consistently fallen behind their high SES counter parts as they have progressed through school. Among the children who were classed as high achieving at five years old, the children from high SES families out perform their low SES counterparts through to 15 years old. The same is true among the children who were classed as low achieving at five years old, where those from low SES families lag behind.

Most disconcerting is that in all countries, the the initially low achieving-high SES group catches up to and perhaps overtakes the initially high achieving-low SES group. Children born into low SES families have not been able to learn the material they need in order to reach their potential. We do not only need policies to help narrow gaps at early ages (such as good quality targeted ECE), we also need policies that address the structural issues that hinder children's educations that exist outside the school (i.e. the urban rural divide in infrastructure, gender bias in Ethiopia, institutional discrimination based on ethnicity, caste or language).

It may seem counter intuitive that the SES gaps in achievement reported in Chapter 7 remain relatively steady, while the achievement trajectories reported here appear to be contributing to a widening of the SES gap in achievement (low SES groups' scores are falling relative to their high SES groups' average). This can be partly explained by regression to the mean also, though in this case when it comes to reporting SES gaps in achievement as they are reported in Chapter 7 (rather than the trajectories). As Bradbury et al. (2015) explain, if there is measurement error in the initial measure of achievement (where it is safe to assume there is) and we use this data to calculate SES gaps in achievement, then we would expect the SES gaps to narrow with age, as groups regress to their longitudinal the mean. However, we do not see a narrowing of the SES gaps in achievement, rather they appear to remain parallel. This can be explained by the divergent achievement trajectories by SES we report here.

In the introduction I discuss how work by Feinstein (2003) was robustly criticised by Jerrim and Vignoles (2013) for not taking regression to the mean into account, and that this triggered some policy changes and also a lively (sometimes tense) debate among UK policy and research circles. Amusingly, perhaps only for me, is that when I use the Young Lives data, I seem to get the same story if I use either the approach used by Feinstein (2003) or Jerrim and Vignoles (2013), the only difference is with the latter there is less regression to the mean (i.e. the slopes of the lines are a bit steeper).

Chapter 9

Early Childhood Education and its Association with Achievement from 5 to 15 years old

9.1 Introduction

The literature examining the relationship between early childhood education (ECE) and children's achievement is predominantly focused on high-income countries. While this research has influenced policy makers in low- and middle-income countries to invest in early childhood programmes the literature on the association between ECE and achievement in those contexts is limited. Based on this literature from high-income countries, we know that ECE is most beneficial to children from low socioeconomic status (SES) backgrounds, is effective if it provides high quality education and its direct effects on achievement seem to fade as children grow up. In the Chapter 1, the Introduction and Review of Literature, I review the research in detail. As more low- and middle-income countries consider implementing or widening ECE programmes, more evidence from similar countries would be beneficial for policy makers facing resource constraints in the design and implementation of ECE programmes.

In this chapter I apply to low- and middle-income country data a similar set of research questions oft applied to data from high-income countries. In this chapter I consider my third research question (RQ3); whether attendance in ECE is associated with relative gains in achievement across childhood, and examine what the estimates for each country reveal about different ECE systems?

Comparative studies on the association between ECE and children's achievement are limited across all countries. I have identified only four, namely Malmberg et al. (2011), Montie et al. (2006), Ojala and Talts (2007), and Rao et al. (2019), two of which focus solely on low- and middle-income countries. This chapter is the only comparative study among low- and middle-income countries that I have come across that looks at the association between ECE and achievement through to 15 years old, other studies have focused on short-term associations. As the Young Lives data includes a broad sample of children, this study also offers a wider view on the association between ECE and achievement than one of a specific programme evaluation or that of a specific city.

In the next sections I define ECE and briefly review the context of each study country's ECE system. This is followed by a discussion on the methodology and measures I use, after which I report and discuss my findings on the association between ECE and achievement, before concluding.

9.2 Early childhood education: a definition

ECE is the education a child receives prior to primary school entry, as is described in the UNICEF website on ECE (UNICEF, 2023). In Anglo-saxon countries it is often referred to as *preschool education*, but international organisations tend to use the term *early childhood education*.

The Organisation for Economic Cooperation and Development (OECD) tends to write about ECE in two categories: that which is provided to children aged 3 to 5 years old (as most children enter primary school at age 6), and that which is provided to children before three years old (see OECD, 2021). In this chapter I focus on the education children receive prior to entering formal schooling from the age of 3 years until primary school entry. In Ethiopia, children start primary school at 7 years old, and in the other three countries it is 6 years old.

9.3 Country contexts: Early childhood education

Around 2006, when the Young Lives children were attending ECE, the ECE context differed markedly between the study countries. In the Country Context Chapter 5 I discuss in detail the ECE context of each Young Lives study country. Here I briefly summarise it.

Ethiopia

In Ethiopia, there was no public ECE provision, so the few children attended ECE. Those who did attended private ECEs and resided mostly in urban areas (UNESCO, 2008). These ECE institution offered, in general, good quality education.

India

In India, ECE provision in both the public and private sector, was poor quality. Public *Anganwadi* centres accounted for the majority of ECE provision in India. While the intention of the *Anganwadi* centres was to provide ECE in reality it seems that for the most part they provided childcare to the children, as evidenced in Citizens' Initiative for the Rights of Children Under Six (CIRCUS) (2006) and Integrated Child Development Services (ICDS) (2007).

Private ECE provision was also available in India. Qualitative research by Woodhead et al. (2009) found that private ECEs in poor areas were generally poor quality, a finding supported by UNESCO (2010), SCERT (2011) and Kaul and Sankar (2009).

Peru

Peru had public provision that was stratified by SES, along with some private ECE institutions. There were two types of ECE institutions for children aged 3 to 5. *Centros de Educación Inicial* (CEIs) provided formal education and were public or private, and catered mostly to urban populations and provided generally good quality education. Public CEIs received support from the state. *Programas no Escolarizados de Educación Inicial* (PRONOEIs) were community based entities that offered a lower-cost and a more flexible approach to ECE, and were usually found in rural and poor urban areas. They received limited public resources and little educational material, and on average provided lower quality education.

Vietnam

The ECE options in Vietnam included state-owned, community-owned and privately-owned institutions. Based on UNESCO (2006), when the Young Lives children were attending ECE there was an age-appropriate play-based kindergarten curriculum in place. All forms of ECE relied on the community for financial support to cover a significant portion, if not all the ECE's costs.The quality of the ECEs therefore varied by the SES of the community.

Public ECE institutions were widely distributed except for in rural and remote areas. Community ECE institutions were usually found in hard to reach or poor rural areas – and accounted for only a small percentage of ECE provision. Private ECE institutions were often found in urban and advantaged areas.

9.4 Data & Methods

I draw on the Young Lives dataset, a multi-country longitudinal birth-cohort dataset that includes information on ECE attendance and achievement data across childhood. (Recall, Young Lives follows approximately 2,000 children in each of four low- and middle-income countries from 6 months old to 15 years old, and followed up at 5, 8 and 12 years old. The countries include Ethiopia, India, Peru and Vietnam.) There are no major policy changes that happened across these countries that allowed for a natural experiment. I discuss the association between attending an ECE programme and maths scores through to 15 years old.

Investigating the link between ECE and achievement scores with longitudinal data has its advantages. First, we can investigate the association over time for the *same* group of children. Second, ECE attendance is measured in real time, removing any measurement error due to recall issues. Third, the association between ECE and achievement scores can be examined in the ECE context in each country.

A challenge in examining the association between ECE and achievement is that attendance and the quality of ECE institution tend to correlate with the family's resources and level of education, as Gambaro et al. (2014) show with UK data. Wealthier parents are more likely to be able to afford high quality ECE either by paying for private education or by living in a neighbourhood where the quality of public institutions is higher.

To account for factors that may confound the relationship between receiving ECE and achievement, I fit an OLS regression which controls for child- and household-level variables associated with achievement, while also accounting for variation between communities.

The regression equation I use for each Young Lives child at age a is as follows:

$$y_a = \beta_0 + \beta_1 P_5 + \beta_2 \mathbf{Z}_1 + \beta_3 L_a + \gamma + \epsilon_a \tag{9.1}$$

where

y = achievement score at age a (5, 8, 12 or 15 years old)

P = receiving early childhood education at 5 years old

$\mathbf{Z}$ = vector of initial conditions; includes child- and household-level characteristics

L = language of the maths test at age a

γ = community-level fixed effects at 1 year old

ϵ = error term.

Because each Young Lives household includes exactly one study child of a specific age, for simplicity I omit the child (i) and household (j) subscripts.

The vector Z represents initial conditions as these may confound the association between participation in ECE and achievement scores. For example, a child from a wealthier family, a firstborn child or a child from a particular caste or ethnicity may be more likely to attend ECE than one from another group, I try to account for these differences by controlling for initial conditions (as other studies do, as discussed in Engle et al., 2011).

Child-level characteristics include height-for-age (at 1 year old), gender, being a firstborn child, the child's mother tongue in Ethiopia and Peru, ethnicity in Vietnam and caste in India (at 1 year old).

Household-level characteristics include three measures of initial SES. Each measures captures different aspects of SES. These three measures include mother's level of education, the household's asset index (both at age 1 year old) and household expenditure (at 5 years old). The asset index and expenditure are both continuous variables, and I group maternal education according to the International Standard Classification of Education (ISCED) categories.

I also account for household size (at 1 year old, continuous) and maternal employment (at 5 years old when the children were at the age to receive ECE) and, a test-level characteristic is the language of the test (within each round). In the Framework Chapter 6 I explain why I select these variables as covariates.

To ensure I compare the same group of children, I use a balanced panel, including children with all pertinent variables across the five rounds. This results in a sample size of 1,303 in Ethiopia, 1,740 in India, 1,649 in Peru and 1,673 in Vietnam for maths scores.

Equation 9.1 is an age specific regression, where the outcome variable changes with age (as does the language of test, where relevant). This is not a panel regression since I do not pool the data. I run a separate regression at each age children are tested (5, 8, 12 and 15 years old) and for each country. This amounts to 16 regressions (four countries and four ages). As I run these separately by country I cannot compare in a statistical manner the magnitude of the coefficients across countries, rather I make descriptive comparisons drawing on contextual information.

The estimate I am interested in is the β_1 coefficient on the ECE variable. It represents a partial correlation between receiving ECE and achievement scores, accounting for child- and household-level factors, and time-invariant community-factors.

Research based on high-income countries shows that this association fades through childhood. Using this specification, I will be able to identify a 'fading' if the size and strength of the β_1 coefficient decreases over the course of childhood.

An advantage of using the Young Lives data is that I am examining the association between ECE and achievement in a policy relevant context that reflects the reality of thousands of children. I cannot however account for all the variation in maths scores, some of which will be due to influences we cannot measure, therefore the error term will be positively correlated with ECE in my analysis and my estimates are likely biased upwards.

While previous studies have found that attending a higher quality ECE institution produces better results (for example Singh and Mukherjee, 2018 in India and Cueto et al., 2016 in Peru), I do not examine whether this is the case here. I believe the selection effect into high-quality ECE will be quite strong exposing this analysis to more bias. Also, my principle aim is to report on the ECE system as a whole in each country and compare and contrast which between countries.

Implications of using fixed effects

As discussed in the context section, there are differences in the ECE provision across regions and urban-rural areas. These differences correlate with other factors that influence achievement scores, such as poor access to nutrition and poor infrastructure (such as road, electricity, healthcare and sanitation). If I do not account for differences between regions or urban-rural areas the coefficient on ECE (β_1) would reflect regional and urban-rural differences and difference in scores depending on ECE attendance. If ECE provision and broader infrastructure differs substantially between rural and urban areas and within regions also, as they do in all study countries, and if I do not take into account these differences, then I would not be able to discern whether any boost in achievement scores is related to ECE or living in an area with better infrastructure.

To account for these differences I use community-level fixed effects. My estimates therefore present average differences in outcomes *within* communities not between them. As I use community-level fixed effects, variations outside the community level, such as urban–rural and regional differences, are accounted for. The communities I use are the initial 20 sites from which approximately 100 children were sampled (as discussed in the Data section above). A description of these different communities (or sites) can be found in Annex H.1.

There are some limitations to using fixed effects. In sites where there is no variation, I am not able to ascertain whether attendance in ECE is associated with an advantage over not attending. In Ethiopia, there are 5 sites of the 20 where no child attended ECE (sites 17, 18 and 20 in Tigray and sites 6 and 7 in Amhara) and two more with a one per cent attendance rate only (site 5 in Amhara and site 9 in Oromia). This means that most of my analysis is based on variation within the remaining 13 sites. In Peru there are two sites where attendance is 100 per cent (Sites 1 and 7 in Tumbes and Cajamarca respectively). In Vietnam there are four sites with 100 per cent attendance rates (sites 13, 14 and 16 in the Red River Delta and site 10 in Lao Cai). As I use fixed effects at the site level, I am not able to ascertain whether attendance in ECE offers an advantage over not attending (or conversely) in these sites.

By using fixed effects I am also removing some exogenous variation that can be valuable for policymaking, such as the variation in the nature or quality of ECE provision. More specifically, the analysis will not show that scores are potentially different in communities with different types of ECE provision. Though, as the variation in ECE provision correlates with factors that also shape achievement scores that differ across communities, I believe the use of fixed effects remains justified.

Annex H.1 reports the percentage of children attending ECE by sites. In Ethiopia the range of attendance is quite wide, ranging from 0 to 99 per cent. In the other three countries, average attendance rates are much higher, ranging from 66 to 99 per cent in India, 58 to 100 per cent in Peru and 79 to 100 per cent in Vietnam.

In Ethiopia, the highest rates of ECE attendance are in the capital city, Addis Ababa (reporting over 90 per cent attendance).

In Peru there is also a division in attendance between urban and rural areas. The four sites with the lowest attendance rates in ECE are rural, and the highest four are urban (two of which are in the captial city, Lima).

Unlike Ethiopia and Peru, India and Vietnam do not reflect a clear divergence in attendance rates between rural and urban areas. In India the two sites with the highest and the lowest attendance rates are in rural areas (lowest are in Kadapa and Srikakulam, and the highest two are also in Srikakulam). In Vietnam, ECE attendance rates do not seem to overlap with rural and urban differences or socioeconomic ones. Among the four sites with the lowest rates of attendance, three are in the Mekong River delta (which is not among the poorest regions in Vietnam).

9.5 Measures

Outcome: achievement

The Young Lives data offer two sets of achievement measures derived from tests administered at age 5 to 15 years old: maths and vocabulary tests.

I rely on the the maths test as it tests a wider range of academic skills compared with receptive vocabulary test scores. Maths tests demand not only comprehension, but also the identification and completion of mathematical expressions. Duncan et al. (2007) conceive of maths skills as involving both conceptual and procedural competencies. They find that maths skills offer a more powerful measure of achievement.

The maths tests are age-specific. Early tests involve multiple choice questions and later ones included a mixture of simple maths operations, multiple choice questions and complex word problems. There is some overlap in the questions across later rounds, though this represents only a small proportion of the questions.

I calculate age-specific z-scores which have a mean of zero and a standard deviation of one.

Early childhood education

When the children were 5 years old a parent or a carer was asked a set of ECE related questions, one of which was whether the child had attended ECE since the age of three years. The answers from some additional questions are provided in Table 9.1. The ECE variable I use is a binary 1 equalling one if the child attended ECE and 0 if not.

I am limited by the ECE data on two counts. First, evidence from high-, middle- and low-income countries shows that ECE benefits disadvantaged children the most and when so I attempted an interaction between the parent's SES and ECE attendance, but cell sizes were quite small or empty. For example, the number of children who have mothers with tertiary education who did not attend ECE ranges from zero in Vietnam to five in Peru. This affects the reliability of the estimates and complicates their interpretation as the situation of a particular child could skew the estimates. Second, the literature shows that for ECE to lead to gains in achievement it has to be of high quality. However, data collected by Young Lives at 5 years old do not differentiate in Peru between children who attended CEIs and PRONOEIs, and in Vietnam between the community and public ECE institutions. Instead they are grouped together into a single 'public' category. In subsequent rounds, parents are asked again whether their child attended ECE and if so which type of ECE (for example, CEI vs PRONOEI). However, given a relatively long recall period the number of children who are reported to have attended ECE differs from the data collected when the children were 5 years old as does the type of ECE they attended. Therefore, I use data collected with the shortest recall period, when the children were 5 years old.

Covariates

In the Framework Chapter 2 I discuss in detail the rationale for selecting each of these covariates, and in the Measures Chapter 6 I discuss how these are measured by Young Lives.

Attending ECE is likely to be shaped by parental SES, particularly in Ethiopia, where most ECE is private. Here I use maternal education, expenditure and the asset index as measures of parental SES. I include them separately in the regression as each represents a distinct aspect of SES.

Height-for-age is a reflection of physical development, and is linked inextricably to achievement if one does not receive sufficient nutrition as a child. There are important differences in performance across genders, as discussed in Balart and Oosterveen (2019). Being firstborn may offer an advantage for two reasons: parents have more disposable income and first-born preferences in society. Alternatively, older children may be expected to engage in domestic and caring duties to the detriment of the education.

In order to take into account that particular populations may be discriminated against I control for a variable that represents a main dividing line in each society. In Peru and Ethiopia this is mother tongue, in Vietnam this is ethnicity and in India this is caste.

Household size may be negatively or positively associated with achievement, larger families may mean more caring or work responsibilities for older children. It may also be associated with more income earners. In high-income countries, larger households are generally associated with lower achievement.

One motivation to enrol children in ECE is to enable their mother to work. Maternal employment may benefit the achievement scores of a child through indirect channels, for example increased income and/or increased decision-making power of the mother. It may also disadvantage the child as they may have to take on additional responsibilities at home, impacting on their achievement scores.

In some countries at some ages, children are given the choice of languages in which to take the test. The difficulty of a test may vary depending on the language of maths test. (The test is administered in one language in India at 12 and 15 years old and in Vietnam at 8, 12 and 15 years old.)

9.6 Estimates and Discussion

Descriptive statistics

Table 9.1 provides information on how many children in the sample ever attended ECE, what type it was, how many hours per week they attended and at what age they started. While in Ethiopia only about a quarter ever attended ECE, in the other three countries well over 80 per cent did, with Vietnam reaching 92 per cent on average. Consistent with the broader context in Ethiopia, most children attended private ECE institutions. In India, almost two thirds attended *Angawadis*, and one third private institutions. In Peru and Vietnam, the large majority attended public ECE institutions, though Young Lives data don't distinguish which type of public institution. Across the four countries, children spent between 21 (in Peru) and 34 (in Ethiopia) hours on average in ECE. They started at around 3.3 years (in India) to 3.9 years old in Ethiopia (where primary school starts at 7 years old).

Table 9.1: Information on early childhood education, by country

	Ethiopia	India	Peru	Vietnam
ECE attendance (%)				
Yes	26	88	86	92
No	74	12	14	8
Type of ECE (%)				
Private	76	35	14	10
Public	7	64	85	87
Other	18	0	1	3
Hours per week in ECE (avg.)				
	34	29	21	32
Age starting ECE (avg. age in years)				
	3.9	3.3	3.5	3.6
Primary school starting age (years)				
	7	6	6	6
Sample size				
	1303	1740	1649	1673

Table 9.2 reports the differences in the various control variables I use in my analysis, by ECE. Children who did not attend ECE tended to have lower SES, with the biggest difference observed in Ethiopia, where no meaningful public ECE provision existed, so only those families who could afford private instruction enrolled their children.

All children, irrespective of attendance were below the UN established average in their height-for-age scores, a fact related to the pro-poor sampling method of Young Lives. Yet, children who did not attend ECE had a more pronounced difference in their their height-for-age scores, except for Vietnam where the difference, although still there, was less pronounced.

In Ethiopia, India and Vietnam the proportion of girls who ever attended ECE is almost the same as that for boys, suggesting there is no clear boy preference in ECE attendance. Peru there is a slight gender gap in favour of boys.

In all countries firstborn children had higher rates of attending ECE than non-firstborn children. Ethiopia had the largest difference in enrolment between firstborn and non-firstborn children, whose rates were 42 and 21 per cent respectively. In all countries, children who did not attend ECE had slightly larger households than those who did.

Table 9.3 shows the average maths score for each group according to whether children ever attended ECE or not at 5 years old. Note that there are a few variables that were included in the regression model as continuous variable but that I have grouped for the purposes of this table (specifically maternal education, the asset index, expenditure, height-for-age and household size).

As mother's education increases so do the children's average scores. Children in the top quartile of the asset index and expenditure also score on average higher than children in the bottom quartile of each measure. The same is the case with height-for-age, where children who were stunted at 1 year old consistently perform worse than their non-stunted counterparts. There is no clear pattern across the countries when it comes to gender differences, being a firstborn child, what the mother's work entailed and the child's household size at one year old.

Estimates

Table 9.4 has the coefficient estimates for ECE (i.e. β_1) by country and by age. These are summarised in Figure 9.1. Full regression estimates of all 16 regressions are available in Annex H.2.

At 5 years old the size of the coefficients for Ethiopia is 0.26 s.d., India 0.12 s.d., Peru 0.19 s.d. and Vietnam 0.20 s.d.. These are in line, or slightly lower than those reported in earlier studies. The size of the association reported in the review article by Engle et al. (2011) is 0.24 s.d., Berlinski et al. (2009) report an effect size of 0.23 s.d., Rao et al. (2019) report an association of about a third of a standard deviation, as do Nores and Barnett (2010). These are described as 'modest' or 'moderate' in size by Engle et al. (2011) and Rao et al. (2019) and others. Compared with the range of –0.14 to 1.68 s.d. (with the mean of 0.24 s.d.) reported in Engle et al. (2011) the ranges reported in Table 9.4 are relatively narrow.

In Ethiopia a 0.26 s.d. is equivalent approximately a 5 percentage point advantage for those who attended ECE (an increase from 61 to 66 per cent correct in the test). In India the equivalent figure for 0.12 s.d. is an advantage of 1.5 percentage points, for Peru (0.19 s.d.) and Vietnam (0.20 s.d.) this is equivalent to a an advantage of about 3 percentage points. (These figures are based on the age specific z-cores for children aged 66 months.)

At 5 years old, in all countries except for India, children who attend ECE have on average a statistically significant higher maths score, than children who did not

Table 9.2: Descriptives of covariates: by country and ECE attendance at five years old

	Ethiopia		India		Peru		Vietnam	
	No	Yes	No	Yes	No	Yes	No	Yes
Mother's education (%)								
None	92	8	14	86	26	74	18	82
Primary and lower secondary	56	44	11	89	20	80	7	93
Upper secondary and above	15	85	8	92	4	96	4	96
Asset index (avg)								
	-0.3	1.0	-0.1	0.0	-0.5	0.1	-0.6	0.1
Expenditure (avg.)								
	98	176	700	831	129	202	267	411
Height for age (z-score)								
	-1.6	-1.2	-1.7	-1.3	-1.6	-1.2	-1.0	-1.1
Child's sex (%)								
Male	73	27	12	88	13	87	8	92
Female	74	26	13	87	16	84	8	92
Firstborn (%)								
No	79	21	13	87	18	82	10	90
Yes	58	42	11	89	9	91	5	95
Mother's work (%)								
None	81	19	11	89	16	84	9	91
Agriculture	99	1	14	86	29	71	8	92
Non-agriculture	44	56	13	87	8	92	7	93
Household size (avg.))								
	6.0	5.5	5.2	5.1	5.8	5.3	4.8	4.6
Sample size (n)								
	958	345	210	1530	236	1413	127	1546

Notes: Where percentages are reported, these are row percentages by country. Within each group they report what percentage of children did and did not attend ECE. Covariates not included in this table are ethnic group and the language of the test, whose categories vary from country to country.

Table 9.3: Descriptives of covariates: average maths score per group at 5 years old (age specific z-scores)

	Ethiopia		**India**		**Peru**		**Vietnam**	
	No	Yes	No	Yes	No	Yes	No	Yes
Mother's education								
None	-0.22	0.23	-0.31	-0.21	-0.56	-0.09	-1.06	-0.54
Primary and lower secondary	-0.03	0.62	-0.04	0.20	-0.22	-0.06	-0.25	0.10
Upper secondary and above	0.29	0.78	0.48	0.49	-0.10	0.32	0.39	0.46
Asset index								
Top quartile	0.22	0.70	0.38	0.41	0.15	0.39	-0.02	0.41
Bottom quartile	-0.21	0.19	-0.42	-0.18	-0.39	-0.15	-0.62	-0.19
Expenditure								
Top quartile	-0.16	0.58	0.29	0.26	0.13	0.39	0.30	0.39
Bottom quartile	-0.19	0.39	-0.40	-0.17	-0.40	-0.18	-0.71	-0.20
Height for age								
Not stunted (>2 s.d.)	-0.11	0.63	0.01	0.15	-0.22	0.17	-0.27	0.16
Stunted (<=2 s.d.)	-0.22	0.48	-0.35	-0.15	-0.37	-0.09	-0.78	-0.18
Child's sex								
Male	-0.18	0.57	-0.07	0.06	-0.29	0.06	-0.52	0.09
Female	-0.13	0.60	-0.19	0.06	-0.26	0.14	-0.27	0.09
Firstborn								
No	-0.15	0.63	-0.29	0.00	-0.24	0.09	-0.36	0.06
Yes	-0.17	0.51	0.02	0.11	-0.35	0.12	-0.51	0.13
Mother's work								
None	-0.19	0.61	-0.04	0.11	-0.37	0.01	0.15	0.15
Agriculture	-0.07	-1.48	-0.26	-0.04	-0.16	-0.16	-0.84	-0.10
Non-agriculture	-0.07	0.57	-0.08	0.12	0.01	0.29	-0.22	0.23
Household size								
Four or fewer	-0.15	0.56	0.05	0.08	-0.21	0.12	-0.44	0.10
More than four	-0.16	0.60	-0.24	0.05	-0.30	0.09	-0.38	0.08
Sample size (n)								
	958	345	210	1530	236	1413	127	1546

Notes: Average maths scores per group, by ECE attendance. Groupings are based on data collected when children are one year old, except for expenditure which is collected when children are 5 years old.

attend ECE. In India ECE is not associated with comparatively higher maths scores. Singh and Mukherjee (2018) also finds this among 12 year old children in India, when they group all types of ECE provision into one variable. In all likelihood, this is due to the lower quality of ECE institutions at the time. Indeed, a draft report states that in Andhra Pradesh the "quality of pre-primary education in the ECCE Centres is not satisfactory" (SCERT, 2011 p. 15). In all other countries, there is a significant portion of ECE provision that is considered high-quality and in Ethiopia almost all is high-quality.

To assess whether this association persists though childhood the magnitude and the statistical strength of the estimate (β_1) would have to persist. However as we see in Table 9.4 and Figure 9.1 both fall as childhood progresses.

In Ethiopia and Peru the association remains through to 8 years old. In Peru the magnitude of the coefficient estimate remains at 0.19 s.d., and the statistical significance remains (albeit slightly weaker). In Ethiopia the association increases from 5 to 8 years old (0.26 to 0.30 s.d.). This coincides with Ethiopia's schooling system, as primary school starts at 7 years old and the benefits of ECE on achievement are likely to be the greatest.

Heckman (2008) and colleagues discuss the concept of dynamic complimentary, where skills beget skills over the course of childhood and therefore benefits of ECE should accrue over time. However, in all study countries the association between ECE and maths skills fades away by 15 years old. By then the advantage in maths scores is not only small (ranging from 0.01 s.d. in Peru and Vietnam to 0.07 in Ethiopia) but also statistically insignificant. It appears the children who did not attend ECE catch up with those who did and eventually reach the same level. This is consistent with findings reported in high-income countries, such as studies cited in Durkin et al. (2022).

Figure 9.1: Predicted gap between children who did and did not attend ECE, by country (with 95 per cent confidence interval)

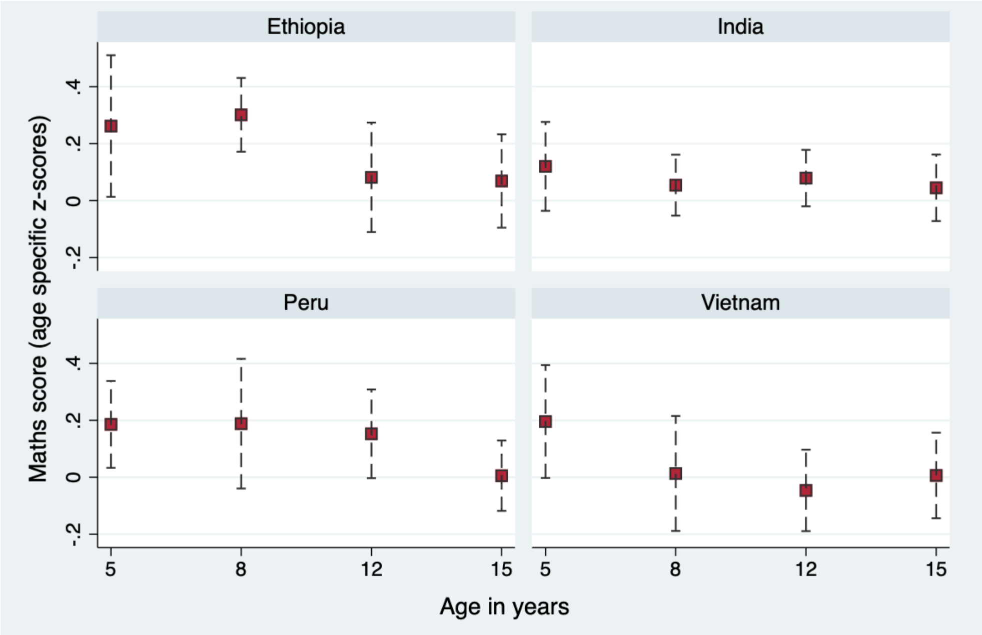


Table 9.4: Regression coefficient estimates for early childhood education, by country and age

	Ethiopia	**India**	**Peru**	**Vietnam**
Age in years				
5	0.26**	0.12	0.19**	0.20*
	(0.01 - 0.51)	(-0.04 - 0.28)	(0.03 - 0.34)	(-0.00 - 0.39)
8	0.30***	0.05	0.19*	0.01
	(0.17 - 0.43)	(-0.05 - 0.16)	(-0.04 - 0.42)	(-0.19 - 0.22)
12	0.08	0.08	0.15*	-0.05
	(-0.11 - 0.27)	(-0.02 - 0.18)	(-0.00 - 0.31)	(-0.19 - 0.10)
15	0.07	0.05	0.01	0.01
	(-0.10 - 0.23)	(-0.07 - 0.16)	(-0.12 - 0.13)	(-0.14 - 0.16)

Outcome: Maths scores (age specific z-scores)

Note: Confidence intervals in parentheses

*** p<0.01, ** p<0.05, * p<0.1

As children grow up, the influence of ECE may lessen compared with all other inputs children receive. Children require a constant flow of inputs to thrive, and these change at different stages in their lives. Given all other inputs along the way, the role of ECE (positive or negative) may become less important over time.

In Peru, the persistence of the association between ECE and maths scores is interesting. It remains constant from 5 to 8 (at 0.19 s.d.), falls slightly at 12 years old (to 0.15 s.d.) and is still statistically significant. In their analysis of 5 year olds, Cueto et al. (2016) also use Young Lives data (but use a two-stage least-squares regression approach and use ECE data with a long recall period) and find there is only an association between ECE and maths scores when the ECE attended is an ECE that offers higher quality education that caters to wealthier and urban children (a CEI). If the child attends a type of ECE that offers lower quality education and caters to rural and poorer children (PRONOEIs) the authors find no association. Perhaps the benefits of attending CEI specifically has brought up the average of those attending ECE education in Peru.

In Vietnam, at 5 years old those who attend ECE have an advantage over those who do not, specifically 0.20 s.d.. Rogers et al. (2019) examine the association between ECE and maths scores using Young Lives data for Vietnam, and their estimates from a similar OLS regression to mine and they also do not find a statistically significant association at 12 years old.

Durkin et al. (2022) and Kottelenberg and Lehrer (2019) use data from high-income countries and find that ECE sometimes produce negative effects on achievement scores. One reason put forward by the authors is that parents changed their behaviour and provided children with less 'academic' support at home. ECE could not make up for the one-to-one attention and support at home, so the children's achievement scores dropped compared with when they were not attending ECE. Young Lives does not have information on parental behaviour at home to test this hypothesis.

When I compare the association of ECE with that of other variables in the regression (see Annex H.2 for full regression estimates) it is clear that initial parental SES remains a strong predictor for maths scores throughout childhood (specifically maternal educaiton and an asset index). In Ethiopia at 15 years old, for example, the asset index (measured at 1 year old) has a coefficient estimate of 0.26 s.d., the same magnitude as the association between ECE attendance and maths scores at 5 years old.

The main points from these findings are that:

- where low quality ECE was widespread, as it was in India, children who attended ECE do not gain a significant advantage over those who did not attend;
- the association between ECE and maths scores is moderately sized at 5 years old in Ethiopia, Peru and Vietnam, but the magnitude and the statistical significance declines through childhood. By the age of 15 years, the magnitude of the association is close to zero and there is no statistically significant difference in maths scores between those who attended and did not attend ECE;
- initial parental SES is a stronger predictor of maths scores than is ECE over the course of childhood;
- there is no evidence of a negative association between attending ECE and maths scores.

9.7 Conclusion

My findings are not consistent with the Heckman curve. While I find attending ECE is associated with an advantage in maths scores, compared with not attending, I find no evidence this is maintained as children reach adolescence. Investments in ECE do not seem to produce observable long-term effects.

Ethiopia's private ECE system makes it an unlikely candidate for scaling up, as most families will not be able to afford private tuition or school fees. Nonetheless, there are clear policy lessons to be learned about the quality of ECE services. Based on my analysis, the ECE system in India would benefit from improvements across the board in the quality of ECE. Both Peru and Vietnam manage to deliver a meaningful proportion of good quality ECE, which is associated with a longer lasting advantage in maths scores for those who attended ECE. Though both countries have an ECE system that is stratified by socioeconomic factors. Whether such stratification perpetuates socioeconomic disparities, merits further research.

The socioeconomic circumstances in which a child grows up are the strongest predictor of outcomes later in life. ECE firmly belongs to one's socioeconomic context. More resources will enable wider coverage and/or better quality of ECE services, which in turn will lead to better maths scores later on and better primary and secondary education in general. Children with more educated and well-off parents are more likely to have a home environment that supports learning, with access to better nutrition, fewer caring responsibilities and a more stable financial situation. I control for SES and other child- and household-level factors that shape achievement and use community fixed effects to examine variations within small, relatively homogeneous communities and geographic areas. My findings confirm that parental SES is key in predicting the child's maths scores, well beyond the association with ECE.

An important policy question is whether investments in ECE should be prioritised. If the benefits diminish over time, even if the services are of relatively high quality, then other investments that yield returns throughout a child's lifetime may be preferable to those that impact simply in the early years. Hence, policy makers may consider more carefully whether to invest scarce resources in ECE. Two observations are relevant here.

First, research on high-income countries confirms that the highest benefits of ECE are obtained by the more disadvantaged children. Further research is needed for low- and middle-income countries. An argument could be made in favour of targeting ECE so as to yield the highest possible returns on ECE investments. Any targeting would need to be tailored to the characteristics of each country's or community's context. In any case, further improvements in the quality of the services rendered by PRONOEIs in Peru and the community schools in Vietnam would likely considerably improve the association between ECE and maths scores.

