Working Paper 127

Young Lives
An International Study of Childhood Poverty

Socio-economic Status and Early Childhood Cognitive Skills

Is Latin America Different?

Florencia Lopez Boo

www.younglives.org.uk

MAY 2014

Working Paper 127

Young Lives
An International Study of Childhood Poverty

Socio-economic Status and Early Childhood Cognitive Skills

Is Latin America Different?

Florencia Lopez Boo

www.younglives.org.uk **MAY 2014**

Socio-economic Status and Early Childhood Cognitive Skills: Is Latin America Different?

Florencia Lopez Boo

First published by Young Lives in May 2014

ISBN: 978-1-909403-41-3

A catalogue record for this publication is available from the British Library.

Printed on FSC-certified paper from traceable and sustainable sources.

Funded by

UKaid from the British people

Ministry of Foreign Affairs of the Netherlands

Irish Aid
An Roinn Gnóthaí Eachtracha agus Trádála
Department of Foreign Affairs and Trade

Young Lives, Oxford Department of International Development (ODID), University of Oxford, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, UK

Tel: +44 (0)1865 281751 • E-mail: younglives@younglives.org.uk

Contents

Summary ii

The Author ii

Acknowledgements ii

1. Introduction 1

2. Data, economic framework and descriptive analysis 3

2.1. Data and methodology 3

2.2. SES and cognitive skills 5

2.3. SES and nutrition 10

3. Regression analysis 11

3.1. The SES gradient in PPVT scores 11

3.2. The SES gradient for changes in PPVT scores across rounds 16

4. Discussion of results, caveats and conclusions 18

References 21

Appendix 1. Additional figures and tables 24

Appendix 2. Validity of the PPVT to compare gaps across countries 32

Summary

This paper documents differences in cognitive development – as measured by a receptive vocabulary test – between children from households with high and low socioeconomic status (SES) in two different phases of childhood (before and after early school years) in four developing countries: Peru, Ethiopia, India and Vietnam. Intercontinental evidence on the timing, pattern, and persistence of these differences is provided. The non-parametric analysis suggests that differences found at the age of 5 persist into the early school years across all four countries, and the conditional analysis shows that the magnitude of within-country SES differences seem to diminish over time (with the exception of the India sample). However, both the magnitude of the gap and the degree of persistence vary. The main result is that Peru stands out, not only as the country with the largest cross-section difference between rich and poor (of around 1.30–1.40 standard deviations), but also as the country with the highest persistence in cognitive development, as shown by the value-added specification. Some channels behind these trends are discussed, but overall, the SES gradient persists even when controlling for a range of important mediators, such as preschool, early nutrition, and years of schooling. Past performance on the Peabody Picture Vocabulary Test (PPVT) is an important determinant of the SES gradient at the age of 8.

The Author

Florencia Lopez Boo is a Senior Economist at the Social Protection and Health Division, Inter-American Development Bank, and a Research Fellow at the Institute for Study of Labor (IZA). She works with Latin American governments on the design and evaluation of early childhood and social protection programmes. She holds a DPhil in Economics from the University of Oxford.

Acknowledgements

I am grateful for the Round 3 data provided by Young Lives. I am also grateful to Marta Favara for her early work on this piece, and to Fiorella Benedetti and Romina Tome for excellent research assistance.

About Young Lives

Young Lives is an international study of childhood poverty, following the lives of 12,000 children in 4 countries (Ethiopia, India, Peru and Vietnam) over 15 years. **www.younglives.org.uk**

Young Lives is funded from 2001 to 2017 by UK aid from the Department for International Development (DFID), co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014, and by Irish Aid from 2014 to 2015.

1. Introduction

Household poverty is one of the main factors associated with the risk that a child will have poor nutrition and stimulation (Walker et al. 2011). Moreover, research from a number of developed (Currie and Thomas 2000; Feinstein 2003; Case and Paxson 2008; Heckman et al. 2010) and developing countries (Hoddinott et al. 2008; Maluccio et al. 2009; Behrman et al. 2009; Walker et al. 2011) find that poor nutritional status and low levels of cognitive development in early childhood are important determinants of outcomes such as fewer years of schooling, less learning while in school, lower cognitive skills and worse health in adolescence (including mental health) and adulthood, a lower probability of employment, lower earnings, lower wage rates, and more criminal activity. This evidence points to the fact that disadvantages found at an early age will result in the intergenerational transmission of poverty and inequalities if not addressed by policymakers before children reach adulthood.

Understanding socioeconomic status (SES) gradients in cognitive skills early in a child's life is therefore a crucial step toward understanding the intergenerational transmission of poverty and inequality. The persistence of a connection between parents and children's outcomes is well documented (Black and Devereux 2010), but the magnitude of the intergenerational correlation varies dramatically across countries and regions. In particular, Hertz et al. (2008) provide a survey of parent–child schooling correlations for a sample of 42 countries using comparable data and find that the correlation is 0.66 SD in Peru, 0.40 in Vietnam, and 0.10 in Ethiopia, which are three of the four countries of interest in this paper.[1] The correlation of wages earned by parents with those earned subsequently by their children shows similar results. Moreover, decompositions have shown that part of the observed intergenerational correlation can be explained by the fact that parental SES strongly predicts the cognitive and non-cognitive skills and health of children (Black and Devereux 2010),[2] – a finding that motivates my interest in the SES–cognitive skills link in the first place.

Until very recently, there were no comparable data on the cognitive development of young children (0–6 years of age) for most developing countries (Harpham 2002). Therefore, not much research exists on differences in cognitive development by SES, when they arise, whom they affect, and how they evolve as children grow up. Two studies carried out in the early years of this century found gradients in cognitive development in Brazil' (Victora et al. 2003) and Guatemala (Stein et al. 2008). More recently, Fernald et al. (2011) and Naudeau et al. (2011) have used single cross-sections of data from low-income countries to study cognitive skills gradients, while related research from Ecuador (Paxson and Schady 2007; Schady 2011) has shown substantial differences in cognitive development at young ages between children of higher and lower socioeconomic status, which increased between 3 and 5 years of age. Paxson and Schady's (2007) analysis from Ecuador included only households in the poorest half of the nationwide distribution of a composite measure of wealth, and it was limited to relatively young families in rural areas. That is why the study

1 Unfortunately, Hertz et al. (2008) do not have data available for India, the fourth country of interest in this paper.

2 Blanden et al. (2007) examine the role of non-cognitive skills and ability for intergenerational income persistence in Britain. Their work demonstrates that covariates can account for approximately half of the estimated intergenerational income elasticity (of 0.32), with a sizeable portion attributable to cognitive and non-cognitive skills via their effect on educational attainment. Moreover, it has been fairly well established that better infant health has a positive causal effect on later adult outcomes (Currie and Moretti 2003).

more related to this research is Schady et al. (2011). In their work, wealth gradients in five Latin American countries (including Peru) are studied and substantial differences in children's performance on the Peabody Picture Vocabulary Test (PPVT) scores are found; these differences seem to persist as children grow up. However, I am not aware of any study providing intercontinental evidence on the magnitude of the differences, as well as the persistence of PPVT scores using longitudinal data in developing countries.

Therefore, the contribution in this paper is the investigation of the relationship between SES and an important component of cognitive development – receptive language ability – among children aged 55 to 102 months in four developing countries (Ethiopia, India, Peru and Vietnam). In these four countries, the Young Lives study has collected extensive comparable data over three rounds of surveys.[3] Specifically, this paper provides intercontinental evidence on the magnitude of SES gradients as well as the persistence of PPVT scores using longitudinal data. It analyses the shape, timing, and the pattern of these differences in two different phases of childhood (before and after early school years). As far as I know, no analysis has been done in this respect.[4] An understanding of which children are most likely to show deficits in cognitive development, at what ages these deficits become apparent, and how these deficits evolve as children grow up is indispensable for the formulation of appropriate policies.

Additional motivation for this work is provided by the fact that Latin America is the most unequal region in the world (Gasparini and Lustig 2011; IDB 2011). One simple and intuitive measure of inequality is the ratios of the income or consumption of households at the 90th percentile of the distribution to those at the 10th percentile. Using data from around 2000 and ranking countries by this ratio, 14 of the 15 countries in the world with the highest levels of inequality are found in Latin America (World Bank 2005). Moreover, much of Latin America's inequality is associated with inequality of opportunities, not just outcomes. That is, a substantial proportion of the inequality in incomes that is observed is determined by socioeconomic status (characteristics such as race, place of birth, wealth, or the education levels of their parents).

In this context, I am most interested in Peru because it is the 21st most unequal country in the world, with a Gini coefficient of 0.48 (World Bank 2005), while none of the other Young Lives countries was anywhere close to that position in the rankings.[5] Not surprisingly, existing analysis of Young Lives data shows that socioeconomic gradients between socioeconomic status (as measured by mother or father's education or wealth) and early childhood skills (as measured by the PPVT) are present at the age of 5 (Engle et al. 2011). Extending the analysis (with the availability of Round 3 data) to the children who were 8 years old in 2009 will be useful for understanding how these differentials evolve, particularly now that these children are in school. It might well be that differentials fade away due to the equalising effect of schools, but this is the empirical question I will attempt to answer.

3 Young Lives, established in 2000, is an international research project on childhood poverty. The study traces the lives of children in Ethiopia, India, Peru, and Vietnam. See http://www.younglives.org.uk/ for more information.

4 An additional, though more methodological, contribution is to see whether previous results from studies in Latin America that were not representative hold for the Peru results in this paper (the Young Lives Peruvian sample is representative of 95 per cent of the population) and with a survey that covers a longer period; there are about four years between Round 2 and 3 of the Young Lives data, while in previous studies there were only two years.

5 Urban India is in 84th place, Vietnam (urban and rural) in 87th, rural India in 119th, and Ethiopia (urban and rural) in 121st place in this ranking.

I will draw on the literature on SES and health (Currie, 2009; Case et al. 2002) and SES and cognitive skills (Victora et al. 2003). In particular, I will build on the work of Paxson and Schady (2007) and Schady et al. (2011) for the descriptive section that uses non-parametric regressions. For the regression analysis, I build on the value-added production function approach in Todd and Wolpin (2007), which allows this research to go beyond all previous empirical studies on this topic.

Results show that, although differences in cognitive skills by SES are present in all countries, they arise more starkly in Peru, with a gap of around 1.30–1.40 standard deviations (SD) between the poorest and the richest quintiles. For Peru, which has norms provided by test developers, these differences imply developmental lags of up to one year at the age of 5. OLS regressions show that SES differences in the PPVT scores seem to fall from the age of 5 to the age of 8; and that, overall, the SES gradient persists even when controlling for a large number of important mediators, such as preschool, early nutrition, and primary school attendance. Using the value-added regressions that exploit the longitudinal data, I find that Peru shows the highest coefficient of past PPVT on current PPVT (0.42). This indicates that *ceteris paribus,* convergence between the rich and the poor will happen at a much slower speed in Peru than in the other three countries, unless accompanied by appropriate policies.

The structure of the paper is as follows. Section 2 presents the analytical approach and some brief descriptive analysis of the data. Section 3 presents the regressions, while Section 4 discusses the results and concludes.

2. Data, economic framework and descriptive analysis

2.1. Data and methodology

The data are from the longitudinal survey carried out by Young Lives. Beginning in 2002, Young Lives surveyed children in two age cohorts. I use data for the 4,525 children in the Young Lives Younger Cohort dataset speaking the majority language of the region or country – Amharic in Ethiopia, Telugu in India, Spanish in Peru,[6] and Vietnamese in Vietnam – following Cueto et al. (2009). In a robustness check, I use the full sample of 6,089 children in the Young Lives Younger Cohort with non-missing information in the relevant variables. It is worth noting that the samples in Ethiopia, India, and Vietnam are not fully representative of each country, as the focus is on poor children; by contrast, the sample in Peru is representative of all but the richest 5 per cent of districts. Moreover, in the case of India, the sample comes from one state: Andhra Pradesh. The sample for all countries contains children from both urban and rural communities, which were selected using a sentinel site methodology to purposively over-sample poor areas. Within the selected sites, children in the right age groups (between 6 and 18 months and between 7.5 and 8.5 years) were sampled randomly. I pool the urban and rural children to obtain more precise estimates. Minority

6 In Peru, the PPVT was translated into Quechua, an indigenous language spoken primarily in rural areas of the highlands, and children were given the option of taking the test in Spanish or Quechua. Twenty-two per cent of children in rural areas, but only 0.1 per cent of children in urban areas, chose to take the test in Quechua. In the robustness tests, I test whether our results are robust to including these children in the analysis.

language speakers might be more likely to come from rural communities. If this is the case, I will have an under-representation of rural communities when using the majority language sample. However, by using the full sample in the robustness checks, this problem is solved.

A major strength of this study is the use of a common measure of child cognitive development: performance on the widely used PPVT. Raw scores on this test, given by the number of items answered correctly, can be used only to compare children of the same age. I therefore present results based on the following common standardisation of PPVT scores separately by country. I constructed internally standardised, age-specific z-scores by subtracting the month-of-age-specific mean of the raw score and dividing by the month-of-age-specific standard deviation. These results can be used to make comparisons within countries, including comparisons between children of different ages. However, internally standardised scores are not informative about differences across countries. I use this method to see SES gaps within countries, and then I compare those SES gaps (in standard deviations) across countries.

