Richer but more unequal? Nutrition and caste gaps

Florencia López Bóo *and Maria Eugenia Canon †

October 21, 2012

Abstract

This paper explores children's cognitive outcomes using novel panel data from India for children 6 months through 8 years. For the first time in a developing country, this allow us to estimate a value-added model of cognitive development at a very young age. We look at the nutrition-cognition link and at the relationship between caste and test scores. We use an instrumental variable approach and find that a 1 standard deviation increase in height-for-age at the age of 5 leads to cognitive test scores that are about a 16 per cent of a SD higher at age 8. Our analysis suggests that the differences in income levels between castes found in adulthood arise early in childhood. After controlling for a wide range of controls; upper caste children show a substantial advantage in vocabulary tests, but most importantly, they show a more pronounced gender inequality than their lower caste counterparts. Compensating low caste children with the average nutritional status of their upper caste counterparts would close around one fifth of the caste cognitive differentials. We also show that UC families discriminate more against girls. Using a sub-sample of the data with the siblings' birthweight in a unique way, we find that family fixed effects explain 1 SD of the overall nutrition-cognition effect.
JEL Classification Codes: J13, J15, J24
Keywords: cognitive outcomes,nutrition, children and test score gaps.

---

*email: florencial@iadb.org. Inter-American Development Bank/IZA. 1300 New York Av., NW (E0605), Washington DC, 20577. The authors wish to thank M.N.Asadullah and G. Kingdon. This research was funded by Young Lives, a longterm international research project investigating the changing nature of childhood poverty. Young Lives is core-funded by the UK Department for International Development (DFID) for the benefit of developing countries. Sub-studies are funded by IDRC (in Ethiopia), UNICEF (in India), the Bernard van Leer Foundation (in India and Peru) and Irish Aid (in Vietnam). The views expressed here are those of the authors. They are not necessarily those of the Young Lives Project, the University of Oxford, DFID or any other funders, neither those of the Inter-American Development Bank or their boards of directors nor the Federal Reserve Bank of St Louis, The Federal Reserve System, or the Board of Governors.

†Federal Reserve Bank of Saint Louis

1 Introduction and Literature Review

The process of skill formation in early childhood is receiving increasing attention in both research and policy circles in developed and developing countries (UNESCO, 2008; Almond and Currie, 2010; Engle et al., 2007, Engle et al., 2011). Presumably, one of the main drivers of this interest is the notion that investment during early childhood can have significant large returns for human capital development (Heckman 1995).

The scientific basis for this claim rests upon different strands of literature. First, an emerging empirical literature finds positive average impacts for programs aimed at improving nutrition, psychosocial and educational inputs during early childhood on a range of short- and long-term outcomes (Nores and Barnett, 2010; Engle et al., 2011). Second, biological research has demonstrated that the human brain experiences most of its growth during early childhood, reaching 90 percent of its adult size in the first five years of life suggesting a critical short window to ensure adequate inputs (Shonkoff and Phillips, 2000). Third, evidence from developmental psychology suggests that measures of cognitive skills are set early on in life (Sameroff et al., 1993).

Though a general consensus may be arising regarding the critical nature of early life (particularly for cognitive skills [1]) and the case for early investments, significant uncertainties prevail regarding what type of interventions are more effective and under which contexts; particularly so for developing countries.

An additional motivation for research on a developing country comes from the well-known evidence on the extent of inappropriate early nutrition (Engle et al, 2011; Walker et al, 2011). Specifically, India is placed 1st in the world-wide rankings of malnutrition (Arnold et al, 2009) with 48% of the children under five exhibiting stunting, according to the Indian National Family Health Survey (NFHS-3). This is 20 times as high as would be expected in a healthy, well-nourished population (according to the international child growth standards).

Probably due to the lack of data on cognitive skills for most developing countries not much research exists neither on the link between nutrition and cognition nor on the socioeconomic status (SES hereafter) "gradients" in cognitive skills; when they arise, whom they affect and how they evolve as children age. [2] [3] Indeed, investments in nutrition have been shown to be one of the most important predictors of later cognitive development in developing countries, however this evidence comes mostly from only one influential long-term longitudinal study in Guatemala (Victora et al. (2008)). [4] Moreover, the most important shortcoming is the very small sample size (a highly nutritious drink was randomly assigned to two villages and a placebo drink to other two villages).

[1] See Heckman, Stixrud and Urzua (2006) on the importance of non-cognitive skills.

[2] Disparities in cognitive skills between socio-economic groups in developed countries is well-documented (Carneiro, Cunha, and Heckman 2003, Heckman 2005, Cunha and Heckman 2007).

[3] Fernald et al (2011) and Naudeau et al (2011) use single cross-sections of data from low-income countries, while related research from Ecuador (Paxson and Schady 2007, 2011) showed substantial differences in cognitive development at young ages between children of high and low socioeconomic status; and those differences seem to persist as children age. A caveat of the latter is that the analysis included only households in the poorest half of the nationwide distribution of wealth, was limited to relatively young families, and only covered rural areas; with two survey measurements only 2 years apart.

[4] There is also evidence of long-run effects of this intervention in Guatemala three-four decades later or on the next generation (e.g., Hoddinott et al. 2008, Behrman et al 2009, Maluccio et al 2009, Martorell et al 2010).