Second, the fact that returns on ECE investments may not last beyond the early years may be related to the many other inputs that become important in the process of learning. It is plausible, however, that ECE will have indirect effects on these inputs. Children who attend ECE may, for example, develop a habit of going to school every day, they may be less likely to drop out later, or they may acquire a good study routine and work ethic that are conducive to learning. These possible effects of ECE are more intangible and harder to observe, and I am not able to identify them in this analysis. However, they merit further investigation. Though my findings challenge the Heckman curve when it comes to achievement scores in four low- and middle-income countries, it does not mean that the curve ought to be dismissed out of hand.

In terms of further research, there is evidence that duration in ECE can matter.Young Lives measures how long children attended ECE, further research can investigate whether the duration of attendance is associated with achievement scores.

Chapter 10

Conclusion

In this chapter I first provide an overview of the answers to the research questions raised in the introduction. Next I discuss the implications for policy and research, offer some recommendations for subsequent cross-country panel studies, discuss avenues for future research and end with a reflection on conducting a cross-country comparison.

10.1 Overview of findings

RQ1: Documenting socioeconomic gaps in achievement

Socioeconomic status (SES) gaps in achievement exist for children before they enter primary school. At 5 years old (prior to school entry in all study countries), gaps in maths scores between the high and the low SES children are 0.93 s.d. in Ethiopia, 0.67 s.d. in India, 0.65 s.d. in Peru and 0.70 s.d. in Vietnam (when using maternal education as a measure of SES). These are comparable to those estimated for high-income countries and for other middle- and low-income nations, as reported by Bradbury et al. (2012, 2015, 2019).

Between 5 and 8 years old the SES gap in maths scores increases in Ethiopia, Peru and Vietnam, but the change is only statistically significant in Peru where it increases from 0.65 s.d. to 1.20 s.d. Between these ages children transition from early childhood education (ECE) to primary school, so advantages accrued from attending ECE are likely to emerge in maths scores taken at the time. In India the SES gap falls slightly (though the change is not statistically significant). This may in part be explained by the generally low quality ECE provided, leaving children inadequately prepared for formal schooling.

Between the ages of 8 to 15 years SES gaps in achievement do not exhibit a statistically significant change. By 15 years old the gaps are not statistically significantly different to the gaps at 5 years old. An unchanging achievement gap between high and low SES children is also reported in the literature for high-income countries and others, including by Bradbury et al. (2015), Heckman and Mosso (2014), Magnuson et al. (2012), and Schady et al. (2015).

The three different measures of SES I use are maternal education, an asset index and expenditure per capita. My findings are consistent across the three measures. SES gaps are similarly sized when I used maternal education and an asset index, and smallest when I used expenditure as a measure of SES (except for Peru, which was the most developed country of the four and where expenditure may have been a more stable of SES).

RQ2: Do achievement trajectories differ by SES?

As Feinstein (2003) does, I plot the trajectories of high and low achieving children at 5 years old, by SES. To address concerns related to regression to the mean I adopt the approach suggested by Jerrim and Vignoles (2013). The trajectories look remarkably similar across countries. We observe a divergence of trajectories within the initially high (and low) achieving groups, by parental SES. In other words, among the children who at 5 years old had performed well in the maths test (i.e. were in the high achieving group) children from low SES families fell behind as they grew up, while those from high SES families continued to score well. Those who had performed poorly at 5 years old (i.e. were in the low achieving group) also saw a divergence in their trajectories, where the children from high SES families outperformed the low SES group and sometimes caught up to the initially high achieving-low SES group. These findings are similar to the one reported using data from high-income countries, such as in Bradbury et al. (2015), Crawford et al. (2017), Feinstein (2003), and Jerrim and Vignoles (2013). A breakdown of the composition of the groups reveal that the bottom groups (low initial achievement-low SES) comprise disproportionately more of children from marginalised communities and the top (high initial achievement-high SES) from more privileged communities.

RQ3: Early childhood education and achievement

When I examine whether attendance at ECE is associated with gains in achievement, over and above parental SES and other demographic factors, I find that where low quality ECE is widespread, as in the case of India, children who attended ECE do not have an advantage over those who did not. I find an association between ECE and maths scores that is moderately sized at 5 years old in Ethiopia, Peru and Vietnam, but its magnitude and statistical significance declines through childhood. By 15 years old, the size of the association is close to nil and there is no statistically significant difference in maths scores between those who attended and did not attend

ECE. I also find that initial parental SES (as measured by maternal education or an asset index) is a stronger predictor of maths scores than is ECE over the course of childhood. Finally, unlike some studies have found, I find no evidence of a negative association between attending ECE and maths scores.

10.2 Implications for policy

I cannot draw specific policy conclusions that would be applicable across all four Young Lives countries. They have their own specific political, economic and cultural contexts that must be taken into consideration when designing policy. The comparative value of how a low- and middle-income country performs against another three countries, however does raise important questions as to why this may be the case. Without a careful comparison of associations between SES and achievement, and between ECE and achievement, we cannot start raising questions as to why these associations differ (or are similar) across different policy regimes.

Socioeconomic status and achievement across countries

One advantage of cross-country comparisons is that they can reveal structural differences that mediate the relationship between SES and achievement. This is most apparent in my analysis of ECE, where my findings vary across countries for the 5 to 12 years age range, and can be explained by the differing ECE systems. However, structural differences cannot explain why SES achievement gaps and achievement trajectories are so consistent across countries, nor why they are so similar to those reported in high-income countries.

These findings highlight that, despite the diversity in policies across countries, there appears to be a deeper, more profound, force at play. Parents from high SES backgrounds across countries, regardless of the policy contexts, are able to transfer privilege to their children (by way of achievement scores), even if at 5 years old they are not high-achievers. This raises an important question as to whether national policies aimed at reducing SES gaps are only able to do so to a small degree.

Early childhood education

Structural differences between ECE systems help to explain some of the differences in my findings. For example, a plausible explanation for the lack of association between ECE attendance and maths scores in India is the generally low quality ECE provision. India's ECE system could benefit from improvements in the quality of its education. The Ethiopian ECE system, which is almost entirely private, is not a scalable one as most families would not be able to afford private ECE.

However, there are lesson to be learnt about the quality of this ECE provision. Both Peru and Vietnam have a meaningful coverage of good quality ECE, and this is associated with a longer-lasting advantage in maths scores. Both ECE systems in Peru and Vietnam are stratified by SES, but whether this stratification perpetuates socioeconomic disparities, as Cueto et al. (2016) contend, would require further research. Based on previous evidence, a strategy of targeting public provision to the more disadvantaged children would likely yield the highest returns in terms of ECE, so long as the education provided is good quality. Any targeting would need to be tailored to each country's context.

10.3 Implications for research

Measures of achievement

Deriving tests of receptive vocabulary skills that are cross-nationally comparable when there are differences in language use within and between countries is challenging. The first challenge is translation: the difficulty of items varies across languages. While the Peabody Picture Vocabulary Test (PPVT) is a well-established test for receptive vocabulary (though designed for a specific population), the Young Lives' experience indicates that translation to different contexts and languages is not always straightforward. Rather than rely on translating an internationally renound vocabulary test, researchers are better served drawing on existing vocabulary tests in the local language (as Young Lives did with the Spanish PPVT) or designing their own receptive vocabulary test with local experts, though these are likely to be quite resource intensive.

Testing concepts that go beyond the understanding of a word, such as maths concepts, provided a more stable and reliable measure of achievement across and within countries. Maths tests test a wider range of academic skills. Maths questions demand an understanding of, not only, the vocabulary in the questions, they also require the child to evaluate a mathematical expression and integrate it into the meaning of the overall question. Vocabulary test scores are sensitive to how a single word is translated and its appropriateness in the testing population. The distribution of the Young Lives maths scores were also comparably less lumpy and less flat, as reported in Annex E. Therefore, when it comes to cross-country studies on achievement that involve multiple languages across and within countries, maths tests will offer a better measure of achievement. In cross-country studies that share the same language, vocabulary tests can be considered. For example, Schady et al., 2015 examined the association between SES and achievement in five Latin American countries and used the Spanish version of the PPVT.

Measure of socioeconomic status

I anticipated that, for the purposes of my research, initial expenditure data is not fit for purpose, as it tends to be a more noisy and volatile measure of SES than an asset index and maternal education. As achievement scores are a reflection of children's long term socioeconomic conditions, an SES measure that does not vary substantially from year to year (or season to season), such as maternal education and the asset index, is more appropriate. My estimates show that both these measures are strongly correlated to maths scores. When I control for all three measures of SES in a regression where maths scores is the outcome variable in Chapter 9, the coefficient on expenditure across all countries and all ages is consistently nil. This is reflective of the long term association between SES and achievement, one that is nearly not impacted by expenditure.

Studies on high-income countries sometimes use an average of multiple years of income or expenditure data to obtain a longer term measure of the family's SES, however Young Lives data does not permit it. The first year expenditure was collected was when the children were 5 years old. To obtain an average expenditure score, I would need data on family expenditure in the first three or four years of the child's life. When researching the association between SES and achievement, measures of SES using an asset index or maternal education are going to produce higher correlations than those using expenditure data.

Reflection on the Young Lives data set

The Young Lives dataset is an unique resource for researchers investigating early predictors of academic achievement and their association with parental SES in low- and middle-income countries. No other dataset offers a similar cross-country comparison, particularly among early primary school students. The Trends in International Mathematics and Science Study (TIMMS) or Programme for International Student Assessment (PISA) only sample students already in school, Young Lives samples in- and out-of-school children and follows the same children over a long period of time. The TIMMS also only starts sampling from 4th grade (Year 5) onwards and PISA samples children at 15 years old, where as Young Lives samples children prior to primary school entry.

Three enhancements to the data set would have been an earlier measure of achievement, more information on the quality of the education provided in ECE centres and a nationally representative sample in all countries covered.

As Grantham-McGregor et al. (2007) show, brains develop at a tremendous speed between birth and age 5, but Young Lives' first achievement score is collected at age 5 which is quite far along in a child's brain development. While measuring achievement at young ages is challenging and current tests are not as precise as those administered to older children, some studies start measuring achievement around 2 or 3 years old, such Feinstein (2003) and Jerrim and Vignoles (2013). Passaretta et al. (2022) compare across three European countries (Germany, the Netherlands and the UK) and find that the main contributors to the SES gaps in achievement took hold prior to school entry. The resources required to design and administer an achievement test appropriate for 3 year olds in the four countries would have been considerable, which is why Young Lives chose not to do so. Nonetheless, evidence on low- and middle-income countries on when SES gaps emerge and when the main contributors to SES gaps take hold is important for policy development.

Young Lives does not provide information on the quality of the ECE institution attended by the children, nor does it disaggregate between the various types of public options that were available in some countries, for example between Centros de Educación Inicial (CEI) and Programas no Escolarizados de Educación Inicial (PRONOEI) in Peru. Previous research has shown that the quality of the education received in ECE is positively correlated children's academic outcomes, particularly the most disadvantaged children, as discussed in Ruhm and Waldfogel (2012). I cannot use the Young Lives data to test whether this is also the case across the four Young Lives countries. Young Lives has started to collect retrospective data on this in subsequent rounds but, given the challenges associated with recall issues the responses given at subsequent ages do not correspond to the data collected when the children were 5 years old.

The Young Lives study did not intend to sample a nationally representative cohort of children. At its inception, it was deemed impossible to overcome the enormous constraints of funding limitations, logistics and research capacity, in order to collect a nationally representative longitudinal sample. Perhaps subsequent cross-country studies can endeavour to do so, while also acknowledging the tremendous cost this would entail, and the risk of attrition. In the absence of that, a sampling approach similar to that of Young Lives can be considered,namely a carefully selection of sites to represent key geographical and socioeconomic groupings in each country. Having more cohort studies in low- and middle-income countries is crucial if research in this area is to expand.

10.4 Further research to build on findings

In response to a critique of his work, Feinstein (2003), writes that "[i]t is my hope that this debate will lead to further comparative work using diverse methods across diverse datasets to establish what differences are due to measurement, what to modelling and what to time and place" (Feinstein, 2015, p. 341). This thesis engages in comparative work across diverse countries, and I have established that the trajectories, even when adjusting for regression to the mean, are remarkably similar to each other and to trajectories reported for other high-income countries (mostly Anglo-Saxon countries). My findings do not identify significant differences in trajectories due to measurement (i.e. when I employ the approach to address regression to the mean suggested by Jerrim and Vignoles (2013)), time (i.e. when I compare shape of the trajectories in this thesis with those of children born at different times in high-income countries such as in Feinstein (2003)) or place (i.e. when I compare across countries).

Further research plotting achievement trajectories by SES in more equal societies, such as Scandinavian countries, would help to establish whether the phenomenon of diverging trajectories, by SES, of initially high and low achieving children, is shared across all types of countries, or whether some countries have been able to mitigate the effects of SES on achievement trajectories.

Another area of research to examine further is whether the ECE systems in Vietnam and Peru, which are stratified by SES, perpetuate or narrow achievement gaps. Where a large portion of children attend public ECE institutions, there is generally one type that caters to the urban and more privileged children that is of higher quality, and another type of ECE provision that offers lower quality education and caters to the more remote areas and less privileged children.

While no association between ECE and achievement at 15 years old is evident in my findings (see Chapter 9), it would be interesting to see whether changes occur late in life. This would involve following the Young Lives children into adulthood exploring whether there are differences in adult outcomes (such as employment, age of first child, mortality and more) that is associated with ECE attendance. Perhaps the children gained skills with ECE that lie dormant at 15 years old, but that prove useful later in life.

In the framework chapter I discuss various mechanisms through which parental SES may support or inhibit children's achievement growth, they include the Investment Theory, the Family Stress Model and the Neurobiological Model. Most of the research underpinning these frameworks is from high-income countries. More work from low- and middle-income countries is needed to establish whether these frameworks are relevant to help understand the association between parental SES and achievement among low- and middle-income countries, or whether alternative models are necessary. If we have the appropriate framework to understand the link between SES and achievement then policy interventions can be adapted accordingly.

Bibliography

Abbink, J. (2011). Ethnic-based Federalism and Ethnicity in Ethiopia: Reassessing the Experiment After 20 Years. *Journal of Eastern African Studies*, *5*(4), 596–618.

Adair, L. S., Popkin, B. M., Akin, J. S., Guilkey, D. K., Gultiano, S., Borja, J., Perez, L., Kuzawa, C. W., McDade, T., & Hindin, M. J. (2011). Cohort Profile: The Cebu Longitudinal Health and Nutrition Survey. *International Journal of Epidemiology*, *40*(3), 619–625.

Admassie, A. (2002). The Implication of Asset Ownership on Child Work in Rural Ethiopia. *Ethiopian Journal of Economics*, *11*(2), 25–58.

Ahmed, A., Hoddinott, J., Quabili, W., Roy, S., & Sraboni, E. (2015). *Which Kinds of Social Safety Net Transfers Work Best for the Rural Ultra Poor in Bangladesh? Operation and Impacts of the Transfer Modality Research Initiative*. International Food Policy Research Institute (IFPRI) & World Food Programme (WFP).

Ahmed, A. U., Sraboni, E., & Shaba, F. K. (2014). *Safety Nets in Bangladesh: Which Form of Transfer is Most Beneficial? Operational Performance of the Transfer Modality Research Initiative*. International Food Policy Research Institute (IFPRI) & World Food Programme (WFP).

Akresh, R., Bagby, E., de Walque, D., & Kazianga, H. (2010). *Child Ability and Household Human Capital Investment Decisions in Burkina Faso* (IZA Discussion Paper No. 5326). IZA.

Alderman, H., & Headey, D. D. (2017). How Important is Parental Education for Child Nutrition? *World Development*, *94*, 448–464.

Alderman, H., Hoddinott, J., & Kinsey, B. (2003). *Long-term Consequences of Early Childhood Malnutrition*. International Food Policy Research Institute (IFPRI).

Alemu, T., Asgedom, G., Liebenberg, J., Mekonnen, A., Seager, J., Tefera, B., & Woldehanna, T. (2003). *Young Lives Preliminary Country Report: Ethiopia*. Young Lives.

Allison, P. (1999). Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research*, *28*(2), 186–208.

Amigo, H., Bustos, P., Zumelzú, E., & Rona, R. J. (2014). Cohort Profile: The Limache, Chile, Birth Cohort Study. *International Journal of Epidemiology*, *43*(4), 1031–1039.

Attanasio, O. P., & Kaufmann, K. M. (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, *109*.

Attawell, K. (2003). *International Longitudinal Research on Childhood Poverty: Practical Guidelines and Lessons Learned from Young Lives*. Young Lives.

Balaj, M., York, H. W., Sripada, K., Besnier, E., Vonen, H. D., Aravkin, A., Friedman, J., Griswold, M., Jensen, M. R., Mohammad, T., Mullany, E. C., Solhaug, S., Sorensen, R., Stonkute, D., Tallaksen, A., Whisnant, J., Zheng, P., Gakidou, E., & Eikemo, T. A. (2021). Parental Education and Inequalities in Child Mortality: A Global Systematic Review and Meta-Analysis. *The Lancet*, *398*(10300).

Balart, P., & Oosterveen, M. (2019). Females show more sustained performance during test-taking than males. *Nature Communications*, *10*(3798).

Banati, P., & Zharkevich, I. (2014). *International Symposium on Cohort and Longitudinal Studies in Low and Middle Income Countries*. UNICEF Office of Research - Innocenti. http://www.unicef-irc.org/files/documents/outcomereport.pdf

Barrón, M. (2008). Exclusion and Discrimination as Sources of Inter-Ethnic Inequality in Peru. *Economía*, *31*(61), 51–80.

Baulch, B., Chuyen, T. T. K., Haughton, D., & Haughton, J. (2007). Ethnic Minority Development in Vietnam. *The Journal of Development Studies*, *43*(7), 1151–1176.

Becker, G. S. (1981). *A Treatise on the Family*. Harvard University Press.

Behrman, J. R., Cheng, Y., & Todd, P. E. (2004). Evaluating Preschool Programs when Length of Exposure to the Program Varies: A Nonparametric Approach. *Review of Economics and Statistics*, *86*(1), 108–132.

Beltrán, A., & Seinfeld, J. (2013). *La Trampa Educativa en el Perú : Cuando la Educación Llega a Muchos Pero Sirve a Pocos*. Universidad del Pacífico (Lima, Peru).

Bendheim–Thoman Center for Research on Child Wellbeing Staff. (2011). *Data User's Guide for the Nine-Year Follow-Up Wave of the Fragile Families and Child Wellbeing Study*. Center for Research on Child Wellbeing (CRCW).

Berger, L. M., Paxson, C., & Waldfogel, J. (2009). Income and child development. *Children and Youth Services Review*, *31*(9), 978–989.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The Effect of Pre-primary Education on Primary School Performance. *Journal of Public Economics*, *93*(1), 219–234.

Bhargava, S. K., Sachdev, H. S., Fall, C. H. D., Osmond, C., Lakshmy, R., Barker, D. J. P., Biswas, S. K. D., Ramji, S., Prabhakaran, D., & Reddy, K. S. (2004). Relation of Serial Changes in Childhood Body-Mass Index to Impaired Glucose Tolerance in Young Adulthood. *New England Journal of Medicine*, *350*(9), 865–875.

Blanden, J., Bono, E. D., Hansen, K., & Rabe, B. (2022). Quantity and Quality of Childcare and Children's Educational Outcomes. *Journal of Population Economics*, *35*, 785–828.

Blanden, J., Katz, I., & Redmond, G. (2012). Family Background and Child Outcomes. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 140–163). Russell Sage Foundation.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. John Wiley & Sons, Inc.

Boyd, W., & Phuong, T. D. (2017). Early Childhood Education in Vietnam: History and Evaluation of Its Policies. In H. Li, E. Park, & C. J. (Eds.), *Early Childhood Education Policies in Asia Pacific. Education in the Asia-Pacific Region: Issues, Concerns and Prospects* (pp. 263–283, Vol. 35). Springer.

Boyden, J., Hardgrove, A., & Knowles, C. (2012). Continuity and Change in Poor Children's Lives: Evidence from Young Lives. In A. Minujin & S. Nandy (Eds.), *Global Child Poverty and Well-Being: Measurement, Concepts, Policy and Action*. The Policy Press.

Boyden, J., & James, Z. (2014). Schooling, Childhood Poverty and International Development: Choices and Challenges in a Longitudinal Study. *Oxford Review of Education*, *40*(1), 10–29.

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2012). Inequality in Early Childhood Outcomes. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 87–119). Russell Sage Foundation.

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2015). *Too Many Children Left Behind: The U.S. Achievement Gap in Comparative Perspective*. Russel Sage Foundation.

Bradbury, B., & Jenkins, S. P. (2001). Conceptual and measurement issues. In B. Bradbury, S. P. Jenkins, & J. Micklewright (Eds.), *The Dynamics of Child Poverty in Industrialised Countries*. Cambridge University Press.

Bradbury, B., Waldfogel, J., & Washbrook, E. (2019). Income-Related Gaps in Early Child Cognitive Development: Why Are They Larger in the United States Than in the United Kingdom, Australia, and Canada? *Demography*, *56*, 367–390.

Breierova, L., & Duflo, E. (2004). *The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?* (NBER Working Paper No. 10513). NBER.

Briones, K. (2017). *How Many Rooms Are There in Your House? Constructing the Young Lives Wealth Index* (Technical Note No. 43). Young Lives.

Briones, K. (2018). *A Guide to Young Lives Rounds 1 to 5 Constructed Files* (Technical Note No. 48). Young Lives.

Brito, N. H., & Noble, K. G. (2014). Socioeconomic Status and Structural Brain Development. *Frontiers in Neuroscience*, *8*, 276.

Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, *32*(1), 3–22.

Bynner, J. (2015). Social Class Differences In Early Cognitive Development. *Longitudinal and Life Course Studies*, *6*(3), 331.

Cameron, C. E., Kim, H., Duncan, R. J., Becker, D. R., & McClelland, M. M. (2019). Bidirectional and Co-developing Associations of Cognitive, Mathematics, and Literacy Skills During Kindergarten. *Journal of Applied Developmental Psychology*, *62*, 135–144.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2010). Early Childhood Education: Young Adult Outcomes From the Abecedarian Project. *Applied Developmental Science*, *6*(1), 42–57.

Caro, D. H., McDonald, J. T., & Willms, J. D. (2009). Socio-economic Status and Academic Achievement Trajectories from Childhood to Adolescence. *Canadian Journal of Education / Revue canadienne de l'éducation*, *32*(2), 558–590.

Cattan, S., Goodman, E., Fitzsimons, A., Ploubidis, A., Phimister, G. B., & Wertz, J. (2022). Early childhood inequalities. *IFS Deaton Review of Inequalities*.

Chaimay, B., Ruagdaraganon, N., Thinkhamrop, B., & Thinkhamrop, J. (2015). Association between Infant Feeding Practices and First Meaningful Words at First Year of Life: A Prospective Cohort Study of Thai Children. *Asia-Pacific Journal of Public Health*, *27*(2), NP1071–NP1084.

Chmielewski, A. K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review*, *84*(3), 517–544.

Chowa, G., Ansong, D., & Masa, R. (2010). Assets and Child Well-being in Developing Countries: A Research Review. *Children and Youth Services Review*, *32*(11), 1508–1519.

Chowa, G., Masa, R. D., Wretman, C. J., & Ansong, D. (2013). The Impact of Household Possessions on Youth's Academic Achievement in the Ghana Youthsave Experiment: A Propensity Score Analysis. *Economics of Education Review*, *33*, 69–81.

Citizens' Initiative for the Rights of Children Under Six. (2006). *Focus on Children Under Six*. Secretariat of The Right to Food Campaign.

Claessens, A., Duncan, G. J., & Engel, M. (2009). Kindergarten Skills and Fifth-Grade Achievement: Evidence from the ECLS-K. *Economics of Education Review*, *28*(4), 415–427.

Connelly, R., & Platt, L. (2014). Cohort Profile: UK Millennium Cohort Study. *International Journal of Epidemiology*, *43*(6), 1719–1725.

Cooper, K., & Stewart, K. (2021). Does Household Income Affect Children's Outcomes? A Systematic Review of the Evidence. *Child Indicators Research*, *14*, 981–1005.

Crawford, C., Macmillan, L., & Vignoles, A. (2017). When and why do initially high-achieving poor children fall behind? *Oxford Review of Education*, *43*(1), 88–108.

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*, 253.

Cueto, S., & León, J. (2012). *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives*. Young Lives.

Cueto, S., León, J., Miranda, A., Dearden, K., Crookston, B. T., & Behrman, J. R. (2016). Does Pre-school Improve Cognitive Abilities Among Children with Early-life Stunting? A Longitudinal Study for Peru. *International Journal of Educational Research*, *75*, 102–114.

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, *97*(2), 31–47.

Das, J., Singh, A., & Chang, A. Y. (2022). Test Scores and Educational Opportunities: Panel Evidence from Five Low- and Middle-Income Countries. *Journal of Public Economics*, *206*, 104570.

Davis, C. E. (1976). The Effect of Regression to the Mean in Epidemiologic and Clinical Studies. *American Journal of Epidemiology*, *104*(5), 493–498.

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.

Deaton, A., & Grosh, M. (2000). Chapter 17: Consumption. In M. Grosh & P. Glewwe (Eds.), *Designing Household Survey Questionnaires for Developing Countries: Lessons From Ten Years of LSMS Experience*. World Bank.

Defo, B. K. (1992). Mortality and Attrition Processes in Longitudinal Studies in Africa: An Appraisal of the IFORD Surveys. *Population Studies*, *46*(2), 327–348.

de Neubourg, E., Borghans, L., Coppens, K., & Jansen, M. (2017). Explaining Children's Life Outcomes: Parental Socioeconomic Status, Intelligence and Neurocognitive Factors in a Dynamic Life Cycle Model. *Child Indicators Research*, *11*(5), 1495–1513.

Dorling, D. (2010). The Return to Elitism in Education: A Society's Attitudes to Innate Intelligence are Closely Correlated with its Levels of Inequality. *Soundings*, *44*, 35–46.

Dräger, J., Washbrook, E., Schneider, T., Akabayashi, H., Keizer, R., Solaz, A., Waldfogel, J., De la Rie, S., Kameyama, Y., Kwon, S. J., Nozaki, K., Sano, S., Sheridan, A., Shikishima, C., & Casoni, V. P. (2023, February). Cross-national Differences in Socioeconomic Achievement Inequality in Early Primary School: The Role of Parental Education and Income in Six Countries. https://research-information.bris.ac.uk/en/publications/cross-national-differences-in-socioeconomic-achievement-inequalit

Duflo, E. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, *50*(4), 1051–1079.

Duncan, G. J., Bergman, L., Duckworth, K., Kokko, K., Lyyra, A.-L., Metzger, M., Pulkkinen, L., & Simonton, S. (2012). Child Skills and Behaviors in Intergenerational Inequality. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 207–234). Russell Sage Foundation.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(5), 1428–1446.

Duncan, G. J., & Magnuson, K. (2011a). The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? : Rising Inequality, Schools, and Children's Life Chances* (pp. 47–69). Russell Sage Foundation.

Duncan, G. J., & Magnuson, K. (2011b). The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems: Online Appendix. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*. Russel Sage Foundation.

Duncan, G. J., & Magnuson, K. (2012). Socioeconomic Status and Cognitive Functioning: Moving from Correlation to Causation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 377–386.

Duncan, G. J., Magnuson, K., & Votruba-Drzal, E. (2014). Boosting Family Income to Promote Child Development. *The Future of Children*, *24*(1), 99–120.

Dunn, L., Padilla, E., Lugo, D., & Dunn, L. (1986). *Manual del Examinador para el Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana (Peabody Picture Vocabulary Test: Hispanic-American Adaptation).* AGS.

Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a Statewide Pre-Kindergarten Program on Children's Achievement and Behavior Through Sixth Grade. *Developmental Psychology*, *58*(3), 470–484.

Dworsky, A. (2014). Children as Self-Informants in Longitudinal Studies: Substantive Findings and Methodological Issues. In G. B. Melton, A. Ben-Arieh, J. Cashmore, G. S. Goodman, & N. K. Worley (Eds.), *The SAGE Handbook of Child Research*. Sage.

Ederer, F. (1972). Serum Cholesterol Changes: Effects of Diet and Regression Toward the Mean. *Journal of Chronic Diseases*, *25*, 277–289.

Edwards, S. (2021). Process quality, curriculum and pedagogy in early childhood education and care. *OECD Education Working Papers No. 247*.

Elliott, L., & Bachman, H. J. (2018). SES Disparities in Early Math Abilities: The Contributions of Parents' Math Cognitions, Practices to Support Math, and Math Talk. *Developmental Review*, *49*, 1–15.

Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., de Mello, M. C., Hidrobo, M., Ulkuer, N., Ertem, I., & Iltus, S. (2011). Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries. *The Lancet*, *378*(9799), 1339–1353.

Ermisch, J., Jäntti, M., Smeeding, T. M., & Wilson, J. (2012). Advantage in Comparative Perspective. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 3–31). Russell Sage Foundation.

Ermisch, J., Peter, F. H., & Spiess, C. K. (2012). Early Childhood Outcomes and Family Structure. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 120–139). Russell Sage Foundation.

Escobal, J., Ames, P., Cueto, S., Penny, M., & Flores, E. (2008). *Young Lives: Peru Round 2 Survey* (Young Lives Country Report). Young Lives.

Escobal, J., & Flores, E. (2008). *An Assessment of the Young Lives Sampling Approach in Peru* (Technical Note No. 3). Young Lives.

Estadística de Calidad Educativa. (2018, March). Serie Hasta 2015: Tendencias. escale.minedu.gob.pe/ueetendencias20002015

Fafchamps, M., & Shilpi, F. (2015). Education and household welfare.

Feinstein, L. (2003). Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica*, *70*(277), 73–97.

Feinstein, L. (2015). Social Class Differences in Early Cognitive Development and Regression to the Mean. *Longitudinal and Life Course Studies*, *6*(3), 331–343.

Feinstein, L., Jerrim, J., Vignoles, A., Goldstein, H., French, R., Washbrook, E., Lee, R., & Lupton, R. (2015). Social Class Differences in Early Cognitive Development Debate. *Longitudinal and Life Course Studies*, *6*, 331–376.

Fernald, L. C. H., Kariger, P., Hidrobo, M., & Gertler, P. J. (2012). Socioeconomic Gradients in Child Development in Very Young Children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences*, *109*(Supplement 2), 17273–17280.

Filmer, D., & Pritchett, L. (1999). The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries. *Population and Development Review*, *25*(1), 85–120.

Filmer, D., & Pritchett, L. (2001). Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application To Educational Enrollments In States Of India. *Demography*, *38*(1), 115–132.

Filmer, D., & Scott, K. (2012). Assessing Asset Indices. *Demography*, *49*(1), 359–392.

Foster, E. M. (2002). How Economists Think about Family Resources and Child Development. *Child Development*, *73*(6), 1904–1914.

Gabel, S. G. (2010). The Development of International Comparative Child and Family Policies. In S. B. Kamerman, S. A. Phipps, & A. Ben-Aryeh (Eds.), *From Child Welfare to Child Well-being: An International Perspective On Knowledge in the Service of Policy Making* (pp. 175–188). Springer.

Galab, S., Reddy, M. G., Antony, P., McCoy, A., Ravi, C., Raju, D. S., Mayuri, K., & Reddy, P. P. (2003). *Preliminary Country Report: Andhra Pradesh, India.* Young Lives.

Galab, S., Reddy, P. P., & Himaz, R. (2008). *Young Lives Round 2 Survey Report Initial Findings: Andhra Pradesh, India.* Young Lives.

Galton, F. (1886). Regression Toward Mediocrity in Hereditary Stature. *Journal of Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Gambaro, L., Stewart, K., & Waldfogel, J. (2014). *An Equal Start?: Providing Quality Early Education and Care for Disadvantaged Children.* Policy Press.

General Statistics Office of Viet Nam & UNICEF. (2007). *Viet Nam Multiple Indicator Cluster Survey.*

Georgiadis, A. (2017). *Parental Background and Child Human Capital Development Throughout Childhood and Adolescence: Evidence From Four Low- and Middle-Income Countries* (Working Paper No. 164). Young Lives.

Giang, K. B., Oh, J., Kien, V. D., Hoat, L. N., Choi, S., Lee, C. O., & Minh, H. V. (2016). Changes and Inequalities in Early Birth Registration and Childhood

Care and Education in Vietnam: Findings from the Multiple Indicator Cluster Surveys, 2006 and 2011. *Global Health Action*, *9*(1), 29470.

Glewwe, P., James, Z., Lee, J., Rolleston, C., & Vu, K. (2021, September). *What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the Young Lives Data from Ethiopia, Peru, India, and Vietnam (RISE Working Paper 21/078)*. RISE.

Goldstein, H., & French, R. (2015). Differential Educational Progress and Measurement Error. *Longitudinal and Life Course Studies*, *6*(3), 351–358.

Gonçalves, H., Assunção, M. C. F., Wehrmeister, F. C., Oliveira, I. O., Barros, F. C., Victora, C. G., Hallal, P. C., & Menezes, A. M. B. (2014). Cohort Profile Update: The 1993 Pelotas (Brazil) Birth Cohort Follow-Up Visits in Adolescence. *International Journal of Epidemiology*, *43*(4), 1082–1088.

Gould, W. (2024, April). Pooling Data and Performing Chow tests in Linear Regression. https://www.stata.com/support/faqs/statistics/pooling-data-and-chow-tests/index.html

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental Potential in the First 5 Years for Children in Developing Countries. *Lancet*, *369*(9555), 60–70.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Wiley.

Hantrais, L. (2007). Contextualization in Cross-National Comparative Research. In L. Hantrais & S. Mangen (Eds.), *Cross-National Research Methodology and Practice* (pp. 3–19). Routledge.

Harpham, T., Huttly, S., Wilson, I., & De Wet, T. (2003). Linking Public Issues with Private Troubles: Panel Studies in Developing Countries. *Journal of International Development*, *15*(3), 353–363.

Haveman, R., & Wolfe, B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature*, *33*(4), 1829–1878.

Havnes, T., & Mogstad, M. (2011). No Child Left Behind: Subsidized Child Care and Children's Long Run Outcomes. *American Economic Journal: Economic Policy*, *3*(2), 97–129.

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, *1*(1), 1–46.

Heckman, J. J. (2008). Schools, Skills and Synapses. *Economic Inquiry*, *46*(3), 289–324.

Heckman, J. J., & Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics*, *6*, 689–733.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Hill, Z. (2004). Reducing Attrition in Panel Studies in Developing Countries. *International Journal of Epidemiology*, *33*(3), 493–498.

Ho, M.-H., Lefresne, F., & DEPP-MIREI. (2014). *Education et Accueil des Jeunes Enfants: Cadrage Européen* (Note d'Information No. 33). Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche.

Hobcraft, J. (2007). *Child Development, the Life Course, and Social Exclusion: Are the Frameworks Used in the UK Relevant for Developing Countries?* Chronic Poverty Research Centre.

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378.

Hoot, J. L., Szente, J., & Mebratu, B. (2004). Early Education in Ethiopia: Progress and Prospects. *Early Childhood Education Journal*, *32*(1), 3–8.

Horta, B. L., Gigante, D. P., Gonçalves, H., dos Santos Motta, J., de Mola, C., Oliveira, I. O., Barros, F. C., & Victora, C. G. (2015). Cohort Profile Update: The 1982 Pelotas (Brazil) Birth Cohort Study. *International Journal of Epidemiology*, *44*.