I use household expenditure at the age of 5 as a proxy of SES.[7] Expenditure is calculated for all sample households in each country (about 2,000); it includes expenditure on food, transport, security, telephone, electricity, water supply, housing, clothes, and footwear. Quintiles of expenditure are created separately for each country on the basis of the aggregation of all sampled households in that country. The distribution of expenditure in the country samples, as well as the distribution of wealth, are presented in Figures A1 and A2 in Appendix 1. The wealth distribution in India and Ethiopia seems shifted to the left, while the Peruvian distribution of wealth looks bimodal. By contrast, all four distributions of expenditure have the usual bell shape of a normal distribution, which emphasises the need to use expenditure data instead. I then sort children into household expenditure quintiles and compare outcomes for children in the top and bottom quintiles. The mean language scores for each expenditure quintile are presented for each country in terms of SDs.

Lastly, I build a balanced panel of children present in Round 2 (at the age of 5) and Round 3 (at the age of 8) to analyse the evolution of the gaps in PPVT scores by SES. Specifically, I inspect age patterns in SES gradients in child development using non-parametric regressions (Fan and Gijbels 1996). I compare two quintiles of the sample: households in the first quintile of household expenditure and those in the fifth quintile. These particular results are presented in Figure 2. The non-parametric regressions smooth out average PPVT scores by age and make patterns more apparent. Similar methods have been used extensively in the literature in economics (Deaton 1997; Case et al. 2002) and, more recently, in medicine (Park et al. 2005; Moscicki et al. 2004; Fleming et al. 2011). Standard errors and confidence intervals for the non-parametric regressions are constructed by bootstrapping (Efron and Tibshirani 1993).

7 The analysis of the associations between PPVT scores and wealth (with wealth indexes calculated separately for urban and rural areas) was done in the previous version of this paper. However, as noted by a reviewer, household expenditure (or consumption) are a better indicator of SES.

In Section 3, I run OLS regressions of the PPVT scores when the children were 5 and when they were 8, introducing potential mediators of the SES–PPVT relationship gradually. Finally, I run a value-added production function, as explained in the next sub-section, for the PPVT scores when the children were 8.

2.2. SES and cognitive skills

In the first column for each country, Table 1 presents differences in average PPVT z-scores between children from the fifth and first expenditure quintiles. The second column presents the *p*-value of the difference.

Table 1. *PPVT z-scores at the ages of 5 and 8, gaps between 1st and 5th expenditure quintiles and p-values of t-tests*

	Ethiopia		India		Peru		Vietnam	
	(Q5-Q1)	*p-value*	*(Q5-Q1)*	*p-value*	*(Q5-Q1)*	*p-value*	*(Q5-Q1)*	*p-value*
PPVT age 5	1	0.000	0.3	0.000	1.3	0.000	1	0
PPVT age 8	1	0.000	0.5	0.000	1.4	0.000	0.7	0

Notes: The calculation of the mean scores gives equal weight to each month of age, within a country.

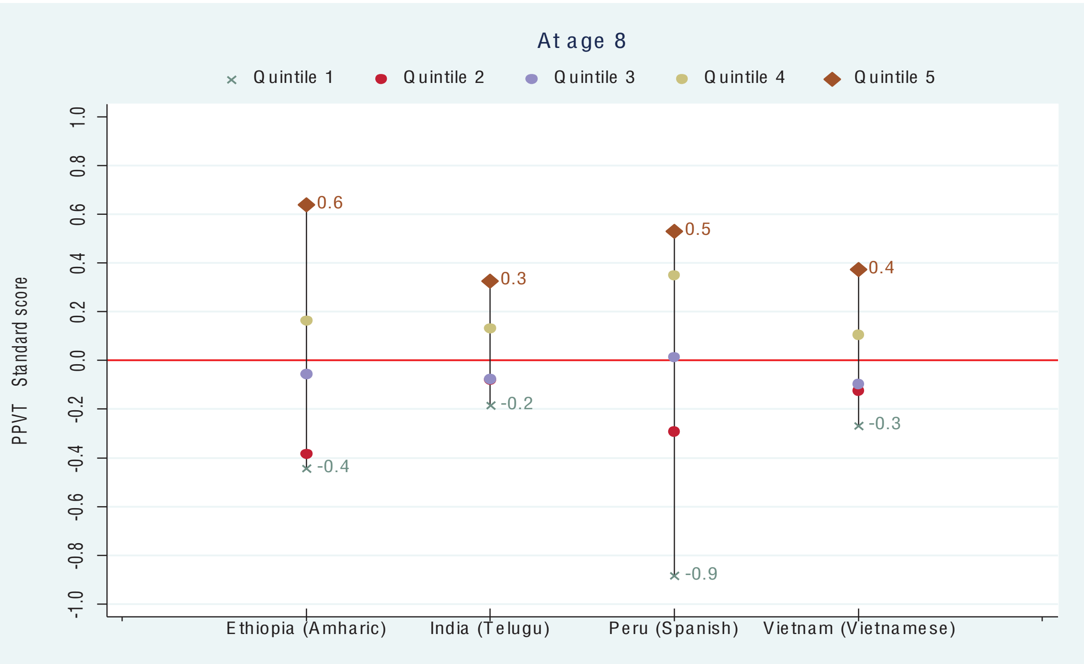
Differences in language development between richer and poorer children within countries are large and statistically significant, both at the ages of 5 and 8.[8] Table 1 and Figure 1 show that these differences are biggest in Peru in both survey rounds. At the age of 5 in Peru, there is a 1.30 SD gap between the first and the fifth quintiles, followed by the gaps in Ethiopia and Vietnam (both with 1 SD gap), and then by the gap in India (0.30 SD).[9] In Round 3, Peru presents the widest gap (1.40 SD), followed by Ethiopia with 1 SD, Vietnam (0.70 SD), and India (0.50 SD). The last two are not significantly different from each other.

The gaps for Peru are larger than those found in similar studies – specifically, in Victora et al. (2003); Stein et al. (2008); Fernald et al. (2011) and Naudeau et al. (2011). Using the same dataset for Peru, Schady et al. (2011) found that the gap is 0.95 SD in urban areas and 0.77 SD in rural ones. Moreover, they find, when looking at other Latin American countries, that differences in internally standardised scores between the first and fourth quartiles in the distribution of wealth are biggest in urban Colombia (1.23 SD) and rural Ecuador (1.21 SD). Although these numbers are not completely comparable because they use quartiles of wealth (and not quintiles of expenditure), they are informative of the fact that what proxy is used for SES – wealth/mother schooling levels is used in Schady et al. (2011) versus household expenditure here – matters for the magnitudes found.

8 *t*-test results available upon request.

9 In the case of India, the caste of the children might be the best indicator of Indian social inequalities rather than the expenditure distinction made in this paper.

Figure 1. *PPVT z-scores at the ages of 5 and 8 by country and expenditure quintile, majority language speakers*



In Figure 2, I investigate age patterns in SES gradients in child development using non-parametric regressions on the full panel sample. I compare two quintiles of the sample: households in the first quintile (solid blue line in the bottom half of the graph) of household expenditure, and households in the fifth quintile (red line with dots and dashes in the upper half of the graphs). The red lines with dashes are the confidence interval for the fifth quintile, and the blue ones the interval for the first quintile. These internally standardised scores suggest that the bulk of the difference between poorer and less poor children is apparent by the age of 5 (around 57 months) in all four countries. Looking at externally standardised scores, I observe that by the time children are 5½ to 6 years of age, the poorest children in

Peru are 35 to 40 points (approximately 2.5 SD, or one year) behind the reference population (see also Schady et al. 2011).[10] Unfortunately, it is not possible to perform this calculation for the other three countries, as there are no externally referenced tests in Vietnamese, Telugu, or Amharic.

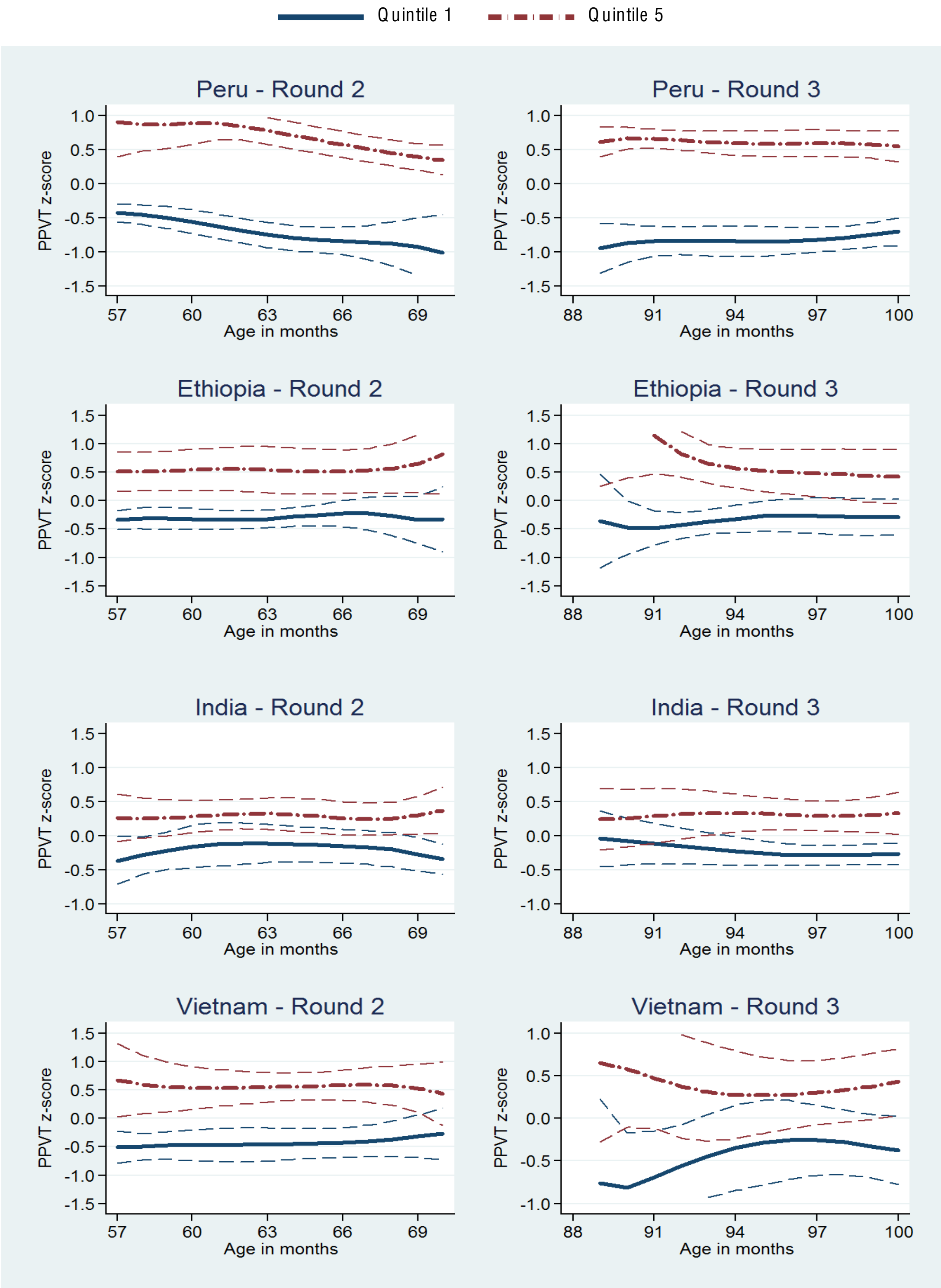
There is yet another peculiarity about Peru: the left panel of Figure 2 shows that the gap immediately starts out wide at the age of 57 months (1.4 SD) and stays at that level throughout that round. In all the other three countries, the gap is also stable, but smaller. In Peru, the scores of both the poorer and the richer children go down slightly after 63 months of age – a pattern worth exploring.

Round 3 data on the same children (shown in the right panel of Figure 2) show that the household expenditure gradients that are apparent among 4–5 year old children continue to be apparent as these children enter the first years of primary school in all countries. On the other hand, the poorest children do not appear to fall further behind, either. For all four countries, the gap between the PPVT score of children living in households in the first household expenditure quintile and those in the fifth quintile seems pretty stable over time once children are in school. Overall, there is no clear pattern of a systematically widening gap, which challenges the findings in Paxson and Schady (2007) for the Ecuadorian data (the only longitudinal data on PPVT scores in a developing country that I am aware of besides the Young Lives data).

As a robustness test, I have pooled the two rounds of data for each country and regress PPVT scores on expenditure quintiles and the interactions of expenditure quintiles with month dummies. The sum of the coefficient of the fifth quintile and its interaction with the coefficient of any given age month is the estimate of the difference in PPVT scores between the fifth and first quintiles (that is, the first quintile is the excluded dummy) at the given age. These coefficients were all statistically significant, positive, and showed the same pattern as the ones described in Figure 2 (results available upon request).

10 The calculation of the relative delays follows the following steps: first, I calculate the average raw score by month of age, separately for children in urban and rural areas and separately for children in the poorest and richest quintile. Second, I use the tables provided by the developers of the PPVT (TVIP in Spanish) test to identify the age in months at which this raw score corresponds to a score of 100 in the reference population. The difference in the age at which children in my sample and children in the reference population attain the same vocabulary level is my estimate of the months delayed. Third, I report the difference in months delayed between children in the first and fifth expenditure quintiles. The calculation of the relative delays gives equal weight to each month of age, within a country and by location (urban or rural).