Divisions by SES in India, are defined according to whether a household belongs to a certain caste. Gang et al (2008) find that differences in educational attainment explain about 25 per cent of the poverty gap between both the Scheduled Caste and Schedule Tribe (the so called Lower Caste) and non-Scheduled-Hindu households (for further information on castes, see Appendix 1). For that reason, in the recent past, the government of India has introduced a range of policy interventions targeting social groups like Scheduled Castes (SCs) and Scheduled Tribes (STs). Some evidence shows that these interventions have been successful (Jenkins and Barr 2006), particularly for STs. The key question is whether these interventions happened early enough in the life cycle for them to be successful. And indeed, most of the existing studies on determinants of school participation and attainment in India today acknowledge socio-religious differences already for primary school-age children (Dreze and Kingdon 2001, Kingdon 2002, Dostie and Jayaraman 2006, Bhalotra and Zamora 2010, Asadullah, Kambhampati, and Lopez Boo 2009). In these studies, low attendance and completion of lower castes (LCs) is explained by a range of factors including rural infrastructure, conditions in the local village economy, the functioning and size of the relevant labour market, household credit-constraints, sex discrimination, and the poor quality and inadequate supply of schools. However, none investigates the extent and nature of caste gaps before age 7 nor the effects of early childhood conditions on the process of skill formation by caste at an early age.

The objective of this paper is therefore to fill an important gap in the literature by investigating the determinants of a child's development of cognitive skills over two phases of childhood: pre-school and school ages. To the best of our knowledge, this is the first study using panel data to assess: firstly, a value-added production function of cognitive skills, second; the causal relation between nutrition (height-for-age) and cognition (receptive vocabulary scores as measured by the PPVT) at early ages in India; and third, the existence and magnitude of socio-economic gradients in early childhood in cognition outcomes; and how they are mediated by nutrition and family income. For that, we will exploit the novel measures of anthropometrics, cognitive outcomes and other rich measures provided in the Young Lives longitudinal data (YL hereafter).[5] This data consists of two cohorts of children (the 'Younger' and the 'Older') surveyed over three rounds four years apart (Round 1 in 2002, Round 2 in 2006 and Round 3 in 2010). We use the Younger Cohort that spans data from 6 months to 8 years of age.

In this context, the closest empirical studies to ours are Sanchez (2009) and Outes et al (2010). Both document a causal relation between nutrition and cognition in Peru. The first paper uses maternal height and exposure to low temperatures during the first months of life as instrumental variables (IV) and finds that one standard deviation increase in HAZ at age 1 causes a 37% increase in PPVT at age 5 in Peruvian highlanders. [6] The second study does a within–sibling investigation combined with IV; and finds an average effect of around 20% using the same data as Sanchez (2009), but this time complemented with

[5] See www.younglives.org.uk

[6] Sanchez (2009) finds that using maternal height as an instrument produces a higher coefficient for the parameter of interest, implying that OLS might be downward biased. Using average minimum temperature as an IV for a sub-sample of communities located in the Highlands produces results in the same direction.

the siblings data. Helmers and Patnam (2011) also investigate early childhood development in India, but they do not focus on caste differentials nor nutrition. They rather demonstrate that parental investment has contemporaneously positive effects on skill levels for all age groups; and that child health at age one influences cognitive abilities at age five, but there is no quantification of the effects.

Furthermore, the methodology used here allows us to go beyond previous empirical studies. We follow the spirit of Todd and Wolpin (2007) and Cunha and Heckman (2007) who model cognitive skills as a function of the child's innate genetic ability and the cumulative effect of present and past home and school investments. This structural production function analysis makes considerable progress in sorting out the causal relationship between nutrition and cognitive outcomes. The challenge in estimating this relationship is that of other inputs being missing. Yet another problem is that unobserved child, parents or household-specific factors may affect both nutrition and cognitive outcomes, which may lead to a correlation even though no causation exists.

Unlike previous research, we attempt to take into account all of these issues in the following way: firstly, the contribution of all unobserved previous inputs and past endowments is resolved with our value-added specification; second; the endogeneity of nutritional status, as measured by height-for-age (HAZ hereafter) will be explicitly considered. We argue that since one important indicator of child malnutrition (HAZ at age one) and one of its major determinants (mother's height) are established well before the age at which the tests were given, the identification problem is ameliorated (although not completely solved). Thirdly, and following Aaronson (1998), we exploit in a novel way a sub-sample of children with information on the birthweight of their siblings in order to perform a robustness check to account for family fixed effects .

We find strong evidence that better cognitive outcomes are related to a better nutritional status in early childhood, evidence which survives numerous specifications and one robustness check in which we find that family fixed effects only explain 1 SD of the overall nutrition-cognition effect. Past test scores are an important determinant of current test scores, which supports the *self-productivity* effects present in Cunha and Heckman (2008). Moreover, we claim that our estimates are the more robust so far in the literature for the early ages given the novelty of the longitudinal data and the estimation of the unrestricted valued added model.