Huffman, M. D., Prabhakaran, D., Osmond, C., Fall, C. H. D., Tandon, N., Lakshmy, R., Ramji, S., Khalil, A., Gera, T., Prabhakaran, P., Biswas, S. K. D., Reddy, K. S., Bhargava, S. K., & Sachdev, H. S. (2011). Incidence of Cardiovascular Risk Factors in an Indian Urban Cohort: Results From the New Delhi Birth Cohort. *Journal of the American College of Cardiology*, *57*(17), 1765–1774.

Integrated Child Development Services. (2007). ICDS in India: An Eyewitness Account (2005-2007).

International Institute for Population Sciences (IIPS) & ORC Macro. (2000). Chapter 6: Mortality, Morbidity, and Immunization. In *National Family Health Survey (NFHS-2), India, 1998/99: Andhra Pradesh.* IIPS.

James, Z. C. (2018). *Language and Learning in Ethiopia* [Doctoral dissertation, University College London]. https://discovery.ucl.ac.uk/id/eprint/10042137/

Jaramillo, A., & Tietjen, K. (2001). *Early Childhood Development in Africa: Can We Do More or Less? A Look at the Impact and Implications of Preschools in Cape Verde and Guinea*. World Bank.

Jenkins, S. P., & Siedler, T. (2007). *Using Household Panel Data to Understand the Intergenerational Transmission of Poverty*. Chronic Poverty Research Centre.

Jerrim, J., & Micklewright, J. (2012). Children's Cognitive Ability and Changes over Age in the Socioeconomic Gradient. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 262–284). Russell Sage Foundation.

Jerrim, J., & Micklewright, J. (2014). Socio-economic Gradients in Children's Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background? *European Sociological Review*, *30*(6), 766–781.

Jerrim, J., & Vignoles, A. (2013). Social Mobility, Regression to the Mean and the Cognitive Development of High Ability Children from Disadvantaged Homes. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *176*(4), 887–906.

Jowell, R. (1998). How Comparative is Comparative Research? *The American Behavioral Scientist*, *42*(2), 168–177.

Kalil, A., Ryan, R., & Corey, M. (2012). Diverging Destinies: Maternal Education and the Developmental Gradient in Time With Children. *Demography*, *49*, 1361–1383.

Kaul, V., & Sankar, D. (2009). *Education For All - Mid Decade Assessment: Early Childhood Care and Education in India*. National University of Educational Planning and Administration (NUEPA).

Kim, P., Evans, G. W., Chen, E., Miller, G., & Seeman, T. (2018). How Socioeconomic Disadvantages Get Under the Skin and into the Brain to Influence Health Development Across the Lifespan. In N. Halfon, C. B. Forrest, R. M. Lerner, & E. M. Faustman (Eds.), *Handbook of Life Course Health Development* (pp. 463–497). Springer International Publishing.

Kim, S. w., Cho, H., & Kim, L. Y. (2019). Socioeconomic Status and Academic Outcomes in Developing Countries: A Meta-Analysis. *Review of Educational Research*, *89*(6), 875–916.

Kinra, S., Radha Krishna, K. V., Kuper, H., Rameshwar Sarma, K. V., Prabhakaran, P., Gupta, V., Walia, G. K., Bhogadi, S., Kulkarni, B., Kumar, A., Aggarwal, A., Gupta, R., Prabhakaran, D., Reddy, K. S., Davey Smith, G., Ben-Shlomo, Y., & Ebrahim, S. (2014). Cohort Profile: Andhra Pradesh Children and Parents Study (APCAPS). *International Journal of Epidemiology*, *43*(5), 1417–1424.

Kottelenberg, M. J., & Lehrer, S. F. (2019). Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Child Care. *Journal of Labor Economics*, *35*(3), 609–653.

Krishnaveni, G. V., Veena, S. R., Hill, J. C., Karat, S. C., & Fall, C. H. D. (2015). Cohort Profile: Mysore Parthenon Birth Cohort. *International Journal of Epidemiology*, *44*(1), 28–36.

Kumra, N. (2008). *An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India* (Technical Note No. 2). Young Lives.

Landers-Potts, M. A., Wickrama, K. A. S., Simons, L. G., Cutrona, C., Gibbons, F. X., Simons, R. L., & Conger, R. (2015). An Extension and Moderational Analysis of the Family Stress Model Focusing on African American Adolescents. *Family Relations*, *64*(2), 233–248.

Leon, J., & Singh, A. (2017). *Equating Test Scores for Receptive Vocabulary Across Rounds and Cohorts in Ethiopia, India and Vietnam* (Technical Note No. 40). Young Lives.

León, J., Miranda, A., & Cueto, S. (2015). *Equating Children's PPVT Scores Across Survey Rounds and Age Cohorts in Peru* (Technical Note No. 32). Young Lives.

Linberg, T., Schneider, T., Waldfogel, J., & Wang, Y. (2019). Socioeconomic Status Gaps in Child Cognitive Development in Germany and the United States. *Social Science Research*, *79*, 1–31.

Lockheed, M. E., Fuller, B., & Nyirongo, R. (1989). Family Effects on Students' Achievement in Thailand and Malawi. *Sociology of Education*, *62*(4), 239–256.

Lohr, S. L. (2010). *Sampling: Design and Analysis* (2nd ed.). Brooks/Cole.

London, J. D. (2011). Contemporary Vietnam's Education System: Historical Roots, Current Trends. In J. D. London (Ed.), *Education in Vietnam*. Institute of Southeast Asian Studies.

Lopez-Boo, F. (2016). Socio-Economic Status and Early Childhood Cognitive Skills. *International Journal of Behavioral Development*, *40*(6), 500–508.

Lu, Y.-C., Kapse, K., Andersen, N., Quistorff, J., Lopez, C., Fry, A., Cheng, J., Andescavage, N., Wu, Y., Espinosa, K., Vezina, G., du Plessis, A., & Limperopoulos, C. (2021). Association Between Socioeconomic Status and In Utero Fetal Brain Development. *JAMA Network Open*, *4*(3), e213526.

Magnuson, K. (2007). Maternal Education and Children's Academic Achievement During Middle Childhood. *Developmental Psychology*, *43*(6), 1497–1512.

Magnuson, K., & Duncan, G. J. (2016). Can Early Childhood Interventions Decrease Inequality of Economic Opportunity? *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(2), 123–141.

Magnuson, K., Sexton, H. R., Davis-Kean, P. E., & Huston, A. C. (2009). Increases in Maternal Education and Young Children's Language Skills. *Merrill-Palmer Quarterly*, *55*(3), 319–350.

Magnuson, K., Waldfogel, J., & Washbrook, E. (2012). SES Gradient in Skills During the School Years. In J. Ermisch, M. Jäntti, & T. M. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 235–261). Russell Sage Foundation.

Malmberg, L.-E., Mwaura, P., & Sylva, K. (2011). Effects of a Preschool Intervention on Cognitive Development among East-African Preschool Children: A Flexibly Time-Coded Growth Model. *Early Childhood Research Quarterly*, *26*, 124–133.

Manji, S., Arnold, C., Gowani, S., Bartlett, K., Kaul, V., Sharma, S., Chaudhary, A. B., & Sharma, S. (2015). *How Are We Doing and How we Get it Right for Children?* (Background paper prepared for the Education for All Global Monitoring Report 2015). UNESCO.

Marion, P. (2018). *A Guide to Young Lives Rounds 2 to 5 Consumption Aggregates* (Technical Note No. 49). Young Lives.

McCaw-Binns, A., Ashley, D., Samms-Vaughan, M., Wilks, R., Ferguson, T., Younger, N., Reece, J.-A., Tulloch-Reid, M., & Foster-Williams, K. (2011). Cohort Profile: The Jamaican 1986 Birth Cohort Study. *International Journal of Epidemiology*, *40*(6), 1469–1476.

Mehta, A. (2007). *Elementary Education in India: Where Do We Stand?* (State report cards 2005-06). National University of Educational Planning and Administration.

Mensch, B. S., Chuang, E. K., Melnikas, A. J., & Psaki, S. R. (2019). Evidence for Causal Links Between Education and Maternal and Child Health: systematic review. *Tropical Medicine and International Health*, *24*(5).

Ministerio de Educacion. (2007). *Plan Estrategico Institucional (2007-2011)*. Ministerio de Educacion (MdE).

Ministry of Education. (2008). *Education statistics Annual Abstract 2000 E.C. (2006/07)*. Ministry of Education, Ethiopia.

Ministry of Education and Training. (2015). *Education for All 2015 National Review Report: Viet Nam*. UNESCO.

Ministry of Finance and Economic Development. (2012). *Ethiopia's Progress Towards Eradicating Poverty: An Interim Report on Poverty Analysis Study (2010/11)*. MoFED.

Ministry of Women and Child Development. (2007). *A Report. Working Group on Development of Children for the Eleventh Five Year Plan (2007–2012)*. Government of India.

Minot, N., Baulch, B., & Epperecht, M. (2003). *Poverty and inequality in Vietnam: Spatial patterns and geographic determinants*. International Food Policy Research Institute (IFPRI).

Mohammed, A., Adel, B.-Y., & Cuong, N. (2016). *Ethnic and Racial Disparities in Children's Education* (WIDER Working Paper No. 2016/87). UNU-WIDER.

Mongkolchati, A., Thinkhamrop, B., Mo-Suwan, L., Chittchang, U., & MD, C. C. (2010). Prevalence and Incidence of Child Stunting from Birth to Two Years of Life in Thai Children: Based on the Prospective Cohort Study of Thai

Children (PCTC). *Journal of The Medical Association of Thailand*, *93*(12), 1368–1378.

Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool Experience in 10 Countries: Cognitive and Language Performance at Age 7. *Early Childhood Research Quarterly*, *21*(3), 313–331.

Morrow, V. (2009). *The Ethics of Social Research with Children and Families in Young Lives: Practical Experiences* (Working Paper No. 35). Young Lives.

Moss, P. (2010). Early Childhood Education and Care. In S. B. Kamerman, S. A. Phipps, & A. Ben-Aryeh (Eds.), *From Child Welfare to Child Well-being : An International Perspective On Knowledge in the Service of Policy Making* (pp. 371–384). Springer.

Mostafa, T. (2012). *Sample design and attrition in MCS*. Centre for Longitudinal Studies (CLS), Institute of Education.

Narea, M. (2015). *Who Cares for Our Children Matters: Early Maternal Employment, Early Childcare, and Child Development in Chile* [Doctoral dissertation, London School of Economics and Political Science]. http://etheses.lse.ac.uk/3069/1/Narea_Who_cares_for_our_children_matters.pdf

Naudeau, S., Martinez, S., Premand, P., & Filmer, D. (2011). Cognitive Development among Young Children in Low-Income Countries. In H. Alderman (Ed.), *No Small Matter : The Impact of Poverty, Shocks, and Human Capital Investments in Early Childhood Development* (pp. 9–47). World Bank.

Nguyen, N. P. (2008). *An Assessment of the Young Lives Sampling Approach in Vietnam* (Technical Note No. 4). Young Lives.

Nguyen, Q. K., & Nguyen, Q. C. (2008). Education in Vietnam: Development History, Challenges, and Solutions (Chapter 4). In B. Fredriksen & J. P. Tan (Eds.), *An African Exploration of the East Asian Education Experience*. World Bank.

Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Libiger, O., Schork, N. J., Murray, S. S., Casey, B. J., Chang, L., Ernst, T. M., Frazier, J. A., Gruen, J. R., Kennedy, D. N., Zijl, P. V., ... Sowell, E. R. (2015). Family Income, Parental Education and Brain Structure in Children and Adolescents. *Nature Neuroscience*, *18*(5), 773–780.

Nores, M., & Barnett, W. S. (2010). Benefits of Early Childhood Interventions Across the World: (Under) Investing in the Very Young. *Economics of Education Review*, *29*(2), 271–282.

Ojala, M., & Talts, L. (2007). Preschool Achievement in Finland and Estonia: Cross-Cultural Comparison between the Cities of Helsinki and Tallinn. *Scandinavian Journal of Educational Research*, *51*(2), 205–221.

Organisation for Economic Co-operation and Development. (2016). *PISA 2015 Results in Focus* (PISA in Focus No. 67). OECD Publishing.

Organisation for Economic Co-operation and Development. (2017). Scaling PISA Data. In *PISA 2015 Technical Report*. OECD Publishing.

Organisation for Economic Co-operation and Development. (2021). *Starting Strong VI: Supporting Meaningful Interactions in Early Childhood Education and Care*. OECD Publishing.

Outes-Leon, I., & Sanchez, A. (2008). *An Assessment of the Young Lives Sampling Approach in Ethiopia* (Technical Note No. 1). Young Lives.

Passaretta, G., Skopek, J., & Van Huizen, T. (2022). Is Social Inequality in School-Age Achievement Generated before or during Schooling? A European Perspective. *European Sociological Review*, *38*(6).

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.

Petrou, S., & Kupek, E. (2010). Poverty and childhood undernutrition in developing countries: a multi-national cohort study. *Social Science & Medicine*, *71*(7), 1366–1373.

Planning Commission (Government of India). (2008). *Eleventh Five Year Plan (2007-2012) Volume II Social Sector*. India Planning Commission.

Platt, L. (2014). *Millennium Cohort Study: Initial Findings from the Age 11 Survey*. Institute of Education.

Purpura, D. J., Logan, J. A. R., Hassinger-Das, B., & Napoli, A. R. (2017). Why do Early Mathematics Skills Predict Later Reading? The Role of Mathematical Language. *Developmental Psychology*, *53*(9), 1633–1642.

Raine, A., Liu, J., Venables, P. H., Mednick, S. A., & Dalais, C. (2010). Cohort Profile: The Mauritius Child Health Project. *International Journal of Epidemiology*, *39*(6), 1441–1451.

Ramirez-Zea, M., Melgar, P., & Rivera, J. A. (2009). INCAP Oriente Longitudinal Study: 40 Years of History and Legacy. *The Journal of Nutrition*, *Symposium: The Development and Legacy of the INCAP Oriente Studies 1969–2009*.

Rao, N., Richards, B., Sun, J., Weber, A., & Alanna Sincovich. (2019). Early childhood education and child development in four countries in East Asia and the Pacific. *Early Childhood Research Quarterly*, *47*, 169–181.

Rea, D., & Burton, T. (2020). New Evidence on the Heckman Curve. *Journal of Economic Surveys*, *34*(2), 241–262.

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. In G. J. Duncan &

R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 91–116). Russell Sage Foundation.

Reichman, N. E., Teitler, J. O., Garfinkel, I., & McLanahan, S. S. (2001). Fragile Families: Sample and Design. *Children and Youth Services Review*, *23*(4), 303–326.

Revollo, P. E., & Scott, D. (2022). *Cognitive and Achievement Tests in the Young Lives Study* (Technical Note No. 53). Young Lives.

Reynolds, S. A., Andersen, C., Behrman, J., Singh, A., Stein, A. D., Benny, L., Cueto, B. T. C. S., Dearden, K., Georgiadis, A., Krutikova, S., & Fernald, L. C. (2017). Disparities in Children's Vocabulary and Height in Relation to Household Wealth and Parental Schooling: A Longitudinal Study in Four Low- and Middle-income Countries. *SSM - Population Health*, *3*, 767–786.

Rhemtulla, M., & Tucker-Drob, E. M. (2012). Gene-by-socioeconomic status interaction on school readiness. *Behavior Genetics*, *42*(4).

Richter, L., Norris, S., Pettifor, J., Yach, D., & Cameron, N. (2007). Cohort Profile: Mandela's Children: The 1990 Birth to Twenty Study in South Africa. *International Journal of Epidemiology*, *36*(3), 504–511.

Richter, L. M., Victora, C. G., Hallal, P. C., Adair, L. S., Bhargava, S. K., Fall, C. H. D., Lee, N., Martorell, R., Norris, S. A., Sachdev, H. S., Stein, A. D., & the COHORTS Group. (2012). Cohort Profile: The Consortium of Health-Orientated Research in Transitioning Societies. *International Journal of Epidemiology*, *41*(3), 621–626.

Rogers, R., Ma, D. H., Nguyen, T., & Nguyen, N. A. (2019). Early Childhood Education and Cognitive Outcomes in Adolescence: a Longitudinal Study from Vietnam. *Education Economics*, *27*(6), 658–669.

Rose, D. (2000). *Researching Social and Economic Change: The Uses of Household Panel Studies*. Routledge.

Rosenzweig, M. R. (2003). Payoffs from Panels in Low-Income Countries: Economic Development and Economic Mobility. *The American Economic Review*, *93*(2), 112–117.

Ruhm, C., & Waldfogel, J. (2012). Long-Term Effects of Early Childhood Care and Education. *Nordic Economic Policy Review*, *1*, 23–51.

Russell, F., & Azzopardi, P. (2019). WASH: a Basic Human Right and Essential Intervention for Child Health and Development. *The Lancet*, *7*(4), PE417.

Rutstein, S. O., & Stavetei, S. (2014). *Making the Demographic and Health Surveys Wealth Index Comparable*. ICF International.

Sabates, R., & Di Cesare, M. (2021). Can Maternal Education Sustain or Enhance the Benefits of Early Life Interventions? Evidence from the Young Lives Longitudinal Study. *Compare*, *51*(5), 651–669.

Sahn, D. E., & Stifel, D. C. (2003). Urban–Rural Inequality in Living Standards in Africa. *Journal of African Economies*, *12*(4), 564–597.

Sampling Expert Panel of the European Social Survey (ESS). (2012). *Sampling for the European Social Survey Round VI: Principles and Requirements.* European Social Survey.

Sánchez, A., & Escobal, J. (2020). Survey Attrition After 15 Years of Tracking Children in Four Developing Countries: The Young Lives Study. *Review of Development Economics*, *24*(4), 1196–1216.

Sangsupawanich, P., Chongsuvivatwong, V., Mo-Suwan, L., & Choprapawon, C. (2007). Relationship between Atopic Dermatitis and Wheeze in the First Year of Life: Analysis of a Prospective Cohort of Thai Children. *Journal of Investigational Allergology and Clinical Immunology*, *17*(5), 292–296.

Santos, I. S., Barros, A. J. D., Matijasevich, A., Zanini, R., Chrestani Cesar, M. A., Camargo-Figuera, F. A., Oliveira, I. O., Barros, F. C., & Victora, C. G. (2014). Cohort Profile Update: 2004 Pelotas (Brazil) Birth Cohort Study. Body Composition, Mental Health and Genetic Assessment at the 6 Years Follow-Up. *International Journal of Epidemiology*, *43*(5), 1437–1437.

Schady, N., Behrman, J., Araujo, M. C., Azuero, R., Bernal, R., Bravo, D., Lopez-Boo, F., Macours, K., Marshall, D., Paxson, C., & Vakis, R. (2015). Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries. *Journal of Human Resources*, *50*(2), 446–463.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton Mifflin.

Sigle-Rushton, W. (2009). Comparative Methods in Research on Gender and Welfare States. *Twenty-First Century Society*, *4*(2), 137–148.

Singh, A. (2014a). *Emergence and Evolution of Learning Gaps across Countries: Panel Evidence from Ethiopia, India, Peru and Vietnam* (Working Paper No. 124). Young Lives. Young Lives.

Singh, A. (2014b). Test score gaps between private and government sector students at school entry age in India. *Oxford Review of Education*, *40*(1), 30–49.

Singh, R., & Mukherjee, P. (2018). Effect of Preschool Education on Cognitive Achievement and Subjective Wellbeing at Age 12: Evidence from India. *Compare: A Journal of Comparative and International Education*, 1–19.

Skopek, J., & Passaretta, G. (2020). Socioeconomic Inequality in Children's Achievement from Infancy to Adolescence: The Case of Germany. *Social Forces*, 1–27.

Socialist Republic of Viet Nam. (2003). *National Education For All Plan of Action (2003-2015).* Socialist Republic of Viet Nam.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What Are We Weighting For? *Journal of Human Resources*, *50*(2), 301–316.

StataCorp. (2017). *Stata Item Response Theory Reference Manual: Release 15.* Stata Press.

State Council of Education Research and Training. (2011). *Draft Position Paper on Early Childhood Care and Education (ECCE).* SCERT (India).

Stein, A. D., Melgar, P., Hoddinott, J., & Martorell, R. (2008). Cohort Profile: The Institute of Nutrition of Central America and Panama (INCAP) Nutrition Trial Cohort Study. *International Journal of Epidemiology*, *37*(4), 716–720.

Tao, F.-B., Hao, J.-H., Huang, K., Su, P.-Y., Cheng, D.-J., Xing, X.-Y., Huang, Z.-H., Zhang, J.-L., & Tong, S.-L. (2013). Cohort Profile: The China-Anhui Birth Cohort Study. *International Journal of Epidemiology*, *42*(3), 709–721.

Tigistu, K. (2013). Professionalism in Early Childhood Education and Care in Ethiopia: What Are We Talking About? *Childhood Education*, *89*(3), 152–158.

Tiruneh, D. T., Hoddinott, J., Rolleston, C., Sabates, R., & Woldehanna, T. (2021). *Understanding Achievement in Numeracy Among Primary School Children in Ethiopia: Evidence from RISE Ethiopia Study* (RISE Working Paper No. 21/071). RISE.

Tough, P. (2012). *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. Houghton Mifflin.

Tu, Y.-K., & Law, G. R. (2010). Re-examining the Associations Between Family Backgrounds and Children's Cognitive Developments in Early Ages. *Early Child Development and Care*, *180*(9), 1243–1252.

Tuan, T., Lan, P. T., Harpham, T., Huong, N. T., Thach, T. D., Tod, B., Dua, T. T., & Ha, N. T. V. (2003). *Young Lives Preliminary Country Report: Vietnam.* Young Lives.

UN. (2018, September). MDG Data. https://millenniumindicators.un.org/unsd/mdg/Data.aspx

UN. (2023, July). Net Enrolment Rate in Primary Education: Methodology Sheet. https://www.un.org/esa/sustdev/natlinfo/indicators/methodology_sheets/education/net_enrolment.pdf

UNICEF. (2012). *The State of the World's Children 2012: Children in an Urban World.* UNICEF.

UNICEF. (2015). *Fairness for Children: A League Table of Inequality in Child Well-Being in Rich Countries.* UNICEF Office of Research – Innocenti.

UNICEF. (2023). Early childhood education. (accessed 3 May 2023).

UNICEF, World Health Organization, & World Bank. (2023, July). Prevalence of stunting, height for age (% of children under 5). https://data.worldbank.org/indicator/SH.STA.STNT.ZS

United Nations Educational Scientific and Cultural Organization. (2006). *Viet Nam: Early Childhood Care and Education (ECCE) programmes*. UNESCO.

United Nations Educational Scientific and Cultural Organization. (2008). *Overcoming Inequality: Why Governance Matters* (EFA Global Monitoring Report). Oxford University Press and UNESCO Publishing.

United Nations Educational Scientific and Cultural Organization. (2009). *EFA Global Monitoring Report 2010: Reaching the Marginalised*. UNESCO Publishing and Oxford University Press.

United Nations Educational Scientific and Cultural Organization. (2010). *Early Childhood Care and Education. Regional Report. Asia and the Pacific*. UNESCO.

United Nations Educational Scientific and Cultural Organization. (2011). ISCED Mappings (2011 versions). https://isced.uis.unesco.org/data-mapping/

United Nations Educational Scientific and Cultural Organization. (2020, November). Glossary. UNESCO UIS. Upper Secondary Education (ISCED 3). http://uis.unesco.org/en/glossary

United Nations Educational, Scientific and Cultural Organization. (2023, July). School enrollment, primary (% net). https://data.worldbank.org/indicator/SE.PRM.NENR

United Nations Inter-agency Group for Child Mortality Estimation. (2023, July). Under-Five Mortality Rate - Total. https://childmortality.org/data

United Nations University World Institute for Development Economics Research. (2022, November). World Income Inequality Database.

Van De Walle, E. (1990). The IFORD Surveys. In J. Vallin, S. D'Souza, & A. Palloni (Eds.), *Measurement and Analysis of Mortality*. Clarendon Press.

Vandell, D. L., Burchinal, M., Vandergrift, N., Belsky, J., Steinberg, L., & National Institute of Child Health Human Development Early Child Care Research Network. (2010). Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development. *Child Development*, *81*(3), 737–756.

Vandemoortele, M. (2014). Measuring Household Wealth with Latent Trait Modelling: An Application to Malawian DHS Data. *Social Indicators Research*, *118*, 877–891.

Vandemoortele, M. (2018). *Inequality in Attainment From Early Childhood to Adolescence: Longitudinal Evidence from Ethiopia* (Working Paper No. 177). Young Lives.

Vandemoortele, M., & Bird, K. (2011). *Viet Nam's progress on economic growth and poverty reduction: Impressive improvements*. Overseas Development Institute (ODI).

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., & Sachdev, H. S. (2008). Maternal and child undernutrition: consequences for adult health and human capital. *Lancet*, *371*, 340–357.

Victora, C. G., & Barros, F. C. (2006). Cohort Profile: The 1982 Pelotas (Brazil) Birth Cohort Study. *International Journal of Epidemiology*, *35*(2), 237–242.

Victora, C. G., Hallal, P. C., Araújo, C. L. P., Menezes, A. M. B., Wells, J. C. K., & Barros, F. C. (2008). Cohort Profile: The 1993 Pelotas (Brazil) Birth Cohort Study. *International Journal of Epidemiology*, *37*(4), 704–709.

Waldfogel, J., Kwon, S. J., Wang, Y., Washbrook, L., Casoni, V. P., Olczyk, M., Schneider, T., Panico, L., Solaz, A., Weinert, S., Volodina, A., de la Rie, S., Keizer, R., Nozaki, K., Yamashita, J., Kameyama, Y., & Akabayashi, H. (2023). Inequalities in Resources for Preschool-Age Children by Parental Education: Evidence from Six Advanced Industrialized Countries. *European Journal of Population*, *39*(37).

Wamani, H., Tylleskär, T., Åstrøm, A. N., Tumwine, J. K., & Peterson, S. (2004). Mothers' Education but not Fathers' Education, Household Assets or Land Ownership is the Best Predictor of Child Health Inequalities in Rural Uganda. *International Journal for Equity in Health*, *3*(9).

Washbrook, E., & Lee, R. (2015). Beyond the Feinstein Chart: Investigating Differential Achievement Trajectories in a US Cohort. *Longitudinal and Life Course Studies*, *6*(3), 359–368.

West, A. (2016). *Nursery Education and Reducing Social Inequalities: An International Comparison* (tech. rep.). Conseil national d'évaluation du système scolaire.

Wetterhall, S. F., & Noji, E. K. (1997). Surveillance and Epidemiology. In E. K. Noji (Ed.), *The Public Health Consequences of Disasters*. Oxford University Press.

Wilson, I., Huttly, S. R. A., & Fenn, B. (2006). A Case Study of Sample Design for Longitudinal Research: Young Lives. *International Journal of Social Research Methodology*, *9*(5), 351–365.

Woldehanna, T., Mekonnen, A., & Alemu, T. (2008). *Young Lives: Ethiopia Round 2 Survey Report*. Young Lives.

Woodhead, M., Ames, P., Vennam, U., Abebe, W., & Streuli, N. (2009). *Equity and Quality? Challenges for Early Childhood and Primary Education in Ethiopia, India and Peru* (Working Paper No. 55). Bernard van Leer Foundation.

World Bank. (2018, September). World Development Indicators (Last updated: 19 Sep 2018). https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

World Bank. (2023, July). GDP per capita (current US$). https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

World Health Organization. (2018, September). WHO Strengthens Sentinel Surveillance for Vaccine Preventable Diseases. https://www.afro.who.int/news/who-strengthens-sentinel-surveillance-vaccine-preventable-diseases

Young Lives. (2002a). *Peru Community Questionnaire (Round 1)*. Young Lives.

Young Lives. (2002b). *Vietnam Community Questionnaire (Round 1)*. Young Lives.

Young Lives. (2006). *Ethiopia Household Questionnaire (Round 2)*. Young Lives.

Young Lives. (2009). *Procedures and Instructions for the Administration of Cognitive Development and Achievement Instruments – Round 3 – Fieldworker's Manual (Peru)*. Young Lives.

Young Lives. (2011a). *Data Management: Cleaning, Storage and Archiving* (Mathods Guide). Young Lives.

Young Lives. (2011b). *Piloting: Testing Instruments and Training Field Teams.* (Methods Guide). Young Lives.

Young Lives. (2011c). *The Longitudinal Survey* (tech. rep.). Young Lives.

Young Lives. (2013a). *Young Lives Cognitive Tests Fieldworker Manual (Round 4 - Ethiopia)*. Young Lives.

Young Lives. (2013b). Younger Cohort Child Cognitive Tests (Round 4).

Young Lives. (2014a). *Young Lives Survey Design and Sampling in Ethiopia: Preliminary Findings from the 2013 Young Lives Survey (Round 4)*. Young Lives.

Young Lives. (2014b). *Young Lives Survey Design and Sampling in India: Preliminary Findings from the 2013 Young Lives Survey (Round 4)*. Young Lives.

Young Lives. (2014c). *Young Lives Survey Design and Sampling in Peru: Preliminary Findings from the 2013 Young Lives Survey*. Young Lives.

Young Lives. (2014d). *Young Lives Survey Design and Sampling in Viet Nam: Preliminary Findings from the 2013 Young Lives Survey (Round 4)*. Young Lives.

Young Lives. (2015). *Section 2: What Can Comparative Country Research Tell Us About Child Poverty?* (A Guide to Young Lives Research, Young Lives), Young Lives.

Young Lives. (2016). *Young Lives Fieldworker Manual (Round 5 – Ethiopia)*. Young Lives.

Young Lives. (2017a). *Education and Learning: Preliminary Findings from the 2016 Young Lives Survey (Round 5): Ethiopia* (Factsheet). Young Lives.

Young Lives. (2017b). *Education and Learning: Preliminary Findings from the 2016 Young Lives Survey (Round 5): United Andhra Pradesh* (Factsheet). Young Lives.

Young Lives. (2017c). *Section 6: Child, Household and Community Surveys*, Young Lives.

Young Lives. (2017d). *Young Lives Survey Design and Sampling (Round 5) United Andhra Pradesh*, Young Lives.

Young Lives. (2018a). *Young Lives Survey Design and Sampling (Round 5) Ethiopia*, Young Lives.

Young Lives. (2018b). *Young Lives Survey Design and Sampling (Round 5) Peru* (tech. rep.). Young Lives.

Young Lives. (2018c). *Young Lives Survey Design and Sampling (Round 5) Viet Nam* (tech. rep.). Young Lives.

Zhang, L., & Lee, K. A. (2011). Decomposing Achievement Gaps Among OECD Countries. *Asia Pacific Education Review*, *12*(3), 463–474.

Appendix A

Annex: Chapter 2

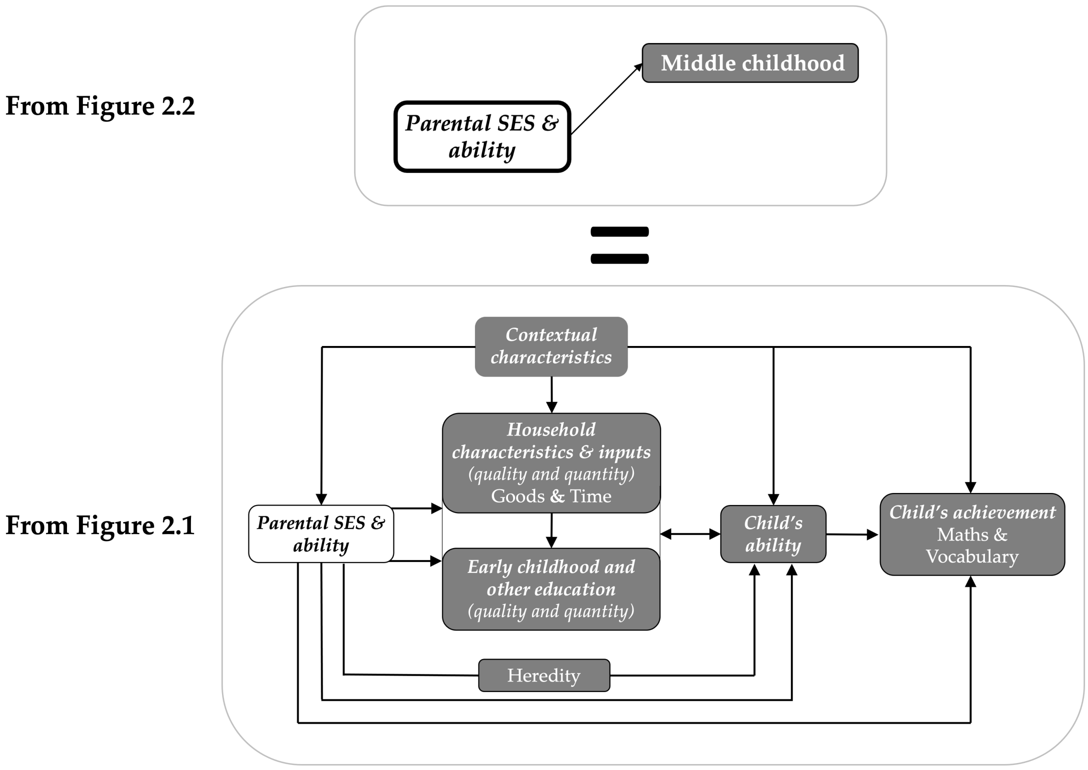


Figure A.1: Linking the amended Haveman and Wolfe (1995) (Figure 2.1) and Ermisch, Jäntti, et al. (2012) (Figure 2.2) frameworks

Figure A.1 illustrates how the two frameworks are linked. To integrate the longitudinal perspective into the amended Haveman and Wolfe (1995) framework (Figure 2.1), the box in Figure 2.2 labelled *Parental SES & household 'ability'* represents the homonymous box in Figure 2.1. All the boxes following the *Parental SES & household characteristics* box in the amended Haveman and Wolfe (1995) framework (Figure 2.1) are intermediary outcomes, and are included in the life stage box. The content of these boxes change according to the life stage, as discussed earlier. So, for example, only preschool education is relevant for early childhood, but primary is also relevant for middle childhood.

Appendix B

Annex: Chapter 3

B.1 Literature Review: Small-n Cross-Country Comparisons

This literature review underpins section 3.4. The aim of this review is to inform my comparative approach.

Here, I review small-n cross-country comparative studies that examine the link between parental socioeconomic status (SES) and academic achievement and early childhood education (ECE) and achievement. I define "small-n" as comparing ten countries or fewer. When selecting these articles, I restricted myself to articles published in peer-reviewed journals, books and Young Lives working papers (whereas, in the subsequent empirical chapters, I rely on published, not grey literature or working papers). I also restricted myself to papers that employ the same regression approach as mine, which is to run separate regression analyses for each country. In fact, only 2 of the 17 pieces of literature I identified did not employ this regression approach (specifically Malmberg et al., 2011; Montie et al., 2006). Authors from both articles pool their data and employ a multi-level model.

To identify this literature, I used Google Scholar and the LSE Library online catalogue using several key words interchangeably in my search, and identified additional references among the references in the initial search results.

Here, I provide a short summary of each piece of literature identified. I summarise: the objective, which countries the researchers examine, their comparative approach, their analytical approach (i.e. the regression approach), and whether they consider limitations or conduct robustness checks.