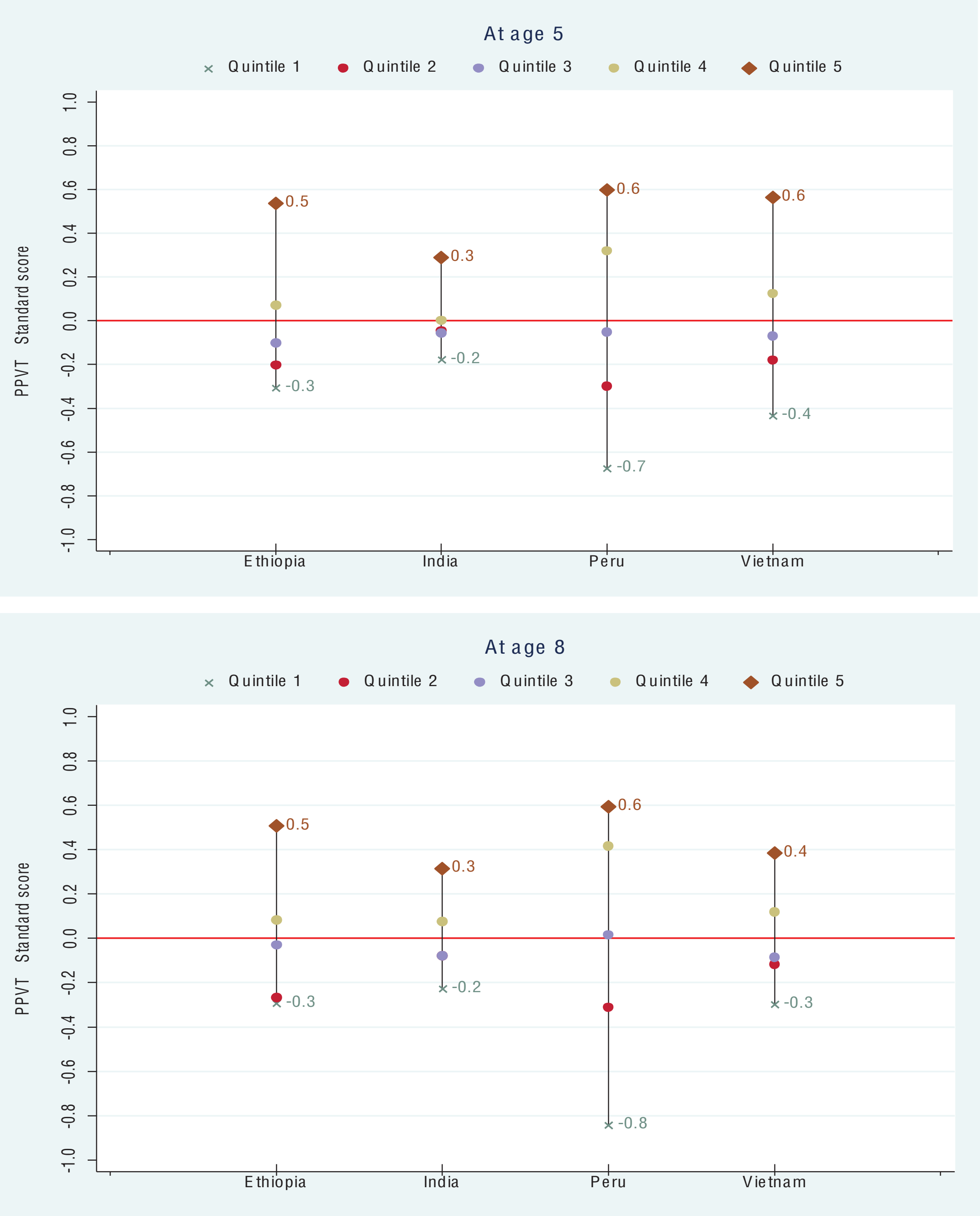
Figure 2. *Panel data analysis of PPVT score age patterns, Young Lives study countries, majority language speakers*



Note: As in all calculations in this paper, regressions are restricted to the panel of children observed in both Rounds 2 and 3. The non-parametric regressions of PPVT score (internally standardised for main language group) on age in months, by household expenditure quintile (Q1 vs. Q5). The bandwidth of the regressions is 5. Data for children 88–91 months of age in Round 3 in Ethiopia have confidence intervals that are too wide because of small sample size at those ages (when the data are restricted to children whose first language is Amharic, the sample from Ethiopia is reduced by half of the total size). The same applies for India for children 57–59 and 88–91months of age.

I conduct an important robustness checks on this main result. Mainly, it is likely that the poorest children in all countries are being dropped from the sample above as a consequence of the strategy of considering only those children who speak the main language of the country. Because of this consideration, I present Figure 3, which replicates Figure 1 except for the full sample of children in this cohort in each country. The ordering of countries in terms of the gaps remains exactly the same. There are only very small changes in magnitude in relation to Figure 1 – in the case of Ethiopia (both survey rounds) and India in Round 2. Replicating Figure 2 for the full sample also gives the same results (available upon request). Regressions analysis will also consider both samples.

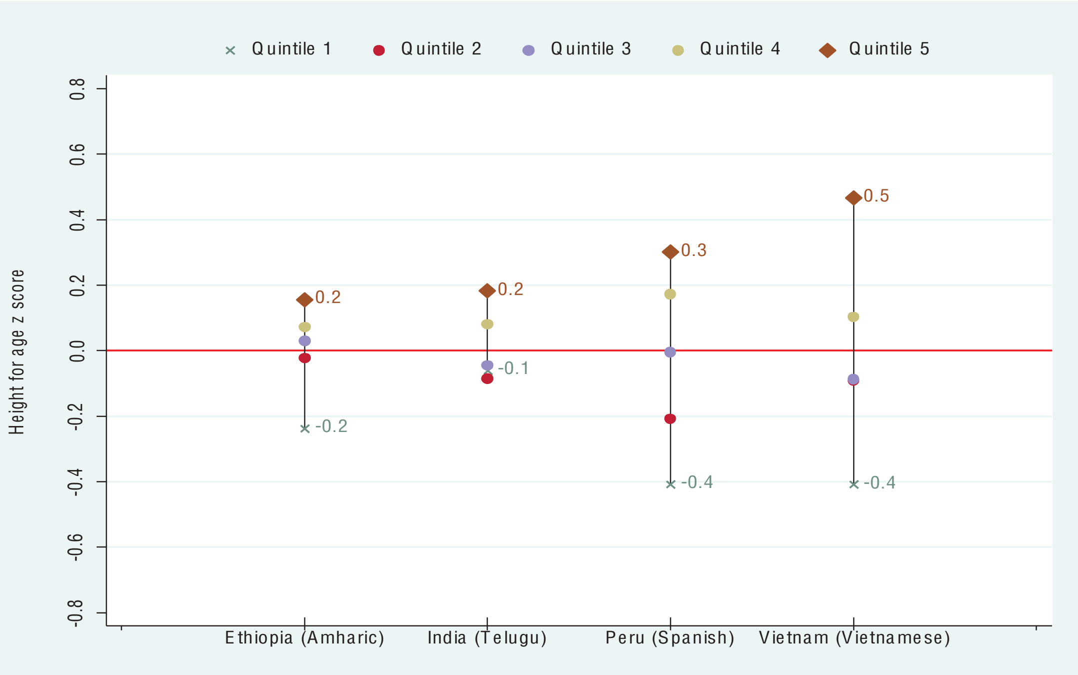
Figure 3. *PPVT z-scores at the ages of 5 and 8 by country and expenditure quintile, full sample*



2.3. SES and nutrition

Looking for important covariates behind the trends found in Subsection 2.2, I look at one variable that has been shown to be the most important determinant of cognitive development: the World Health Organization (WHO) z-score[11] of height-for-age at the age of 5 (HAZ at the age of 5), a proxy for chronic malnutrition, known to be strongly related with mental functioning, particularly during the first five years of life (Alderman 2000; WHO 2000; among many others). The HAZ distribution in Round 2 (Figure 4) shows that the biggest gap in HAZ between the richest and the poorest at the age of 5 is found among Vietnamese children (0.90 SD). The Peruvian children follow closely, with a gap of 0.70 SD.[12] This shows that this variable must be closely examined in the regression analyses.

Figure 4. *Height-for-age z-scores at the age of 5, by country and expenditure quintile*



11 A z-score is a standardised score.

12 Given what is stated in the medical literature (WHO 2000) in the regressions analysis, I include HAZ at the age of 5, as it is at this stage that nutritional status builds up.

3. Regression analysis

3.1. The SES gradient in PPVT scores

In this section, I estimate production functions under different assumptions, i.e. the non-value-added regressions or contemporaneous specification that produces valid estimates under certain assumptions (see Todd and Wolpin 2007) in this section; and a value-added regression in Section 3.2.

The value-added regression analysis is inspired by the educational production function literature that commonly adopts these specifications when data on lagged inputs are missing or incomplete (Todd and Wolpin 2007). In its most basic form, the value-added specification relates an achievement outcome measure to contemporaneous measures of school and family inputs and to a lagged (baseline) achievement measure. The goal is then to identify the relation between current cognition (at the age of 8) and past cognition (at the age of 5) to explore the level of persistence of the performance in the PPVT. In Equation (1), I show a production function of cognitive skills (θ_t), which depend on the child's past cognitive skills (θ_{t-1}), family inputs (i_t), child's nutritional status at the beginning of the period (H_t), the family's characteristics (X_t) and the child's ability (μ_t)

$$\theta_t = f(\theta_{t-1}, i_t, H_t, X_t, \mu_t) \qquad (1)$$

Then, if lagged cognitive skills (PPVT score at the age of 5) is a sufficient statistic for input histories and unobserved ability, estimating Equation (2) below would result in consistent estimates of the production function of cognitive skills today (PPVT score at the age of 8). I therefore assume that the production function has a value-added form, and that for child *i* in time *t* it can be written as:

$$\theta_{it} = \alpha + \beta H_{it} + \gamma X_{it} + \psi \theta_{it-1} + \delta \mu_{it} + \varepsilon_{it} \qquad (2)$$

In Equation (2) I have added as it is standard an error term (ε_{it}) and a constant (α).

The coefficient of interest in Section 3.2 is ψ.

Even if I estimate a production function for cognitive skills, it is worth clarifying that SES enters the production function not as an input directly affecting cognitive outcomes but as an indirect determinant affecting the productivity of inputs and the shape of the production function. Currie (2009) suggests potential mechanisms through which SES may affect child outcomes. She looks at the link between SES and child health. However, a very similar framework can be applied to the links between SES and child cognitive achievement, and that is the framework I follow here.

To further investigate the relationship between SES and PPVT score across ages and countries, I run an OLS regression of PPVT at the age of 5. Table 2 presents the main results, while Table A1 in Appendix 1 is the complete table. Results for the age of 5 are in the first four columns, and those for the age of 8 are in the last five columns. Both types of regressions control for standard predetermined variables at the child level (such as gender, age, and birth order) and household level (region, urban, size, expenditure quintile, education of the mother and father, and mother's height) measured at the relevant age. Standard errors are clustered at the sentinel site level. All these controls have the expected sign, with the urban dummy, and parental education co-variates showing high correlations with the PPVT score.

Table 2. *OLS Regressions 4 YL countries, majority language sample.*

PERU	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.63***	0.62***	0.60***	0.59***	0.54***	0.53***	0.51***	0.54***	0.50***
	(0.099)	(0.103)	(0.099)	(0.102)	(0.105)	(0.108)	(0.101)	(0.107)	(0.106)
Preschool		0.25***		0.22***		0.21**			0.17*
		(0.062)		(0.063)		(0.088)			(0.088)
HAZ age 5			0.12***	0.12***			0.14***		0.13***
			(0.030)	(0.031)			(0.029)		(0.030)
In school								0.57*	0.52*
								(0.281)	(0.288)
Constant	2.03**	2.11**	3.50***	3.50***	-2.06*	-1.84*	-0.51	-2.64**	-0.91
	(0.847)	(0.852)	(0.961)	(0.978)	(1.059)	(1.052)	(1.050)	(1.080)	(1.072)
Observations	1,562	1,562	1,562	1,562	1,562	1,562	1,562	1,562	1,562
R-squared	0.367	0.373	0.379	0.384	0.358	0.362	0.372	0.359	0.376
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

ETHIOPIA	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.51***	0.47**	0.51***	0.47**	0.41***	0.37***	0.40***	0.41***	0.37***
	(0.168)	(0.164)	(0.169)	(0.165)	(0.092)	(0.087)	(0.093)	(0.086)	(0.080)
Preschool		0.34*		0.34*		0.34***			0.31**
		(0.166)		(0.168)		(0.101)			(0.106)
HAZ age 5			0.17**	0.15*			0.07		-0.00**
			0.074	0.074			0.047		(0.048)
In school								0.34***	0.31***
								(0.083)	(0.085)
Constant	0.15	0.27	0.59	0.73	-0.97	-0.78	0.47	-0.93	0.65
	(0.804)	(0.822)	(0.986)	(0.959)	(1.233)	(1.209)	(1.447)	(1.259)	(1.379)
Observations	403	403	403	403	403	403	403	403	403
R-squared	0.339	0.347	0.340	0.348	0.493	0.502	0.494	0.505	0.513
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