Using a specification that incorporates these features, we find that the predicted PPVT score gaps (by gender) would be reduced by 21 percent (0.09 points or 9 percent of a standard deviation) if we compensate LC children with low HAZ with the average level of HAZ observed for UC children. It is interesting to note that the gap between UC and LC for boys is closed by relatively less when equalizing HAZ (0.09 points; or one sixth of the gap). While for girls about one-third of the gap (very similar in absolute terms: 0.087 points) would be closed by leveling nutritional status. For the income equalization, figures are much lower. Our simulations show evidence of pro-male discrimination in both castes. However, as UC families seem to be discriminating more against their girls (gender gap in the PPVT at age 8 is 41 percent of one standard deviation for UC, while it is 13 percent for the LC), any policy directed to level the playing field in nutrition or other inputs will have to take into account this discriminatory parental behavior and direct more resources to girls in UC homes and to boys in LC homes.

What type of interventions can be effective in India then? It seems that nutritional interventions (acting on HAZ) as well as stimulation interventions (acting on the PPVT at age 5) could be promising given the magnitudes of the effects found in this paper. However, differential policies towards LC children are needed as well as policies that acknowledge differential treatment of boys vis a vis girls.

This paper proceeds as follows. The second section describes the methodology for modeling the production function and considers its empirical application and challenges. The third section gives details on the data and variables used, while the fourth section presents descriptive statistics. The fifth section shows estimates of the cognitive skills production function, and we use the estimated production function to evaluate caste disparities in test scores. The last section concludes.

2 Methodology

2.1 Economic Framework

As a guide to our empirical analysis, we set up a model of household production and time allocation. The model follows the spirit of Todd and Wolpin (2003) and Cunha and Heckman (2007), but we focus on the choice of parental investment to produce child's human capital as well as child's nutrition. We assume that parents maximize utility derived from consumption (c), leisure (l) and their child's cognitive skills (θ our measure of the child's human capital):

$$max_{h,i,ih} U(c, l, \theta) \tag{1}$$

subject to: a production function of cognitive skills (θ_t) that depends on the stock of cognitive skills at the beginning of the period (θ_{t-1}),parental investment, i.e. family inputs, (i_t), the child's nutrition status at the beginning of the period (H_t), family's characteristics (X_t), and the child's ability (μ_t):

$$\theta_t = f(\theta_{t-1}, i_t, H_t, X_t\ , \mu_t), \tag{2}$$

a production function for next period nutrition status

$$H_{t+1} = g(H_t, ih_t, X_t, , \mu_t), \tag{3}$$

where ih_t, are family nutrition inputs. A time constraint, where h are the total number of hours worked and l is leisure

$$l_t = 1 - h_t, \tag{4}$$

and a budget constraint.

$$wh_t = c_t + p_i i_t + p_h ih_t. \tag{5}$$

Given wages and prices, parents choose how much to work in the market, how much to invest in their child's production of cognitive skills and their child's nutrition. We assume that investment in nutrition only affect the child's nutrition status at the end of the period.

2.2 Empirical Strategy

Our goal is to estimate (2), asses caste differentials and identify the causal relation between nutrition and cognition in early childhood in India. In doing so, we start by taking a step backward and estimate a contemporaneous version of the production function of cognitive skills. We will use this specification as a benchmark and focus, when data permitting, in a more rich value-added specification.

2.2.1 The contemporaneous specification

The contemporaneous specification relates cognitive skills with family inputs in period t (i_{it}), nutrition at the beginning of period t (H_{it}), other child, family and community factors in period t (X_{it}), and the child's ability (μ_{it}). Empirically, we also allow for an additive error, which results in:

$$\theta_{it} = \alpha i_{it} + \gamma H_{it} + \delta X_{it} + \beta \, \mu_{it} + \varepsilon_{it} \tag{6}$$

The problems of the contemporaneous specification are widely well known. As Todd and Wolpin (2003) point out, this specification would be justified if either only contemporaneous inputs matter for the production of current cognitive skills; or inputs are unchanging over time, so that current inputs capture the entire history of inputs. Additionally, we need to assume that contemporaneous inputs are unrelated to unobserved ability and the additive error (which accounts for omitted – i.e. unobserved– inputs, and measurement error). This last assumption is inconsistent with the model we just specified, or any other economic model of optimizing behavior in the spirit of Becker and Tomes (1986), where parents care about their child's human capital.

Even though, the contemporaneous specification might be informative when very limited data is available; more flexible specifications allows us to estimate the production function under milder assumptions. We now describe a more preferable value-added specification.

2.2.2 The value-added specification

In its most common form, the value-added production function of cognitive skills adds to the contemporaneous specification a relation of current skills to a lagged cognitive skills measure. It can be written as:

$$\theta_{it} = \alpha i_{it} + \gamma H_{it} + \delta X_{it} + \varsigma \theta_{it-1} + \beta \, \mu_{it} + \varepsilon_{it} \tag{7}$$

Note that if lagged cognitive skills is a sufficient statistic for input histories and unobserved ability, estimating (7) would give us consistent estimates of the production function of cognitive skills. However, we need to assume that the rate of decline (if $\varsigma < 1$) must be the same for all the inputs, and the impact of ability must be geometrically declining at the same rate as inputs.