I have grouped the papers according to the following categories, listing the references included in each category. Those that:

- neither consider the countries' context nor the limitations of a comparative study – Das et al. (2022), Duncan et al. (2012), Georgiadis (2017), Jerrim and

Micklewright (2012, 2014), and Lopez-Boo (2016);

- do consider each country's context but not the limitations – Jaramillo and Tietjen (2001), Lockheed et al. (1989), Ojala and Talts (2007), and Singh (2014a);

- do not consider countries' contexts but do consider the limitations – Fernald et al. (2012);

- consider both the countries' contexts and the limitations – Blanden et al. (2012), Bradbury et al. (2012, 2015, 2019), Crookston et al. (2014), Linberg et al. (2019), Magnuson et al. (2012), and Schady et al. (2015).

B.2 Works that neither consider countries' context nor the limitations of a comparative study

Duncan et al. (2012) investigate: across four countries, do children from better-educated parents perform better in tests (maths and reading); does higher achievement in maths and reading predict completed schooling of the child; how much of intergenerational inequality is explained by these relationships? They use five data sets from four different countries – the United Kingdom, the Unites States, Sweden and Finland. Data from the latter three are from cities. The data sets do not overlap identically on children's age, nor the gap between measurements, nor are achievement measures captured in all studies. Limitations related to city versus country comparisons and complex data points are not discussed. The only limitation they discuss is regarding "bias owing to the limited variability in [the] three community data sets" (Duncan et al., 2012).

The authors take the same approach as I do, running separate regression analyses for each country. They do not provide any country contexts. Rather they discuss mechanisms linking parental education and middle childhood or adolescent skills, and middle childhood or adolescent skills with adult achievement.

When interpreting regression coefficients of achievement, they use a standardised outcome in order to have unitary standard deviations. Descriptive comparisons are made regarding the magnitude of the coefficients, how they change over time and their statistical significance. When interpreting regression coefficients where parental education predicts maths and reading scores, the authors provide a descriptive interpretation without incorporating country contexts, for example: "[t]here is little indication that Swedish gradients are flatter than gradients in the United States or United Kingdom" (Duncan et al., 2012, p. 223).

Jerrim and Micklewright (2012) compare the socioeconomic gradient in children's cognitive ability at ages 10 and 15 years across nine countries. Their data is cross-sectional at each age. They compare different children who completed two different surveys administered in different years. They compare the gradients at the two ages and whether they differ significantly across countries, then measure changes in these gradients across countries. They aspire to comment on why there may be similarities and differences and "especially whether the changes seem related to institutions that kick in between the two ages" (p. 263). Country contexts are not provided, with the exception of a line on Germany to illustrate a point. Nor are limitations of cross-country analysis discussed. A discussion of measurement issues across countries do not arise, rather discussion on measurement focuses on how best to measure parental SES and how to standardise test scores for analysis. Coefficients are juxtaposed, but authors mostly focus on comparing averages at each age, rather than on country-specific trajectories, for example: "[t]he differences average 0.88 of a national standard deviation at age ten..." (p. 275). Authors are not able to draw any conclusions related to their research questions as the results they obtain "are not very robust to changes in the specification of the simple regression models that [they] use to explore socioeconomic gradients" (p. 280). Given the instability of results and the number of countries investigated by authors do not focus on country specific trajectories and write, "[a]s far as individual countries are concerned, it is not always easy to summarize..." (p. 275). While I am not faced with the data limitations they face (I have panel data and a rich set of control variables, as well as a rich set of variables on parent's SES), I find their desire for rigour appealing.

Jerrim and Micklewright (2014) conduct a similar analysis to their 2012 chapter, comparing ten countries with the aim of exploring and explaining instability in their results when regression specifications change. For simplicity, I will not delve into the detail of this article. Their comparative approach is similar to that of league tables, where country contexts are not taken into consideration and coefficients are juxtaposed only to in search of a pattern across countries.

The Young Lives working paper by Georgiadis (2017) investigates the association between child development indicators (achievement, height-for-age and non-cognitive skills) and a wide range of parental background dimensions across different stages of childhood. Georgiadis states that "[t]o his knowledge, there is no other study, to date, from developing countries, that has simultaneously examined the associations of such a wide set of background factors with children's human capital outcomes and how these associations may differ across national and cultural contexts and over children's life cycles" (pp. 6–7 – ironic given two other papers making similar claims, specifically Singh, 2014a and Lopez-Boo, 2016). He uses the Young Lives data (Rounds 2 and 3). His analytical approach, like mine, entails running separate regressions in each country for each age group. The comparative strategy is descriptive and long, containing multiple regression analyses and figures, with no country context provided nor taken into account. The author does not consider the limitations of a cross-country study nor of his methods. For example, he directly compares outcomes from children in different cohorts as if they were from the same cohort; he also compares raw vocabulary test scores without discussing potential problems of doing so across countries.

Lopez-Boo (2016) investigates the relationship between socioeconomic background and receptive vocabulary among Young Lives children, as well as the potential mediators of that relationship. She claims her article to be the "first longitudinal multicontinent comparison of SES gradients in cognitive development for young children in the developing world over critical periods of their lives" (p. 506), as well as "among the first to explore the relative importance of different mediators in different settings and ages with a validated measure of child development" (p. 506). She uses Rounds 2 and 3 and selects only those children from the younger cohort who took their PPVT tests in the majority language. Lopez-Boo conducts a mediation analysis. Like I do, she runs separate regressions for each country. No context is provided and on the whole she conducts a descriptive interpretation of results across countries. For example stating that "although differences in early language development by SES are present in all countries, they arise more starkly in Peru" (p. 506) or "there is heterogeneity across countries and even within countries (across ages) in the relative importance of different mediators" (p. 506) or the "SES gradient persists but is also significantly reduced when controlling for a large number of highly important variables in all countries, except India (age 5) and Vietnam (age 8) where the gradient disappears" (p. 506). Only when discussing results on ECE is some context provided in parenthesis: "in Ethiopia (where preschool attendance is lowest) and Vietnam (where preschool attendance is highest but probably heterogeneous in quality)" (p. 507). The problem with this type of context provision is she does not ground it in a broader context, so it can be perceived as conveniently selected contextual information to support results. She does not discuss limitations of cross-country analysis, only limitations of the data she utilises.

Das et al. (2022) examine associations between SES, test scores at 12 years old and college attendance at ∼22 years old in five countries. While this aim does not align with my research question, as part of their analysis they examine when SES gaps in achievement emerge. For this, they report (in their Table 2, Panel B) SES gaps in maths scores from eight to 15 years old. In their broader discussion and specifically about the reported SES gaps in maths scores, the authors do not discuss country context nor do they discuss limitations of a cross-country comparison. They do engage in robustness testing (around the measure of SES they use).

B.3 Works with country context, not limitations

Ojala and Talts (2007) compare between Helsinki (Finland) and Tallinn (Estonia) how children's learning achievements differ or are similar across nine areas of learning. Rather than test scores, they use teachers' evaluations at the end of ECE. For context, the authors provide a rich description of values underlying each country's approach to ECE, their practical approaches, and how they compare with each other. This context is also taken into account when results are interpreted, for example using the following phrases: "...our results showed national as well cross-national, variation.... Beside differences, the results also showed similarities. ... There were probably many factors for explaining cross-cultural differences...." (p. 217), which they go on to explain. This rich contextual information is a useful example for how I can provide a context on ECE and then incorporate it into my interpretation. The authors do not discuss the limitations of cross-country comparisons.

In his Young Lives working paper, Singh (2014a) considers the age at which gaps in math scores emerge between different groups, how they evolve over time and what the proximate determinants of this divergence are. Singh states that his paper is "the first analysis of the emergence and evolution of gaps in cognitive achievement across countries, from the age of 5 to 15 years, using internationally comparable child-level panel data" (p. 2) – although Crookston et al. (2014) did something very similar in the *BMC Pediatrics* journal in the same year). To conduct analysis on this age range, he has to group the younger cohort (born in 2001) with the older cohort (born in 1994). This approach can be complicated by cohort effects, especially if one doesn't account for them. To measure achievement, he uses IRT scores, and in doing so he implicitly considers the limitations of comparing test results across countries. He does not consider this nor other limitations of comparative analysis explicitly. He employs various models, including one where a separate regression analysis is run for each country (albeit only at one age); his longitudinal analysis pools all countries. His comparative approach is descriptive, for example interpreting results as follows: "The difference between the two sets of countries seems to be a difference in the intercepts and not the slopes" (p. 13). He provides a half-paragraph of country contexts – specifically on enrolment and dropout rates across countries – and does not incorporate it into the interpretation of results.

Jaramillo and Tietjen (2001) compare the association between ECE and children's achievement in Cape Verde and Guinea. For each country, they provide a short paragraph on the following main topics: ECE access and distribution; ECE provision; ECE personnel; ECE students. Then, they give descriptive tables on the characteristics of ECE centres in each country. The authors also run separate regression analyses for each country. They are not clear about the specifics of their regression model, nor do they correctly interpret the coefficients. For example, when interpreting the coefficient for gender authors write "[p]reschool will increase a girl's score in Guinea by more than three points, but the small increase in Cape Verde is not statistically significant" (p. 25), rather it should say that girls perform better than boys controlling for all other variables, including ECE. Further, there is no discussion on the potential role of unobserved effects and how they may differ between countries. The authors do not consider limitations of cross-country comparisons, rather they focus on limitations of their study design but do not link it to complications that may arise when interpreting results from distinct countries. They draw causal conclusions where, given their methods and data, a causal link cannot be demonstrated, for example: "…preschool improve[s] all children's cognitive development scores…" (p. 29) or "[c]hildren who participate in more than one year of preschool score higher on cognitive development tests and gain greater language skills" (p. 30).

Lockheed et al. investigate whether the role of family background on achievement "can be adequately tested in the Third World by relying on conventional measures of class, borrowed from industrialized societies" (p. 240). They report on two separate studies that use different data, different research questions and different variables and methods. In the Thailand study, they first examine the association between family background and students' mathematics achievement, then gains in achievement, using a panel data set. They also investigate how family background is associated with mediating variables: students' educational expectations, perceived parental encouragement, attitudes and effort. They find that "material facets of social class help explain variations in these mediating motivational factors" (p. 253). The Malawi study examines the association of family background, as defined by material facets of social class, with achievement among a sample of 105 primary school students. This data set is cross-sectional. This study shows material measures of wealth such as housing status, access to electricity and child labour explain large portions of the variance in achievement. They conclude that, "[b]y utilizing narrow and generally Western indicators of class (parents' occupational and educational statuses), previous analysts may have artificially underestimated the effects of social-class background on students' achievement"(p. 254).

In terms of incorporating the countries' contexts Lockheed et al. provide scant information for the Thailand study, but more for the Malawi study. For example, in the case of Thailand, the authors cite research that shows that "parents' aspiration for their children's education was the most important predictor of the educational achievement of sons, while parental landholding... was the most important predictor of their daughter's educational attainment" (p. 242). When a result is unexpected in Malawi, they draw more substantially on context:

> Curiously, the percentage of families that came from farming backgrounds (as estimated by the headmasters) was positively related to achievement. The direction of this effect is difficult to explain. It may indicate that parents who own their land, compared to farmhands on estates or unskilled urban workers, provide stronger encouragement to their children to achieve in school. This variable may also be a stronger proxy for attendance at rural schools, which are typically smaller and display a lower student-to-teacher ratio than do urban schools (p. 251).

This can be seen as conveniently selected contextual information. When authors discuss the limitations of their study, they mostly focus on individual study limitations, rather than the limitations of comparing across countries. No robustness checks are conducted.

B.4 Works without country context, with limitations

Fernald et al. (2012) examine gaps in cognitive development in young children across four developing countries: India, Indonesia, Peru and Senegal. They use data from a baseline survey of households in a water and sanitation trial, collected from 2008 to 2009 from children aged three to 23 months. These are cross-sectional data. First, the authors conduct regression analyses for each country, then conduct both nonparametric and semiparametric analyses to examine whether gaps are wider or narrower among older children than among their younger counterparts. Second, they do a mediational analysis, investigating the contribution of 'home environment' to the relationship between parental wealth or education. Authors provide descriptive statistics for each country in table format, including information on each country's GNI, life expectancy, prevalence of stunting in children, prevalence of wasting in children, adult literacy rate, HDI rank, U5MR. Their interpretation is descriptive and does not incorporate any of the country context. Fernald et al. acknowledge that their study is severely limited by not incorporating "economic, cultural, political, or historical information about any of the countries [so they are, therefore,] unable to comment on any country-related issues that may explain" their findings (p. 17276).

Fernald et al. find that, across countries, "[b]eing in the fifth wealth quintile compared with the first quintile conferred a significant advantage to EASQ scores" (p. 17275) and that gaps among older children are wider than among younger children for India and Indonesia, not in the other countries. They also find that having a highly educated mother compared with having one with no formal education is associated with significantly better performance on the EASQ and that gaps among older children are wider than among younger children for India, Indonesia and Peru. In their mediational analysis, they find that parental home stimulation variables explain between 18 (Indonesia) to 37 (Senegal) per cent of wealth effect, and 12 (Senegal) to 31 (India) per cent of the education effect on EASQ scores. They do conduct a robustness check on their age categories.

Fernald et al. use the cross-sectional data in a longitudinal manner: examining whether gaps are wider among older children, when compared to their younger counterparts. Their terminology implies they are using longitudinal data: "differences increased with age" and "[w]ealth and education gradients also increased with age" (p. 17275). The authors acknowledge the limitation of being unable to examine the causal mechanisms underpinning this link and suggest using longitudinal data to do this – even though longitudinal data would also not provide an opportunity to establish causal links.

B.5 Works that consider both the countries context and the limitations

Blanden et al. (2012) compare the relationship between parents' SES and achievement in the UK and Australia. They assess whether differences are significant and whether the relationship between SES and achievement diminishes or increases over the child's early life course. The authors' ultimate goal is to identify policies that "appear to be most effective at improving social mobility" (p. 143).

Blanden et al. justify their country selection, then briefly discuss policy differences in early childhood. The comparisons are descriptive and include transition matrices of cognitive test scores of the top and bottom SES quintile groups and transition matrices in test scores comparing where children who were in the best quartile at age five are at age seven. The authors also juxtapose information on mean cognitive outcomes by age and parental education, compare persistence of low and high scores, and replicate Feinstein's curves (see Feinstein (2003)) for each country.

Based on these descriptive comparisons, Blanden et al. provide plausible explanations for these differences, drawing on the context provided earlier in the chapter using their cautious terminology: "It is possible that…" (p. 159); "an alternative explanation is one of..." (p. 159); "... appear to be..." (p. 143); "...a great deal of research remains to be done in this area..." (p. 160); "[a]lthough these findings are consistent with the overall body of research...we can only speculate here on the reason for these differences" (p. 160). Finally, authors do acknowledge a limitation of cross-country comparison when they state that their article "is a comparison of two countries, so it is difficult to contextualise the differences between them" (p. 160).

Magnuson et al. (2012) investigate whether, in the US and the UK, early SES gradients in skills widen, remain constant or diminish as children move through the school year. They employ same analytical strategy as mine and run separate regression analyses for each country and at each age. In terms of their comparative strategy, explain that they chose two similar countries in terms of inequality and social mobility. Next, they provide a very brief context on the education systems in each country. When interpreting results, they juxtapose regression coefficients and descriptively interpret differences and similarities. Magnuson et al. first interpret country regressions separately, then discuss similarities and differences between countries. Country contexts are only lightly drawn upon when discussing comparisons. An example of a comparative result is as follows: "Our findings confirm that large and meaningful gaps in both cognitive skills and behaviors are apparent in school entry.... Our results ... suggest that such sorting [in UK secondary schooling] is disequalizing, increasing the gradient present when children left primary school, and to a greater extent than occurs in the United States" (pp. 256–257). Limitations of cross-country comparisons are briefly discussed by authors, stating that: while "[school] factors differ across countries, we can not attribute all cross-country difference to the influence of schools" (p. 256); and that "[a]lthough results also differ by country with regards to gradients in socioemotional development, here we hesitate to draw strong conclusions given the differences in measurement across the two countries" (p. 257). This chapter provides another good model for a small-n comparative study.

Schady et al. (2015) compare the association between SES and children's vocabulary scores across five Latin American countries. Their aim is descriptive and their analytical approach is like mine – a separate regression analysis for each country. They provide brief context for the countries, focusing on comparing and contrasting GDP per capita, grades or completed schooling of adults and the Gini coefficient. When interpreting the results, the authors use this context, albeit lightly, to thicken the descriptive comparisons between countries.

Schady et al. find a large gap between children from wealthy and poor backgrounds. Three of their data sets are longitudinal and, from these, they find that the gap does not widen. They also discuss the limitations of their study, such as using different data sets, some of which are not nationally representative and using measures of SES that differ across countries, looking at a single outcome, and the need for further econometric or qualitative work to better understand the mechanisms through which wealth affects children's achievement, and how this varies from country to country.

Crookston et al. (2014) use Young Lives data to examine the relationship between SES, child growth and changes in cognitive achievement scores in adolescents. Their aim, like mine, is descriptive. The also employ the same analytical approach as I do, running separate regression analyses for each country. They provide a very brief country context (half a paragraph long) that discusses difference between GNIs, differences in underweight prevalence and primary school completion rates across countries. Subsequently, they descriptively discuss the differing importance of wealth, maternal and paternal schooling, child growth, child's height-for-age on scores. They do not draw on any country-specific information to try to explain the differences. They also discuss several limitations of this cross-country study.

Bradbury et al. (2015) examine the U.S. achievement gap in comparison against those of Australia, Canada and the UK. They selected these countries purposively for their similarities. Their aim is descriptive. They take into account countries' context, especially that of the US, when interpreting results. They do not provide a separate country context section, but rather integrate the context into their interpretation. Most of their analysis involves separate regression analyses for each country. They also discuss limitations of their cross-country study, specifically of comparing measures across countries and discuss how they address these.

Bradbury et al. (2012) document the emergence of inequality in the early years (at five years old) across four countries: Australia, Canada, the UK and the US. They run separate regression analyses for each country. The approach of incorporating the countries' context is rather comprehensive, providing some pertinent context at the beginning of their chapter, then integrating it into the interpretation of results and the conclusion.

First, Bradbury et al. highlight the similarities between countries that "often look to each other for policy models and reforms" (p. 89). Then, to facilitate a discussion comparing countries across key areas of interest, they summarise the similarities by tabulating inequality, child poverty, per capita social expenditure on children aged six as proportion of median working-age income and public expenditure as share of total health expenditure. This initiates a discussion on the performance of these countries according to these areas, which leads on to a detailed discussion on the differing policy approaches to early childhood care and education. Each country receives approximately a paragraph of attention. Before delving into the analysis, Bradbury et al. provide some descriptive statistics for each country illustrating key demographic differences between the countries that cannot be overlooked when comparing across them. Authors acknowledge that "[t]hese differences are intrinsic features of the countries in question, and it is not clear how to interpret results that adjust them away" (p. 103) – a challenge that I sympathise with. When it comes to family characteristics, they again explain that these "are an intrinsic feature of the countries, and it is not clear that estimates should adjust for them" (p. 103). They therefore choose both to "adjust these differences away" by controlling for various racial-ethnic-nativity and family characteristics, as well as "not to adjust the differences away", and simply examine the SES gradient without any controls. They do the former because, "in a descriptive sense, it is useful to know to what extent the SES gradients change if these factors are held constant" (p. 103).

Bradbury et al. acknowledge the limitations of cross-country analyses (e.g. comparability of measures across countries) and are cautious in their interpretations, highlighting that "[a]lthough ascribing the variation in outcomes to particular policies or institutions is difficult, our results do complement other indicators of social inequality and mobility, and offer a starting point to reflect on the particular accomplishments and challenges in each country" (p. 112). They also recognise that "children experience very different policy contexts across the four countries in four policy domains that determine the amount of time parents have for nonmarket activities associated with family life. . . " (p. 113). Further, while providing a plausible storyline explaining some differences in their results, they emphasise that "exploring the role of these policy contexts in early inequalities is an important challenge for future research" (p. 113), and offer a hypothesis to test in future research. The authors conclude with an emphasis on the value of their type of research: "[o]ur analysis is descriptive, but good description is the first step to informed policy discussion and hypotheses about causal relationships." (p. 114).

Bradbury et al. (2019) use income groupings defined in absolute terms (using the US income distribution) to compare achievement gaps across four countries. The aim is to ascertain how much of the SES achievement gap is a result of differences in the income distribution. The authors find that the same differences in absolute differences translate into wider gaps in children's test scores in the US, compared with the UK, Australia and Canada. They show that this can be explained in difference with non-monetary differences between groups (e.g. maternal education, family structure, maternal age, parental nativity, maternal employment patters and exposure to ECE). This article is written in a conversational and accessible manner and integrates the country contexts neatly into the discussion both early in the article and also when interpreting results. The authors discuss some limitations of cross-country studies and conduct a robustness test related to cross-country comparisons (i.e. whether the taxation system in each country may skew their results).

Linberg et al. (2019) compares raw and adjusted SES gaps in achievement between Germany and the US (for children at 6/7 years old). The authors produce some unexpected findings, where raw language gaps are larger in Germany than in the US (1.25 s.d. in Germany and 0.86 s.d. in the US) while the raw gap in maths scores are similar (0.96 and 0.99 s.d. in the US and Germany respectively). When including additional controls, they find that the "main drivers of inequalities differ between countries"(p. 16), specifically the ECE regime in the US contributing to gaps, while in Germany it does not appear to be associated with a change in the SES gaps in achievement. Throughout the article, the authors intersperse relevant contextual information, discuss the limitations of a comparative study and use also run separate regression analyses for each country and juxtapose them in the results tables.

B.6 Other comparative studies that do not run separate regressions for each country

I do not include these studies in my methods chapter on cross-country comparisons as they do not involve separate regressions for each country and, therefore, offer little guidance on how to interpret my results. They are the only small-n country-comparative studies I identified where the authors do not run a separate regression analyses for each country. Rather, Montie et al. (2006) and Malmberg et al. (2011) run pooled analyses of ten countries.

Montie et al. (2006) offer a comparative study examining the links between ECE and children's language performance at age seven across 10 countries. The authors focus on process and structural characteristics of ECE. In the discussion, they do not provide nor take into consideration country context. They do discuss the limitations of cross-country comparisons both in the discussion as well as in a separate section. They do not provide each country's context nor incorporate it into results interpretation.

Malmberg et al. (2011) examine the role of the Madrasa Resource Center (MRC) on young children's cognitive development in three East African countries – specifically Kenya, Zanzibar and Uganda. They use a panel data set and their sample size is much smaller than most studies (321 across all three countries). They also investigate only a specific kind of ECE. Limitations of the study are discussed, although not the limitations of a cross-country study. The author pool their data on all countries and employ a multilevel model. In their comparison, they provide a half a paragraph of information on non-MRC ECE centres in each country and a few lines on differing age of entry into ECE across countries. This is not taken into consideration in the results. In the interpretation of results, they provide one line of context to try to explain a specific result: "A plausible explanation for the Ugandan attrition is mass-migration from preschool into primary school" (p. 130). They find that both linear and quadratic modelling of achievement fit the data well, there were some differences between countries, and classroom quality was more strongly associated in the MRC group than in the non-MRC group.

Appendix C

Annex: Chapter 4

C.1 Child-focused panel studies from low- and middle-income countries – Table

Table C.1: Child-focused panel studies from low- and middle-income countries

Name	Cohort size	Completed rounds	Sampling	Following rule	Attrition	Notable features	References
High-income countries							
MCS, UK	18,827[1]	9 months, 3, 5, 7, 11 years	Probability sample (multistage stratified random) of births between September 2000 and January 2002 with child living in UK at age 9 months and eligible to receive child benefit.	Remained in UK	31% (11 years)	A multi-disciplinary study that covers a diverse set of topics including achievement and non-cognitive development, income and poverty and demographic characteristics. Ongoing.	Connelly and Platt (2014) and Platt (2014)
Fragile Families, US	4,898	Birth, 1, 3, 5, 9 years	Probability sample (multistage stratified random) of non-marital and marital live hospital births between 1998 and 2000 in US cities with 200,000 people or more[2]	Remained in US, alive and were not adopted	31% (9 years)	Research questions include: What are the capabilities of unmarried parents, especially fathers? How do children fare and how does family structure and stability affect child wellbeing? Ongoing	CRCW. 2011, Reichman et al. (2001)
Low- and middle-income countries							

Continued on next page

Table C.1 – *Continued from previous page*

Name	Cohort size	Completed rounds	Sampling	Following rule	Attrition	Notable features	References
1. TRMI evaluation, Bangladesh	5,000	April 2012, June 2013, April 2014	Cluster RTC (multistage cluster sampling), where treatment villages belonged to each of the five alternative modalities of transfer. Within villages selected, all households with a child aged 6–24 months, poor, not receiving other benefits	n/a	3% (2 years)	Programme evaluation. Completed.	Ahmed et al. (2015) and Ahmed et al. (2014)
2. PIDI evaluation, Bolivia	1,501	November 1995–May 1996, November 1997–May 1998	Probability sample (mix of simple and stratified random) of 3 subsamples of children aged 6–72 months in each round[3]	n/a	51 % (2 years)[4]	Programme evaluation. Completed.	Behrman et al. (2004)

Continued on next page

Table C.1 – *Continued from previous page*

Name	**Cohort size**	**Completed rounds**	**Sampling**	**Following rule**	**Attrition**	**Notable features**	**References**
3. Pelotas 1982 Birth Cohort, Pelotas, Brazil	5,914	Birth, 2, 4, 23, 30 years[5]	Census of live hospital births in 1982 in Pelotas.	Remaining close to Pelotas	37% (30 years)	Inspired by the British perinatal study, now studies a wide range of demographic, socioeconomic, physiological, non-cognitive, achievement outcomes and more. Ongoing.	Horta et al. (2015) and Victora and Barros (2006)
4. Pelotas 1993 Birth Cohort, Pelotas, Brazil	5,249	Birth, 11, 15, 18 years[6]	Census of live hospital births in 1993 in Pelotas	Remaining close to Pelotas	15% (11 years) 22% (18 years).	——	Gonçalves et al. (2014) and Victora, Hallal, et al. (2008)
5. Pelotas 2004 Birth Cohort,[7] Pelotas, Brazil	4,231	Birth, 3 months, 1, 2, 4, 6 years	Census of live hospital births in 2004 in Pelotas	Remaining close to Pelotas	12% (6 years).	——	Horta et al. (2015) and Santos et al. (2014)

Continued on next page

Table C.1 – *Continued from previous page*

Name	**Cohort size**	**Completed rounds**	**Sampling**	**Following rule**	**Attrition**	**Notable features**	**References**
6. ELPI, Chile	15,175	6 months-5 years, 2.5–7 years	Probability sample (multistage stratified) of children aged 6 months–5 years in 2010 in Chile	Remained in Chile[8]	15% (2 years)	Collects data on achievement, socio-emotional and anthropometric assessments to caregiver and child. Ongoing.	Narea (2015)
7. Limache Birth Cohort Study, Limache, Chile	1,232	Birth, 1, 22-28, 32-38 years[9]	Probability sample (multistage stratified) of children aged 6 months–5 years in 2010 in Limache	Remained in vicinity	35% (10 years)[10]	Assesses whether improvement in living conditions has had an impact on the pattern of chronic diseases in adults. Ongoing.	Amigo et al. (2014)
8. China-Anhui Birth Cohort Study, China	16,766	Prenatal: <12 weeks;[11] post-natal: 0–6, 7–12 months[12]	Non-probability sample (judgment) of 6 cities from 3 regions of Anhui province, then a census of women pregnant November 2008–October 2010[13]	Singleton live births in same hospital as antenatal appointment	20% (when children turned 12 months)	Examines effects of maternal environmental exposures on birth outcomes and children's development. Ongoing.	Tao et al. (2013)

Continued on next page

Table C.1 – *Continued from previous page*

Name	**Cohort size**	**Completed rounds**	**Sampling**	**Following rule**	**Attrition**	**Notable features**	**References**
9. INCAP Nutrition Trial Cohort Study, Guatemala	1,992	0–7, 11–26, 26–41, 29–44 years[14]	Cluster RTC; non-probability sample (judgement) of 2 pairs of villages, where treatment is nutritional intervention; census of children eligible to have supplement intakes during trial, aged < 7 years in 1969 and births 1969 to September 1977	Remained in Guatemala	36% (11 years) 40% (26 years)	Collects data three generations: index children, their parents and their offspring. Ongoing.	Ramirez-Zea et al. (2009) and Stein et al. (2008)
10. APCAPS, Andhra Pradesh, India	4,338	Birth, 13–18, 19–22, 20–25 years	Cluster RTC; non-probability sample (judgement) of 15 intervention and 14 control villages, where the treatment is a nutritional intervention; census of births between 1978 and 1990 in villages	Remained in sites	75% (18 years) 60% (22 years) 69% (25 years)	A follow-up of the Hyderabad Nutrition Trial (1987-1990) to study long-term effects of early-life under-nutrition on risk of cardiovascular disease. Ongoing.	Kinra et al. (2014)

Continued on next page

Table C.1 – *Continued from previous page*

Name	Cohort size	Completed rounds	Sampling	Following rule	Attrition	Notable features	References
11. Mysore Parthenon Birth Cohort, Mysore, India	830[15]	Birth, 1, 2, 3, 4, 5 years, then every 6 months until 16 years	Census of all pregnant women between June 1997 and August 1998 in 1 hospital in Mysore	Delivered baby at hospital and remained in surrounding area	35% (9.5 years) 34% (13.5 years)	Examines long-term effects of maternal glucose tolerance and nutritional status on children's cardiovascular diseases risk factors. Ongoing.	Krishnaveni et al. (2015)
12. New Delhi Birth Cohort, New Delhi, India	8,181	Birth–1 year (every 3 months); 1–20 years (every 6 months); 29–32 years; 37–40 years	Census of all live births conceived between 1969 and 1977 in a defined area of New Delhi	Remained in New Delhi	55% (11 years) 86% (40 years)	Predominantly middle-class. Ongoing.	Bhargava et al. (2004), Huffman et al. (2011), and Richter et al. (2012)

Continued on next page

Table C.1 – *Continued from previous page*

Name	Cohort size	Completed rounds	Sampling	Following rule	Attrition	Notable features	References
13. Jamaican 1986 Birth Cohort Study, Jamaica	8,567	6 weeks–3 months, 11–12, 15–16, 18–20 years	Census of all live births between September and October 1986 in Jamaica	Remained or moved into 2 of 4 south-east region parishes	83% (11 years) 91% (18 years)	Collects data on health, nutrition, social environment, school, jobs, emotions, behaviours, achievement, demographic data, social behaviour, health status and biological samples.[16] Ongoing.	McCaw-Binns et al. (2011)
14. Mauritius Longitudinal Child Health Study, Mauritius	1,795	3, 8 11, 17, 23, 28, 35, 40 years	Non-probability sample (judgement and convenience) of two major towns. Census of all 3-year-old children in 1969 in these towns	Remained in vicinity	32% (8 years) 57% (40 years)	Aims to identify early risk factors for later psychopathology, as well as forms of early primary prevention. Ongoing.	Raine et al. (2010)

Continued on next page

Table C.1 – *Continued from previous page*

Name	Cohort size	Completed rounds	Sampling	Following rule	Attrition	Notable features	References
15. CLHNS, Cebu, Philippines	3,080	Birth to 2 years (every 2 months), 8, 11, 15, 19, 22, 24, 29 years	Probability sample (cluster) of women pregnant between May 1983 and April 1984 in Cebu[17]	Remained in sample area	30% (11 years) 41% (24 years)	Original aim to study infant feeding patterns, factors affecting feeding decisions and how feeding patterns affect the infant, mother and household; now covers a wide range of health-related topics specific to each stage of the life cycle. Ongoing.	Adair et al. (2011) and Richter et al. (2012)
16. BT20, Soweto, Johannesburg, South Africa	3,273 eligible	Birth, 6 months, 1, 2, 3, 4, 5, 7.5, 11.5, 13, 14, 15, 16, 17.5, 18.5, 19.5, 22-24[18] years	Census of all live singleton births between April and June 1990 in Soweto	Live in Gauteng province[19]	25% (11 years) 32% (24 years)	Under-enrolment of white and middle class children. Ongoing.	Richter et al. (2007) and Richter et al. (2012)

Continued on next page

Table C.1 – *Continued from previous page*

Name	**Cohort size**	**Completed rounds**	**Sampling**	**Following rule**	**Attrition**	**Notable features**	**References**
17. PCTC, Thailand	4,245	Prenatal: 28–38 weeks; postnatal: birth, 6, 12, 18, 24 months	Non-probability (judgement) sample of sites, then census of all births October 2000–September 2002 to women intending to stay in study sites[20]	Remained in geographic region	n/a	Examines biological, psychosocial and other aspects of development from the perinatal period to childhood. Completed.	Chaimay et al. (2015), Mongkolchati et al. (2010), and Sangsupawanich et al. (2007)
18. The IFORD surveys,[21] West African cities[22] (focus on Yaoundé)	9,774 (Yaoundé)	Birth, 1, 4, 8, 12, 16, 20, 24 months	Census of hospital births over one calendar year (1978 for Yaoundé) in the city	n/a	38% (2 years, Yaoundé)	Designed to measure levels and patterns of mortality during the first 2 years of life, and its determinants; criticism is measurement of morality is problematic given high attrition rates. Completed.	Defo (1992) and Van De Walle (1990)

Notes: This table was compiled in August 2015

Table notes:

[1] To the three year olds' sample, 692 "new families" were added (assuming one child per new family), as they were eligible at nine months old but had not been identified in the Child Benefit Register because of a recent home move. The attrition calculations include these.

[2] With an additional four purposively sampled cities.

[3] Three subsamples include (i) children randomly selected from those enrolled in the PIDI; (ii) children not enrolled in the PIDI selected from a stratified random sample of households in an area providing the PIDI; and (iii) children in appropriate age range selected from a stratified random sample of comparable sites but with no PIDI provision.

[4] Calculated from sample sizes without imposing secondary eligibility criteria, rather initial eligibility criteria.

[5] Subsamples at 1, 15, 18, 19 and 26 years of age.

[6] Subsamples were seen at 4, 6, 9 and 13 years.

[7] Currently, recruitment of the 2015 cohort is underway.

[8] Personal correspondence with Marigen Narea (Sep 2015).

[9] Data from birth and first year of life acquired from clinical records.

[10] Between 22–28 and 32–38 years.

[11] If women had her first antenatal visit during the first trimester, 13–28 weeks, or >28 weeks.

[12] Not yet completed at time of Tao, 2013 RN387 publication: 1.5–3.0 year visit, 4.5–6.0 years, 9–15 (boys) and 8–14 (girls).

[13] The study invited all women in their first or second trimester visiting a maternity hospital or centre to participate. Also, after the first phase (i.e. after 12 months), the study will only follow up on children from Ma'anshan during the preschool period until they reach adolescence.

[14] Subsample at 21–35 years.

[15] The Mysore Parthenon Birth Cohort is included in the table since it had a sampling frame population of 1,233 pregnant women; the sample included only 830.

[16] A new cohort study started in 2011.

[17] The initial round included only live singleton births; the survey later re-incorporated twins.

[18] Information on 22–24 years old round based on personal correspondence with Shane Norris (Dec 2015).

[19] The following rule was initially those who remained in the area until the child was six months old. Then it was those who live in a 400 km^2 area. Now it is those who remain in the state.

[20] Data were collected over one calendar year. Start date varied by site.