INDIA	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.34**	0.34**	0.30*	0.30*	0.45***	0.45***	0.41***	0.44***	0.41***
	(0.157)	(0.155)	(0.161)	(0.160)	(0.116)	(0.116)	(0.108)	(0.116)	(0.108)
Preschool		0.09		0.09		0.01			0.01
		(0.102)		(0.100)		(0.088)			(0.087)
HAZ age 5			0.12***	0.12***			0.12***		0.12***
			(0.033)	(0.033)			(0.023)		(0.023)
In school								0.26**	0.19
								(0.119)	(0.120)
Constant	0.61	0.56	1.48*	1.44*	-0.42	-0.43	0.38	-0.62	0.22
	(0.693)	(0.706)	(0.814)	(0.821)	(1.136)	(1.136)	(1.067)	(1.122)	(1.050)
Observations	1,191	1,191	1,191	1,191	1,191	1,191	1,191	1,191	1,191
R-squared	0.133	0.134	0.145	0.146	0.202	0.202	0.216	0.203	0.216
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

VIETNAM	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.53***	0.52***	0.42***	0.41***	0.40***	0.39***	0.32***	0.37***	0.30***
	(0.107)	(0.106)	(0.117)	(0.117)	(0.112)	(0.111)	(0.103)	(0.106)	(0.099)
Preschool		0.00		0.00		0.00			0.00
		(0.003)		(0.003)		(0.003)			(0.003)
HAZ age 5			0.18***	0.18***			0.12***		0.12***
			(0.050)	(0.050)			(0.034)		(0.034)
In school								0.92***	0.89***
								(0.184)	(0.182)
Constant	-0.53	-0.53	1.16	1.15	0.22	0.23	1.37	-0.48	0.66
	(0.949)	(0.950)	(0.737)	(0.741)	(1.244)	(1.245)	(1.160)	(1.181)	(1.121)
Observations	1,149	1,149	1,149	1,149	1,149	1,149	1,149	1,149	1,149
R-squared	0.182	0.182	0.203	0.204	0.141	0.142	0.152	0.149	0.159
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

To explore the factors mediating the link between SES and cognitive development, I include sequentially a set of relevant controls in Table 2. In the first specification (first and fifth columns), only predetermined controls at the child and household level are included. In a second specification (the second and sixth columns), I include the preschool attendance dummy, as this is one important variable that could explain the gaps found, if richer children go to better private schools and poorer children do not attend school at all or go to (bad) government preschools. In the third specification (the third and seventh columns), HAZ at the age of 5 is included, as nutrition during the first five years of life is an important determinant of cognition (Walker et al. 2011). In the fourth specification (only presented for PPVT at the age of 8 regressions), I include an indicator of whether the child is attending school or not (this indicator is not present at the age of 5, as children are too young to be in formal school).

The coefficient of interest, *Expq5*, is that of the fifth household expenditure quintile (with the first quintile as the omitted category). This coefficient shows that *ceteris paribus*, at the age of 5 richer children in Peru will have a PPVT score that is 0.63 SD higher than poorer children. This figure is 0.51 SD in Ethiopia, 0.34 SD in India, and 0.53 SD in Vietnam. At the age of 8, these figures are 0.54, 0.41, 0.45, and 0.40, respectively. This shows that Peru exhibits the largest gradient both at the age of 5 and the age of 8; see columns (1) and (5).

When controlling for preschool attendance, results remain unaltered for the age 5 regressions (with the exception of the coefficient in Ethiopia, which decreases by 4 percentage points – although that does not alter their ranking position). Preschool attendance is significant only for Peru and Ethiopia and it is associated with a 0.25 SD and 0.34 SD higher PPVT score, respectively At the age of 8, however, the inclusion of preschool alters the ranking, with Ethiopia falling to last place and Vietnam to third place. This is due to Ethiopia's coefficient decreasing by 4 percentage points. The coefficient on preschool attendance is still significant only for Peru and Ethiopia (with coefficients of 0.21 SD and 0.34 SD, respectively).

When adding HAZ at the age of 5, results for the PPVT at the age of 5 regressions remain unaltered in relation to the benchmark in column (1) for Peru and India.[13] HAZ is significant for Peru, India, and Vietnam, and ranges from 0.12 SD (India and Peru) to 0.18 SD (Vietnam). Vietnam, whose coefficient on the fifth household expenditure quintile has decreased by 12 percentage points, becomes the third in the ranking, while Ethiopia becomes the second. At the age of 8, the inclusion of HAZ does not alter the ranking either, although the coefficient of Peru decreases to 0.51 SD. The coefficient on HAZ is significant in all countries, and of a similar magnitude to four years previously, at the exception of Vietnam. Adding whether the child is at school at the age of 8 – in column (8) – does not change the ranking results, and this coefficient is significant (and large) in all countries.

Probably the most important specification is the one in columns (4) and (9), in which all the potential mediators are included. The ranking shows that Peru is followed by Ethiopia, Vietnam, and India at the age of 5; while at the age of 8, it is followed by India, Ethiopia, and Vietnam. Moreover, even if the individual inclusion of each of these controls does not change the magnitude of the coefficient of interest, the inclusion of all of them does indeed decrease the magnitude of the SES coefficient.

13 I have added the interaction of expenditure quintiles with HAZ to see if that pattern provides some hints on the channel nutrition–cognition; however, although these interactions are all significant, they are all very close to zero.

Overall, the SES gradient persists even when controlling for a large number of important mediators. Peru has the largest coefficients on the fifth quintile of expenditure in all specifications across ages. Moreover, these finding suggests that preschool might be one important channel that mediates the SES gradients in Ethiopia, but less so in the other countries, while early nutrition seems to be an important channel in Vietnam.

Comparisons of the coefficient of the fifth consumption expenditure quintile between these regressions – using the preferred specifications in columns (4) and (9) – would suggest that SES differences in cognitive achievement decrease over time (with the exception of India, where it increases by 11 percentage points). The decline is 9 percentage points in Peru and Vietnam, and 10 percentage points in Ethiopia.

Table 3 replicates Table 2, but includes all observations in the country samples (that is, not only the majority language) and a dummy that indicates whether the language used by child during test administration was different from the majority language of the country. Main results remain unchanged, although in the regression of the PPVT score at the age of 5, the coefficient on the fifth quintile increases substantially in India, from 0.30 to 0.40 SD in the preferred specification in column (4).

Table 3. *OLS Regressions 4 YL countries, full sample*

PERU

	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.65***	0.63***	0.61***	0.59***	0.57***	0.55***	0.52***	0.56***	0.50***
	(0.081)	(0.085)	(0.081)	(0.084)	(0.094)	(0.096)	(0.089)	(0.097)	(0.092)
No majority language	0.32**	0.32**	0.33**	0.32**	-0.32***	-0.32***	-0.31***	-0.32***	-0.31***
	(0.135)	(0.113)	(0.137)	(0.116)	(0.097)	(0.109)	(0.097)	(0.099)	(0.108)
Preschool		0.26***		0.23***		0.21**			0.18*
		(0.058)		(0.059)		(0.092)			(0.094)
HAZ age 5			0.10***	0.10***			0.10***		0.10***
			(0.027)	(0.028)			(0.026)		(0.027)
In school								0.40*	0.34
								(0.212)	(0.229)
Constant	2.67***	2.71***	2.98***	2.99***	-1.70**	-1.51**	-1.43**	-2.08***	-1.61**
	(0.381)	(0.374)	(0.352)	(0.347)	(0.620)	(0.601)	(0.588)	(0.661)	(0.622)
Observations	1,745	1,745	1,745	1,745	1,745	1,745	1,745	1,745	1,745
R-squared	0.358	0.365	0.367	0.373	0.408	0.413	0.417	0.409	0.422
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

ETHIOPIA

	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.47***	0.42***	0.47***	0.42***	0.48***	0.42***	0.47***	0.42***	0.37***
	(0.108)	(0.098)	(0.108)	(0.098)	(0.126)	(0.115)	(0.125)	(0.113)	(0.098)
No majority language	-0.36***	-0.17*	-0.36***	-0.17*	-0.85***	-0.62***	-0.85***	-0.83***	-0.63***
	(0.089)	(0.094)	(0.089)	(0.095)	(0.235)	(0.193)	(0.235)	(0.231)	(0.191)
Preschool		0.41***		0.41***		0.49***			0.45***
		(0.115)		(0.115)		(0.105)			(0.113)
HAZ age 5			0.10***	0.10**			0.09***		0.06**
			0.035	0.035			0.024		0.022
In school								0.38***	0.36***
								(0.110)	(0.113)
Constant	0.33	0.13	0.40	0.18	1.74***	1.55***	2.49***	1.79***	2.43***
	(0.327)	(0.324)	(0.448)	(0.436)	(0.517)	(0.494)	(0.756)	(0.554)	(0.716)
Observations	1,061	1,061	1,061	1,061	1,061	1,061	1,061	1,061	1,061
R-squared	0.245	0.257	0.245	0.257	0.355	0.372	0.356	0.378	0.394
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

INDIA

	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.40***	0.40***	0.35**	0.35**	0.49***	0.49***	0.45***	0.48***	0.45***
	(0.135)	(0.134)	(0.136)	(0.135)	(0.124)	(0.123)	(0.121)	(0.125)	(0.121)
No majority language	0.35***	0.35***	0.31***	0.31***	-0.06	-0.06	-0.10	-0.06	-0.10
	(0.085)	(0.083)	(0.076)	(0.074)	(0.116)	(0.114)	(0.119)	(0.117)	(0.117)
Preschool		0.03		0.02		0.01			0.01
		(0.101)		(0.102)		(0.077)			(0.076)
HAZ age 5			0.13***	0.13***			0.11***		0.11***
			(0.032)	(0.032)			(0.022)		(0.023)
In school								0.24*	0.17
								(0.123)	(0.122)
Constant	-0.14	-0.17	0.03	0.01	-0.01	-0.02	-0.00	-0.22	-0.16
	(0.470)	(0.415)	(0.502)	(0.453)	(0.702)	(0.718)	(0.671)	(0.695)	(0.672)
Observations	1,454	1,454	1,454	1,454	1,454	1,454	1,454	1,454	1,454
R-squared	0.181	0.181	0.195	0.195	0.189	0.189	0.201	0.189	0.201
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

VIETNAM

	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expq5	0.53***	0.53***	0.39***	0.39***	0.41***	0.41***	0.31***	0.38***	0.29***
	(0.104)	(0.104)	(0.110)	(0.111)	(0.105)	(0.105)	(0.093)	(0.099)	(0.089)
No majority language	-0.69***	-0.68***	-0.61***	-0.60***	-1.05***	-1.04***	-0.99***	-0.69***	-0.64***
	(0.144)	(0.145)	(0.127)	(0.127)	(0.251)	(0.251)	(0.275)	(0.141)	(0.164)
Preschool		0.00		0.00		0.00			0.00
		(0.003)		(0.003)		(0.003)			(0.002)
HAZ age 5			0.18***	0.18***			0.12***		0.12***
			(0.052)	(0.052)			(0.034)		(0.034)
In school								0.97***	0.95***
								(0.165)	(0.157)
Constant	1.03**	1.04**	1.31***	1.32***	2.20***	2.22***	2.39***	1.32*	1.53**
	(0.392)	(0.395)	(0.367)	(0.371)	(0.660)	(0.668)	(0.659)	(0.684)	(0.685)
Observations	1,180	1,180	1,180	1,180	1,180	1,180	1,180	1,180	1,180
R-squared	0.199	0.200	0.223	0.224	0.165	0.165	0.177	0.177	0.189
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

3.2. The SES gradient for changes in PPVT scores across rounds

In this section, a *value-added* specification is estimated, taking advantage of the longitudinal data. Given my interest in the pattern of the PPVT scores over time and SES, the coefficient of interest is now that of the lagged PPVT scores, and again the fifth household expenditure quintile.

The first columns of each country in Table 4 show the same specification as in column (9) of Table 2. The second column of each country shows value-added regression of the PPVT score at the age of 8 on the PPVT score at the age of 5 with child and household controls. Here the two coefficients of interest are positive, as expected. Given the magnitude and significance of PPVT score at age 5, we can say that lagged PPVT score has indeed an independent and significant effect on PPVT score at the age of 8, above and beyond the effect of SES.

The coefficient on lagged PPVT score is largest for Peru (0.42) and Vietnam (0.25). This means that an increase of the PPVT score at the age of 5 of one SD will be associated with an increase in PPVT score at the age of 8 of 0.42 SD in Peru and of 0.25 SD in Vietnam, 0.20 SD in India, and 0.13 SD in Ethiopia. These coefficients are interpreted in the education literature as 'persistence' or depreciation rates of human capital. In this sense, Peru has the highest persistence, and policies aimed at increasing performance on vocabulary tests should be more effective (on average) in Peru than elsewhere.