2.2.3 Endogeneity of Child Nutritional Status

A common problem in the production function approach to studying child outcomes relates to endogeneity of particular regressors, such as nutrition or home inputs. Particularly, since parental taste for child quality and a child's genetic ability are unobserved, Ordinary Least Squares (OLS) estimations of the nutrition-cognition nexus, as well as the parental investment-cognition nexus are likely to be biased.

If child's ability is malleable, μ_{it} represents factors such as innate ability and also motivation and child's effort, which are out of parents' control but are influenced by home environment as well as by genetics (Rosenzweig and Wolpin 1988). According to our model the level of investments in each period will be responsive to the child's ability. This introduces an endogeneity bias that can be either positive or negative. Consider the case of family inputs: on the one hand, it could happen that parents observing that their child is of high ability expect a higher return to their investment and so invest more in him. In this case it would exist a positive correlation between children's ability and the amount of inputs. On the other it could be that parents who observe that their child is of low ability try to compensate this by investing more inputs. This type of behavior would induce a negative sample correlation between child's ability to learn and the level of inputs. Note that for inputs in the production of nutrition, it is only the component of μ_{it} – that remains fixed over time– the one that introduces the bias.

Another challenge in the estimates arises from the fact that parental preferences might affect the level of investment. For example, parents with a strong preference for child investments will provide their children with inputs that improve both child nutritional status and cognitive skills. This could thus lead to an *omitted variable bias* and upwardly bias the estimates of the coefficient on nutrition.

The rich set of controls included for family and non-family characteristics will deal with factors that are common across families. Yet parents could favour one sibling based on observable features. To take this into account we include birth order and gender (i.e., a child can be favoured because he is the first born or because of his gender).

To solve the problems in the coefficient of nutritional status we assume, as in the model we described, that nutritional status at the end of the previous period is the one that affects the child's performance in the cognitive test. Because parental preferences might contaminate the levels of investments in both nutrition and cognitive skills at ages 5 and 8, we instrument nutritional status at age 5 with nutritional status at age 1. Although we cannot reject that height-for-age at age 1 could be related with unobserved factors related to cognitive outcomes, we expect the relation between the first two to be weaker.

On a similar basis, we instrument with mothers' height (Thomas, Strauss, and Henriques 1990). Thomas, Strauss and Henriques (1990) for instance find in their Brazilian study that parental height has a large positive effect on child height and, more importantly, on child survival rates even after controlling for many observable characteristics. The result on child survival suggests that the effect of mother's height on child height-for-age involves more than just inherited stature among healthy individuals, supporting the claim that mother's height reflects, in part, inherited disease susceptibility.

In addition, for a subsample of children in the third round, the birthweight of one of their siblings was recorded. Birth-weight is a measure of innate endowments and it should not be contaminated by parental investments on the basis of revealed innate ability but it should be correlated, if any, with parental preferences in a similar way as height for age at age 1. As a robustness check, we add the sibling's average birthweight as a regressor. If unobserved family characteristics associated with children's nutrition are sibling invariant, variation within families can be used to identify children's nutritional status influence on children's performance on test scores (see Aaronson, 1998). Even though, this specification would be preferred, we only have this information for 20% of our total sample.

3 Data

3.1 General description

We use data from three rounds of the Indian survey of the Young Lives (YL) project.[7] In Round 1, in 2002, 2,000 children aged 6 to 18 months from the so-called 'Younger Cohort' were surveyed. Following up, Round 2 tracked the same children and surveyed them in 2006 at age 5 and Round 3 surveyed them in 2010 at age 8. The attrition rate was only 0.9 per cent, which is very low for a study of this size. In terms of the representativeness, despite a few biases (see Kumra (2008) in a note comparing the YL survey to DHS), it is shown that the YL sample in Andhra Pradesh covers the diversity of children in the country.

The stratified cluster sample is both region and caste representative; and the estimation used in this study incorporates the YL survey design by using regions as the stratification variable and the sentinel sites as the clustering variable.

3.2 Description of key variables

Table 1 shows key variables of interest. A brief introduction to the main variables is given below.

3.2.1 Cognitive skills

We focus on one test that measures different aspects of cognitive abilities: the Peabody Picture Vocabulary Test (PPVT hereafter) which is a a test of vocabulary recognition that has been widely used as a general measure of cognitive development. It measures vocabulary knowledge. A concern might be related to the effect of the language in which the tests were provided. Actually, the questionnaires and the manuals for the field supervisors were translated into Telugu.[8] But still, it is hard to assume that people responding to a vocabulary

[7] Data is actually from Andhra Pradesh (AP hereafter), which is divided into 23 administrative districts, which are each subdivided into a number of mandals or sentinel sites, dependent upon the size of the district. There are 1,125 mandals and around 27,000 villages in AP. Generally, there are between 20 and 40 villages in a mandal, although in tribal mandals there can be as many as 200 villages. Villages are normally composed of a main village site with a small number (two to five) of associated hamlets. Tribal villages tend to have a large number of dispersed hamlets. AP has three distinct agro-climatic regions: Coastal Andhra, Rayalaseema and Telangana. The sampling scheme adopted for YL was designed to identify interregional variations with the following priorities: (1) a uniform distribution of sample districts across the three regions to ensure full representation; (2) the selection of one poor and one non-poor district from each region; and (3) when selecting poor districts and mandals, consideration was given to issues which might impact upon childhood poverty, including the presence or non-presence of the AP District Poverty Initiative Programme (APDPIP).