[21] In French, l'Enquête sur la mortalité infantile et juvénile (EMIJ).

[22] Yaoundé, Cameroon; Ouagadougou, Burkina Faso; Lomé, Togo; Cotonou, Benin; Bamako, Mali.

C.2 Weights for Peru

As the probability of selecting a child in each site is known, Escobal and Flores (2008) calculate the site-level weights by multiplying the probability of a site being selected with that of a child being selected within the site, as follows:

$$f_{\text{exp}} = P_T/P_{d_i} \times \mathrm{NH}_{e_i} \,/\, \mathrm{NC}_{d_i} \tag{C.1}$$

where

P_T = population in the country

P_{d_i} = total population in the selected district

NH_{e_i} = number of eligible households in the district

NC_{d_i} = number of eligible children in the district

Table C.2: Peru weights by site

Site no.	District	Population	No. of eligible households	No. of eligible children	Weight
1	Tumbes	90,625	5,350	100	522
2	Piura	22,279	1,462	100	580
3	Piura	11,564	523	101	392
4	Amazonas	7,697	478	101	538
5	San Martin	16,194	1,237	101	662
6	San Martin	66,997	3,045	102	386
7	Cajamarca	141,588	7,950	107	434
8	La Libertad	124,766	7,070	102	482
9	Ancash	9,585	476	103	414
10	Ancash	55,732	2,306	105	332
11	Huánuco	10,773	757	101	609
12	Lima	713,018	39,943	100	495
13	Lima	380,480	21,245	102	475
14	Lima	324,107	18,205	103	468
15	Junín	24,376	1,839	105	605
16	Ayacucho	7,392	1,064	108	1091
17	Ayacucho	17,068	1,052	102	524
18	Apurimac	15,282	1,099	105	577
19	Arequipa	10,329	310	102	255
20	Puno	189,275	10,150	102	456

Source: Escobal and Flores (2008).

Appendix D

Annex: Chapter 5

Appendix E

Annex: Chapter 6

E.1 Standardising test scores

In this section, I discuss the range of standardisation options available to me, and why I chose not to use them. In the main text I discuss four key concepts a standardisation approach should reflect to enable comparison across rounds: an established meaning, a common metric, distance from the mean score and the dispersion of scores from the mean.

Rank

A rank score orders all the children in the sample according to their score and assigns them with a rank. If there are 100 children the rank score will range from 1 to 100. If the average rank of a one group is 70 and that of the other group is 30, we know the children from the first group performed better than those from the second group. However, this measure is not useful for comparison across different ages for various reasons. First, we cannot explain by *how much* the first group performed better than the second (e.g. what do 40 rank units mean?). Second, a rank score does not reflect differences in distances between scores. For example, a gap of 40 units in a country with 100 children is different to the same gap in a sample of 800 children. Third, based on a rank score we do not immediately know whether the group has scored above or below the average score. Fourth, the dispersion of scores is not reflected in a rank score. Take a two country samples with 100 children in each sample. A gap of 40 units in one country, where the scores range from 40 to 80 is larger than a gap of 40 in another country where the scores range from 5 to 95.

Percentile

Percentile scores are similar to a rank scores, but are bound between 1 and 100, with 50 representing the median score. (To obtain the percentile score, we first we order the children according to their score, then calculate how many scores fall at or below a child's score, and divide that number by the total number of scores, then multiply by 100.) If the average percentile of one group is 90 and for another group its 10, we know the children in the first group performed better than those from the second. We also know that, on average, the top group performed the same as or better than 90 per cent of the sample, and the bottom group performed the same as or better than 10 per cent of the sample. The gap between the two is 80 percentiles - so 80 per cent of the sample's scores lie between these two scores. Also, distance of 80 percentiles between two groups represents the same thing in a sample of 100 and 800 children.We also know immediately from a child's percentile score her position relative to those of her peers, and relative to the median score (i.e. 50).

For my analysis, the main limitation to using a percentile score is that a percentile score does not take the dispersion of the the scores into account. As with the rank score, a 40 percentile gap is different in a sample where scores range between 40 to 80 compared to one where the scores range from 5 to 95.

Percent

A percent score is bound between 0 and 100. (To calculate a percent score, we divide the number of items answered correctly by the total number of items and multiply by 100.) Given its ubiquity in schools and the media, a percent score is well understood by the general population. A gap of 40 percentage points the same in a population of 100 and 800 children. However, it does not take the mean score into account. A gap of 40 percentage points between 30 and 70 per cent does not tell me whether one or both groups scored above or below the average. And as with a rank score and percentile score, a percent score does not take. Again the relative size of a 40 percentage point gap differs in a sample where scores range between 40 to 80 compared to one where the scores range from 5 to 95.

IRT score

An IRT score is commonly used in the educational testing literature to compare between samples and over time. The Programme for International Student Assessment (PISA) for example use an IRT model to scale their results (see the PISA 2015 Technical report OECD, 2017).

There are many similarities between an IRT and a z-score. Both scores have an established meaning in the educational testing literature, in both cases the unit that represents the distance between groups is the same regardless of the sample's size. We know, from the score, where it is situated relative to the sample mean (e.g. a score of 0.40 means the group's average score lies 0.40 standard deviations above the mean). And finally, the dispersion of the scores is taken into account as the scores are standardised to have standard deviation of one.

An advantage of an IRT score is it incorporates information from the children's responses to test items. For example, if only a few children answer an item correctly, this item receives a higher weight, than an item that most children answer correctly. Some argue this is because the former is assumed to be more difficult than the later. This provides a more informative measure of educational achievement, compared to a z-score that applies an equal weight to easy and difficult questions. Some authors, such as Bradbury et al. (2012) use IRT scores for PPVT results (though without discussion or justification). Others who use an IRT score, such as Leon and Singh (2017), claim it provides a more robust score for comparisons across time.

An IRT score is calculated by fitting a latent trait model. Figure E.1 is a simplified prototypical path diagram of this model. In this example, the maths tests consists of four items. The latent variable, maths ability, and this is estimated based on children's responses to each item. Here I assume the responses are binary: correct or incorrect.

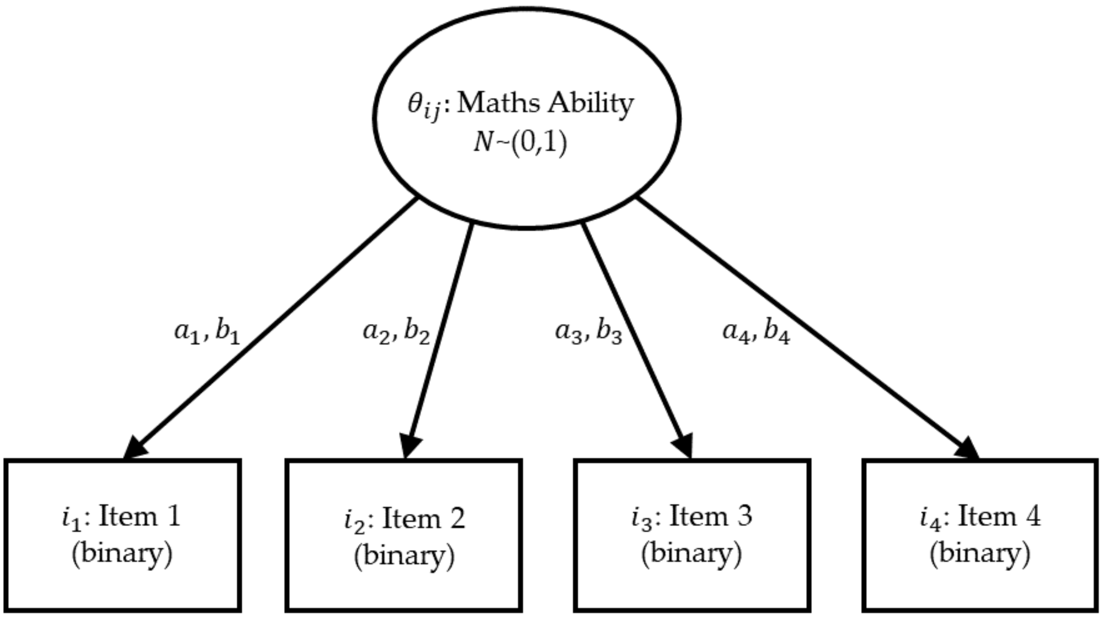


Figure E.1: Simplified prototypical path diagram for maths ability IRT model

Following notation from the Stata 15 IRT manual (StataCorp, 2017), let θ represent the latent variable - achievement in maths. Y_{ij} is the outcome for item i from child j, and y_{ij} is the observed value of the latent variable. If a child answers an item correctly, then $y_{ij} = 1$, and incorrectly then $y_{ij} = 0$. The probability of a child j with a maths achievement θ_j answering item i correctly is estimated by:

$$Pr(Y_{ij} = 1|a_i, b_i, \theta_i) = \frac{exp(a_i(\theta_j - b_i))}{1 + exp(a_i(\theta_j - b_i))} \tag{E.1}$$

where b_i, represents is the difficulty parameter of item i and the a_i represents its ability to discriminate between children with high and low maths achievement. Equation E.1 represents a general IRT equation. The 1-parameter logistic (1PL), 2-PL and 3-PL models are variants of this equation.

Two assumptions of such a one-trait model are: first, conditional independence of the items given the latent trait, so given a child's maths achievement, answering one item correctly does not affect how a child performs on another item; and, second, the latent trait has a standard normal distribution, and therefore has a mean of zero and a variance of one.

An IRT score from a 1-PL model, estimates the item difficulty (b_i) parameter. It is also referred to as a Rasch score. Most educational testing models that use IRT use the 1PL model. It is often referred to as the Rasch model. (A 2-PL model estimates the item difficulty and discrimination parameters a_i. A 3-PL model allows for the possibility of guessing on a test, and estimates a parameter that corresponds to this possibility of guessing.)

The Young Lives data has *anchor* items that are shared across rounds. Anchor items may cause floor (or ceiling) effects, as an item for a 8 year old will be easy for a 12 and 15 year old, therefore limiting how much variation there is at the bottom of the maths achievement distribution. This is the case, regardless of which standardisation approach I use.

These anchor items are sometimes used to assess progress in maths or vocabulary scores, and are sometimes used to claim more robust comparisons over time. By using the anchor items we assume these items behave the same way in every round, but this may not be the case (e.g. a term may popular or pervasive over the intervening time and the test will discriminate between children's vocabulary achievement differently than as it did in the previous test). To measure progress over time, one would pool all rounds of data, link the anchor items so they share the same item code, and fit the IRT model to the data. Stata has a *vce (cluster child_ID)* option, to account for within child correlations across rounds. However, the standard error estimated in an IRT model are those that correspond to each child's estimated score. These are not reported on or used when IRT is used with educational testing. As I document gaps in educational achievement over childhood, rather than progress in educational achievement, if I were to use an IRT score, I would not need to exploit the 'anchor' items in the Young Lives data.

An IRT score does not take the child's age into account. To do so I would have to add another level to the model and produce a score that is further abstracts from the child's raw score.

Calculating an IRT score, also often demands that the researcher examine the differential item functioning (DIF) of each item, and the select out those who do not measure the latent ability well. While there is a wide literature and a statistical justification for this, I find that disregarding an item from a test simply because it does not perform well problematic (in a statistical sense), particularly if there is a theoretical reason to include the question in a test.

E.2 Cleaning maths scores

Issues shared across rounds

In Round 2 the data shows that the answer key the Young Lives team used was incorrect for item 6. This happens in subsequent rounds as well. The correct answer for item 6 is option 1 (or 'Left'). (From the Round 2 Fieldwork manual you can infer that questions 1-7 have 3 options (left=1, middle=2, and right=3) and questions 8-15 have four options (top left=1, top right=2, bottom left=3, bottom right=4).) Young Lives also codes as missing several children's test scores, perhaps excluding those who they consider experienced inadequate testing conditions. I do not do this. When I recalculated Round 2 results the number of children who had a different score in Ethiopia, India, Peru and Vietnam were 454, 1439, 1846, and 491, respectively. The the correlation (pairwise and rank) between my score and the Young Lives score and mine, for the four countries, was 0.99, 0.99, 0.96 and 0.99 respectively.

In Rounds 3 to 5 missing variables were often incorrectly coded. The missing-item codes include: did not know (-77/77); refused to answer (-79/79); N/A (-88/88), and; missing (-99/99). As a rule of thumb, I re-coded numbers that resembled missing-item codes. Table E.1 documents all the mistakes I identified, and re-coded as missing. In subsequent cleaning, I coded items with refused to answer (-79/79), N/A (-88/88), and missing (-99/99) codes as missing, and items with did not know (-77/77) as incorrect. We don't know anything about a child's educational achievement if they refuse to answer a question, while we do know if they say they don't know. (In India Round 5 data, the Young Lives team, however, who used -77 as a missing-item code, and I recoded it accordingly.)

In Rounds 4 and 5 there are open ended questions that can have various options. For example a correct answer for 0.75 is also 3/4 or .75. In Table E.2 I include all the options I include as correct (in bold are those that I classify as correct, that the Young Lives team has not).

Round 3

In Ethiopia's Round 3 data, after cleaning and recalculating the score, there are 21 children who have different scores to the Young Lives score. Thirteen children answered 'don't know' (77) in items 1 to 9, and answers to all subsequent questions are missing. Young Lives gave them a 'missing' score. However, when a child has a missing code for items 1 to 9, but has answered 'don't know' (-77) for items 10 to 29, Young Lives has allocated the child with a score of 0. It is inconsistent to consider a 'don't know' for items 1 to 9 as missing, but not missing for items 10 to 29. I code any 'don't know' (-77/77) as an incorrect answer, regardless of the item number. This results in 13 children with a score of 0, rather than the Young Lives score of 'missing'.

There are two children who have answered maths test questions, but Young Lives have allocated them with a missing score, while I have not. This may have been explained by the following rule employed by the Young Lives team. Cueto and León state that "we considered times up to a limit of 14 minutes; any time above that limit was coded as missing" (2012, p. 12). (The fieldworker manual stated that the test should be discontinued at 8 minutes.) But both children completed the test in less than 14 minutes. Moreover, when I looked in detail at the data on the test duration, they too were dirty, with tests ending at times before they began, or children taking the 8 minute test for hours (e.g. 300+ minutes).

Then there is one child who has all items missing, but Young Lives has allocated the child with a score of 0. Based on my calculations the child had a missing score. Finally, there are five children who have different scores to those I have calculated, ranging from 8 to 2 fewer points. I checked manually, and the raw answers in the data reflect the score I calculated. I, unfortunately, cannot explain this discrepancy.

In India's Round 3 data, after cleaning and recalculating the score, there are 26 children who have different scores to the Young Lives score. Nine children had codes of 'don't know'(77/-77) for all items. Young Lives gave them a 'missing' score, and I gave the, a score of 0. Then there is one child who has all items missing, but Young Lives has allocated the child with a score of 0, I gave it a missing score. There are 16 children who have different scores to those I have calculated (ranging from 8 to 1 fewer points). I checked manually, and the raw answers in the data reflect the score I calculated. I cannot explain this discrepancy.

In Peru's Round 3 data, after cleaning and recalculating the score, there are 72 children who have different scores to the Young Lives score. Thirty-seven children answered maths test questions, but Young Lives have allocated them with a missing score, while I have not (most of these took their tests within the 14 minutes). These scores ranged from answering 2 to 26 items correctly.

Thirty-five children have different scores to those I have calculated (ranging from 10 fewer to 2 more points). I checked manually, and the raw answers in the data reflect the score I calculated. I cannot explain this discrepancy.

In Vietnam's Round 3 data, three children were included as duplicates (specifically vn090082, vn170052 and vn200047), I removed them. After cleaning and recalculating the maths score, I found that there was a discrepancy of 56 children between Young

Lives' score and mine. Thirty-two children had different scores (4 to 1 fewer points) - which I cannot explain. One child who answered maths test questions, was given a missing score by the Young Lives team (there is no data on how long the child took to complete the test). Twenty-one children answered 'don't know' in items 1 to 9, and had missing codes for the remaining items. In those cases, the Young Lives team allocated them with a missing score, while I gave them with a score of 0 (as I mark 'don't know' responses as incorrect responses). There are several children with missing codes for items 1 to 9. So this does not appear to be the case of the 'don't know' code being used as a missing code, rather it seems that the child did not know the answer. Similarly, two children answered 'don't know' to all items, and Young Lives gave them a missing score, while I gave them a score of 0. Again, there are other children with missing codes for all items, so it appears that the child genuinely did not know the answers to all the questions.

Round 4

In Ethiopia's Round 4 data, Young Lives allocated 9 children a score of 0, when all their observations were coded as 9999. There are no instructions in the 'Cognitive Test Fieldworker Manual' to use the code 9999 as an additional code for missing values. (In other parts of the manual 9999 used to indicate a child's height or weight is not measured.) Although these children had been administered the questionnaire, but not the maths test, there were many more children for whom this was the case, and the Young Lives team has not allocated them with a 0 score. I concluded that the data had not been cleaned properly to include 9999 as a missing value. After cleaning, ten children have different scores to the Young Lives scores, and the correlation (pairwise and rank) between Young Lives' scores and mine is 1.00.

In India's Round 4 data, after cleaning and recalculating the maths score, there are 3 children who have different scores to the Young Lives score. One where I coded the correct option as "3/4*1/2", another that has several don't know answers (-77), and one where all the items are missing, but the Young Lives score has allocated a score of 0 to. The correlation (pairwise and rank) between Young Lives' scores and mine is 1.

In Peru's Round 4 data, after cleaning and recalculating the maths score, there are 2 children who have different scores to the Young Lives score. One that included a correct answer option, and one whose answers were coded as 'did not know' (77) throughout. The correlation (pairwise and rank) between Young Lives' scores and mine is 1.

In Vietnam's Round 4 the data shows that the answer key the Young Lives team used was incorrect for item 7. The question asks "A garden has 14 rows. Each row has 20 plants. The gardener then plants 6 more rows with 20 plants in each row.How many plants are now there altogether?”. The correct answer is 400, or option C. The Young Lives team has marked A as the correct answer. After cleaning, 1,523 children have different scores to the Young Lives scores, and the correlation (pairwise and rank) between Young Lives’ scores and mine is 0.99.

Round 5

In Round 5, marking was not consistent across countries. In Peru ‘30 scoops’ not ‘30’ was coded as correct, and in India ‘30’ was considered correct. So I applied the same rules consistently to all countries. For this example, all answers that included 30 I coded as correct.

In Ethiopia’s Round 5 data, after cleaning and recalculating the maths score, there are 4 children who have different scores to the Young Lives score. Young Lives allocated a score of 0 to a child who had all items missing (‘MISS’ or 99). I coded this child’s score as missing. The other differences reflect the additional answers I deem as correct (see bold items in Table E.2).

In India’s Round 5 data, The Young Lives team applied the ‘don’t know’ (77/-77) as if it were a missing variable, so I treat it as such. After cleaning and recalculating the maths score, there are six children who have different scores to the Young Lives score. These reflect the additional answers I deem as correct (see bold items in Table E.2).

In the Young Lives Rounds 4 and 5 data, there is a formatting problem with two items whose answers have decimal points. Where the correct answer is, 18.03, for example, this was is formatted to have eight, rather than two, decimal places (e.g. 18.03000069). An answer key with 18.03 would mark all correct answers as incorrect. I reformatted the responses to have two decimal places. (This also happened with an answer 17.43 that was reformatted from 17.43000031). This is exactly what happened in Peru’s Round 5 data. The Young Lives team did not pick up on this formatting discrepancy, and coded all answers to items 5 and 6 as incorrect. When I corrected the formatting. I corrected this error, and included additional correct answers in the answer key (see bold items in Table E.2). I also coded as missing, the 41 children the young lives allocated a score of 0 to, despite all items have missing values. This resulted in a change in the score of 1,348 children. And a correlation (pairwise and rank) between Young Lives’ scores and mine of 0.99.

Vietnam Round 5 data were perhaps the cleanest data set. I identified only one additional missing value and correct answer option. After cleaning, only one item differed in score, and the (pairwise and rank) correlation between the scores was 1.00.

Table E.1: Maths scores: missing-item codes recoded as missing

Country	**Round**	**Codes**
All countries	3	-7900, -79,-7700, -7, -777, -773, -76, -8800, -9900, -8880, 8800, -800, -87, -9
Ethiopia	4	-9999, 9999, -999, 999, -90, -993, -991, -8, -99.00, -9999, 9999, 99.00, _99, -69, -9.899999619, -9.89
	5	Bl[]nk/Miss/, -, –99, :-99, _77, _99, -96,, -77, -79, -99, -88, -9999, -9. , -98, -96, -9, -999, +=, =, bank missing, bl;ank mising, Blank, blank, BLANK, Blank /missing, blank /missing, blank miisingaq, Blank miss, blank miss, blank missig, blank missing, BLANK MISSING, Blank missing, blank missing, blank missng, BLANK OR MISS, BLANK/ MISS/, Blank/ missing, blank/ missing, Blank/mising, BLANK/MISS, blank/missed, Blank/missed, blank/missing, blankmissing, Empety/Miss/, MISS, Miss, MISS /BLANK/, miss blank, Miss blank, MISS(BLANK), MISS/ BLANK, MISS/BLANK, MISS/LEFT BLANK/, MISSED, missed, Missed, MISSEDU, MISSING, missing, Missing, missing blank, MISSING BLANK, Missing blank, MISSING BLLANK, Missng /blank, missng/balnk, MSS, NISSING, no answer, No answer, Missing blank, MISSING BLLANK, Missng /blank, missng/balnk, MSS, NISSING, no answer, No answer, x
Peru	4	–99, -888
India	5	=77, =88, -7, –7, -, -77-77, –77, -, BLANK/MISS, blank/miss, BLANL/MISS. -988
Vietnam	5	-

Notes: Note that Young Lives did not clean up missing values

Table E.2: Correct answers: various correct answers for items in Round 4-5 maths tests

Country	Round	Item	Correct answer	Answer options
Ethiopia	4	18	0.75	3/4, 0.75, 18/24, 18\24, .75, 9/12
	5	5	3	03, 3.
		6	17.43	1.7.43, **3417417.17.43**
		7	18.03	18..03, 18.0.3, 1803/100
		8	0.75	.75, 1/4 or 18/24, 18.24, 18/.24, 3/4, 9/12, 9/12=3/4, 18/24, **6/8**
		15	30	1/30 or30, **Hartuu 30**
India	4	17	0.75	3/4, 18/24, 0.75, 18/24 ., 18/24,3/4, 18/24., 18/24.., 3/4 ,., 3/4 ., **3/4*1/1**, 3/4., 6/8
	5	3	3	**03,0**
		6	0.75	18/24, 18/24 3/4, 18/24=3/4, 18/24=3/4=0.75, 18/24=9/12=3/4, 18/24=3/4, 18\24, 3/4, 3/4.1/1, 6/8, 9/12, **3\4**, **18/24=6/8=3/4**, **18/ 24**
		13	30	30 H0URS, 30 KG, 30 kg, 30 scoops, 30/1, 30KG, 30KGS, 30kg, 30kgs
		30	120	**1‘20**
Peru	4	17	0.75	3/4, 18/24, 0.75, **9/8 x 2/3 = 18/24**, 9/12, 6/8
	5	7	0.75	18/24, 3/4, 6/8, 9/12
		13	30	30 CUCHARADAS, 30 CUCHARADAS DE 1/5 KG., 30 CUCHARADAS DE HARINA, 30 CUCHARADITAS, 30 CUCHARAS, 30 KG CUCHARADAS, 30/1, 30/1 CUCHARADAS, 30CUCHARAS, NECESITAN 30 CUCHARADAS, SE NECESITA 30 CUCHARADAS SE NECESITA 30 CUCHARADAS DE HARINA, SE NECESITA 30 CUCHARADITAS PARA 6 KG, SE NECESITAN 30 CUCHARADAS, SE NECESITAN 30 CUCHARADAS PARA LLENAR UNA BOLSA DE 6KG, SE NECESITAN 30 CUCHARADITAS, SE NECESITARIA 30 CUCHARADAS, UNAS 30 CUCHARADAS, **SE NECESITA 30 CUCHARADAS DE HARINA**, **30 KG**, **SE NECESITA 30 CUCHARADAS, SE NECESITAN 30 CUCHARADAS DE HARINA**
		27	28	28 KM/H, **28/1**
		30	120	120 (ZEDS) , 120 ADICIONALE, 120 ZEDS, S/ 120, S/.12, **S/. 120**
Vietnam	5	7	0.75	18/24, 3/4, 6/8, 9/12, ‘0.75
		25	28	28 KM/H, 28 km/h, 28KM/H, 28km/h
		28	120	120 zeds
		30*	11	11H, 11HSNG, 11h, 11h, 11hsng, 11htra, 1100
		31	40	40 CM, 40 cm, 40CM, 40cm, **30 d___n 40cm**

Notes: In **bold** are answers that I classify as correct, that the Young Lives team has not.

* Had special characters, which when removed, included these options.

E.3 Cleaning vocabulary scores

The vocabulary scores are more complicated than the maths scores. Recall that in Rounds 2 and 3 all countries used a PPVT test, then in Rounds 4 and 5 Ethiopia, India and Vietnam stopped using the PPVT test and had a shorter vocabulary test. Meanwhile in Rounds 4 and 5, Peru persisted with the Spanish version of the PPVT test (TVIP).

As I discussed above, calculating the raw score of the PPVT test is rather complicated. Fieldworkers were tasked with calculating the raw scores. First she had to establish the baseline, then the ceiling according to a set of criteria discussed above. (Simply establishing the ceiling item is rather complicated and prone to human error.) They then needed to count all those below the baseline as correct, and those above the ceiling as incorrect, then sum all correct items in between the ceiling and the baseline, add this number to the baseline and report the raw score. The Young Lives team then reported the fieldworker's raw scores as the final raw score, without making the necessary corrections.

I therefore recalculated the PPVT scores mechanically, using Stata 15. By doing so, I identified numerous errors in the Young Lives data. Ceilings and baselines that had been identified incorrectly, calculated. Raw scores had been incorrectly recoded (for example, sometimes a raw score of 25 really was 125, the number '1' was missing from the beginning of the number.

Issues shared across rounds

In the Round 4 and 5 data for Ethiopia, the Young Lives team coded the correct answer for item 8 (jumping) as 4, however the correct answer is 4. I have had to correct this error (all but 3 respondents were affected by this change).

In Round 5's Peruvian data, one child was allocated a score of 102 when they had no responses in the data set. For example, sometimes a raw score of 25 really was 125, and was missing a 1. Other times, the first set of 8 with 6 errors were not identified, so the next set was counted (incorrectly). This amounted to 103 observations in the Peruvian Round 5 data.

E.4 Test score distributions

Below are the distributions of the test scores. While the sample size of the balanced panel slightly vary, depending on the achievement variable used (i.e. maternal education, the asset index and expenditure) the distributions are similar in shape. I therefore only report the distributions where maternal education and the achievement score is available in the data. Note that the highest marker on the y-axis differs for vocabulary and maths scores.

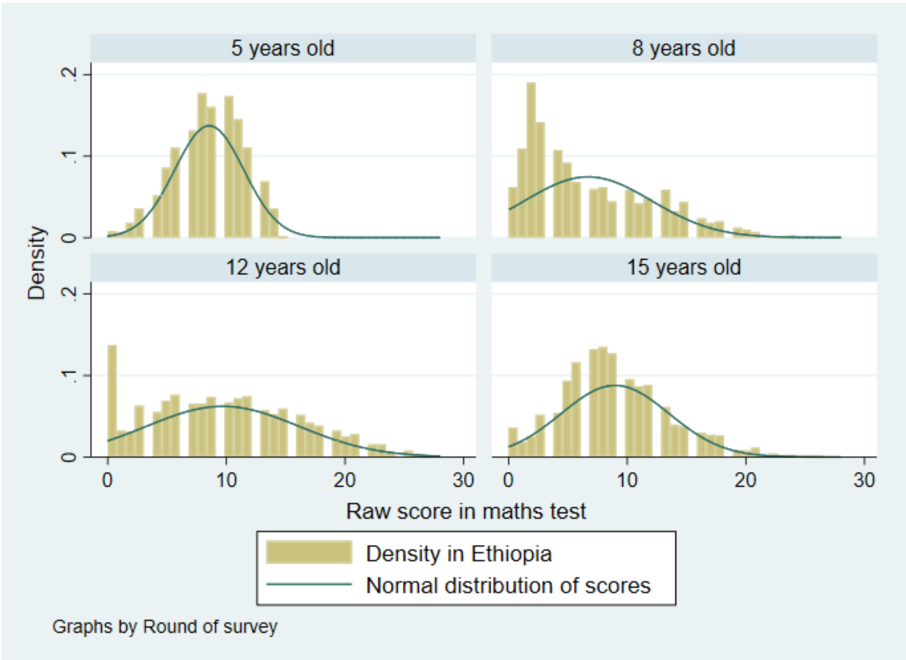


Figure E.2: **Ethiopia**: Distribution of maths scores for all observations with maternal education data at 5 years old, by round

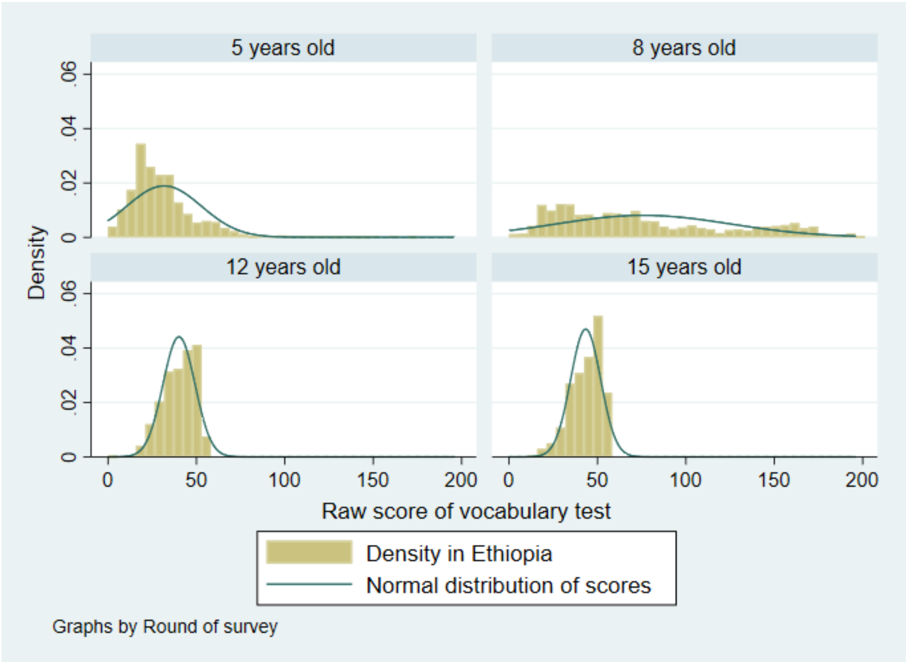


Figure E.3: **Ethiopia**: Distribution of vocabulary scores for all observations with maternal education data at 5 years old, by round

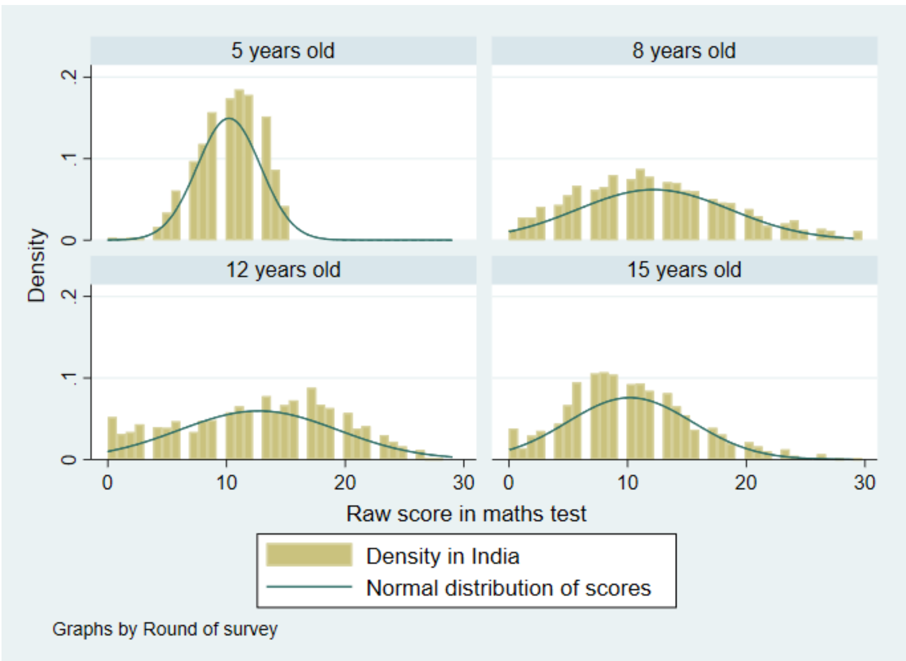


Figure E.4: **India**: Distribution of maths scores for all observations with maternal education data at 5 years old, by round

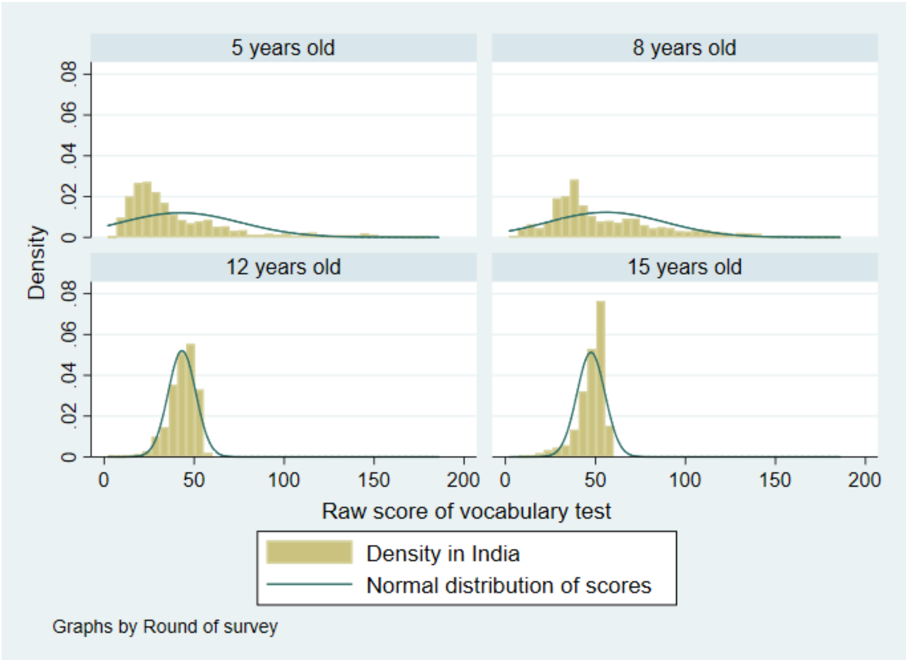


Figure E.5: **India**: Distribution of vocabulary scores for all observations with maternal education data at 5 years old, by round