Table 4. *Value Added regressions 4 YL countries, majority language sample.*

	PERU			ETHIOPIA		
	(1)	(2)	(3)	(4)	(5)	(6)
PPVT age 5 (z-score)		0.42***	0.43***		0.13***	0.16***
		(0.039)	(0.038)		(0.040)	(0.041)
Expq5	0.50***	0.28***		0.37***	0.31***	
	(0.106)	(0.088)		(0.080)	(0.084)	
Observations	1,562	1,562	1,562	403	403	403
R-squared	0.376	0.487	0.478	0.513	0.524	0.513
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Expenditure	Yes	Yes	No	Yes	Yes	No

	INDIA			VIETNAM		
	(1)	(2)	(3)	(4)	(5)	(6)
PPVT age 5 (z-score)		0.20***	0.20***		0.25***	0.25***
		(0.028)	(0.027)		(0.034)	(0.036)
Expq5	0.41***	0.35***		0.30***	0.20**	
	(0.108)	(0.104)		(0.099)	(0.083)	
Observations	1,191	1,191	1,191	1,149	1,149	1,149
R-squared	0.216	0.249	0.239	0.159	0.210	0.203
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Expenditure	Yes	Yes	No	Yes	Yes	No

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

The coefficient on the fifth quintile dummy is still significant, but decreases substantially, and is now largest in India and Ethiopia (0.35 and 0.31 SD, respectively), showing that in terms of *changes* in PPVT scores, the gradient is more pronounced in these two countries. Peru's coefficient (0.28 SD) is then followed by Vietnam (0.20 SD). This means that being richer increases the chances of moving up in the ranking of the distribution of the PPVT scores more in India and Ethiopia, although the coefficients are not that different from the Peruvian one. In the third columns for each country, I take out the expenditure quintile dummies and the lagged PPVT coefficient's magnitude stay nearly the same in the four countries. The latter implies that, as opposed to what happened with the expenditure coefficient (i.e. it decreased in magnitude substantially when including lagged PPVT score), the lagged PPVT coefficient does not decrease in magnitude when including expenditure. This reinforces the message of lagged PPVT score having an independent effect on PPVT score at the age of 8.

Another interesting comparison is to look at the SES coefficient in columns (1) and (2). We observe that the SES coefficient falls by 22, 6, 6, and 10 percentage points in Peru, Ethiopia, India, and Vietnam, respectively.

Overall, the interpretation is that in specification (1) the coefficient of SES reflects the full gradient by the age of 8. The inclusion of PPVT at the age of 5 in column (2), picks up part of the gradient by the age of 8 and the SES coefficient reflects any addition to the gradient between the ages of 5 and 8. Thus, in Peru, around 40 per cent of the gradient by the age of 8 has been determined by the age of 5 and the rest between the ages of 5 and 8. The value-added specification hence suggests that the gradient actually increases with age, but in the other three countries the majority of the gradient has been already formed by the age of 8. The interesting differences is that in the three countries the gradient is much smaller compared to Peru by the age of 5 (as seen in Section 2) but they do increase between the ages of 5 and 8.

In the context of the value-added specifications, a Peruvian child who finds himself below the mean (z-score between −2 and 0) will improve less than children in the other countries, while a child above the mean will see his z-score decrease in relative terms but not as much as in the other three countries. In other words, convergence between groups is going to be (*ceteris paribus*) around twice as slow in Peru as in the other three countries.

However, the coefficient of lagged PPVT score should be interpreted with caution and not in a way that reflects a causal relationship because, by construction, it is correlated with fixed unobserved child characteristics. This suggests that the estimated coefficient will conflate the effect of unobserved innate abilities of the child.

Last but not least, I reject the assumption of lagged PPVT score being a sufficient statistic for all historical inputs, because if it were true, one would not expect that factors such as lagged HAZ would have a significant association with PPVT score at the age of 8. Thus, it seems that in both Peru and India, there are *additional* effects of nutritional status at the age of 5 – a finding that deserves further investigation.

Table 5 replicates Table 4 for the full sample. The ranking of countries in terms of the coefficients remain similar, although the coefficient for past PPVT score decreases in Peru, but not in the other countries. The dummy that indicates whether the language used by the child during the administration of the test was different from the majority language is negative and significant in all specifications, except the India one.

Table 5. *Value Added regressions 4 YL countries, full sample*

	PERU			ETHIOPIA		
	(1)	(2)	(3)	(4)	(5)	(6)
PPVT age 5 (z-score)		0.37***	0.39***		0.16***	0.18***
		(0.039)	(0.039)		(0.026)	(0.026)
Expq5	0.50***	0.31***		0.37***	0.30***	
	(0.092)	(0.076)		(0.098)	(0.094)	
No majority language	-0.31***	-0.45***	-0.43***	-0.63***	-0.60***	-0.59***
	(0.108)	(0.125)	(0.124)	(0.191)	(0.198)	(0.193)
Observations	1,745	1,745	1,745	1,061	1,061	1,061
R-squared	0.422	0.511	0.503	0.394	0.413	0.403
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Expenditure	Yes	Yes	No	Yes	Yes	No
	INDIA			**VIETNAM**		
	(1)	(2)	(3)	(4)	(5)	(6)
PPVT age 5 (z-score)		0.20***	0.21***		0.24***	0.25***
		(0.022)	(0.020)		(0.032)	(0.034)
Expq5	0.45***	0.38***		0.29***	0.20**	
	(0.121)	(0.118)		(0.089)	(0.075)	
No majority language	-0.10	-0.16	-0.12	-0.64***	-0.55***	-0.56***
	(0.117)	(0.116)	(0.121)	(0.164)	(0.188)	(0.185)
Observations	1,454	1,454	1,454	1,180	1,180	1,180
R-squared	0.201	0.233	0.222	0.189	0.237	0.230
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Expenditure	Yes	Yes	No	Yes	Yes	No

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

4. Discussion of results, caveats and conclusions

This paper uses unique comparable data from the four Young Lives countries to show that there are important differences in early language development between children in richer and poorer households that persist as children enter the early school years. It seems that differences in income levels and in other measures of well-being that are apparent in adulthood arise early in children's lives. Although gradients in cognitive development arise in all countries, they do arise more starkly in Peru, *vis à vis* the other three countries.

The main contribution of this study is that it is the first multicontinent comparison of SES gradients in cognitive development for young children in the developing world over critical periods of their lives, based on a common outcome measure for four countries. The second contribution is the use of the longitudinal nature of the data to analyse how deficits in receptive language ability observed at young ages evolve as children enter the early school years and how persistent these are.

Conditional analysis show that SES differences seem to decline over time and that there is no mediator (either preschool, early nutrition, or schooling) that changes the SES cognition gradients – although the magnitude of the gradient decreases when all these covariates are included. Still, preschool seems an important mediator in Ethiopia, particularly at the age of 5, while early nutrition seems to be an important mediator in Vietnam.

Peru has both the largest gradient and the highest PPVT score persistence in the value-added specification. The latter finding is important from the point of view of policy making, as these factors might be preventing a faster catch-up between disadvantaged and better-off children. There is no difference in this persistence between boys and girls, but there are differences between urban and rural households in Peru and Vietnam, and between the rainforest (*Selva*) region and other regions in Peru (see Appendix Tables A2 and A3).

One possible issue that could be investigated (although I am not aware that the data exist) is how much of the higher persistence in Peru is explained by the correlation between SES of the family of a child and the quality of the school this child attends. A cross-country study (Heynemann and Loxley 1983) found that this correlation was 0.25 for Peru and 0.06 for India (with no data for either Vietnam or Ethiopia).[14] If new data of this kind become available, they might provide a good basis for investigating the patterns found in this paper further.

One important caveat in this analysis is the concern over whether it is legitimate to compare gaps in standardised scores across countries because the PPVT may not be comparable when it is translated into different languages. Of the four Young Lives study PPVT tests, only the Peru PPVT has been validated. This might imply that differences in the size of gaps in Peru pick up more variation in true language ability than in other countries where the test has not been validated. For this reason it would be problematic to compare the sizes of coefficients in regressions, even if it still makes sense to talk about factors that matter and ones that do not for language development. Appendix 2 addresses this point and provides justification of why this comparison exercise, even if not perfect, could still be informative. Correlations of the PPVT with other tests and important covariates are shown to provide for a 'validation' of the test, and a summary of the process of adaptation and translation of the test is given.

Another caveat is a more general problem with the value-added specifications, as they can be very sensitive to measurement error, as extensively discussed in the literature (Andrabi et al. 2011). The degree of measurement error is likely to be different across the countries. In Peru, where the test has been most rigorously designed, it is likely to be less of an issue than in the other countries. The same pattern of differences in 'persistence' could then reflect measurement error, with Peru being least affected (therefore having the highest coefficient on lagged test score). In order to address this issue, I have used an alternative measure of cognitive development available in the data, (Cognitive Development Assessment-Quantitative at the age of 5), to instrument the PPVT at the age of 5 and found that the cofficient on lagged PPVT score indeed increases in all countries; however the ranking does not change at all (results available upon request).

Moreover, the analysis for Peru in relation to other countries (and continents) is particularly timely, as the Humala administration is making Early Childhood Development (ECD) a priority in its social policy agenda, and has already created a new programme of integrated

14 According to the Younger Cohort of the Young Lives data, 98 per cent of the Vietnamese children, 83 per cent of the Peruvian children, 73 per cent of the Indian children, and the 68 per cent of Ethiopian children attended government schools.

ECD services for poor households. The results in this paper would directly inform the debate in Peru and, more generally, in many countries in Latin America that are considering similar policy reforms. They reinforce the importance of programmes directed at poor children in developing countries, which has been emphasised in a prominent recent survey (Engle et al., 2011), but with much more direct evidence. A number of interventions have been shown to improve the cognitive development of young children in Latin America and the Caribbean, including preschool in Argentina (Berlinski et al. 2009), parenting interventions in Jamaica (Walker et al. 2011), nutritional supplements in Guatemala (Pollitt et al. 1993), and cash transfers in Nicaragua (Macours et al. 2012). Extending the coverage of these and other programmes should be a high policy priority in Latin American countries.

References

Alderman, H. (2000) 'Anthropometrics' in Margaret Grosh and Paul Glewwe (eds) *Designing HH Survey Questionnaire for Developing Countries: Lessons from 10 Years of LSMS Experience*, Oxford: Oxford University Press.

Andrabi, T., J. Das, A. Ijaz Khwaja and T. Zajonc (2011) 'Do Value-Added Estimates Add Value? Accounting for Learning Dynamics', *American Economic Journal: Applied Economics* 3.4: 2954.

Behrman, J.R., M.C. Calderon, S. Preston, J. Hoddinott, R. Martorell and A.D. Stein (2009) 'Nutritional Supplementation in Girls Influences the Growth of their Children: Prospective Study in Guatemala', *American Journal of Clinical Nutrition* 90.5: 1372–79.

Berlinski, S., S. Galiani and P. Gertler (2009) 'The Effect of Pre-Primary Education on Primary School Performance', *Journal of Public Economics* 93.1–2: 219–34.

Black, S.E. and P.J. Devereux (2010) *Recent Developments in Intergenerational Mobility*, Working Paper 15889, Cambridge, MA: National Bureau of Economic Research.

Blanden, J., P. Gregg and L. Macmillan (2007) 'Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education', *The Economic Journal* 117.519: C43–C60.

Case, A. and C. Paxson (2008) 'Stature and Status: Height, Ability, and Labor Market Outcomes', *Journal of Political Economy* 116.3: 499–532.

Case, A., D. Lubotsky and C. Paxson (2002) 'Economic Status and Health in Childhood: The Origins of the Gradient', *American Economic Review* 92.5: 1308–34.

Cueto, S., J. Leon, G. Guerrero and I. Munoz (2009) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 2 of Young Lives*, Technical Note 15, Oxford: Young Lives.

Cueto, S., J. Leon, G. Guerrero and I. Munoz (2012) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives*, Technical Note 25, Oxford: Young Lives.

Currie, J. (2009) 'Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development', *Journal of Economic Literature* 47.1: 87–122.

Currie, J. and E. Moretti (2003) 'Mother's Education and the Intergenerational Transmission of Human Capital: Evidence From College Openings', *The Quarterly Journal of Economics* 118.4: 1495–532.

Currie, J. and D. Thomas (2000) 'School Quality and the Longer-Term Effects of Head Start', *Journal of Human Resources* 35.4: 755–74.

Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore, MD: Johns Hopkins University Press for the World Bank.

Efron, B. and R.J. Tibshirani (1993) *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Engle, P.L., L.H. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M.C. de Mello, M. Hidrobo, N. Ulkuer I. Ertem, and S. Iltus (2011) 'Strategies for Reducing Inequalities and

Improving Developmental Outcomes for Young Children in Low- and Middle-income Countries', *Lancet* 378.9799: 1339–53.

Fan, J. and I. Gijbels (1996) *Local Polynomial Modeling and its Applications*, London: Chapman & Hall.

Feinstein, L. (2003) 'Inequality in the Early Cognitive Development of British Children in the 1970 Cohort', *Economica* 70.277: 73–97.

Fernald, L.C.H., A. Weber, E. Galasso and L. Ratsifandrihamanana (2011) 'Socioeconomic Gradients and Child Development in a Very Low-income Population: Evidence from Madagascar', *Developmental Science* 14.4: 832–47.

Fleming, S., M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. Maconochie, L. Tarassenko and D. Mant (2011) 'Normal Ranges of Heart Rate and Respiratory Rate in Children from Birth to 18 Years of Age: A Systematic Review of Observational Studies', *The Lancet* 377.9770: 1011–18.

Gasparini, L. and N. Lustig (2011) *The Rise and Fall of Income Inequality in Latin America*, Working Paper 0118, La Plata: CEDLAS.

Harpham, T. (2002) *Measuring Child Poverty and Health: A New International Study*, Working Paper 2, Oxford: Young Lives.