[8] About 85 per cent of the Andhra Pradesh population identifies Telegu as its mother tongue (the second most commonly spoken language in India), another 7.5 per cent speak Urdu, and about 3 per cent speak Hindi. In the YL data, only 4.2 per cent of children speak a minority language.

Table 1: Structure of key variables in the Younger Cohort of YL data

	Cognitive Skills Test	Nutritional status	Inputs
Age 1 Round 1	-	HAZ	NA
Age 5 Round 2	PPVT	HAZ	◦ Family income ◦ Whether child attended pre-school ◦ Whether pre-school public/private/religious or NGO
Age 8 Round 3	PPVT	HAZ	◦ Family income ◦ Whether child attends formal schooling ◦ Whether school public/private/religious or NGO

Note 1: HAZ stands for height-for-age z-score and is based on 2006 WHO standards.

test in different languages could be compared. For this reason, the analysis for the PPVT is restricted to the children who answered the PPVT test in Telugu (90 per cent of the 5 year-olds in the Younger Cohort who we will follow in Round 3). [9]

3.2.2 Nutritional status

The rationale for the use of HAZ is that deficit in the height-for-age measure corresponds to the inability to reach the genetic potential in terms of height. This is viewed as a longer term measure of deprivation than weight-for-height, which is more sensitive to short-term or seasonal variations in food availability. Height is also said to have a strong relationship with mental function and mortality.

3.2.3 Control variables

Control variables refer to the caregiver, father and home characteristics as well as geographical dummies. Caregiver characteristics are age, caste and education. Father's education and home characteristics such as the wealth index (Filmer and Pritchett 1999) and household size, are also included. The wealth index has three components: housing quality, consumer durables and services. Income is included in an alternative version of the regressions. Geographical dummies included are: Coastal Andhra and Rayalaseema (with Telangana being the base category) and whether the household is located in a urban or rural area.

4 Descriptive statistics

In Figure 1 we investigate age patterns in the SES gradients in child development in the top panel; while in the two bottom panels we investigate these by gender (UC in the left and LC in the right). We use seven–month moving averages of the PPVT at age 5 and age 8 and we split the sample in two: children in the upper caste and those in the lower caste. These internally-standardized scores suggest that; first, the bulk of the difference between castes is already apparent by age 5 and stands around 0.20 of one SD; second, these 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 Round 3 and are larger in magnitude (0.4 to 0.6SD) –particularly for UC– suggesting a widening gap; third; there are clear differences by gender that depend on the caste and the age: UC females fare much better than their males counterparts at age 5, but they do fare worse at age 8; while lower caste-females perform worse than lower caste males, regardless of the age. Fourthly; the gender gap for LC is smaller than the gender gap for UC.

In Figure 2 we do the same analysis but for height for age z-scores. These differ from the internally standardized PPVT z-scores, as they are externally-standardized scores, that is, YL use the WHO stands with respect to a reference group of healthy children for the standardization. [10] The deficits are very

[9] Throughout the paper we obtained consistent results using this sample as compared to (i) the full sample or (ii) the full sample with inclusion of a statistical control for whether the respondent spoke the language of the test since birth.

[10] see www.who.int/growthref/

important: by the time children are 5, lower caste children are 2 SD behind the reference population (which places almost all of them in the category of "stunted") [11], while the upper caste are 1 to 1.5 SD behind a healthy population. The nutritional status of upper SES children seems to be worsening over time in Round 2, while in round 3, these gaps remain significant but there is neither worsening nor catch up.

When we look at these differences by caste and by gender in both rounds, we see not significant differences by gender for the UC. This is not at all the case for LC, for which girls appear better nourished than boys at almost every age (t-tests available upon request). This finding is probably related to a well-known fact in the demographic literature on male-to-female ratios in child mortality over the first year of life.[12] However, it is surprising that we only find this for the LC.

The means and SDs of all relevant variables are presented in Table 2. The first column presents results for the full sample, the second column for LCs (composed of SCs and STs), the third column for Backward Classes (or BC) [13] and the fourth column for UCs. The last two columns show p-values for the difference of means between LCs and BCs, and UCs and LCs, respectively (the latter being the most relevant comparison for this paper).

Overall, and based on the analysis of Table 2 and the existing literature, we find that Andhra Pradesh has achieved progress on many indicators since the mid-1990s. It is worth noting that in AP around one out of three children under five are stunted, which is less than the national average of almost 50 %. However, even though LCs and BCs have become wealthier and increasingly urban, significant differences remain based on sector (rural versus urban), caste and region.

On average, in both waves, we have approximately 33 per cent of LC children (20 percent are SC and 13 percent are ST), 52 per cent are BC and 15 per cent are UC. The table shows that in the PPVT, UCs do significantly better at age 5 (17 per cent higher PPVT scores) and age 8 (24 per cent higher PPVT scores) than LCs. The level of stuntedness increased over time for both the UCs and the LCs up to 37 per cent at age 5 (and 32 at age 8), but remained about 20 percentage points lower for UCs (at 18 per cent when they were 1 year-old, and 28 and 22 per cent four and eight years later, respectively).