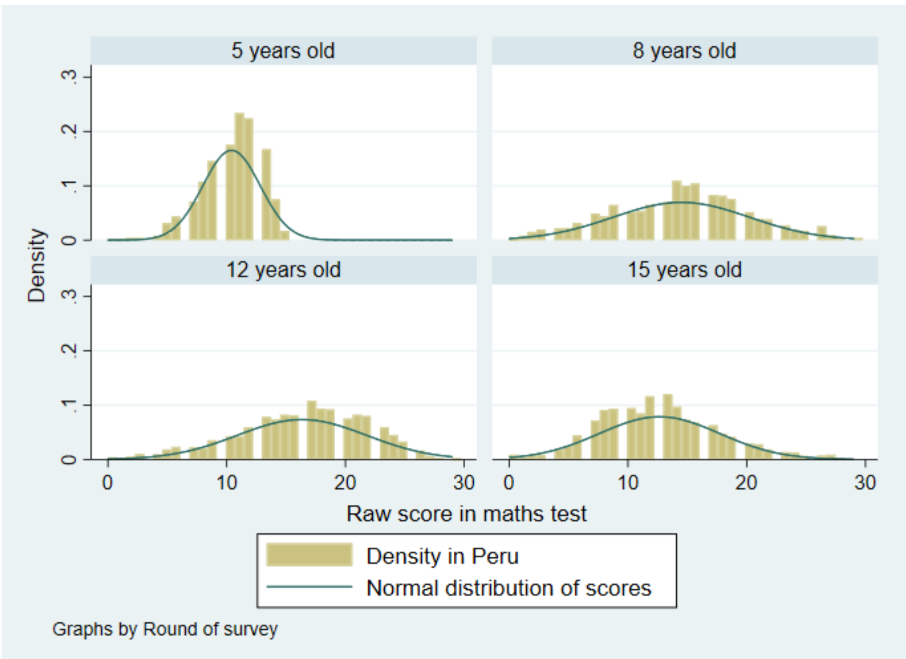


Figure E.6: **Peru**: Distribution of maths scores for all observations with maternal education data at 5 years old, by round

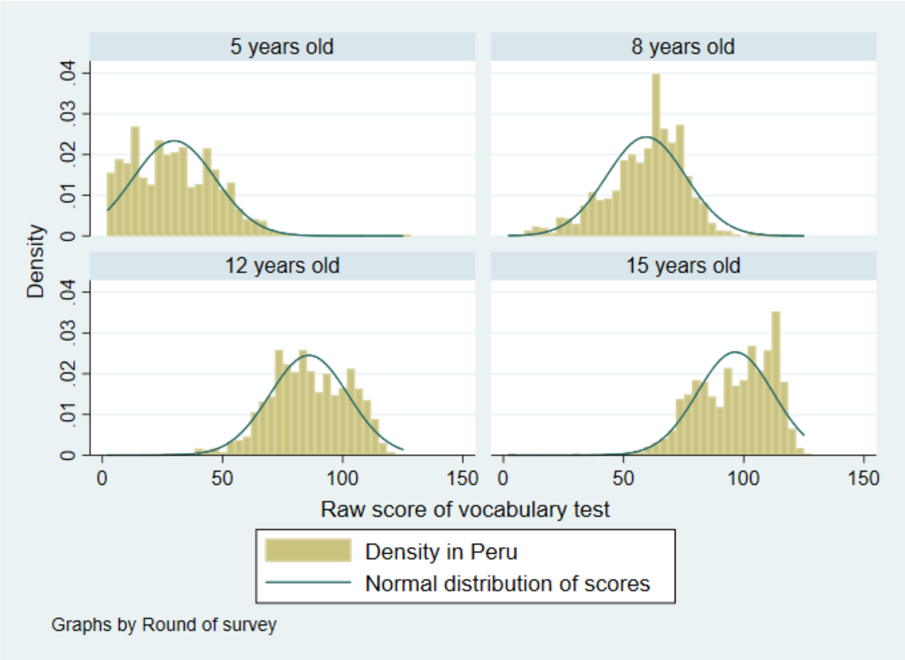


Figure E.7: **Peru**: Distribution of vocabulary scores for all observations with maternal education data at 5 years old, by round

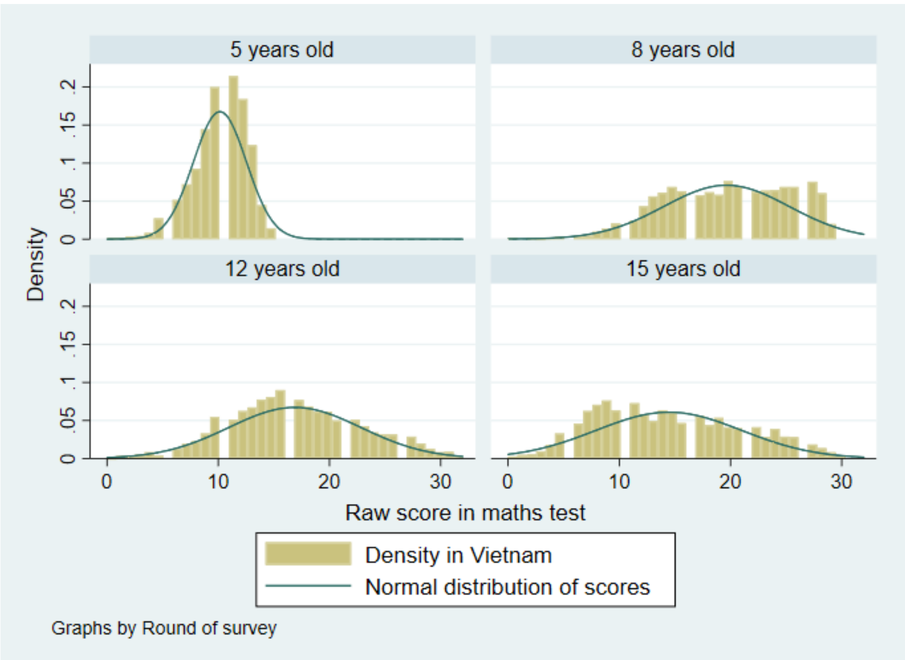


Figure E.8: **Vietnam**: Distribution of maths scores for all observations with maternal education data at 5 years old, by round

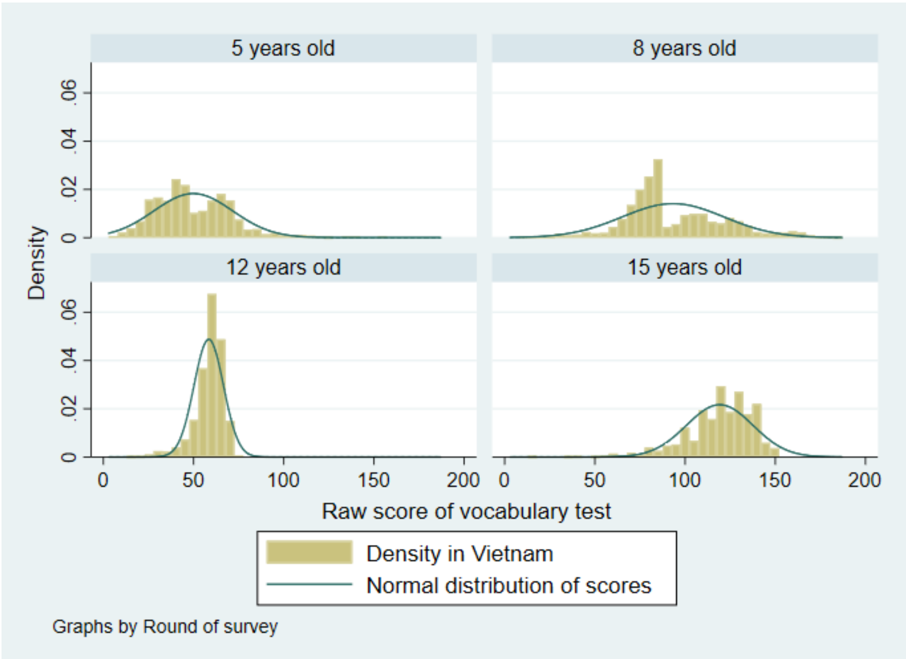


Figure E.9: **Vietnam**: Distribution of vocabulary scores for all observations with maternal education data at 5 years old, by round

E.5 Tables with statistics on test score distributions: mean and standard deviations

Table E.3: Socioeconomic gaps in maths scores: mean and standard deviations from a balanced panel

	Ethiopia	India	Peru	Vietnam
Mother's education				
5 years old				
Mean	8.6	10.2	10.4	10.2
s.d.	2.9	2.7	2.4	2.4
8 years old				
Mean	6.7	12.2	14.5	19.6
s.d.	5.4	6.4	5.7	5.6
12 years old				
Mean	9.7	12.7	16.3	16.9
s.d.	6.4	6.7	5.4	5.9
15 years old				
Mean	9.0	10.2	12.7	14.5
s.d.	4.5	5.3	5.1	6.6
Sample size	1508	1787	1622	1729
Asset index				
5 years old				
Mean	8.6	10.2	10.3	10.2
s.d.	2.9	2.7	2.4	2.4
8 years old				
Mean	6.8	12.1	14.4	19.6
s.d.	5.4	6.4	5.8	5.6
12 years old				
Mean	9.7	12.6	16.2	16.9
s.d.	6.4	6.7	5.5	6.0
15 years old				
Mean	9.0	10.1	12.6	14.5
s.d.	4.5	5.3	5.1	6.5
Sample size	1599	1848	1712	1717
Expenditure				
5 years old				
Mean	8.6	10.2	10.4	10.1
s.d.	2.9	2.7	2.4	2.4
8 years old				
Mean	6.8	12.1	14.5	19.6
s.d.	5.4	6.4	5.8	5.7
12 years old				
Mean	9.7	12.5	16.3	16.8
s.d.	6.4	6.7	5.4	6.0
15 years old				

Table E.4: Socioeconomic gaps in vocabulary scores: mean and standard deviations from a balanced panel

	Ethiopia	**India**	**Peru**	**Vietnam**
Mother's education				
15 years old				
Mean	31.7	42.4	29.9	49.8
s.d.	21.1	33.3	17.1	21.9
8 years old				
Mean	76.0	56.0	59.4	93.3
s.d.	49.4	32.5	16.4	28.3
12 years old				
Mean	40.1	43.3	86.0	58.4
s.d.	9.0	7.7	16.3	8.2
15 years old				
Mean	43.3	47.5	96.5	119.2
s.d.	8.5	7.8	15.8	18.4
Sample size	1410	1713	1536	1533
Asset index				
5 years old				
Mean	31.7	42.4	29.6	49.8
s.d.	21.1	33.6	17.1	21.9
8 years old				
Mean	76.5	55.9	59.1	93.1
s.d.	49.5	32.6	16.5	28.3
12 years old				
Mean	40.2	43.2	85.7	58.3
s.d.	9.0	7.7	16.3	8.3
15 years old				
Mean	43.4	47.4	96.3	119.1
s.d.	8.5	7.8	15.8	18.5
Sample size	1496	1773	1618	1528
Expenditure				
5 years old				
Mean	31.7	42.4	29.7	49.4
s.d.	21.0	33.6	16.9	21.8
8 years old				
Mean	76.6	55.6	59.2	92.7
s.d.	49.6	32.5	16.4	28.1
12 years old				
Mean	40.2	43.1	85.9	58.1
s.d.	9.0	7.7	16.2	8.3
15 years old				

Table E.5: Early childhood education: mean and standard deviations from a balanced panel for **raw** gaps in maths and vocabulary scores

	Ethiopia	**India**	**Peru**	**Vietnam**
Maths				
5 years old				
Mean	8.6	10.2	10.3	10.2
s.d.	2.9	2.7	2.4	2.4
8 years old				
Mean	6.7	12.1	14.5	19.6
s.d.	5.4	6.4	5.7	5.6
12 years old				
Mean	9.7	12.6	16.3	16.9
s.d.	6.4	6.7	5.4	5.9
15 years old				
Mean	9.0	10.1	12.6	14.5
s.d.	4.5	5.3	5.1	6.5
Sample size	1606	1844	1327	1728
Vocabulary				
5 years old				
Mean	31.6	42.4	29.7	49.7
s.d.	20.9	33.6	16.8	21.9
8 years old				
Mean	76.5	55.9	59.2	93.2
s.d.	49.4	32.7	16.7	28.4
12 years old				
Mean	40.2	43.1	85.6	58.3
s.d.	9.0	7.7	16.4	8.3
15 years old				
Mean	43.4	47.4	96.0	119.0
s.d.	8.4	7.8	15.9	18.6
Sample size	1501	1770	1252	1547

Table E.6: Early childhood education: mean and standard deviations from a balanced panel for **adjusted** gaps in maths and vocabulary scores

	Ethiopia	**India**	**Peru**	**Vietnam**
Maths				
5 years old				
Mean	8.6	10.3	10.4	10.2
s.d.	2.9	2.6	2.4	2.4
8 years old				
Mean	6.8	12.9	14.6	19.8
s.d.	5.3	6.5	5.7	5.5
12 years old				
Mean	10.0	13.3	16.3	17.1
s.d.	6.3	6.6	5.4	5.8
15 years old				
Mean	9.3	10.8	12.7	14.6
s.d.	4.4	5.2	5.1	6.5
Sample size	1366	1417	1277	1606
Vocabulary				
5 years old				
Mean	31.9	44.0	29.8	50.5
s.d.	21.2	34.2	16.8	21.9
8 years old				
Mean	77.3	58.9	59.4	94.3
s.d.	49.6	33.3	16.7	28.2
12 years old				
Mean	40.7	44.0	85.8	58.7
s.d.	8.7	7.5	16.3	7.7
15 years old				
Mean	44.0	48.2	96.1	119.8
s.d.	8.0	7.4	16.0	18.1
Sample size	1253	1355	1204	1422

E.6 Asset index

Table E.7: Asset index: Country specific criteria by asset index components

	Ethiopia	**India**	**Peru**	**Vietnam**
Housing quality				
Good-quality wall	Brick/concrete Mud and bricks/stones Stone	Brick/concrete Stone Concrete blocks	Brick/concrete Concrete blocks	Brick/concrete Concrete blocks
Good-quality roof	Concrete/cement Galvanised/corrugated iron Tiles/slates	Concrete/cement Galvanised/corrugated iron Tiles/slates	Concrete/cement Galvanised/corrugated iron Tiles/slates	AC (asbestos cement) roofing sheets Asbestos sheets Concrete/cement Galvanised/corrugated iron Tiles/slates
Good-quality floor	Concrete/cement/tile Laminated material Marble stone	Concrete/cement/tile Laminated material Stone (granite/marble) Polished stone Stone/brick	Cement/tile Laminated material Stone (granite/marble) Polished stone Parquet	Concrete/cement/tile Laminated material Stone (granite/marble) Polished stone Stone/brick

Continued on next page

Table E.7 – *Continued from previous page*

	Ethiopia	**India**	**Peru**	**Vietnam**
Crowding definition	Bottom third	Bottom third	Bottom third	Bottom third

Continued on next page

Table E.7 – *Continued from previous page*

	Ethiopia	**India**	**Peru**	**Vietnam**
Access to services				
Safe drinking water sources	Piped into own dwelling/yard/plot Piped into neighbour's dwelling/yard/plot Public standpipe/tube well Tube well in own dwelling/yard/plot	Bore well Bought water (delivery/bottled) Piped into own dwelling/yard/plot Piped into neighbour's or relatives' dwelling/yard/plot Protected spring water/well Public standpipe/tube well Tube well in own dwelling/yard/plot Water tank (community/ protected)	Piped water to the house/plot (not communal) (public network) Well/tube well with hand pump	Piped into own dwelling/yard/plot

Continued on next page

Table E.7 – *Continued from previous page*

	Ethiopia	**India**	**Peru**	**Vietnam**
Safely managed sanitation service	Flush toilet/septic tank Pit latrine (household/communal)	Flush toilet/septic tank Pit latrine (household/communal) Toilet in health post	Flush toilet/septic tank Pit latrine (household)	Flush toilet/septic tank Pit latrine (household)
Adequate fuel for cooking	Gas/electricity Kerosene/paraffin	Gas/electricity Kerosene/paraffin	Gas/electricity Kerosene/paraffin	Gas/electricity Kerosene/paraffin
Electricity	Household has electiricty	Household has electiricty	Household has electiricty	Household has electiricty

Continued on next page

Table E.7 – *Continued from previous page*

	Ethiopia	**India**	**Peru**	**Vietnam**
Consumer Durables				
	Radio, television, bicycle, motorbike, automobile, landline phone, mobile phone, table and chair, sofa, bedstead	Radio, television, bicycle, motorbike, automobile, landline phone, mobile phone, refrigerator, fan	Radio, television, bicycle, motorbike, automobile, landline phone, mobile phone, refrigerator, stove, blender, iron, record player	Radio, television, bicycle, motorbike, automobile, landline phone, mobile phone, refrigerator, fan

Source: Adapted from Appendix 1 in Briones (2017)

Appendix F

Annex: Chapter 7

F.1 Regression results

Table F.1: Equation 7.1 estimates: Maths scores as the outcome variable, and maternal education as the measure of SES

	Ethiopia	**India**	**Peru**	**Vietnam**
Mother with no education		(*base category*)		
Mother with primary to lower-secondary education	0.45*** (0.21 - 0.68)	0.41*** (0.27 - 0.54)	0.24** (0.03 - 0.44)	0.38*** (0.18 - 0.57)
Mother with upper-secondary education or above	0.93*** (0.59 - 1.28)	0.67*** (0.53 - 0.81)	0.65*** (0.42 - 0.88)	0.70*** (0.43 - 0.98)
Round 2		(*base category*)		
Round 3	0.07 (-0.19 - 0.34)	0.00 (-0.14 - 0.14)	-0.31*** (-0.53 - -0.09)	-0.17 (-0.48 - 0.15)
Round 4	0.04 (-0.25 - 0.33)	-0.07 (-0.19 - 0.05)	-0.35** (-0.62 - -0.08)	-0.24* (-0.51 - 0.02)
Round 5	0.26** (0.03 - 0.48)	-0.07 (-0.18 - 0.03)	-0.16 (-0.42 - 0.10)	-0.23 (-0.52 - 0.05)
Mother with primary to lower-secondary education * Round 3	0.19* (-0.03 - 0.41)	0.04 (-0.14 - 0.22)	0.29** (0.06 - 0.52)	0.20 (-0.04 - 0.43)
Mother with primary to lower-secondary education * Round 4	0.07 (-0.17 - 0.31)	0.08 (-0.09 - 0.26)	0.33** (0.07 - 0.60)	0.22* (-0.01 - 0.46)
Mother with primary to lower-secondary education * Round 5	-0.03 (-0.21 - 0.14)	0.06 (-0.10 - 0.21)	0.11 (-0.14 - 0.37)	0.17 (-0.06 - 0.40)
Mother with upper-secondary education or above * Round 3	0.44*** (0.14 - 0.75)	0.08 (-0.11 - 0.26)	0.54*** (0.29 - 0.80)	0.28* (-0.05 - 0.60)
Mother with upper-secondary education or above * Round 4	0.20 (-0.08 - 0.47)	0.25*** (0.08 - 0.42)	0.54*** (0.23 - 0.86)	0.49*** (0.22 - 0.76)
Mother with upper-secondary education or above * Round 5	-0.03 (-0.35 - 0.29)	0.24*** (0.08 - 0.41)	0.30* (-0.01 - 0.62)	0.49*** (0.17 - 0.80)
Constant	-0.09 (-0.40 - 0.22)	-0.17*** (-0.25 - -0.10)	-0.36*** (-0.57 - -0.14)	-0.25* (-0.51 - 0.01)
Observations	5,740	7,132	6,748	6,984
R-squared	0.18	0.11	0.14	0.12

*** p<0.01, ** p<0.05, * p<0.1. All estimates adjust for the language of the test, and their interaction with survey rounds . Robust confidence intervals are reported in parentheses. Standard errors are clustered at each country's community level.

Table F.2: Equation 7.1 estimates: Maths scores as the outcome variable, and the asset index as the measure of SES

	Ethiopia	**India**	**Peru**	**Vietnam**
1st quartile (asset index)		(*base category*)		
2nd quartile (asset index)	-0.10*	0.16***	0.23***	0.22***
	(-0.22 - 0.02)	(0.04 - 0.28)	(0.09 - 0.38)	(0.08 - 0.36)
3rd quartile (asset index)	0.30**	0.30***	0.40***	0.31**
	(0.04 - 0.55)	(0.15 - 0.46)	(0.26 - 0.55)	(0.07 - 0.54)
4th quartile (asset index)	0.80***	0.62***	0.64***	0.53***
	(0.46 - 1.14)	(0.44 - 0.79)	(0.51 - 0.77)	(0.27 - 0.79)
Round 2		(*base category*)		
Round 3	-0.14	-0.08	-0.10	-0.17
	(-0.42 - 0.15)	(-0.23 - 0.06)	(-0.23 - 0.03)	(-0.49 - 0.15)
Round 4	-0.14	-0.10	-0.10	-0.21
	(-0.42 - 0.14)	(-0.24 - 0.05)	(-0.25 - 0.04)	(-0.46 - 0.05)
Round 5	0.11	-0.08	-0.04	-0.25*
	(-0.12 - 0.34)	(-0.20 - 0.04)	(-0.17 - 0.09)	(-0.53 - 0.03)
2nd quartile * Round 3	0.30***	0.17**	0.13	0.11
	(0.13 - 0.47)	(0.02 - 0.32)	(-0.05 - 0.31)	(-0.07 - 0.29)
2nd quartile * Round 4	0.34***	0.10	0.16*	0.11
	(0.20 - 0.49)	(-0.05 - 0.26)	(-0.03 - 0.34)	(-0.12 - 0.34)
2nd quartile * Round 5	0.22**	0.02	0.01	0.14
	(0.04 - 0.39)	(-0.13 - 0.18)	(-0.17 - 0.19)	(-0.04 - 0.31)
3rd quartile * Round 3	0.43***	0.17	0.27***	0.17
	(0.15 - 0.71)	(-0.05 - 0.39)	(0.08 - 0.45)	(-0.15 - 0.50)
3rd quartile * Round 4	0.35***	0.12	0.23**	0.20
	(0.10 - 0.61)	(-0.09 - 0.33)	(0.04 - 0.42)	(-0.10 - 0.50)
3rd quartile * Round 5	0.19*	0.12	0.06	0.22
	(-0.03 - 0.41)	(-0.06 - 0.31)	(-0.13 - 0.25)	(-0.07 - 0.50)
4th quartile * Round 3	0.55***	0.08	0.30***	0.35**
	(0.23 - 0.86)	(-0.15 - 0.30)	(0.12 - 0.47)	(0.05 - 0.65)
4th quartile * Round 4	0.30	0.16	0.22**	0.35**
	(-0.06 - 0.67)	(-0.05 - 0.36)	(0.02 - 0.42)	(0.06 - 0.63)
4th quartile * Round 5	0.15	0.16	0.16	0.40***
	(-0.16 - 0.47)	(-0.05 - 0.36)	(-0.06 - 0.37)	(0.13 - 0.66)
Constant	-0.17	-0.21***	-0.27***	-0.17
	(-0.47 - 0.14)	(-0.32 - -0.10)	(-0.38 - -0.16)	(-0.39 - 0.05)
Observations	5,740	7,132	6,748	6,984
R-squared	0.25	0.07	0.12	0.12

*** p<0.01, ** p<0.05, * p<0.1. All estimates adjust for the language of the test, and their interaction with survey rounds . Robust confidence intervals are reported in parentheses. Standard errors are clustered at each country's community level.

Table F.3: Equation 7.1 estimates: Maths scores as the outcome variable, and expenditure as the measure of SES

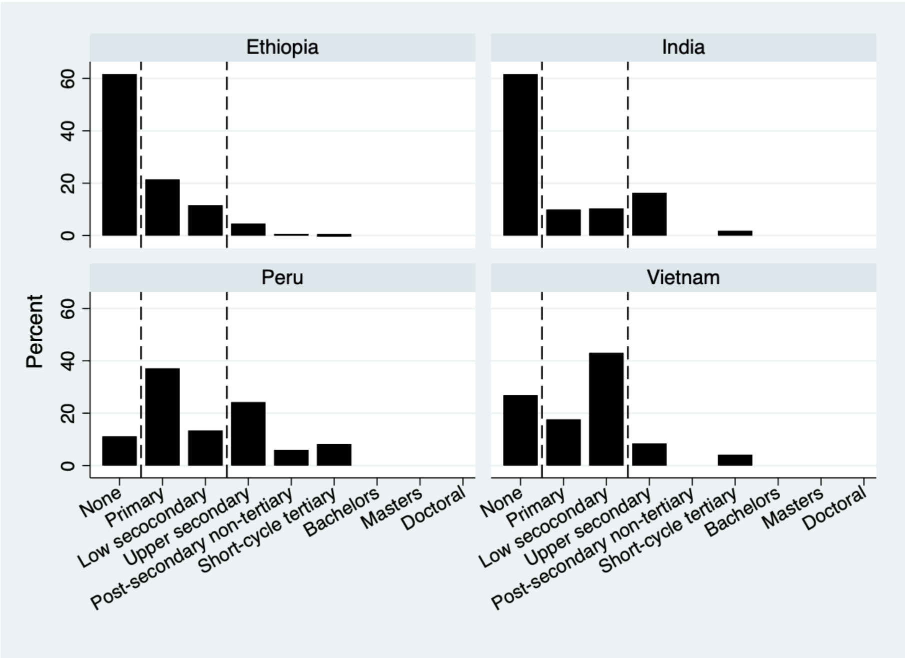
	Ethiopia	**India**	**Peru**	**Vietnam**
1st quartile (expenditure)		(*base category*)		
2nd quartile (expenditure)	0.01	0.18**	0.21***	0.18**
	(-0.14 - 0.16)	(0.03 - 0.33)	(0.07 - 0.34)	(0.04 - 0.32)
3rd quartile (expenditure)	0.21*	0.32***	0.39***	0.28**
	(-0.01 - 0.44)	(0.18 - 0.46)	(0.21 - 0.57)	(0.06 - 0.51)
4th quartile (expenditure)	0.55***	0.46***	0.65***	0.55***
	(0.27 - 0.83)	(0.31 - 0.60)	(0.52 - 0.78)	(0.30 - 0.79)
Round 2		(*base category*)		
Round 3	-0.01	0.13	-0.09	-0.16
	(-0.25 - 0.23)	(-0.08 - 0.34)	(-0.23 - 0.04)	(-0.49 - 0.17)
Round 4	-0.01	0.05	-0.01	-0.20
	(-0.25 - 0.24)	(-0.13 - 0.23)	(-0.17 - 0.15)	(-0.47 - 0.08)
Round 5	0.12	0.06	0.02	-0.18
	(-0.06 - 0.31)	(-0.09 - 0.21)	(-0.11 - 0.15)	(-0.47 - 0.10)
2nd quartile * Round 3	0.14	-0.09	0.06	0.11*
	(-0.06 - 0.33)	(-0.30 - 0.11)	(-0.09 - 0.22)	(-0.01 - 0.22)
2nd quartile * Round 4	0.09	-0.11	-0.08	0.13
	(-0.16 - 0.33)	(-0.31 - 0.08)	(-0.26 - 0.09)	(-0.08 - 0.34)
2nd quartile * Round 5	0.14	-0.17*	-0.17**	0.01
	(-0.06 - 0.35)	(-0.34 - 0.01)	(-0.33 - -0.02)	(-0.16 - 0.18)
3rd quartile * Round 3	0.26**	-0.18*	0.22**	0.20
	(0.06 - 0.46)	(-0.39 - 0.02)	(0.04 - 0.40)	(-0.07 - 0.48)
3rd quartile * Round 4	0.12	-0.13	0.06	0.19
	(-0.09 - 0.33)	(-0.32 - 0.06)	(-0.13 - 0.25)	(-0.08 - 0.46)
3rd quartile * Round 5	0.17*	-0.13	0.02	0.15
	(-0.03 - 0.38)	(-0.30 - 0.03)	(-0.15 - 0.19)	(-0.10 - 0.41)
4th quartile * Round 3	0.38***	-0.16	0.29***	0.27*
	(0.11 - 0.65)	(-0.37 - 0.04)	(0.12 - 0.47)	(-0.03 - 0.57)
4th quartile * Round 4	0.16	0.03	0.14	0.27*
	(-0.10 - 0.42)	(-0.16 - 0.23)	(-0.06 - 0.34)	(-0.03 - 0.57)
4th quartile * Round 5	0.15	0.00	0.10	0.29*
	(-0.09 - 0.39)	(-0.18 - 0.19)	(-0.11 - 0.32)	(-0.02 - 0.59)
Constant	-0.03	-0.20***	-0.29***	-0.17
	(-0.36 - 0.30)	(-0.33 - -0.07)	(-0.40 - -0.17)	(-0.41 - 0.06)
Observations	5,740	7,132	6,748	6,984
R-squared	0.16	0.03	0.11	0.10

*** p<0.01, ** p<0.05, * p<0.1. All estimates adjust for the language of the test and their interaction with survey rounds. Robust confidence intervals are reported in parentheses. Standard errors are clustered at each country's community level.

F.2 Analysis based on vocabulary scores

Descriptives

Figure F.1: Distribution of maternal education (in **vocabulary** sample)



Note: Vertical dashed lines on the left and right indicate cutoff lines for low and high SES households, respectively.

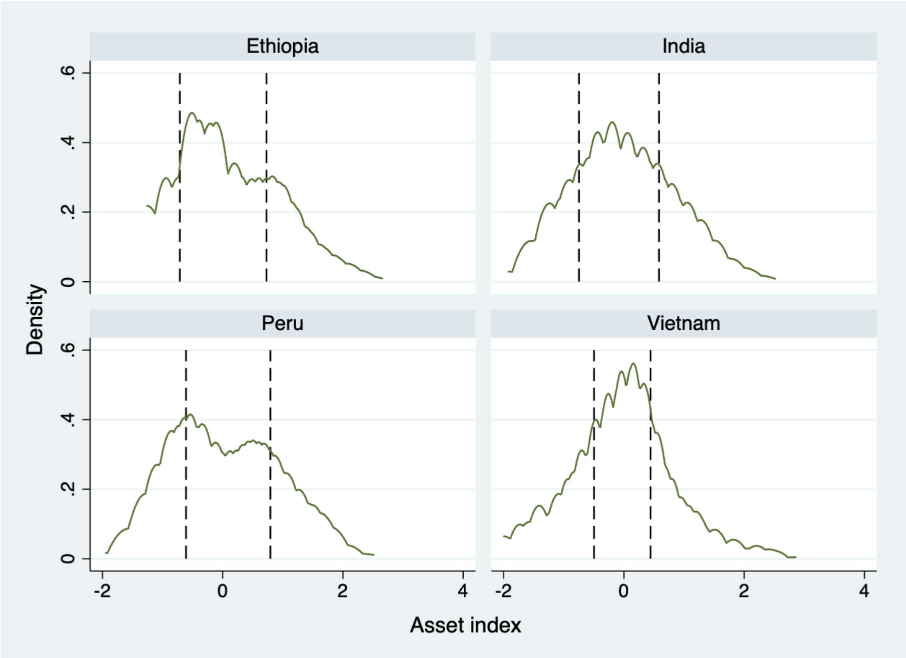
Table F.4: Socioeconomic status group size (**vocabulary** score sample)

	Ethiopia		India		Peru		Vietnam	
	Low	High	Low	High	Low	High	Low	High
Group size (n)								
Maternal education	919	80	1091	321	178	615	417	195
Asset index	423	323	470	381	538	326	488	371
Expenditure	373	372	443	442	401	400	389	388

**Education: none versus upper secondary and above; asset index and expenditure: bottom and top quartile groups*

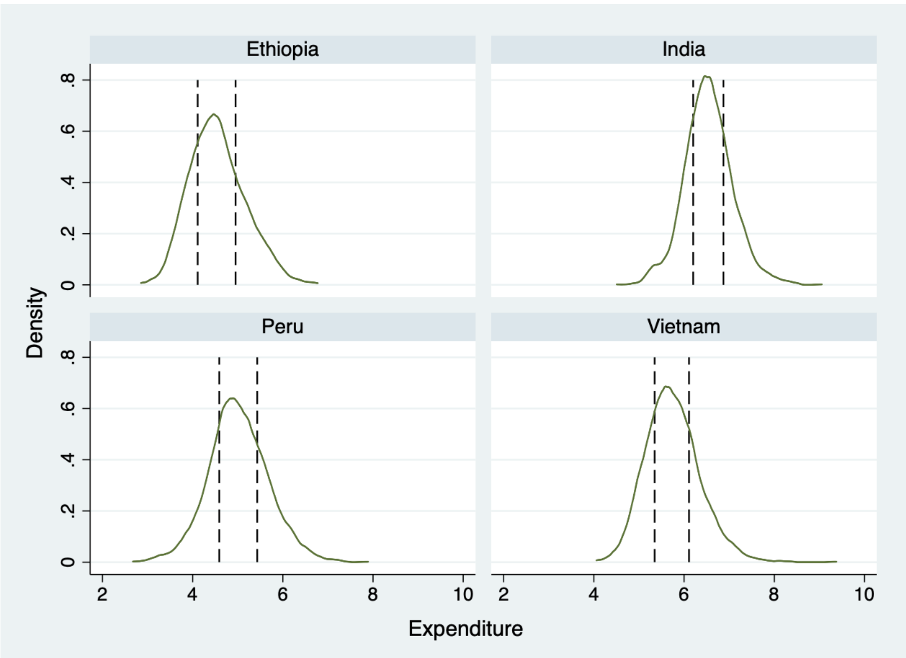
Why are the quartile groups in the maths (and also in the vocabulary samples) different sizes? As low and high SES groups are classified based on quartile groups, we would expect each group to be approximately the same size. This is roughly the case across all countries. The reasons they are not exactly the same size is because, when Stata 17 put the children in separate bins (i.e quartile groups), it groups observations with the same original variable into the same bin (a sensible approach). The differences in group sizes is a result of groups of children with the same value of the asset index around the 25th and 75th percentile of the asset index distribution. Given the number of children are roughly the same in each group, and

Figure F.2: Distribution of asset index (in **vocabulary** sample)



Note: Vertical dashed lines on the left and right indicate cutoff lines for low and high SES households, respectively.

Figure F.3: Distribution of expenditure (in **vocabulary** sample)



Note: Vertical dashed lines on the left and right indicate cutoff lines for low and high SES households, respectively.

the groups are quite large, I do not believe this will have a substantive impact on my results. Nor would the truncation in Ethiopia affect my results, as my aim is to identify those in the lowest quartile group, not evaluate any variation inside the bottom quartile group.