Heckman, J., S. Moon, R. Pinto, P.A. Sazvelyev and A. Yavitz (2010) 'The Rate of Return to the High Scope Perry Preschool Program', *Journal of Public Economics* 94.1–2: 114–28.

Hertz, T., T. Jayasundera, P. Piraino, S. Selcuk, N. Smith and A. Verashchagina (2008) 'The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends', *The B.E. Journal of Economic Analysis & Policy* 7.2: 1–48.

Heynemann, S. and S. Loxley (1983) 'The Effect of Primary School Quality on Academic Achievement across Twenty-nine High- and Low-income Countries', *American Journal of Sociology* 88.6: 1162–94.

Hoddinott, J., J.A. Maluccio, J.R. Behrman, R. Flores and R. Martorell (2008) 'Effect of a Nutrition Intervention during Early Childhood on Economic Productivity in Guatemalan Adults', *The Lancet* 371.9610: 411–16.

IDB (Inter-American Development Bank) (2011) *Estrategia social para la equidad y productividad: América Latina y el Caribe*, Documento de Estrategia, http://www.iadb.org/en/publications/publication-detail,7101.html?id=32906 (accessed 15 Nov 2013).

Macours, K., N. Schady and R. Vakis (2012) 'Cash Transfers, Behavioral Changes, and the Cognitive Development of Young Children: Evidence from a Randomized Experiment', *American Economic Journal: Applied Economics* 4.2: 247–73.

Maluccio, J.A., J. Hoddinott, J.R. Behrman, R. Martorell, A.R. Quisumbing and A.D. Stein (2009) 'The Impact of Experimental Nutritional Interventions on Education into Adulthood in Rural Guatemala', *Economic Journal* 119.537: 734–63.

Moscicki, A., S. Shiboski, N.K. Hills, K.J. Powell, N. Jay, E.N. Hanson, S. Miller, L. Canjura-Clayton, S. Farhat, J.M. Broering and T.M. Darragh (2004) 'Regression of Low-grade Squamous Intraepithelial Lesions in Young Women', *The Lancet* 364.9446: 1678–83.

Naudeau, S., S. Martínez, P. Premand and D. Filmer (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*, Washington, DC: World Bank.

Park, Y., D.J. Hunter, D. Spiegelman, L. Bergkvist, F. Berrino, P.A. van den Brandi, J.E. Buring, G.A. Colditz, J.L. Freudenheim, C.S. Fuchs, E. Giovannucii, R.A. Goldbohm, S. Graham, L. Harnack, A.M. Hartman, D.R. Jacobs Jr, I. Kato, V. Krogh, M.F. Leitzmann, M.L. McCullough, A.B. Miller, P. Pierinen, T.E. Rohan, A. Schatzkin, W.C. Willet, A. Wolk, A. Zeleniuch-Jacquotte, S.M. Zhang, and S.A. Smith-Warner (2005) 'Dietary Fiber Intake and Risk of Colorectal Cancer: A Pooled Analysis of Prospective Cohort Studies', *Journal of the American Medical Association* 294.22: 2849–857.

Paxson, C. and N. Schady (2007) 'Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting', *Journal of Human Resources* 42.1: 49–84.

Pollitt, E., K.S. Gorman, P. Engle, R. Martorell and J. Rivera (1993) 'Early Supplementary Feeding and Cognition: Effects over Two Decades', *Monograms in Social Research in Child Development* 58.7: 1–99.

Schady, N. (2011) 'Parental Education, Vocabulary, and Cognitive Development in Early Childhood: Longitudinal Evidence from Ecuador', *American Journal of Public Health* 101.12: 2299–307.

Schady, N., J. Behrman, M.C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson and R. Vakis (2014) *Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries*, IDB Working Paper No. 482, New York: Inter-American Development Bank (forthcoming in *Journal of Human Resources).*

Stein, A.D., M. Wang A. DiGirolamo, R. Grajeda, U. Ramakrishnan, M. Ramirez-Zea, K. Yount and R. Martorell (2008) 'Nutritional Supplementation in Early Childhood, Schooling, and Intellectual Functioning in Adulthood. A Prospective Study in Guatemala', *Archives of Pediatrics and Adolescent Medicine* 162.7: 612–18.

Todd, P.E. and K.I. Wolpin (2007) 'The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps', *Journal of Human Capital* 1.1: 91–136.

Victora, C.G., A. Wagstaff, J.A. Schellenberg, D. Swatkin, M. Claeson and J-P Habicht (2003) 'Applying an Equity Lens to Child Health Mortality: More of the Same Is Not Enough', *The Lancet* 362: 233–41.

Walker, S.P., S.M. Chang, M. Vera-Hernández and S. Grantham-McGregor (2011) 'Early Childhood Stimulation with Stunted Children Benefits Adult Competence and Reduces Violent Behavior', *Pediatrics* 127.5: 849–57.

WHO (World Health Organization) (2000) *WHO/NCHS Reference Standards: WHO Working Group: Use and Interpretation of Anthropometric Indicators on Nutritional Status*. Geneva: WHO Press.

World Bank (2005) *World Development Indicators*, Washington, DC: World Bank, http://data.worldbank.org/sites/default/files/wdi05fulltext.pdf (accessed 15 Nov 2013).

Appendix 1. Additional figures and tables

Figure A.1. *Distribution of household expenditure in Young Lives study countries (majority language speakers)*

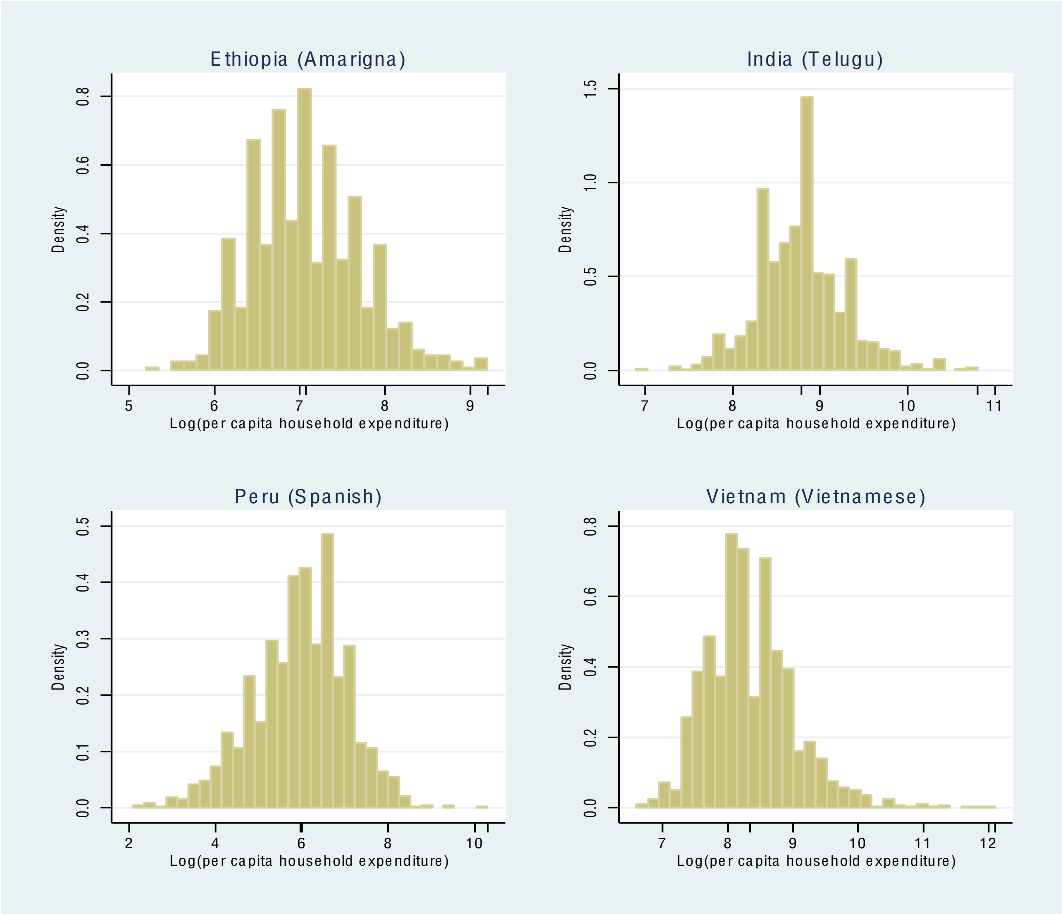


Figure A2. *Distribution of wealth in Young Lives study countries (majority language speakers)*