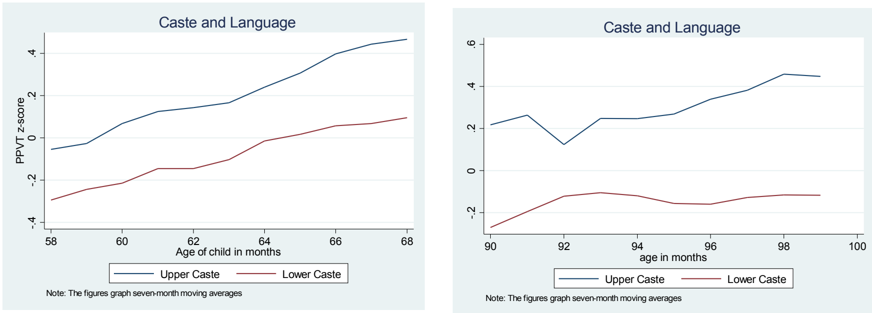
There is also a remarkable increase in urbanisation rates among LCs, who went from about 10 per cent to 35 per cent. Households are larger in BCs; while LCs are poorer than BCs, who themselves are poorer than UCs (as per the wealth index and also the income aggregate). Over time, all castes and cohorts are becoming richer, but as inequality did not decrease substantially, it seems that UCs end up benefiting more from growth.

[11] Children with a HAZ score below 2SD of the mean of the reference group are defined as stunted.

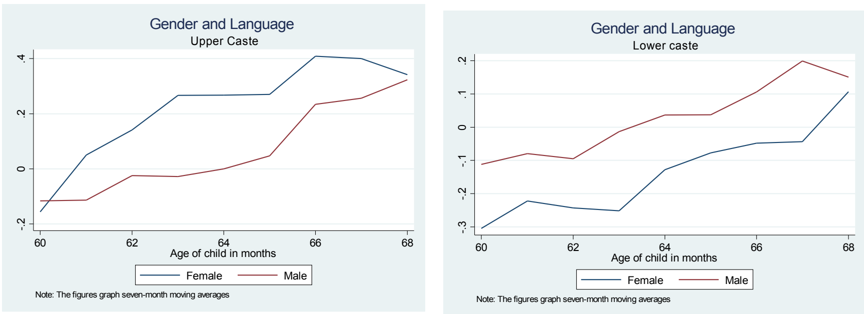
[12] One study using DHS's surveys finds that girls are almost 10 percentage points less likely to die in the first year of life than boys (Baird et al 2010). Also, the World Health Organization (2006) estimates that the male-to-female ratio in neonatal and early neonatal mortality in developing countries is 1.3.

[13] There is no consensus in the literature on whether to explicitly treat this category as a separate social group. Jenkins and Barr (2006) and Dreze and Kingdon (2001) consider SC and ST as separate from Backward Castes on the grounds that completion rates are much lower than for other groups. We have therefore separated out this group and explicitly controlled for BC membership in the results section. We have also further split the lower caste group between SC and ST.

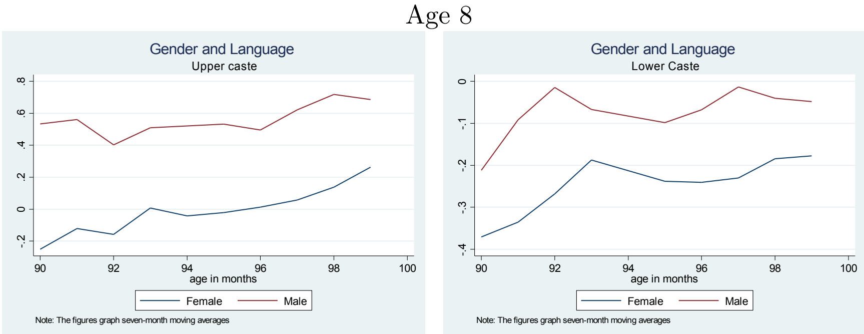
Figure 1: Panel data analysis of PPVT age patterns, age 5 and age 8



Age 5

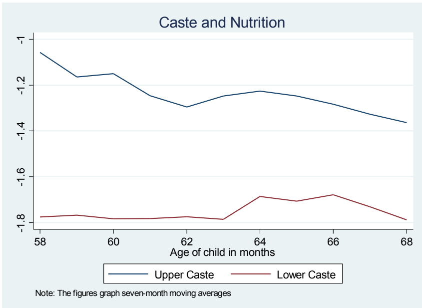


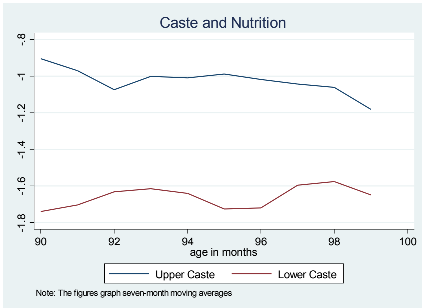
Age 8



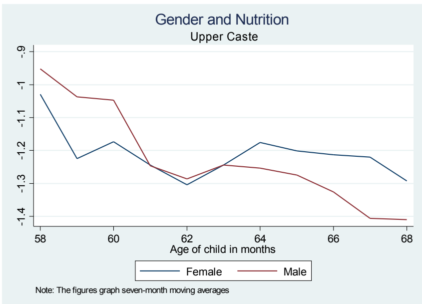
Source: Author's calculations based on YL data.