Gaps in attainment for each measure of socioeconomic status

Figure F.4: Socioeconomic gaps in vocabulary score, based on maternal education at age 1 year old

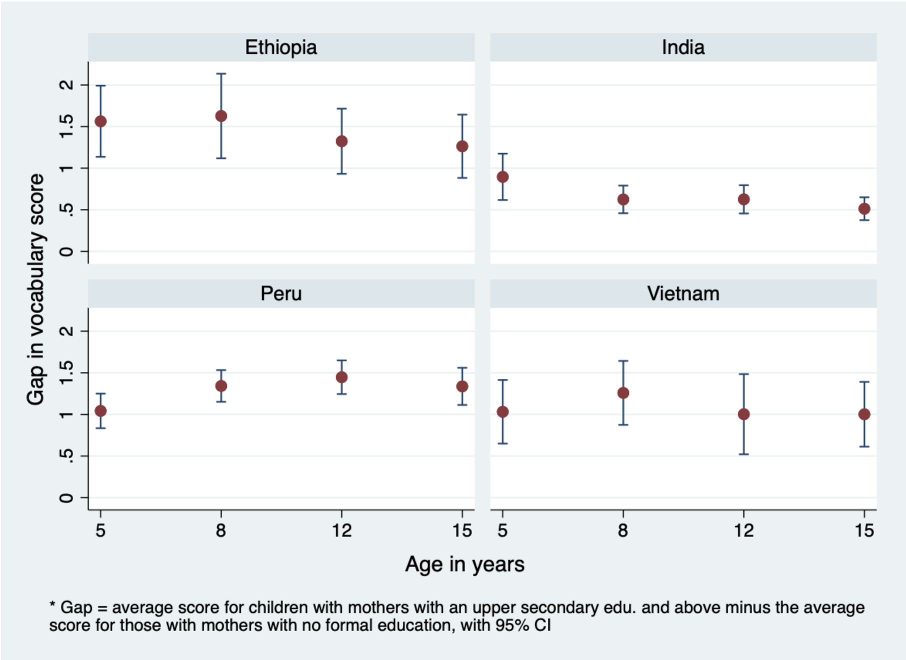


Figure F.5: Socioeconomic gaps in vocabulary score, based on asset index quartile groups at age 1 year old

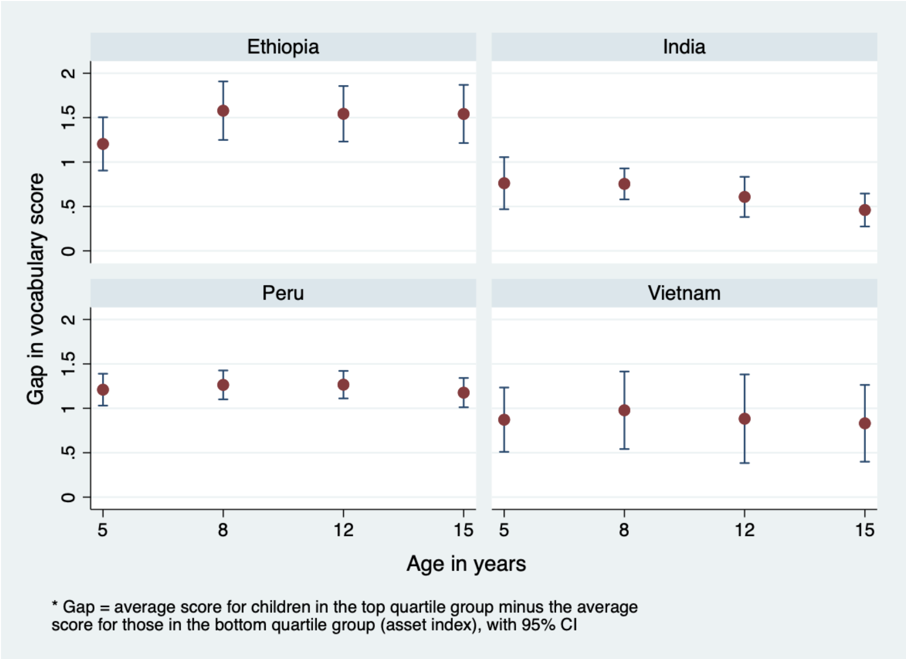


Figure F.6: Socioeconomic gaps in vocabulary score, based on real expenditure per capita quartile groups at age 5 year old

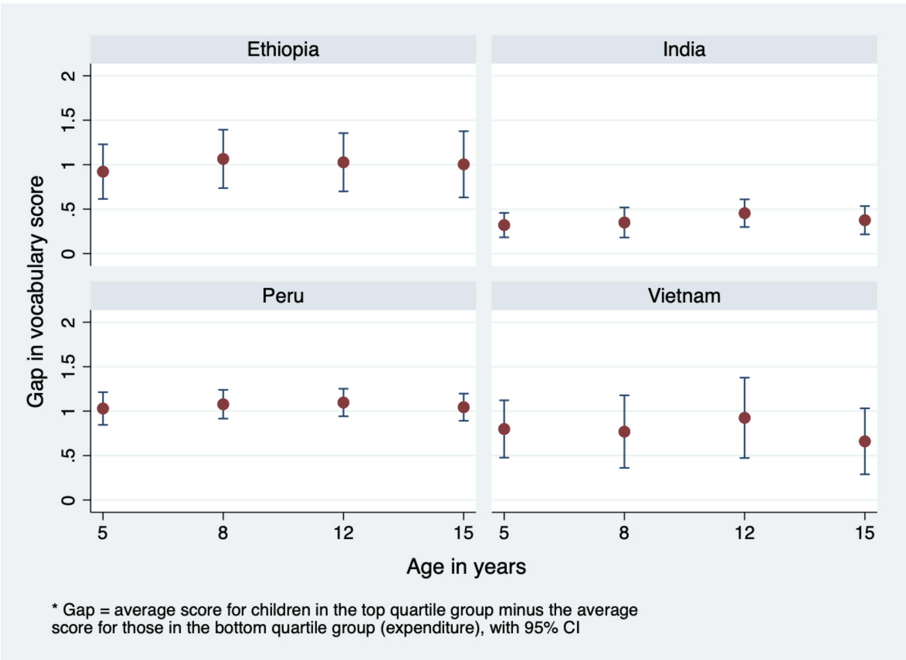


Table F.5: SES gaps in vocabulary scores (age-specific z-scores, s.d.), with 95% CI

	Ethiopia	**India**	**Peru**	**Vietnam**
Mother's education				
5 years old	1.56	0.90	1.04	1.03
	(1.14-1.99)	(0.62-1.17)	(0.83-1.25)	(0.65-1.41)
8 years old	1.63	0.62	1.34	1.26
	(1.12-2.13)	(0.46-0.79)	(1.15-1.53)	(0.87-1.64)
12 years old	1.32	0.63	1.45	1.00
	(0.93-1.72)	(0.46-0.8)	(1.25-1.65)	(0.52-1.48)
15 years old	1.26	0.51	1.34	1.00
	(0.88-1.64)	(0.37-0.65)	(1.11-1.56)	(0.61-1.39)
Asset index				
5 years old	1.20	0.76	1.03	1.23
	(0.9-1.5)	(0.47-1.06)	(1.03-1.39)	(0.51-1.23)
8 years old	1.58	0.75	1.10	1.41
	(1.25-1.91)	(0.58-0.93)	(1.1-1.43)	(0.54-1.41)
12 years old	1.54	0.61	1.11	1.38
	(1.23-1.86)	(0.38-0.83)	(1.11-1.42)	(0.38-1.38)
15 years old	1.54	0.46	1.01	1.26
	(1.21-1.87)	(0.27-0.65)	(1.01-1.34)	(0.4-1.26)
Expenditure				
5 years old	0.92	0.32	1.03	0.80
	(0.61-1.23)	(0.18-0.46)	(0.85-1.21)	(0.48-1.12)
8 years old	1.06	0.35	1.08	0.77
	(0.74-1.39)	(0.18-0.52)	(0.92-1.24)	(0.36-1.18)
12 years old	1.03	0.45	1.10	0.92
	(0.7-1.36)	(0.3-0.61)	(0.94-1.25)	(0.47-1.38)
15 years old	1.00	0.38	1.04	0.66
	(0.63-1.38)	(0.22-0.53)	(0.89-1.2)	(0.29-1.03)

Note: SES gap when using maternal education as a measure for SES is the average score for children with mothers with an upper secondary education and above minus the average score for those with mothers with no formal education. The SES gap using the asset index and expenditure as a measure for SES is the average score for children in the top quartile group minus the average score for those in the bottom quartile group.

F.3 Sensitivity analysis

Missing values

I conduct a robustness check to see how my results would change if I took missing test scores data into account. I test the worst case scenario and I assume all children with missing test scores received a poor score of minus one of a standard deviation (-1 s.d.). Sample sizes are now 1,648 for Ethiopia, 1,788 for India, 1,704 for Peru and 1,748 for Vietnam for the maths scores. Sample sizes are now 1,528 for Ethiopia, 1,787 for India, 1,683 for Peru and 1,644 for Vietnam for the vocabulary scores.

Maths scores

Figure F.7: Sensitivity analysis for missing values: Socioeconomic gaps in maths score, based on maternal education at age 1 year old

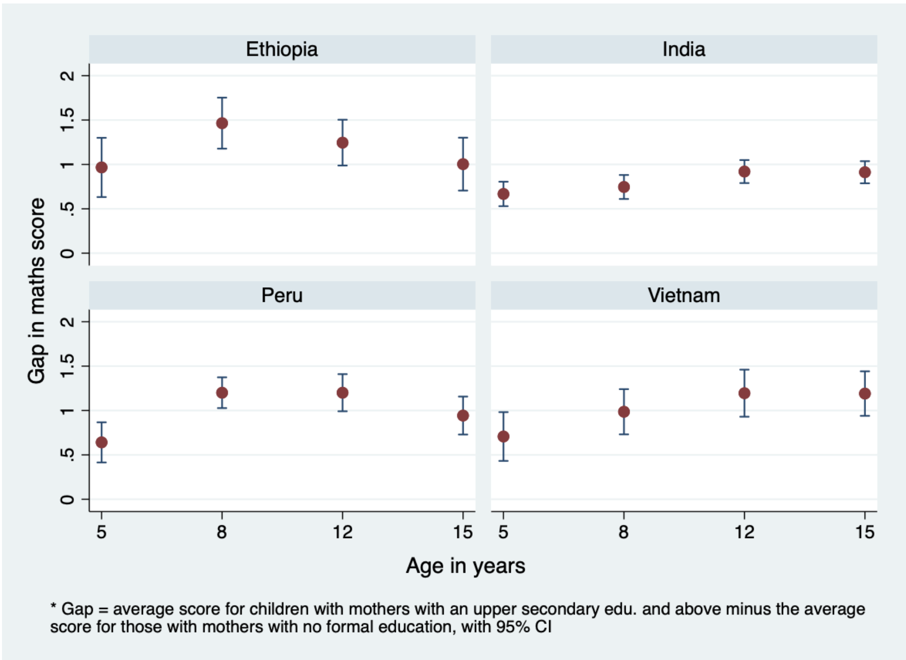


Figure F.8: Sensitivity analysis for missing values: Socioeconomic gaps in maths score, based on asset index quartile groups at age 1 year old

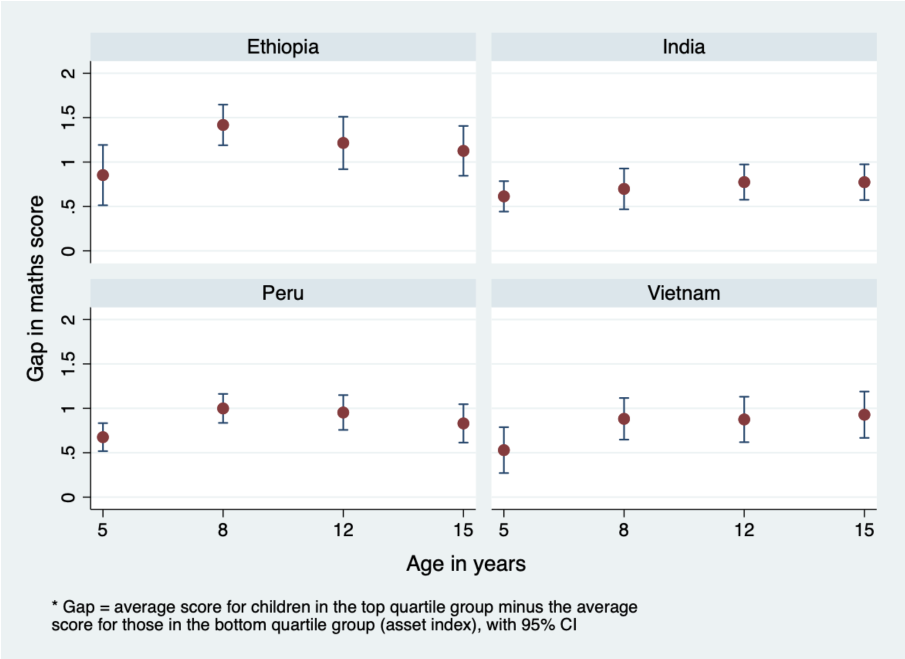
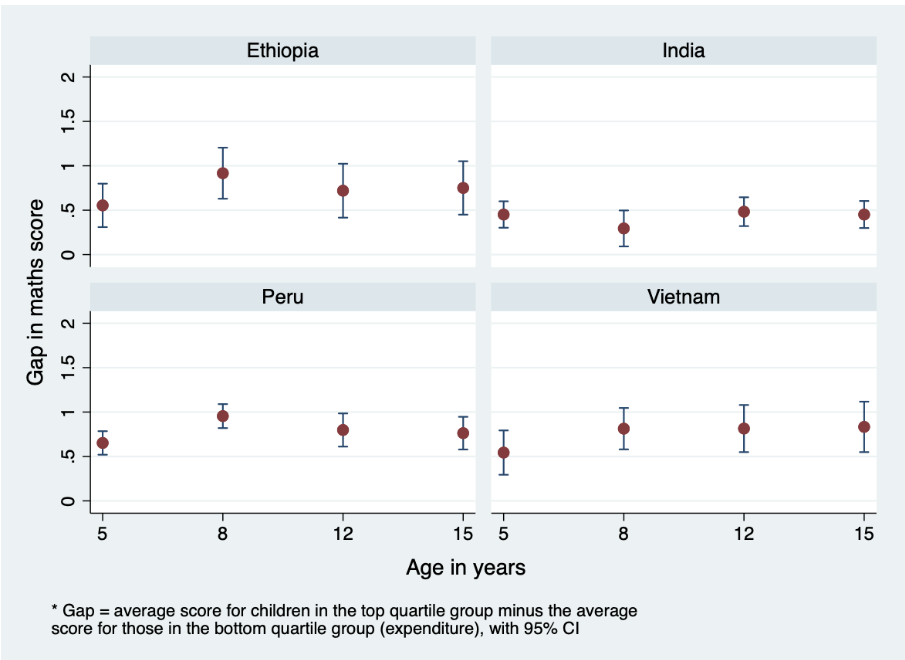


Figure F.9: Sensitivity analysis for missing values: Socioeconomic gaps in maths score, based on real expenditure per capita quartile groups at age 5 year old



Vocabulary scores

Figure F.10: Sensitivity analysis for missing values: Socioeconomic gaps in vocabulary score, based on maternal education at age 1 year old

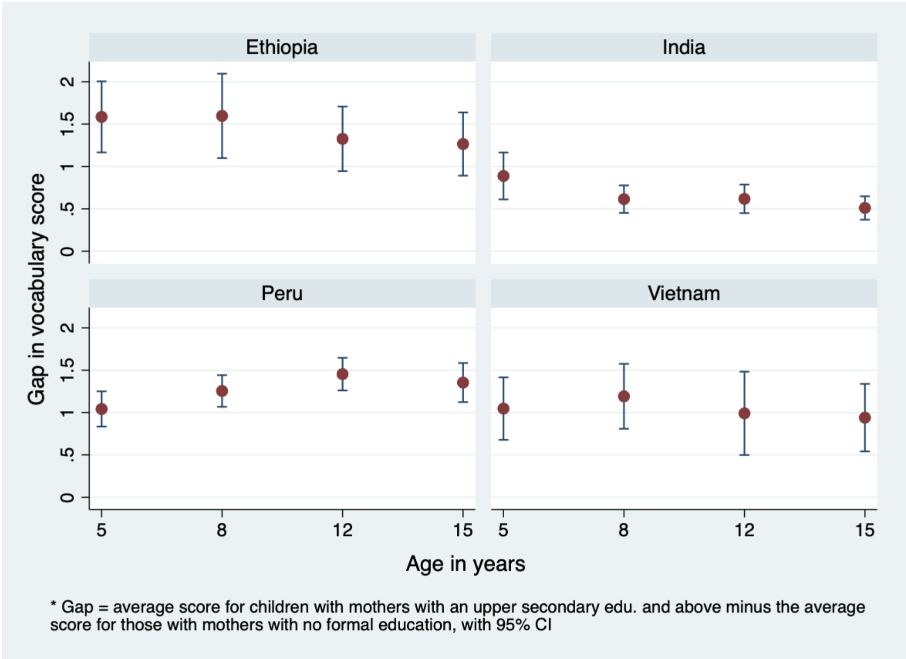


Figure F.11: Sensitivity analysis for missing values: Socioeconomic gaps in vocabulary score, based on asset index quartile groups at age 1 year old

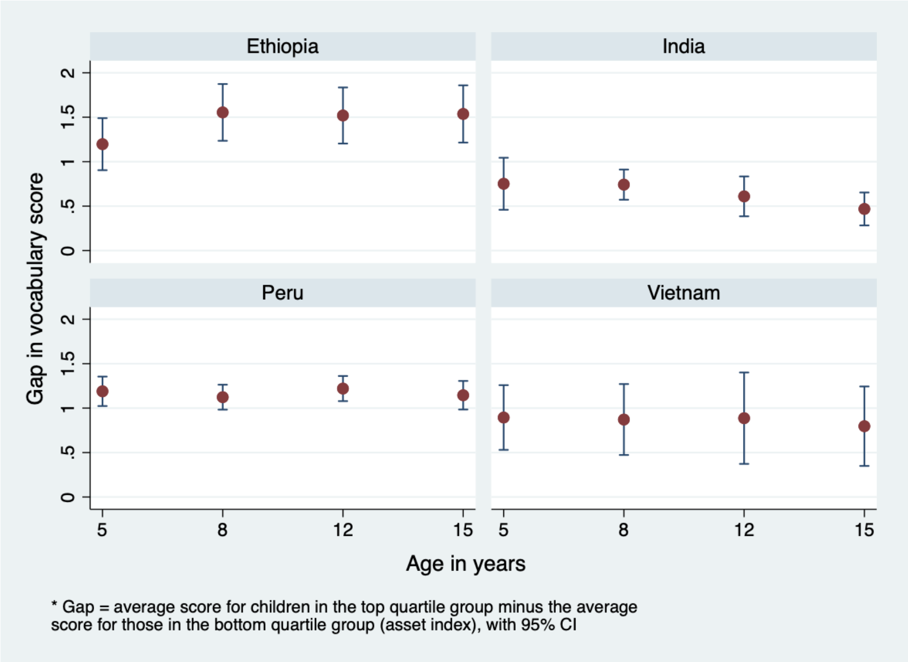
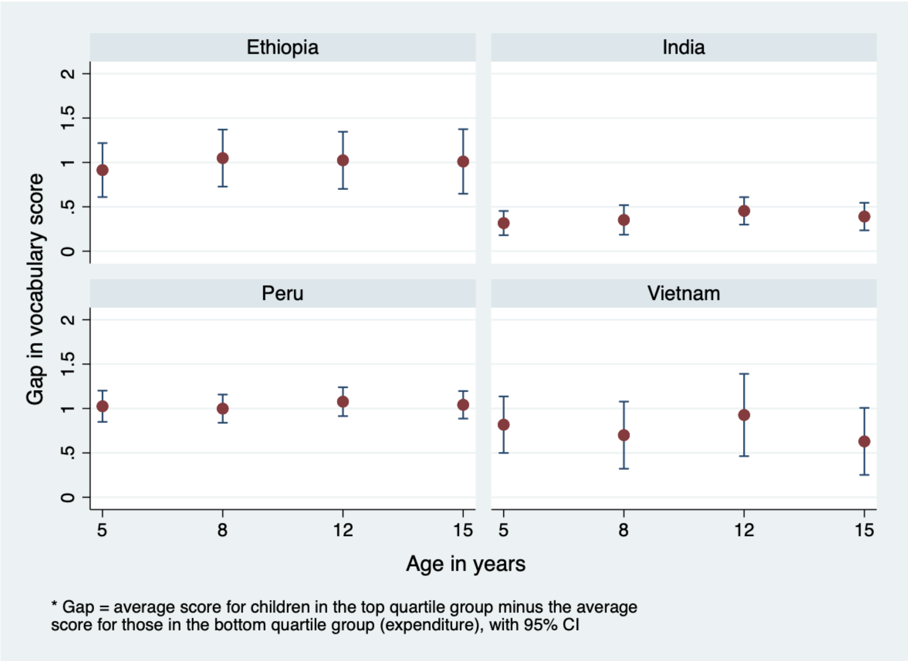


Figure F.12: Sensitivity analysis for missing values: Socioeconomic gaps in vocabulary score, based on real expenditure per capita quartile groups at age 5 year old



Parental education

I conduct some sensitivity analysis, where I use the mother and father's average education level as a measure of SES, and the results are similar (see Figures F.13 and F.14, and Table F.7). Indeed, the correlation between mother's and father's education, as reported in Table F.6, are reasonably high, and range from 0.60 in India to 0.71 in Ethiopia. These correlations point to a sizeable portions of parents engaging in assortative mating.

Table F.6: Correlation between measures parental education, for available sample

	Ethiopia	**India**	**Peru**	**Vietnam**
Mother's & father's education	0.71	0.60	0.67	0.65
Mother's & average of mother's and father's education	0.91	0.87	0.92	0.90
Father's & average of mother's and father's education	0.94	0.92	0.91	0.91
n	1705	1982	1574	1899

Figure F.13: Sensitivity analysis for mother's education: Socioeconomic gaps in **maths** score, based on parental education at age 1 year old

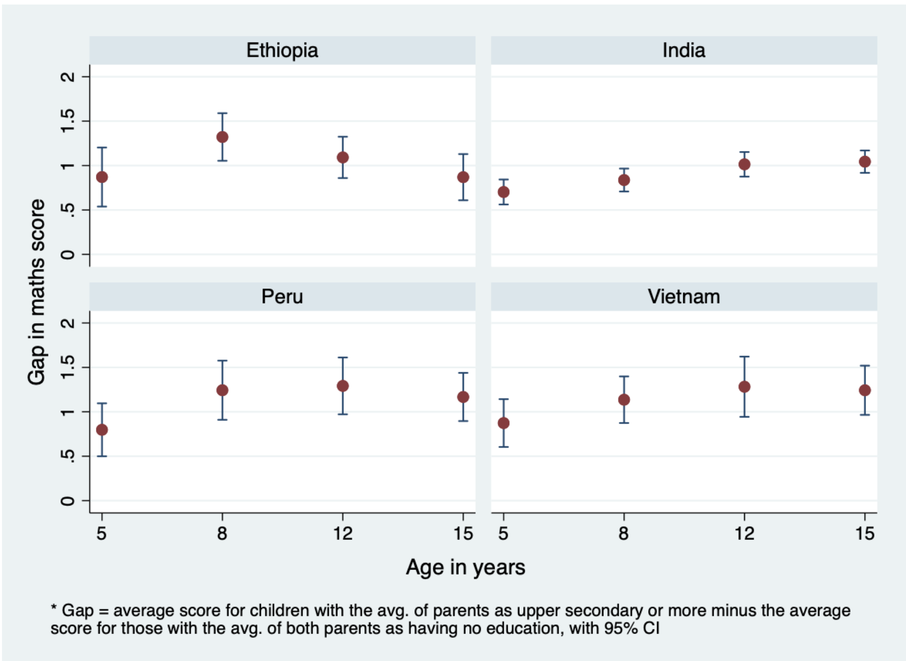


Figure F.14: Sensitivity analysis for mother's education: Socioeconomic gaps in **vocabulary** score, based on parental education at age 1 year old

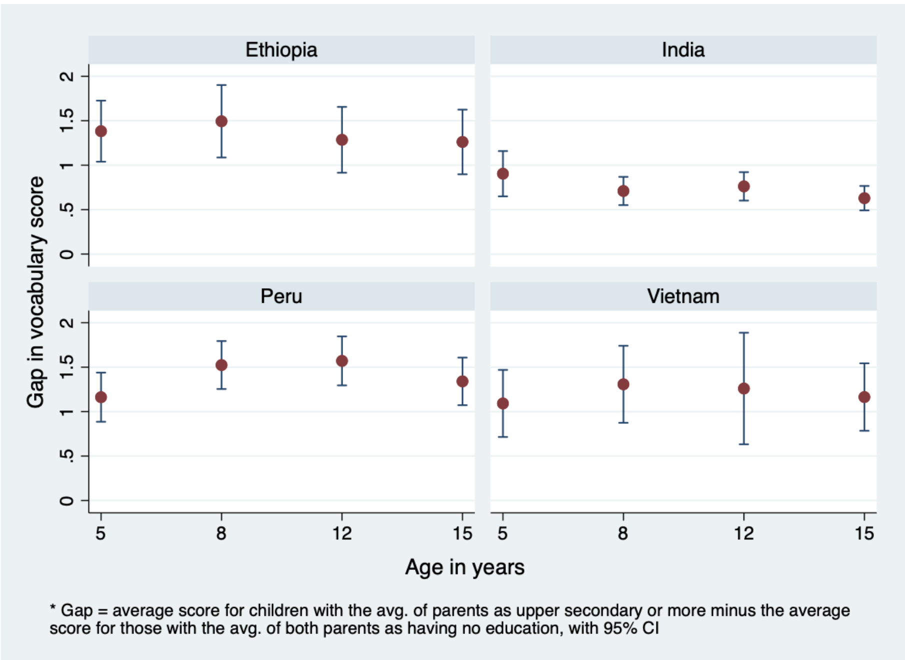


Table F.7: Sensitivity analysis for missing data: SES gaps when SES is measured using an average of mother's and father's education (age-specific z-scores, s.d.), with 95% CI

	Ethiopia	**India**	**Peru**	**Vietnam**
Maths scores				
5 years old	0.87	0.70	0.80	0.87
	(0.54-1.2)	(0.56-0.84)	(0.5-1.1)	(0.6-1.14)
8 years old	1.32	0.84	1.24	1.14
	(1.05-1.59)	(0.71-0.97)	(0.91-1.58)	(0.87-1.4)
12 years old	1.09	1.01	1.29	1.28
	(0.86-1.32)	(0.88-1.15)	(0.97-1.61)	(0.94-1.62)
15 years old	0.87	1.04	1.17	1.24
	(0.61-1.13)	(0.92-1.17)	(0.9-1.44)	(0.97-1.52)
Vocabulary scores				
5 years old	1.38	0.90	1.16	1.09
	(1.04-1.73)	(0.65-1.16)	(0.89-1.44)	(0.71-1.47)
8 years old	1.49	0.71	1.52	1.31
	(1.09-1.9)	(0.55-0.87)	(1.25-1.79)	(0.87-1.74)
12 years old	1.29	0.76	1.57	1.26
	(0.92-1.66)	(0.6-0.92)	(1.3-1.85)	(0.63-1.89)
15 years old	1.26	0.63	1.34	1.16
	(0.9-1.63)	(0.49-0.77)	(1.07-1.61)	(0.78-1.54)

Note: SES gap when using the average of maternal and paternal education as a measure of SES is the average score for children with parents with an upper secondary education and above minus the average score for those with parents with no formal education.

Appendix G

Annex: Chapter 8

G.1 Distribution of maths and vocabulary scores at 5 years old

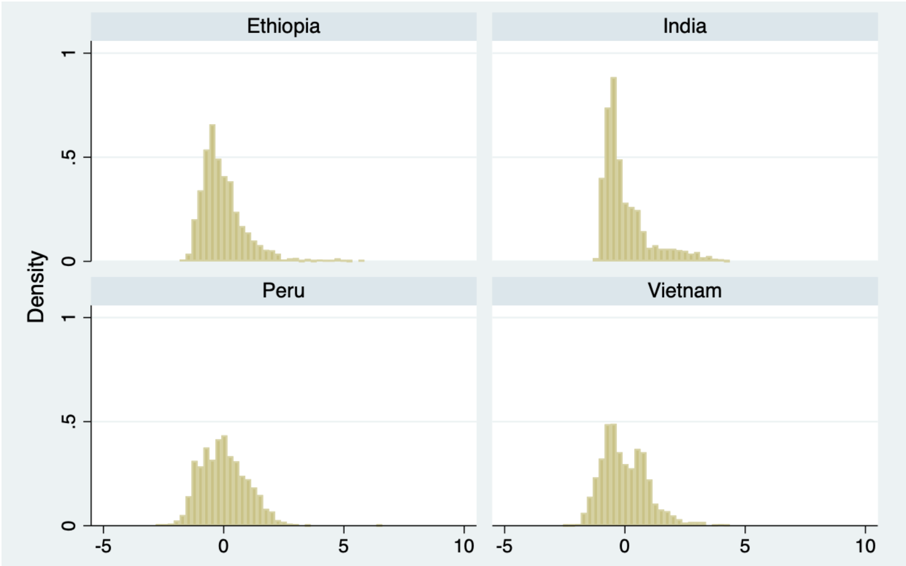


Figure G.1: Distribution of vocabulary scores at 5 years old

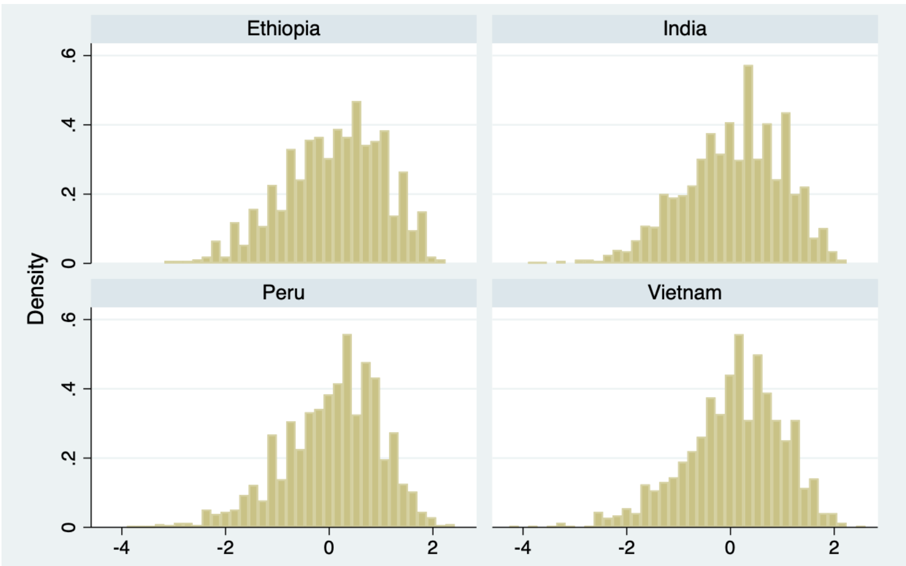


Figure G.2: Distribution of maths scores at 5 years old

When calculating quartile groups with the asset index, I would expect each group to be approximately the same size. This is not the case across all countries. The reasons they are not exactly the same size is because, when Stata 15 put the children in separate bins (i.e quartile groups), it groups observations with the same original variable into the same bin (a sensible approach). The differences in group sizes is a result of groups of children with the same value of the asset index around the 25th and 75th percentile of the asset index distribution. Ethiopia's asset index distribution is truncated at the bottom (see Figures G.2 and G.1 for the distribution). This has resulted in a large bottom quartile (with 619 children) compared to the top quartile with 347 children (from a sample of 1,477 children), as is reported in Table 8.2.

In Ethiopia the asset index is truncated on the left. I checked against the asset index that the Young Lives team calculates using an additive rather than an IRT approach, and find that Ethiopia's asset index is also truncated on the left, indicating that it is not the methodology that causes the truncation. In 2001 Ethiopia was the poorest country in the study, and it appears that the information collected to compile the asset index did not capture variation amongst the lowest SES children in sample.

G.2 Descriptive tables

Table G.1: Children not present in survey, Round 5

	Ethiopia	**India**	**Peru**	**Vietnam**
Quartile groups				
Bottom	50	24	52	21
Second	53	16	48	5
Third	30	39	49	7
Top	54	32	43	29
Attrition	*187*	*111*	*192*	*62*

G.3 Sensitivity analysis

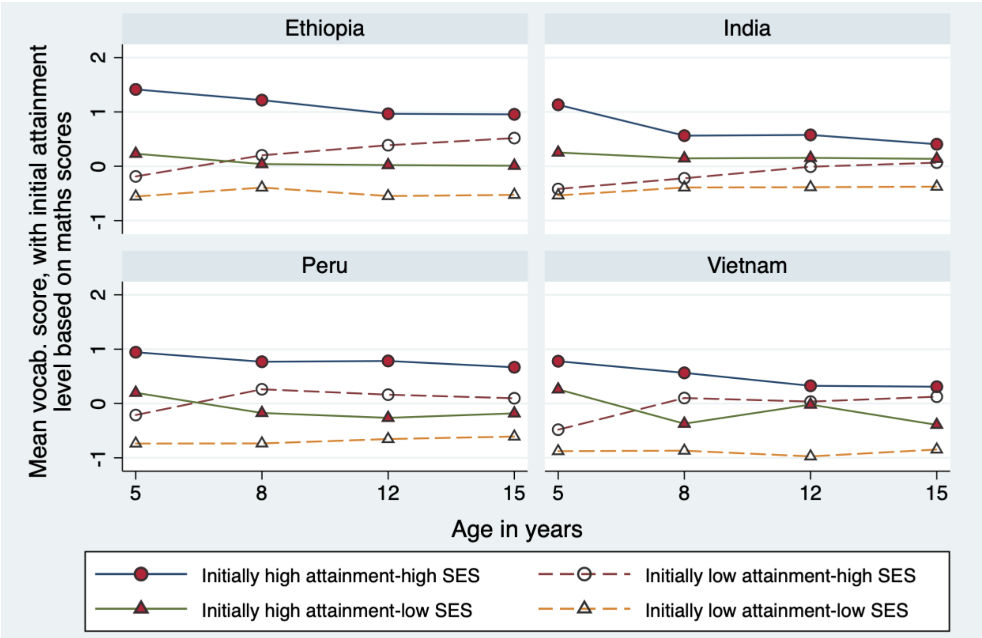
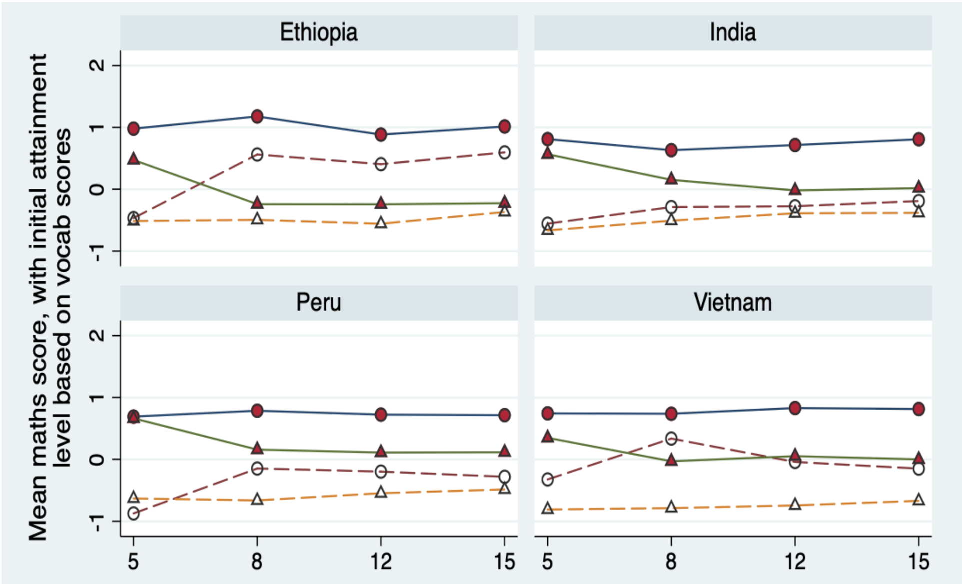
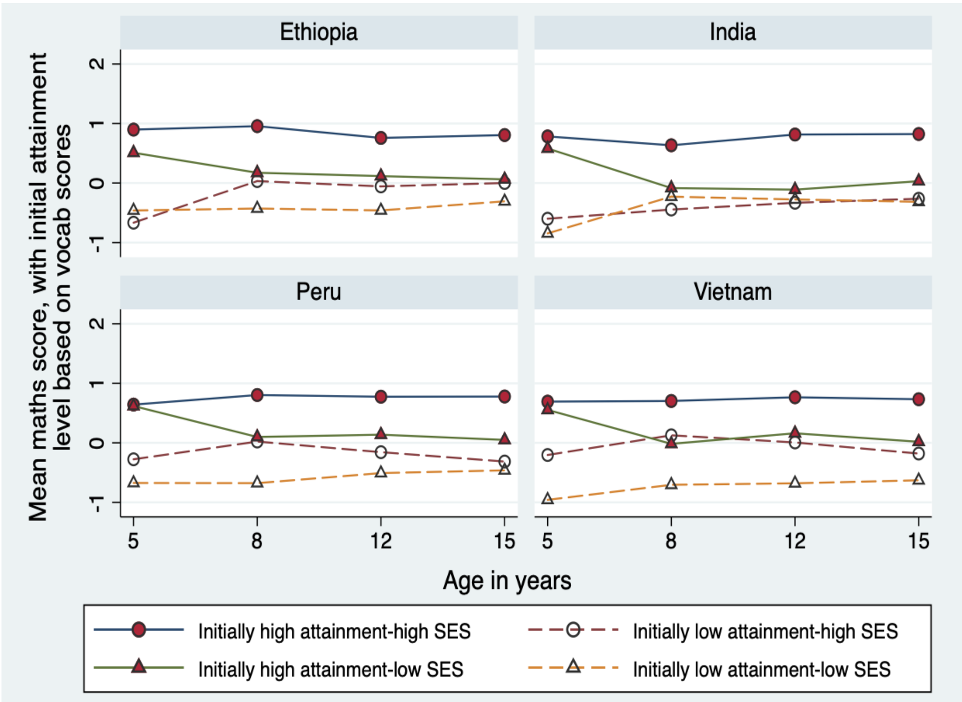


Figure G.3: Trajectories in vocabulary scores from five to 15 years old by initial attainment (defined using maths scores at five years old), using mother's education at one year old as a measure of SES

Notes: Initial attainment based on the top and bottom quartile groups in maths scores. SES based on maternal education. Ethiopia, India and Vietnam: low SES = none, high SES = lower secondary and above. Peru: low SES = none or primary, high SES = upper secondary and above.