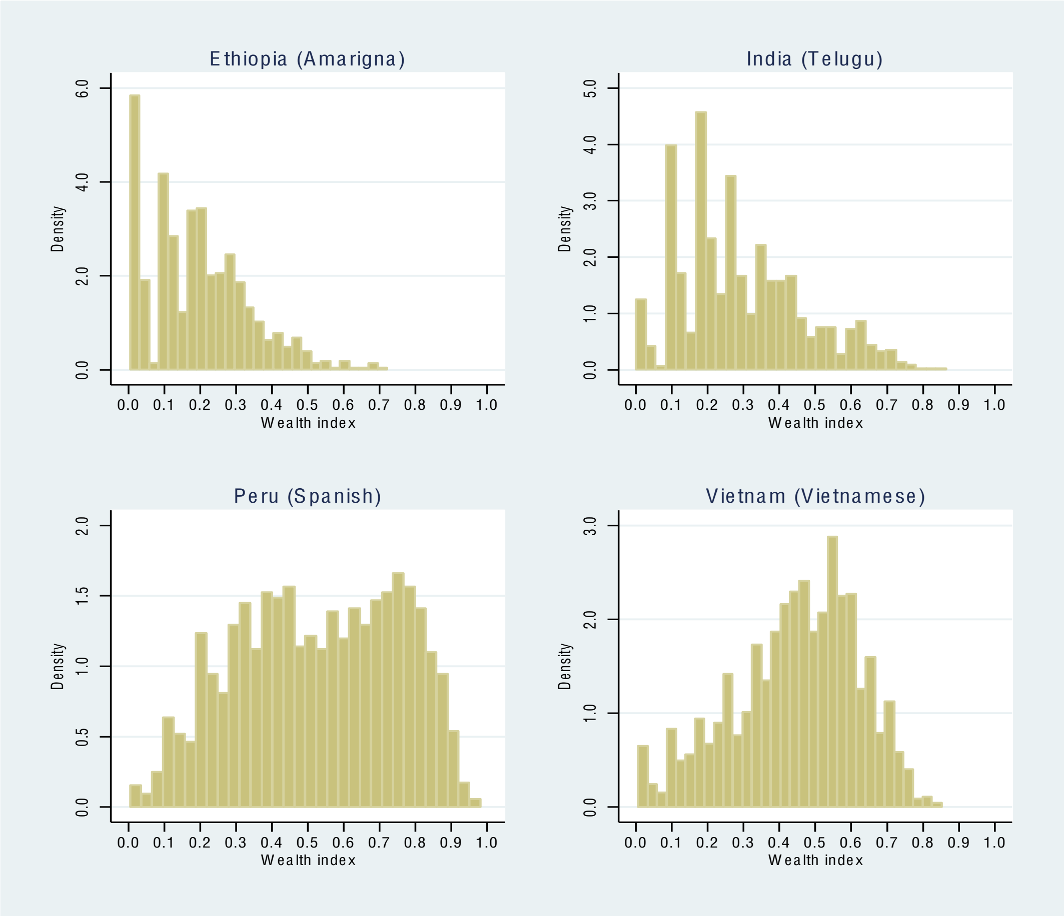


Table A1. *OLS Regressions 4 YL countries, majority language sample.*

PERU	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.05	-0.05	-0.04	-0.04	-0.07	-0.07	-0.06	-0.06	-0.05
	(0.043)	(0.042)	(0.041)	(0.040)	(0.044)	(0.044)	(0.042)	(0.045)	(0.043)
Age (in months)	-0.06***	-0.06***	-0.06***	-0.06***	0.01	0.01	0.01	0.01	0.01
	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)
R1firstborn	0.01	0.06	0.04	0.08	0.58***	0.62***	0.61***	0.58***	0.64***
	(0.197)	(0.172)	(0.198)	(0.176)	(0.165)	(0.171)	(0.190)	(0.167)	(0.197)
R1lastborn	0.07*	0.08**	0.07*	0.08**	-0.02	-0.01	-0.01	-0.02	-0.00
	(0.036)	(0.035)	(0.037)	(0.036)	(0.041)	(0.040)	(0.044)	(0.040)	(0.041)
Region 1	0.10	0.05	0.08	0.04	-0.12	-0.17	-0.14	-0.13	-0.19*
	(0.123)	(0.120)	(0.120)	(0.117)	(0.095)	(0.099)	(0.094)	(0.095)	(0.095)
Region 2	0.05	0.00	0.07	0.02	-0.06	-0.10	-0.04	-0.06	-0.08
	(0.088)	(0.089)	(0.087)	(0.088)	(0.077)	(0.085)	(0.077)	(0.077)	(0.084)
Urban	0.44***	0.44***	0.40***	0.41***	0.37***	0.37***	0.33***	0.37***	0.33***
	(0.077)	(0.071)	(0.075)	(0.071)	(0.093)	(0.090)	(0.088)	(0.092)	(0.085)
HHSIZE	-0.04***	-0.04***	-0.04***	-0.04***	-0.03**	-0.03**	-0.02*	-0.03**	-0.02*
	(0.010)	(0.011)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Expq2	0.12	0.12	0.11	0.11	0.34***	0.34***	0.33***	0.34***	0.33***
	(0.086)	(0.087)	(0.089)	(0.090)	(0.077)	(0.077)	(0.078)	(0.078)	(0.078)
Expq3	0.27***	0.26***	0.25***	0.24***	0.33***	0.32***	0.31***	0.33***	0.30***
	(0.081)	(0.084)	(0.080)	(0.083)	(0.088)	(0.088)	(0.086)	(0.090)	(0.088)
Expq4	0.47***	0.45***	0.44***	0.43***	0.52***	0.51***	0.50***	0.52***	0.48***
	(0.080)	(0.085)	(0.082)	(0.086)	(0.088)	(0.091)	(0.084)	(0.090)	(0.088)
Expq5	0.63***	0.62***	0.60***	0.59***	0.54***	0.53***	0.51***	0.54***	0.50***
	(0.099)	(0.103)	(0.099)	(0.102)	(0.105)	(0.108)	(0.101)	(0.107)	(0.106)
Caregiver's edu	0.05***	0.05***	0.05***	0.04***	0.05***	0.05***	0.05***	0.05***	0.05***
	(0.011)	(0.011)	(0.010)	(0.010)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Father's edu	0.05***	0.04***	0.04***	0.04***	0.05***	0.05***	0.05***	0.05***	0.05***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.008)	(0.009)	(0.009)	(0.008)
Mother's height	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.01*	0.00	-0.01*
	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Preschool		0.25***		0.22***		0.21**			0.17*
		(0.062)		(0.063)		(0.088)			(0.088)
HAZ age 5			0.12***	0.12***			0.14***		0.13***
			(0.030)	(0.031)			(0.029)		(0.030)
In school								0.57*	0.52*
								(0.281)	(0.288)
Constant	2.03**	2.11**	3.50***	3.50***	-2.06*	-1.84*	-0.51	-2.64**	-0.91
	(0.847)	(0.852)	(0.961)	(0.978)	(1.059)	(1.052)	(1.050)	(1.080)	(1.072)
Observations	1,562	1,562	1,562	1,562	1,562	1,562	1,562	1,562	1,562
R-squared	0.367	0.373	0.379	0.384	0.358	0.362	0.372	0.359	0.376
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

Table A1. *OLS Regressions 4 YL countries, majority language sample* continued.

ETHIOPIA									
	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.06	-0.05	-0.06	-0.05	0.04	0.05	0.04	0.02	0.04
	(0.084)	(0.086)	(0.084)	(0.086)	(0.078)	(0.082)	(0.078)	(0.076)	(0.081)
Age (in months)	0.00	0.00	-0.00	-0.01	-0.00	-0.00	-0.02	-0.00	-0.02
	(0.009)	(0.009)	(0.013)	(0.013)	(0.008)	(0.008)	(0.012)	(0.008)	(0.012)
R1firstborn	1.11***	1.14***	1.13***	1.15***	1.40***	1.43***	1.43***	1.38***	1.43***
	(0.317)	(0.320)	(0.325)	(0.328)	(0.245)	(0.240)	(0.257)	(0.227)	(0.233)
R1lastborn	0.05	0.08	0.05	0.08	0.03	0.06	0.03	0.03	0.06
	(0.065)	(0.069)	(0.066)	(0.070)	(0.082)	(0.085)	(0.084)	(0.082)	(0.088)
Region 1	0.26*	0.20	0.27*	0.21	-0.04	-0.11	-0.04	-0.08	-0.14
	(0.147)	(0.146)	(0.152)	(0.152)	(0.176)	(0.164)	(0.180)	(0.174)	(0.169)
Region 2	-0.51***	-0.43***	-0.51***	-0.43***	-0.34*	-0.26	-0.35*	-0.38*	-0.32*
	(0.063)	(0.070)	(0.063)	(0.071)	(0.188)	(0.156)	(0.188)	(0.187)	(0.155)
Urban	0.26***	0.10	0.25***	0.10	0.95***	0.79***	0.94***	0.89***	0.75***
	(0.065)	(0.121)	(0.068)	(0.125)	(0.149)	(0.145)	(0.146)	(0.148)	(0.148)
HHSIZE	-0.01	-0.01	-0.01	-0.01	-0.04**	-0.03**	-0.04**	-0.03**	-0.03**
	(0.017)	(0.016)	(0.018)	(0.018)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)
Expq2	-0.02	-0.02	-0.02	-0.02	-0.06	-0.06	-0.05	-0.04	-0.04
	(0.071)	(0.068)	(0.072)	(0.070)	(0.085)	(0.086)	(0.085)	(0.090)	(0.091)
Expq3	-0.11	-0.10	-0.11	-0.10	-0.04	-0.03	-0.04	-0.03	-0.02
	(0.095)	(0.087)	(0.096)	(0.088)	(0.090)	(0.092)	(0.090)	(0.089)	(0.091)
Expq4	0.11	0.09	0.11	0.09	0.10	0.08	0.10	0.12	0.10
	(0.107)	(0.100)	(0.109)	(0.101)	(0.077)	(0.078)	(0.076)	(0.081)	(0.080)
Expq5	0.51***	0.47**	0.51***	0.47**	0.41***	0.37***	0.40***	0.41***	0.37***
	(0.168)	(0.164)	(0.169)	(0.165)	(0.092)	(0.087)	(0.093)	(0.086)	(0.080)
Caregiver's edu	0.01**	0.01**	0.01**	0.01**	0.01*	0.01	0.01*	0.01*	0.01*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Father's edu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother's height	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.01	0.00	0.00
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)
Preschool		0.34*		0.34*		0.34***			0.31**
		(0.166)		(0.168)		(0.101)			(0.106)
HAZ age 5			0.17**	0.15*			0.07		-0.00**
			0.074	0.074			0.047		(0.048)
In school								0.34***	0.31***
								(0.083)	(0.085)
Constant	0.15	0.27	0.59	0.73	-0.97	-0.78	0.47	-0.93	0.65
	(0.804)	(0.822)	(0.986)	(0.959)	(1.233)	(1.209)	(1.447)	(1.259)	(1.379)
Observations	403	403	403	403	403	403	403	403	403
R-squared	0.339	0.347	0.340	0.348	0.493	0.502	0.494	0.505	0.513
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

Table A1. *OLS Regressions 4 YL countries, majority language sample* continued.

INDIA									
	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.05	-0.05	-0.06	-0.06	-0.24***	-0.24***	-0.25***	-0.23***	-0.25***
	(0.047)	(0.047)	(0.047)	(0.047)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)
Age (in months)	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00
	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
R1firstborn	0.22	0.21	0.23	0.22	-0.23	-0.23	-0.22	-0.23	-0.22
	(0.421)	(0.416)	(0.434)	(0.429)	(0.975)	(0.977)	(0.880)	(0.975)	(0.883)
R1lastborn	-0.05	-0.05	-0.04	-0.04	-0.18***	-0.18***	-0.17***	-0.18***	-0.17***
	(0.067)	(0.066)	(0.066)	(0.066)	(0.044)	(0.044)	(0.042)	(0.044)	(0.042)
Region 1	0.49*	0.49*	0.46*	0.46*	0.21**	0.21**	0.18**	0.21**	0.18**
	(0.266)	(0.266)	(0.263)	(0.263)	(0.084)	(0.083)	(0.077)	(0.083)	(0.077)
Region 2	0.06	0.06	0.03	0.03	0.58***	0.58***	0.55***	0.58***	0.55***
	(0.071)	(0.073)	(0.080)	(0.081)	(0.125)	(0.125)	(0.115)	(0.125)	(0.116)
Urban	0.27	0.27	0.25	0.25	0.42***	0.42***	0.40***	0.42***	0.40***
	(0.224)	(0.223)	(0.228)	(0.227)	(0.092)	(0.091)	(0.084)	(0.092)	(0.083)
HHSIZE	-0.04*	-0.03*	-0.03	-0.03	-0.04***	-0.04***	-0.04**	-0.04***	-0.03**
	(0.019)	(0.018)	(0.019)	(0.019)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Expq2	0.09	0.10	0.08	0.08	0.17	0.17	0.16	0.17	0.16
	(0.106)	(0.107)	(0.105)	(0.106)	(0.118)	(0.118)	(0.118)	(0.118)	(0.118)
Expq3	0.15	0.14	0.14	0.13	0.16*	0.16*	0.15*	0.16*	0.15*
	(0.142)	(0.143)	(0.144)	(0.144)	(0.085)	(0.085)	(0.082)	(0.085)	(0.082)
Expq4	0.21	0.20	0.19	0.18	0.31***	0.31***	0.28**	0.30***	0.28**
	(0.121)	(0.123)	(0.122)	(0.123)	(0.106)	(0.106)	(0.104)	(0.107)	(0.103)
Expq5	0.34**	0.34**	0.30*	0.30*	0.45***	0.45***	0.41***	0.44***	0.41***
	(0.157)	(0.155)	(0.161)	(0.160)	(0.116)	(0.116)	(0.108)	(0.116)	(0.108)
Caregiver's edu	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.01***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Father's edu	0.02**	0.02**	0.02**	0.02**	0.01**	0.01**	0.01**	0.01**	0.01**
	(0.008)	(0.008)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Mother's height	-0.01	-0.01	-0.01**	-0.01**	0.00	0.00	-0.00	0.00	-0.00
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)
Preschool		0.09		0.09		0.01			0.01
		(0.102)		(0.100)		(0.088)			(0.087)
HAZ age 5			0.12***	0.12***			0.12***		0.12***
			(0.033)	(0.033)			(0.023)		(0.023)
In school								0.26**	0.19
								(0.119)	(0.120)
Constant	0.61	0.56	1.48*	1.44*	-0.42	-0.43	0.38	-0.62	0.22
	(0.693)	(0.706)	(0.814)	(0.821)	(1.136)	(1.136)	(1.067)	(1.122)	(1.050)
Observations	1,191	1,191	1,191	1,191	1,191	1,191	1,191	1,191	1,191
R-squared	0.133	0.134	0.145	0.146	0.202	0.202	0.216	0.203	0.216
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

Table A1. *OLS Regressions 4 YL countries, majority language sample* continued.

VIETNAM									
	PPVT age 5				PPVT age 8				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04
	(0.070)	(0.070)	(0.068)	(0.067)	(0.057)	(0.057)	(0.057)	(0.055)	(0.055)
Age (in months)	-0.03***	-0.03***	-0.02***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
R1firstborn	0.20	0.20	0.10	0.10	0.13	0.13	0.06	0.11	0.05
	(0.366)	(0.366)	(0.325)	(0.326)	(0.157)	(0.156)	(0.133)	(0.153)	(0.129)
R1lastborn	-0.10	-0.10	-0.07	-0.07	-0.14*	-0.14*	-0.12*	-0.14*	-0.12*
	(0.059)	(0.059)	(0.057)	(0.057)	(0.066)	(0.066)	(0.064)	(0.069)	(0.067)
Region 1	0.11	0.11	0.21	0.21	-0.01	-0.01	0.06	0.01	0.07
	(0.165)	(0.163)	(0.155)	(0.153)	(0.154)	(0.154)	(0.153)	(0.149)	(0.149)
Region 2	0.39	0.39	0.44	0.44	0.52***	0.52***	0.56***	0.53***	0.56***
	(0.282)	(0.282)	(0.282)	(0.282)	(0.140)	(0.140)	(0.146)	(0.139)	(0.145)
Urban	0.74***	0.74***	0.70***	0.70***	0.36*	0.36*	0.34	0.37*	0.34*
	(0.167)	(0.168)	(0.152)	(0.153)	(0.209)	(0.208)	(0.196)	(0.209)	(0.196)
HHSIZE	-0.03	-0.03	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01
	(0.028)	(0.028)	(0.026)	(0.026)	(0.021)	(0.021)	(0.019)	(0.020)	(0.018)
Expq2	0.30**	0.30**	0.26**	0.26**	0.11	0.11	0.08	0.09	0.07
	(0.119)	(0.119)	(0.124)	(0.125)	(0.124)	(0.124)	(0.124)	(0.115)	(0.116)
Expq3	0.25*	0.24*	0.20	0.20	0.08	0.08	0.05	0.06	0.03
	(0.122)	(0.122)	(0.125)	(0.125)	(0.130)	(0.130)	(0.125)	(0.127)	(0.121)
Expq4	0.30**	0.30**	0.21	0.22	0.32**	0.32**	0.26**	0.29**	0.24*
	(0.135)	(0.136)	(0.135)	(0.136)	(0.129)	(0.130)	(0.122)	(0.122)	(0.116)
Expq5	0.53***	0.52***	0.42***	0.41***	0.40***	0.39***	0.32***	0.37***	0.30***
	(0.107)	(0.106)	(0.117)	(0.117)	(0.112)	(0.111)	(0.103)	(0.106)	(0.099)
Caregiver's edu	0.02*	0.02*	0.02*	0.02*	0.02	0.02	0.02	0.02	0.02
	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)
Father's edu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Mother's height	0.01*	0.01*	0.00	0.00	0.01**	0.01**	0.01	0.01**	0.01
	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Preschool		0.00		0.00		0.00			0.00
		(0.003)		(0.003)		(0.003)			(0.003)
HAZ age 5			0.18***	0.18***			0.12***		0.12***
			(0.050)	(0.050)			(0.034)		(0.034)
In school								0.92***	0.89***
								(0.184)	(0.182)
Constant	-0.53	-0.53	1.16	1.15	0.22	0.23	1.37	-0.48	0.66
	(0.949)	(0.950)	(0.737)	(0.741)	(1.244)	(1.245)	(1.160)	(1.181)	(1.121)
Observations	1,149	1,149	1,149	1,149	1,149	1,149	1,149	1,149	1,149
R-squared	0.182	0.182	0.203	0.204	0.141	0.142	0.152	0.149	0.159
Child and HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