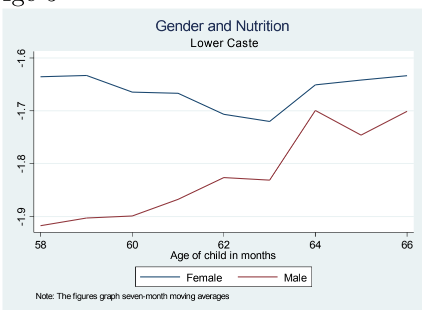
Figure 2: Panel data analysis of HAZ age patterns, age 5 and age 8





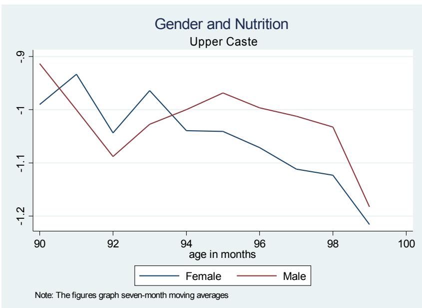
Age 5

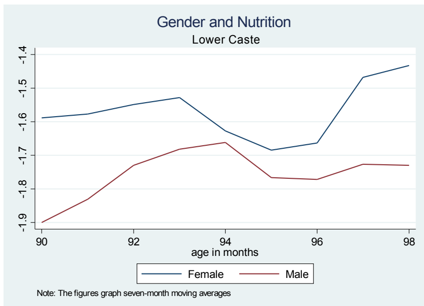




Age

8





Source: Author's calculations based on YL data.

Table 2: Means and t-tests of main variables by caste

	Total	LC	BC	UC	t-test (p-value) LC-BC	LC-UC
Child						
Cognitive Scores						
R2 PPVT	39.23	41.00	35.55	47.86	0.00	0.01
R2 Peabody PVT (z-sc)	-0.09	-0.03	-0.20	0.18	0.00	0.01
R3 PPVT	57.72	55.59	55.83	68.68	0.89	0.00
R3 Peabody PVT (z-sc)	-0.02	-0.09	-0.08	0.33	0.89	0.00
Individual characteristics						
R1 Child Age (m)	11.88	12.04	11.71	12.12	0.11	0.78
R2 Child Age (m)	64.32	64.23	64.30	64.54	0.76	0.32
R3 Child Age (m)	95.54	95.47	95.44	96.01	0.89	0.09
R1 Last-born	0.45	0.50	0.42	0.41	0.01	0.04
SC	0.20	0.62	0.00	0.00	0.00	0.00
ST	0.13	0.38	0.00	0.00	0.00	0.00
BC	0.52	0.00	1.00	0.00		
UC	0.15	0.00	0.00	1.00		
R3 Coastal Andhra	0.38	0.37	0.37	0.43	0.82	0.19
R3 Rayalaseema	0.30	0.28	0.28	0.43	0.92	0.00
R3 Telangana	0.32	0.35	0.36	0.15	0.75	0.00
R2 Pre-school	0.86	0.82	0.88	0.85	0.00	0.33
R2 Private presc	0.21	0.10	0.21	0.42	0.00	0.00
Nutritional status						
Birth-weight (gr).	2724	2678	2722	2784	0.41	0.11
R1 Stunted	0.29	0.35	0.29	0.18	0.01	0.00
R1 HAZ	-1.37	-1.50	-1.39	-1.03	0.18	0.00
R2 Stunted	0.37	0.39	0.39	0.28	0.97	0.00
R2 HAZ	-1.68	-1.72	-1.73	-1.40	0.79	0.00
R3 Stunted	0.32	0.37	0.31	0.23	0.01	0.00
R3 HAZ	-1.51	-1.61	-1.52	-1.24	0.20	0.00
R3 Younger Sib. Birth-wgt.	2798	2780	2862	2650	0.24	0.07
R3 Older Sib. Birth-wgt.	2783	2760	2787	2800	0.77	0.70
R2 Mother Height	151.18	149.84	151.88	151.66	0.00	0.00
Household						
R1 Urban	0.18	0.09	0.18	0.38	0.00	0.00
R2 Urban	0.32	0.38	0.25	0.43	0.00	0.15
R3 Urban	0.31	0.35	0.25	0.45	0.00	0.02
R1 hhsize	5.38	5.15	5.54	5.36	0.00	0.23
R2 hhsize	5.46	5.26	5.60	5.43	0.01	0.28
R3 hhsize	5.37	5.20	5.53	5.23	0.01	0.85
R1 Wealth Index	0.31	0.23	0.31	0.45	0.00	0.00
R2 Wealth Index	0.33	0.28	0.34	0.44	0.00	0.00
R3 Wealth Index	0.49	0.42	0.51	0.60	0.00	0.00
CG and Dad						
R1 CG Age	23.68	23.71	23.54	24.07	0.54	0.33
R1 CG Edu	2.64	1.93	2.33	5.22	0.07	0.00
R2 CG Edu	2.88	2.08	2.56	5.68	0.03	0.00
R1 DAD Edu	5.20	4.93	5.13	6.00	0.05	0.00
R2 DAD Edu	4.75	4.09	4.48	7.07	0.15	0.00

Source: Young Lives-India, Prefix R1, R2, R3 indicated the value comes from Round 1, 2 or 3. Means are taken on children observed in the three rounds with the cognitive skills measure not missing (N= 1456). LCs are: Scheduled Caste (SC) and Scheduled Tribes (ST). Other Backward Classes (BC) include Muslims, while UCs are those classified in the YL data as Other Castes. HAZ=height-for-age z-score and CG=caregiver. Child nutrition variables have number of obs slightly less than 1456 and birthweight variable only has 574 obs.