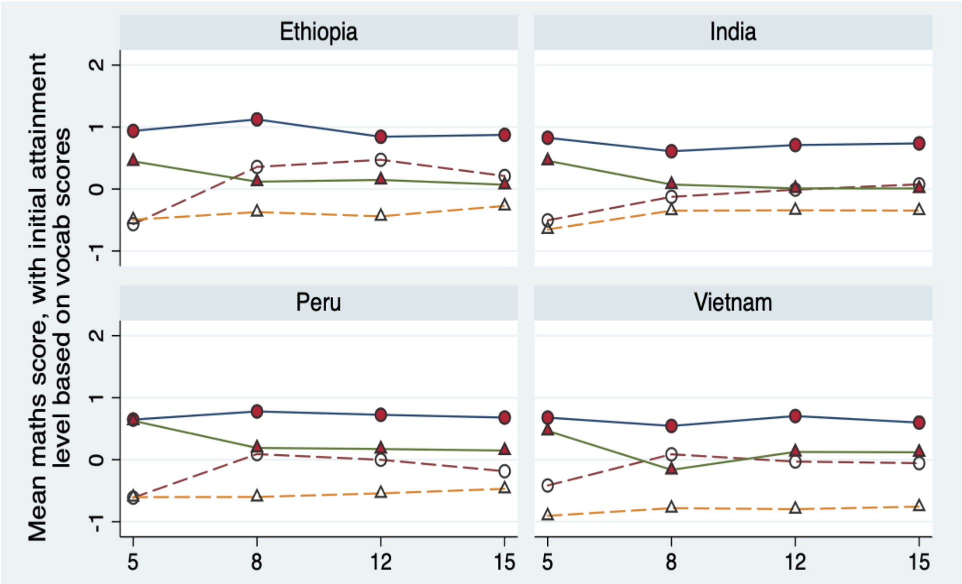


(a) SES measured with the asset index at one year old

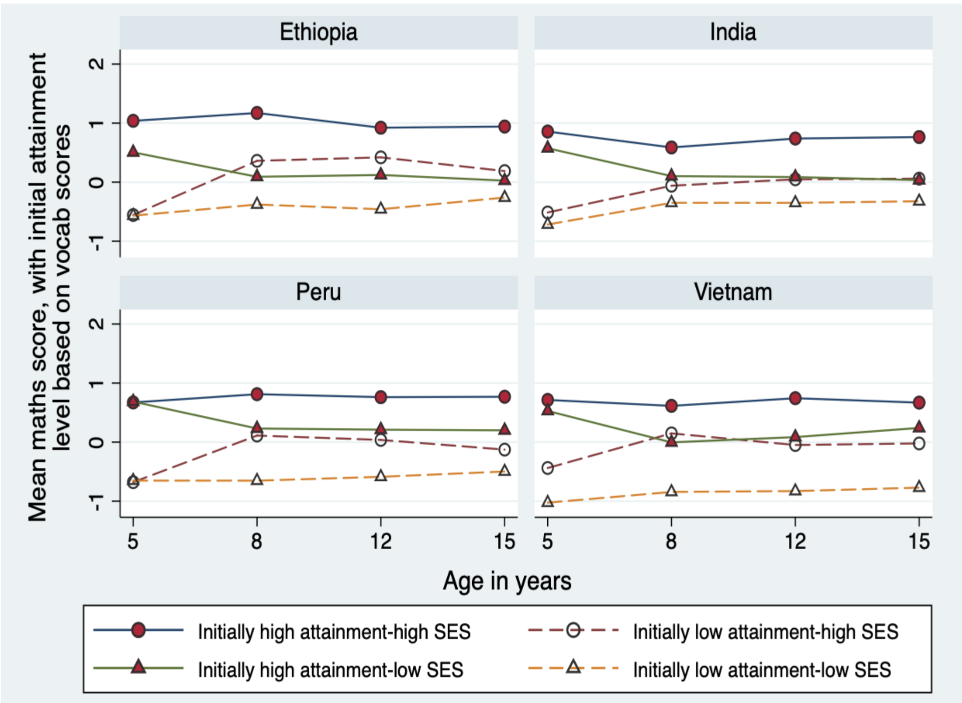


(b) SES measured with expenditure per capita at five years old

Figure G.4: Trajectories in maths scores from five to 15 years old by initial attainment (defined using vocabulary scores at five years old) for the top and bottom SES groups
Notes: Initial attainment groups represent the top and bottom quartile groups. For the asset index and expenditure, low SES comprises of the bottom quartile group, and high SES the top quartile group.



(a) Tercile groups



(b) Quintile groups

Figure G.5: Trajectories in maths scores from five to 15 years old by initial attainment (defined using vocabulary scores at five years old) for the top and bottom tercile and quintile groups of vocabulary scores

Notes: SES based on maternal education. Ethiopia, India and Vietnam: low SES = none, high SES = lower secondary and above. Peru: low SES = none or primary, high SES = upper secondary and above.

Table G.2: Group sizes for Figure G.3 (Trajectories in vocabulary scores from five to 15 years old by initial attainment (defined using maths scores at five years old), using mother's education at one year old as a measure of SES)

	Ethiopia	**India**	**Peru**	**Vietnam**
Initially high attainment - high SES	177	158	145	119
Initially low attainment - high SES	32	57	49	43
Initially high attainment - low SES	43	82	69	46
Initially low attainment - low SES	127	158	144	173
N	*1,514*	*1,774*	*1,645*	*1,570*

Notes: Initial attainment based on the top and bottom quartile groups in maths scores. SES based on top and bottom quartile groups of the asset index at 1 year old.

Table G.3: Group sizes for Figure G.4 (Trajectories in maths scores from five to 15 years old by initial attainment (defined using vocabulary scores at five years old) for the top and bottom SES groups)

	Ethiopia	**India**	**Peru**	**Vietnam**
(a) Asset index				
Initially high attainment - high SES	178	166	223	170
Initially low attainment - high SES	27	46	23	47
Initially high attainment - low SES	68	75	53	69
Initially low attainment - low SES	221	163	272	190
(b) Expenditure per capita				
Initially high attainment - high SES	177	140	195	173
Initially low attainment - high SES	47	72	29	36
Initially high attainment - low SES	46	96	45	50
Initially low attainment - low SES	132	139	203	164
Taken from sample, N	*1,477*	*1,721*	*1,715*	*1,612*

Notes: Initial attainment based on the top and bottom quartile groups in vocabulary scores. SES is based on the (a) asset index and (b) expenditure per capita. Low SES means the bottom quartile group, and high SES the top quartile group.

Table G.4: Group sizes for Figure G.5

	Ethiopia	**India**	**Peru**	**Vietnam**
(a) Tercile groups				
Initially high attainment - high SES	258	257	328	266
Initially low attainment - high SES	67	102	47	104
Initially high attainment - low SES	111	165	77	140
Initially low attainment - low SES	283	282	343	298
(b) Quintile groups				
Initially high attainment - high SES	118	119	152	125
Initially low attainment - high SES	13	23	13	13
Initially high attainment - low SES	29	58	24	28
Initially low attainment - low SES	103	131	168	132
N	*1,477*	*1,721*	*1,716*	*1,612*

Notes: Initial attainment based on the top and bottom quartile groups in (a) maths and (b) vocabulary scores. SES based on top and bottom quartile groups of the asset index at 1 year old.

Appendix H

Annex: Chapter 9

H.1 Community clusters

Table H.1: Young Lives sites in Ethiopia by region, with percentage of children attending early childhood education*

No.	Zone	Description	%
Capital city			
1	Addis Ababa	Overcrowded area in the centre of the city	99
2	Addis Ababa	Industrial area in the southern part of the city	94
3	Addis Ababa	Slum area in the city	92
Amhara			
4	North Wello	Tourist town with some extremely poor neighbourhoods	41
5	North Wello	Poor rural community	1
6	South Gondar	Rural area near Lake Tana	0
7	South Gondar	Rural food-insecure area	0
Oromia			
8	East Shewa	Rural area near Lake Ziway	4
9	Arsi	Drought-prone rural area	1
10	North Shewa	Fast-growing town	39
11	East Shewa	Relatively rich rural area on the outskirts of Debrezeit town	17
Southern Nations, Nationalities and Peoples Region			
12	Gurage	Densely populated rural area growing enset ("false banana")	6
13	Wolayita	Densely populated town	7
14	Hawassa City Administration Zone	Fast-growing business and tourist town	86
15	Sidama	Coffee-growing rural area	12
16	Hadiya	Poor and densely populated rural community	12
Tigray			
17	Southern Tigray	Drought-prone rural area highly dependent on government support	0
18	Eastern Tigray	Extremely poor rural area dependent on the Productive Safety Net Programme and other government support	0
19	Eastern Tigray	Small, very poor urban town	8
20	Eastern Tigray	Model rural area known for its success in soil and water conservation	0

Source: Adapted from Alemu et al. (2003)

Notes: Percentage of children from the maths sample attending early childhood education in each site, author's calculations

Table H.2: Young Lives sites in India by region, with percentage of children attending early childhood education*

No.	Zone	Description	%
Coastal Andhra (Andhra Pradesh)			
1	West Godavari	Urban area in a well-developed coastal region	80
2	West Godavari	Tribal mandal in a well-developed coastal district	90
3	Srikakulam	Town in the north	97
4	Srikakulam	Tribal mandal in the north	85
5	Srikakulam	Rural mandal in the north	99
6	Srikakulam	Rural mandal in the north	76
7	Srikakulam	Rural mandal with a mix of tribes and non-tribes in the north	99
Rayalaseema (Andhra Pradesh)			
8	Kadapa	Rural mandal in the heart of the region where agriculture is the main occupation	86
9	Kadapa	Remote rural mandal in a forested part of the region	66
10	Anantapur	Urban site, which is a district headquarters	95
11	Anantapur	Poor rural mandal affected by Naxalite movements	93
12	Anantapur	Poor rural area spread across hilly areas and affected by Naxalite movements	83
13	Anantapur	Rural mandal bordering the neighbouring state	95
Telangana			
14	Karimnagar	Medium-sized town in the north with people of mixed religion	95
15	Karimnagar	Rural area in the north affected by Naxalite movements	90
16	Mababubnagar	Rural tribal mandal in the forest areas of the south	86
17	Mababubnagar	Rural mandal in the south with people moving in seasonal migration	89
18	Mababubnagar	Rural mandal in the south with high incidence of child labour and seasonal migration	88
19	Mababubnagar	Very poor mandal in the south	87
State capital			
20	Hyderabad	Densely crowded area in the state capital	81

Source: Adapted from Kumra (2008) and Young Lives (2014b)

Notes: Percentage of children from the maths sample attending early childhood education in each site, author's calculations

Table H.3: Young Lives sites in Peru by region, with percentage of children attending early childhood education*

No.	Zone	Description	%
1	Tumbes	Small city on the northern coast	100
2	Piura	Poor coastal rural area	91
3	Piura	Very poor rural area in the northern Andean highlands	88
4	Amazonas	Very poor rural area in the north of the country	89
5	San Martin	Poor rural area	58
6	San Martin	Medium-sized city	78
7	Cajamarca	Medium-sized city in the northern Andean highlands	100
8	La Libertad	Shanty town on the outskirts of a medium-sized city on the northern coast	90
9	Ancash	Poor rural area in the central Andean highlands	70
10	Ancash	Medium-sized city in the central Andean highlands	95
11	Huánuco	Very rural area in the centre of the Andean highlands	76
12	Lima	Large urban district located in the north of the city	99
13	Lima	Large urban district located in the eastern part of the city	91
14	Lima	Large urban district located in the south of the capital city	98
15	Junín	Poor rural area in the Amazon	61
16	Ayacucho	Very rural poor community in the southern-centre of the Andean highlands	70
17	Ayacucho	Poor rural area in the southern-centre of the Andean highlands	90
18	Apurimac	Poor rural area in the southern Andean highlands	94
19	Arequipa	Small city on the southern coast	93
20	Puno	Medium-sized city in the southern Andean highland	72

Source: Adapted from Young Lives (2014c)

Notes: Percentage of children from the maths sample attending early childhood education in each site, author's calculations

Table H.4: Young Lives sites in Vietnam by region, with percentage of children attending early childhood education*

No.	Zone	Description	%
South Central Coast			
1	Phu Yen	Inland flood-prone rural community with high rate of poverty in 2002 but has improved since and is now not so poor	94
2	Phu Yen	Coastal community with average rate of poverty	92
3	Phu Yen	Very poor mountainous community with mostly ethnic minority groups	88
4	Phu Yen	Relatively prosperous coastal community, with shrimp farming	89
Mekong River Delta			
5	Ben Tre	Poor flood-prone coastal area with difficult transport links	87
6	Ben Tre	Inland area with a slightly above-average poverty rate	85
7	Ben Tre	Inland flood-prone area, with difficult transportation but a relatively low poverty rate	79
8	Ben Tre	Relatively prosperous inland area with good transport links	82
North-East			
9	Lao Cai	Among the poorest mountainous communities in Lao Cai province, with mostly ethnic minority groups, very difficult transportation and little infrastructure	81
10	Lao Cai	Very poor mountainous area, with mostly ethnic minority groups and underdeveloped infrastructure	100
11	Lao Cai	Poor mountainous area with mixed ethnic groups	94
12	Lao Cai	Very poor mountainous area, with mixed ethnic groups and underdeveloped infrastructure	91
Red River Delta			
13	Hung Yen	Prosperous rural area, with high population density and good infrastructure	100
14	Hung Yen	Poor rural area, near a major city and with good infrastructure	100
15	Hung Yen	Rural rice-producing community, with good infrastructure	99
16	Hung Yen	Poor rural area, with a high population density and good transport infrastructure	100
Cities			
17	Da Nang	Urban neighbourhood with mostly blue-collar labour and average infrastructure	96
18	Da Nang	Mostly prosperous urban area with very good access to services	97
19	Da Nang	Relatively poor suburb, with quite poor environmental conditions and transportation	96
20	Da Nang	Newly developed urban and fishing community, with average infrastructure and poor environmental conditions	96

Source: Adapted from Nguyen (2008) and Young Lives (2014d)

Notes: Percentage of children from the maths sample attending early childhood education in each site, author's calculations

H.2 Full regression estimtates

Ethiopia

Table H.5: Full regression estimates for Ethiopia. Outcome is maths age-specific z-scores

VARIABLES	(1) 5 years	(2) 8 years	(3) 12 years	(4) 15 years
Attended preschool education from 3yrs (yes)	0.26**	0.30***	0.08	0.07
	(0.01 - 0.51)	(0.17 - 0.43)	(-0.11 - 0.27)	(-0.10 - 0.23)
Mother's education (ISCED categories, 1 year old)	0.08	0.10***	0.10**	0.03
	(-0.02 - 0.19)	(0.03 - 0.18)	(0.02 - 0.18)	(-0.08 - 0.14)
Asset index IRT (1 year old)	-0.06	0.15***	0.27***	0.26***
	(-0.16 - 0.04)	(0.07 - 0.24)	(0.14 - 0.40)	(0.16 - 0.36)
Monthly real consumption per capita - base 2006 (5 years old)	0.00***	0.00	0.00	0.00*
	(0.00 - 0.00)	(-0.00 - 0.00)	(-0.00 - 0.00)	(-0.00 - 0.00)
Height-for-age (z-score)	0.04***	0.05***	0.03*	0.00
	(0.01 - 0.07)	(0.03 - 0.08)	(-0.00 - 0.05)	(-0.03 - 0.03)
Child's sex (female)	0.03	0.01	0.03	-0.01
	(-0.06 - 0.13)	(-0.09 - 0.10)	(-0.08 - 0.15)	(-0.11 - 0.09)
First born	-0.07	-0.16**	-0.09	0.00
	(-0.19 - 0.05)	(-0.30 - -0.03)	(-0.21 - 0.03)	(-0.17 - 0.17)
Child's ethnicity (vs. Other)				
Agew	-0.80	-0.52*	-0.68***	-0.78***
	(-2.12 - 0.51)	(-1.07 - 0.03)	(-1.14 - -0.22)	(-1.08 - -0.47)
Amhara	0.04	-0.09	-0.12	0.02
	(-0.12 - 0.21)	(-0.38 - 0.20)	(-0.35 - 0.11)	(-0.27 - 0.31)

Continued on next page

Table H.5 – continued from previous page

	(1)	(2)	(3)	(4)
VARIABLES	5 years	8 years	12 years	15 years
Gurage	0.01	-0.13	-0.08	0.38
	(-0.14 - 0.17)	(-0.50 - 0.24)	(-0.39 - 0.23)	(-0.15 - 0.90)
Hadiva	0.87***	0.07	-0.33***	1.96***
	(0.73 - 1.00)	(-0.12 - 0.27)	(-0.54 - -0.13)	(1.73 - 2.19)
Kambata	0.13	0.86***	1.07***	1.19***
	(-0.07 - 0.34)	(0.60 - 1.11)	(0.84 - 1.30)	(0.83 - 1.55)
Oromo	-0.01	-0.09	-0.12	0.03
	(-0.10 - 0.08)	(-0.35 - 0.18)	(-0.39 - 0.16)	(-0.33 - 0.39)
Sidama	-0.15	0.09	0.06	0.20
	(-0.49 - 0.20)	(-0.27 - 0.44)	(-0.15 - 0.27)	(-0.09 - 0.50)
Tigrian	-0.19**	-0.17	-0.10	0.49***
	(-0.34 - -0.05)	(-0.40 - 0.06)	(-0.44 - 0.24)	(0.17 - 0.80)
Wolvata	-0.50***	0.00	-0.16**	0.03
	(-0.69 - -0.31)	(-0.25 - 0.26)	(-0.30 - -0.02)	(-0.21 - 0.28)
Language of maths test (vs. Amharic)				
Hadiya	-0.31**	-0.15***	-0.67***	-1.07***
	(-0.58 - -0.04)	(-0.26 - -0.05)	(-0.87 - -0.48)	(-1.18 - -0.96)
Oromiffa	0.02	-0.02	-0.02	-0.34***
	(-0.17 - 0.21)	(-0.39 - 0.35)	(-0.31 - 0.27)	(-0.57 - -0.10)
Wolayta	0.34***	-0.52***	0.29**	-1.02**
	(0.14 - 0.54)	(-0.82 - -0.22)	(0.00 - 0.58)	(-1.92 - -0.11)
Sidamigna	-1.83***	-1.08***	-0.39***	0.03
	(-2.02 - -1.65)	(-1.40 - -0.76)	(-0.56 - -0.22)	(-0.07 - 0.13)
Tigrinya	-0.07	-0.37***	0.03	-1.28***
	(-0.31 - 0.17)	(-0.57 - -0.16)	(-0.20 - 0.26)	(-1.68 - -0.87)
Other	-0.46***	-1.19***		
	(-0.66 - -0.27)	(-1.54 - -0.85)		
Household size	0.00	-0.01	-0.00	0.02
	(-0.03 - 0.03)	(-0.05 - 0.03)	(-0.03 - 0.03)	(-0.01 - 0.05)

Continued on next page

Table H.5 – continued from previous page

	(1)	(2)	(3)	(4)
VARIABLES	5 years	8 years	12 years	15 years
Mother works (vs doesn't work)				
in agriculture	0.01	0.08	-0.00	-0.05
	(-0.19 - 0.21)	(-0.07 - 0.22)	(-0.18 - 0.17)	(-0.15 - 0.06)
in non-agriculture	-0.01	-0.00	0.09**	-0.05
	(-0.17 - 0.16)	(-0.10 - 0.10)	(0.01 - 0.17)	(-0.15 - 0.05)
Constant	0.08	0.76***	0.78***	0.50***
	(-0.28 - 0.44)	(0.39 - 1.14)	(0.41 - 1.15)	(0.20 - 0.81)
Observations	1,303	1,303	1,303	1,303
R-squared	0.23	0.44	0.33	0.34

Notes: With community fixed effects. Confidence intervals in parentheses

*** p<0.01, ** p<0.05, * p<0.1

India

Table H.6: Full regression results for India. Outcome is maths age-specific z-scores

	(5)	(6)	(7)	(8)
VARIABLES	5 years	8 years	12 years	15 years
Attended preschool education from 3yrs (yes)	0.12	0.05	0.08	0.05
	(-0.04 - 0.28)	(-0.05 - 0.16)	(-0.02 - 0.18)	(-0.07 - 0.16)
Mother's education (ISCED categories, 1 year old)	0.11***	0.14***	0.18***	0.17***
	(0.07 - 0.14)	(0.10 - 0.19)	(0.13 - 0.22)	(0.13 - 0.21)
Asset index IRT (1 year old)	0.11**	0.16***	0.18***	0.13***

Continued on next page

Table H.6 – continued from previous page

	(5)	(6)	(7)	(8)
VARIABLES	5 years	8 years	12 years	15 years
	(0.02 - 0.20)	(0.06 - 0.26)	(0.08 - 0.28)	(0.05 - 0.21)
Monthly real consumption per capita - base 2006 (5 years old)	0.00**	0.00*	0.00***	0.00***
	(0.00 - 0.00)	(-0.00 - 0.00)	(0.00 - 0.00)	(0.00 - 0.00)
Height-for-age (z-score)	0.07***	0.09***	0.06***	0.03**
	(0.04 - 0.11)	(0.05 - 0.13)	(0.03 - 0.08)	(0.01 - 0.06)
Child's sex (female)	-0.01	-0.07	-0.02	-0.21***
	(-0.11 - 0.08)	(-0.16 - 0.03)	(-0.11 - 0.08)	(-0.30 - -0.11)
First born	0.08*	0.05	0.08	0.09*
	(-0.01 - 0.17)	(-0.06 - 0.16)	(-0.02 - 0.18)	(-0.01 - 0.19)
Household size	0.01	0.01	0.02*	0.02**
	(-0.01 - 0.03)	(-0.01 - 0.03)	(-0.00 - 0.03)	(0.00 - 0.05)
Language of maths test (vs. Telugu & English)				
Other	-0.17	-0.09		
	(-0.45 - 0.11)	(-0.35 - 0.18)		
Child's ethnicity (vs. scheduled caste)				
Scheduled tribe	0.22***	-0.13	0.04	0.07
	(0.06 - 0.37)	(-0.32 - 0.06)	(-0.13 - 0.20)	(-0.08 - 0.22)
Backwards caste	0.09	0.03	0.12	0.11
	(-0.06 - 0.23)	(-0.13 - 0.19)	(-0.07 - 0.32)	(-0.03 - 0.24)
Other	-0.43***	-0.07	-0.38***	-1.59***
	(-0.56 - -0.30)	(-0.24 - 0.10)	(-0.55 - -0.22)	(-1.71 - -1.46)
Hindu	0.18**	0.13	0.24**	0.30***
	(0.04 - 0.32)	(-0.04 - 0.30)	(0.00 - 0.47)	(0.09 - 0.51)
Muslim	0.16	-0.26**	-0.16	-0.08
	(-0.08 - 0.40)	(-0.50 - -0.02)	(-0.43 - 0.12)	(-0.22 - 0.07)
Buddhist	0.92**	0.13	-0.60	-0.76*
	(0.09 - 1.76)	(-0.71 - 0.97)	(-1.48 - 0.28)	(-1.56 - 0.05)

Continued on next page

Table H.6 – continued from previous page

VARIABLES	(5) 5 years	(6) 8 years	(7) 12 years	(8) 15 years
Christian	0.52	-0.02	0.14	0.55
	(-0.13 - 1.16)	(-0.45 - 0.42)	(-0.33 - 0.61)	(-0.32 - 1.42)
Mother works (vs doesn't work):				
agriculture	-0.01	0.05	0.03	0.04
	(-0.14 - 0.12)	(-0.12 - 0.22)	(-0.09 - 0.16)	(-0.12 - 0.20)
non-agriculture	-0.02	-0.06	-0.02	-0.04
	(-0.16 - 0.12)	(-0.19 - 0.07)	(-0.13 - 0.10)	(-0.18 - 0.10)
other sector	0.31	-0.11	-0.20	0.26
	(-0.70 - 1.32)	(-0.51 - 0.30)	(-0.87 - 0.46)	(-0.47 - 1.00)
Constant	0.12	-0.40***	-0.67***	-0.51***
	(-0.10 - 0.35)	(-0.65 - -0.15)	(-0.92 - -0.43)	(-0.75 - -0.28)
Observations	1,740	1,740	1,740	1,740
R-squared	0.19	0.32	0.28	0.27

Notes: With community fixed effects. Confidence intervals in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Peru

Table H.7: Full regression results for Peru. Outcome is maths age-specific z-scores

VARIABLES	(9) 5 years	(10) 8 years	(11) 12 years	(12) 15 years
Attended preschool education from 3yrs (yes)	0.19**	0.19*	0.15*	0.01
	(0.03 - 0.34)	(-0.04 - 0.42)	(-0.00 - 0.31)	(-0.12 - 0.13)

Continued on next page

Table H.7 – continued from previous page

	(9)	(10)	(11)	(12)
VARIABLES	5 years	8 years	12 years	15 years
Mother's education (ISCED categories, 1 year old)	0.11***	0.13***	0.14***	0.14***
	(0.07 - 0.15)	(0.09 - 0.18)	(0.10 - 0.18)	(0.09 - 0.18)
Asset index IRT (1 year old)	0.14***	0.16***	0.17***	0.17***
	(0.05 - 0.23)	(0.09 - 0.23)	(0.11 - 0.23)	(0.10 - 0.24)
Monthly real consumption per capita - base 2006 (5 years old)	0.00	0.00***	0.00	0.00
	(-0.00 - 0.00)	(0.00 - 0.00)	(-0.00 - 0.00)	(-0.00 - 0.00)
Height-for-age (z-score)	0.06***	0.07***	0.04	0.05*
	(0.02 - 0.11)	(0.05 - 0.09)	(-0.01 - 0.09)	(-0.01 - 0.10)
Child's sex (female)	0.05	-0.13**	-0.05	-0.20***
	(-0.03 - 0.12)	(-0.23 - -0.03)	(-0.16 - 0.05)	(-0.30 - -0.10)
First born	-0.10**	0.10*	-0.02	-0.03
	(-0.17 - -0.02)	(-0.02 - 0.21)	(-0.13 - 0.09)	(-0.13 - 0.07)
Household size	-0.00	-0.01	0.00	0.01
	(-0.02 - 0.02)	(-0.04 - 0.01)	(-0.02 - 0.03)	(-0.01 - 0.03)
Mother works (vs doesn't work):				
agriculture	-0.08	-0.12	-0.01	-0.07
	(-0.29 - 0.14)	(-0.28 - 0.04)	(-0.23 - 0.22)	(-0.29 - 0.15)
non-agriculture	0.10	0.05	0.08	0.11*
	(-0.06 - 0.26)	(-0.08 - 0.18)	(-0.03 - 0.18)	(-0.00 - 0.23)
Child's ethnicity (vs. white)				
Mestizo	0.10	0.12	0.16*	0.13
	(-0.11 - 0.30)	(-0.06 - 0.30)	(-0.00 - 0.33)	(-0.11 - 0.37)
Native of the Amazon	-0.30**	-0.33***	-0.36***	-0.12
	(-0.56 - -0.04)	(-0.53 - -0.13)	(-0.57 - -0.15)	(-0.42 - 0.17)
Black	0.77**	0.07	-0.02	0.21

Continued on next page

Table H.7 – continued from previous page

VARIABLES	(9) 5 years	(10) 8 years	(11) 12 years	(12) 15 years
	(0.20 - 1.33)	(-0.23 - 0.37)	(-0.41 - 0.38)	(-0.30 - 0.72)
Asiatic	0.40***	0.56***	0.95***	0.90***
	(0.16 - 0.64)	(0.35 - 0.76)	(0.72 - 1.18)	(0.62 - 1.17)
Language of maths test (vs. Spanish)				
Quechua	0.29***	-0.39***	-0.69	
	(0.10 - 0.48)	(-0.57 - -0.21)	(-1.54 - 0.15)	
Spanish & Quechua	-0.12	-0.19***	-0.06	-0.10
	(-0.51 - 0.27)	(-0.25 - -0.12)	(-0.18 - 0.07)	(-0.39 - 0.18)
Other	-0.37***			
	(-0.49 - -0.26)			
Aimara		0.50	-1.80***	
		(-0.37 - 1.38)	(-1.95 - -1.64)	
Constant	-0.64***	-0.63***	-0.96***	-0.77***
	(-0.96 - -0.31)	(-0.99 - -0.27)	(-1.28 - -0.65)	(-1.07 - -0.48)
Observations	1,649	1,649	1,649	1,649
R-squared	0.14	0.30	0.27	0.22

Notes: With community fixed effects. Confidence intervals in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Vietnam

Table H.8: Full regression results for Vietnam. Outcome is maths age-specific z-scores

VARIABLES	(13) 5 years	(14) 8 years	(15) 12 years	(16) 15 years
Attended preschool education from 3yrs (yes)	0.20*	0.01	-0.05	0.01

Continued on next page

Table H.8 – continued from previous page

	(13)	(14)	(15)	(16)
VARIABLES	5 years	8 years	12 years	15 years
	(-0.00 - 0.39)	(-0.19 - 0.22)	(-0.19 - 0.10)	(-0.14 - 0.16)
Mother's education (ISCED categories, 1 year old)	0.12***	0.10***	0.17***	0.12***
	(0.07 - 0.17)	(0.06 - 0.15)	(0.13 - 0.21)	(0.07 - 0.18)
Asset index IRT (1 year old)	0.08*	0.10**	0.16***	0.17***
	(-0.00 - 0.17)	(0.00 - 0.20)	(0.05 - 0.26)	(0.07 - 0.27)
Monthly real consumption per capita - base 2006 (5 years old)	-0.00	0.00	0.00	0.00
	(-0.00 - 0.00)	(-0.00 - 0.00)	(-0.00 - 0.00)	(-0.00 - 0.00)
Height-for-age (z-score)	0.06***	0.09***	0.04**	0.04
	(0.02 - 0.09)	(0.05 - 0.12)	(0.01 - 0.08)	(-0.01 - 0.08)
Child's sex (female)	-0.00	0.04	0.12***	0.16***
	(-0.10 - 0.09)	(-0.03 - 0.11)	(0.04 - 0.19)	(0.08 - 0.23)
First born	0.01	0.02	-0.04	-0.00
	(-0.08 - 0.10)	(-0.03 - 0.07)	(-0.15 - 0.07)	(-0.10 - 0.09)
Household size	0.00	0.01	-0.03**	-0.01
	(-0.01 - 0.02)	(-0.01 - 0.04)	(-0.06 - -0.00)	(-0.02 - 0.01)
Language of maths test (vs. Other)				
Vietnamese	-0.16			
	(-0.40 - 0.08)			
Mother works (vs doesn't work):				
agriculture	0.02	0.10	0.04	0.06
	(-0.16 - 0.19)	(-0.08 - 0.28)	(-0.11 - 0.20)	(-0.14 - 0.26)
non-agriculture	-0.03	0.10*	0.11	0.10
Child's ethnicity (vs. Other)				
	(-0.21 - 0.15)	(-0.01 - 0.20)	(-0.07 - 0.29)	(-0.04 - 0.24)
Kinh	0.41**	0.75***	0.14	-0.03

Continued on next page

Table H.8 – continued from previous page

	(13)	(14)	(15)	(16)
VARIABLES	5 years	8 years	12 years	15 years
	(0.00 - 0.82)	(0.55 - 0.94)	(-0.21 - 0.49)	(-0.21 - 0.14)
H'Mong	0.05	-0.20	-0.27	-0.51***
	(-0.44 - 0.54)	(-0.59 - 0.20)	(-0.65 - 0.12)	(-0.75 - -0.27)
Ede	-1.84***	1.59***	-0.94***	0.92***
	(-2.25 - -1.43)	(1.36 - 1.81)	(-1.16 - -0.73)	(0.76 - 1.09)
Bana	-0.47***	-0.04	0.63***	0.32***
	(-0.79 - -0.16)	(-0.18 - 0.10)	(0.38 - 0.88)	(0.19 - 0.45)
Nung	0.06	0.45**	0.12	-0.12
	(-0.40 - 0.53)	(0.11 - 0.79)	(-0.39 - 0.63)	(-0.50 - 0.25)
Tay	0.41**	0.84***	0.18	-0.09
	(0.01 - 0.82)	(0.51 - 1.17)	(-0.18 - 0.54)	(-0.27 - 0.09)
Dao	0.05	0.08	0.24	-0.17***
	(-0.37 - 0.47)	(-0.16 - 0.31)	(-0.14 - 0.62)	(-0.29 - -0.05)
Constant	-1.26***	-0.61***	-0.38**	-0.18
	(-1.81 - -0.72)	(-0.94 - -0.28)	(-0.73 - -0.03)	(-0.54 - 0.18)
Observations	1,673	1,673	1,673	1,673
R-squared	0.34	0.33	0.31	0.26

Notes: With community fixed effects. Confidence intervals in parentheses

*** p<0.01, ** p<0.05, * p<0.1