Table A2. *Value Added regressions: urban vs. rural households*

	PERU		ETHIOPIA		INDIA		VIETNAM	
	Urban (1)	Rural (2)	Urban (3)	Rural (4)	Urban (5)	Rural (6)	Urban (7)	Rural (8)
PPVT age 5 (z-score)	0.49***	0.33***	0.14**	0.13**	0.20**	0.20***	0.13*	0.27***
	(0.043)	(0.048)	(0.046)	(0.045)	(0.082)	(0.027)	(0.058)	(0.039)
Female	0.01	-0.11	0.11	-0.07	-0.48***	-0.17**	-0.01	-0.03
	(0.050)	(0.084)	(0.100)	(0.070)	(0.111)	(0.063)	(0.069)	(0.066)
Age (in months)	0.04***	0.04***	-0.01	-0.02*	-0.06**	0.01	-0.05**	-0.02**
	(0.007)	(0.008)	(0.021)	(0.011)	(0.021)	(0.009)	(0.020)	(0.007)
R1firstborn	0.48*	0.92***	0.77***	1.78***		-0.22	-0.42	0.06
	(0.272)	(0.200)	(0.161)	(0.175)		(0.901)	(0.284)	(0.096)
R1lastborn	-0.06	-0.01	-0.00	0.14	-0.07	-0.16***	-0.25***	-0.06
	(0.049)	(0.085)	(0.123)	(0.097)	(0.129)	(0.044)	(0.078)	(0.071)
Region 1	0.11**	-0.04	-0.17		-0.21	0.13**	0.50	0.01
	(0.045)	(0.135)	(0.163)		(0.202)	(0.062)	(0.370)	(0.146)
Region 2	0.23**	-0.17***	-0.46**	0.04	0.53**	0.59***	0.37	0.42***
	(0.082)	(0.038)	(0.171)	(0.150)	(0.239)	(0.131)	(0.389)	(0.114)
HHSIZE	0.01	-0.03	-0.02*	-0.05*	-0.06	-0.02*	0.04*	-0.02
	(0.009)	(0.019)	(0.012)	(0.022)	(0.056)	(0.012)	(0.019)	(0.019)
Expq2	0.11	0.30***	-0.33*	0.12	0.28	0.14	0.26	0.01
	(0.122)	(0.058)	(0.152)	(0.081)	(0.321)	(0.119)	(0.251)	(0.101)
Expq3	0.04	0.17	-0.16	0.09	0.19	0.12	0.25	-0.02
	(0.101)	(0.125)	(0.136)	(0.086)	(0.241)	(0.100)	(0.158)	(0.106)
Expq4	0.15	0.43***	0.01	0.13	0.12	0.26**	0.35**	0.20*
	(0.124)	(0.124)	(0.056)	(0.175)	(0.250)	(0.097)	(0.121)	(0.095)
Expq5	0.09	0.33	0.24*	0.15	0.20	0.38***	0.33	0.21**
	(0.132)	(0.241)	(0.100)	(0.205)	(0.218)	(0.122)	(0.250)	(0.090)
Caregiver's edu	0.01	0.04***	0.01	0.00	0.05**	0.01	0.04	0.01
	(0.011)	(0.008)	(0.004)	(0.004)	(0.017)	(0.004)	(0.029)	(0.008)
Father's edu	0.01	0.05***	0.00	0.00	0.04**	0.01	-0.00	0.00
	(0.010)	(0.010)	(0.003)	(0.003)	(0.015)	(0.004)	(0.004)	(0.004)
HAZ age 5	0.06*	0.12**	-0.00	-0.00**	0.05	0.11***	0.12	0.06*
	(0.031)	(0.042)	(0.001)	(0.001)	(0.074)	(0.028)	(0.074)	(0.032)
Mother's height	-0.00	-0.01	0.01	-0.01	0.02	-0.01	-0.01	0.01
	(0.004)	(0.007)	(0.009)	(0.005)	(0.015)	(0.004)	(0.004)	(0.005)
In school	1.06***	-0.01	0.32	0.36***	0.52**	0.06		0.69***
	(0.191)	(0.165)	(0.210)	(0.074)	(0.195)	(0.074)		(0.162)
Constant	-4.26***	-2.48	-0.35	2.25	2.07	-0.32	4.99**	-0.47
	(1.039)	(1.501)	(1.839)	(1.734)	(2.760)	(1.117)	(1.890)	(1.195)
Observations	957	605	364	259	203	988	253	896
R-squared	0.441	0.309	0.175	0.194	0.279	0.221	0.167	0.204

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

Table A3. *Value Added regressions: Peru, by region*

	Costa	Sierra	Selva
PPVT age 5 (z-score)	0.47***	0.43***	0.27***
	(0.040)	(0.051)	(0.043)
Female	0.07*	-0.16**	-0.01
	(0.037)	(0.064)	(0.114)
Age (in months)	0.03***	0.05***	0.03
	(0.006)	(0.007)	(0.016)
R1firstborn	0.05	0.87***	0.82***
	(0.104)	(0.140)	(0.170)
R1lastborn	0.05	-0.12*	-0.08
	(0.045)	(0.062)	(0.091)
Urban	0.10	0.24**	0.14*
	(0.102)	(0.106)	(0.072)
HHSIZE	-0.00	-0.01	-0.03
	(0.009)	(0.019)	(0.025)
Expq2	0.19**	0.36***	0.04
	(0.084)	(0.086)	(0.099)
Expq3	0.16**	0.24*	-0.03
	(0.066)	(0.134)	(0.147)
Expq4	0.30***	0.34**	0.19***
	(0.078)	(0.149)	(0.052)
Expq5	0.23**	0.30	-0.03
	(0.096)	(0.176)	(0.078)
Caregiver's edu	0.01	0.04***	0.03
	(0.013)	(0.010)	(0.015)
Father's edu	0.02	0.04***	0.03
	(0.009)	(0.012)	(0.020)
HAZ age 5	0.08**	0.08**	0.14**
	(0.039)	(0.032)	(0.044)
Mother's height	-0.01	-0.00	-0.02
	(0.005)	(0.007)	(0.012)
In school	1.06***	0.70*	0.15
	(0.111)	(0.347)	(0.210)
Constant	-3.17***	-5.93***	0.02
	(1.030)	(1.278)	(1.349)
Observations	632	687	243
R-squared	0.447	0.508	0.375

Notes: OLS regressions. Sex, age, birth-order, region, urban dummy, household size, expenditure quintiles, education of caregiver and father and mother's height are included as controls in all columns. Robust standard errors in parentheses. Significance at 1, 5 and 10 percent denoted by ***, ** and *.

Appendix 2. Validity of the PPVT to compare gaps across countries

As mentioned in the last section of the paper, one important caveat in this analysis is the concern over whether it is legitimate to compare gaps in standardised scores across countries because the PPVT test may not be comparable when it is translated into different languages. As Cueto et al. (2012) point out:

On the one hand this test provides very interesting data and it has been used for several Young Lives papers, usually performing in ways that would be predicted by theory. It must be remembered, however, that it has limitations, arising mostly from the fact that it is an instrument developed in English in the USA. While local teams have worked to adapt the test to its local language and we have analysed DIF (differential item functioning), some bias is likely to remain. Again, we suggest not using the scores for comparisons across different language groups.[15]

In this Appendix, I provide some justification on why the PPVT seems to provide the only longitudinal information available to make some sort of comparisons across the four Young Lives country studies.

We start by showing in Table A4 that the PPVT measures do not seem to be that noisy but a real measure of skills as they are well correlated with measures of SES (such as expenditure) and other cognitive measures such as the Cognitive Development Assessment-Quantitative (CDA-Q) at the age of 5 and maths scores at the age of 8.

Table A4. *Correlations of PPVT scores at ages 5 and 8 with expenditure, and other cognitive tests*

	PERU		ETHIOPIA		INDIA		VIETNAM	
	Age 5	Age 8	Age 5	Age 8	Age 5	Age 8	Age 5	Age 8
Ln expenditure per capita	0.48	0.52	0.38	0.43	0.12	0.20	0.34	0.25
CDA-Q test – raw score	0.45		0.57		0.43		0.47	
CDA-Q test – Rasch score	0.44		0.58		0.45		0.42	
Maths raw score original		0.57		0.64		0.46		0.37
Maths raw score corrected		0.57		0.64		0.44		0.38
Maths Rasch score corrected		0.57		0.63		0.42		0.38

Still, the correlations with expenditure seem much lower in India. Moreover, Figure 4 of this paper shows that HAZ and the PPVT score have similar patterns across and within countries.

More importantly, in Table A5 I replicate the same regressions presented in Table 2 (specifications as in fourth column for the age of 5 and as ninth column for the age of 8) for

15 For more on the process of adaptation and translation of the PPVT in each country, see Young Lives Technical Notes 15 and 25 at http://www.younglives.org.uk/publications/TN.

the available tests in the survey (CDA-Q for Round 2 and maths score for Round 3) and examine the ranking in relation to the magnitude of the coefficient of the fifth expenditure quintile on the PPVT. The rankings are identical when taking the PPVT and the CDA-Q at the age of 5, and this strongly supports the validity of my estimates. However, at the age of 8 the ranking does change as Peru, Ethiopia and Vietnam are somehow 'clustered' with high coefficients and India being the outlier. It has to be mentioned that a maths test is very different from a vocabulary test, particularly at this age, when children are in school.

Table A5. *Ranking of countries in relation to the fifth expenditure quintile for PPVT (ages 5 and 8), CDA-Q (age 5) and maths score (age 8)*

	PPVT (age 5)	Ranking	PPVT (age 8)	Ranking	Raw CDA-Q (age 5)	Ranking	Raw maths (age 8)	Ranking
Peru	0.59***	1	0.50***	1	1.11***	1	2.49***	3
Ethiopia	0.47**	2	0.37***	3	0.84*	2	2.59***	1
India	0.30*	4	0.41***	2	0.67**	4	1.84**	4
Vietnam	0.41***	3	0.30***	4	0.70	3	2.53***	2

Note: The same ranking results are obtained when using the Rasch score of CDA-Q instead of the raw CDA-Q score.
The same ranking results are also obtained when using maths raw score corrected or the maths Rasch score.
All age 5 regressions follow the specification of the fourth column Table 2, while all age 8 regressions follow the specification of the ninth column of Table 2.

Socio-economic Status and Early Childhood Cognitive Skills: Is Latin America Different?

This paper documents differences in cognitive development – as measured by a receptive vocabulary test – between children from households with high and low socioeconomic status (SES) in two different phases of childhood (before and after early school years) in four developing countries: Peru, Ethiopia, India and Vietnam. Intercontinental evidence on the timing, pattern, and persistence of these differences is provided. The non-parametric analysis suggests that differences found at the age of 5 persist into the early school years across all four countries, and the conditional analysis shows that the magnitude of within-country SES differences seem to diminish over time (with the exception of the India sample). However, both the magnitude of the gap and the degree of persistence vary. The main result is that Peru stands out, not only as the country with the largest cross-section difference between rich and poor (of around 1.30–1.40 standard deviations), but also as the country with the highest persistence in cognitive development, as shown by the value-added specification. Some channels behind these trends are discussed, but overall, the SES gradient persists even when controlling for a range of important mediators, such as preschool, early nutrition, and years of schooling. Past performance on the Peabody Picture Vocabulary Test (PPVT) is an important determinant of the SES gradient at the age of 8.

Young Lives
An International Study of Childhood Poverty

Young Lives
An International Study of Childhood Poverty

About Young Lives

Young Lives is an international study of childhood poverty, involving 12,000 children in 4 countries over 15 years. It is led by a team in the Department of International Development at the University of Oxford in association with research and policy partners in the 4 study countries: Ethiopia, India, Peru and Vietnam.

Through researching different aspects of children's lives, we seek to improve policies and programmes for children.

Young Lives Partners

Young Lives is coordinated by a small team based at the University of Oxford, led by Professor Jo Boyden.

- *Ethiopian Development Research Institute, Ethiopia*
- *Pankhurst Development Research and Consulting plc*
- *Save the Children (Ethiopia programme)*
- *Centre for Economic and Social Sciences, Andhra Pradesh, India*
- *Save the Children India*
- *Sri Padmavathi Mahila Visvavidyalayam (Women's University), Andhra Pradesh, India*
- *Grupo de Análisis para el Desarollo (GRADE), Peru*
- *Instituto de Investigación Nutricional, Peru*
- *Centre for Analysis and Forecasting, Vietnamese Academy of Social Sciences, Vietnam*
- *General Statistics Office, Vietnam*
- *University of Oxford, UK*

Contact:
Young Lives
Oxford Department of International Development,
University of Oxford,
3 Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@younglives.org.uk
Website: www.younglives.org.uk