The UC mothers have around two and three times more years of schooling than BCs and LCs, respectively, averaging a total of about 6 years of education completed in round 2 (the latest available for this variable). For fathers, the differences, though significant, are not so ample: UC fathers have about one (in round 1) to three (in round 2) years more education than BC fathers, and two to four years more than LC. It sees as the UC father did complete some studies over the time from the first to the second survey.

Another important predictor of children's success is parental nutrition, and a good proxy at hand in YL data is caregiver's height: as expected, LC mothers are, on average, 2 cm. shorter and 7 kilos lighter than UC mothers. [14] [15]

It is also shown that that more UCs than LC go to pre-school, and moreover go to private and NGO-run pre-schools, which are of better quality.

Overall, disparities between castes are particularly important in cognitive tests, nutrition outcomes, wealth, caregivers' and fathers' level of education and some parental inputs. In general, we find a significant advantage amongst UCs in inputs and background, suggesting that these can be one source of the disparities found in cognitive and nutritional outcomes.

5 Results

5.1 Contemporaneous specification

[14] 98.5 per cent of the caregivers are the biological mothers in this sample.

[15] In terms of inputs not considered in this paper because they refer to round 1 it is found that that more UCs than LCs: (i) were born in a hospital or with a medically trained person, (ii) had been given iron folate tablets/syrup during the antenatal visits, (iii) had better level of antenatal care (LCs having a low-medium level of care), and (iv) had timely immunisation.

Table 3: Production function of cognition, contemporanous equation

VARIABLES	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)
Lag HAZ				0.1141***	0.2174***
				(0.006)	(0.017)
Scheduled caste	0.0592**	0.1202***	0.1282***	0.1067***	0.1068***
	(0.021)	(0.032)	(0.031)	(0.036)	(0.038)
Scheduled tribe	-0.1131***	-0.1031***	-0.0899***	-0.1023**	-0.0975**
	(0.017)	(0.029)	(0.024)	(0.040)	(0.038)
Upper castes	0.4192***	0.3222***	0.2459***	-0.1051**	-0.1221***
	(0.032)	(0.069)	(0.066)	(0.049)	(0.045)
Scheduled caste* Male dummy		-0.1128**	-0.1142***	-0.1507***	-0.1572***
		(0.045)	(0.038)	(0.042)	(0.045)
Scheduled tribe* Male dummy		-0.0363	0.0261	-0.0039	-0.0279
		(0.051)	(0.037)	(0.044)	(0.046)
Upper castes* Male dummy		0.1471*	0.1722**	0.1725***	0.1577***
		(0.073)	(0.080)	(0.048)	(0.043)
Male dummy		0.2461***	0.2324***	0.2737***	0.2912***
		(0.019)	(0.019)	(0.013)	(0.013)
income			0.0040***	0.0021***	0.0019***
			(0.000)	(0.000)	(0.000)
Age of child (months)				0.0271***	0.0282***
				(0.005)	(0.005)
First born dummy				-0.1673	-0.1481
				(0.488)	(0.428)
Last born dummy				-0.1129***	-0.1046***
				(0.016)	(0.013)
Urban dummy				0.2328***	0.2365***
				(0.045)	(0.042)
Household size				-0.0286***	-0.0264***
				(0.004)	(0.003)
Caregiver's education level				0.0420***	0.0415***
				(0.002)	(0.002)
Constant	-0.0853***	-0.2104***	-0.4089***	-2.3465***	-1.7061***
	(0.020)	(0.022)	(0.027)	(0.400)	(0.527)
Observations	1,457	1,457	1,457	1,409	1,396
R-squared	0.027	0.043	0.084	0.253	0.243

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: When taking income out of the regression, results remain unaltered. Also, see Appendix 6 for results that include wealth instead of income and show no change in the main coefficients

Before presenting our estimates of the effect of nutrition on cognition, we present some cross sectional regressions showing, first, case differentials by gender and second, a significant effect of nutrition on cognition. Table 3 presents results for a cross-section regression of PPVT at age 8 on caste dummies, household and child characteristics. These regressions present initial evidence that nutrition status matters for the production of cognition; and that the belonging of a child to a caste changes his chances of having more or less vocabulary according to his gender. Column 1 of the OLS specification shows that (without any controls) being an UC children implies a 42 per cent of 1SD higher PPVT at age 8 than the BC (the base category), while column 2 shows that the interaction with gender matters for SC and UC, with UC boys outperforming UC girls and SCs showing just the opposite trend. If we compare column (2) and column (3), we see these are caste-specific effects regardless of the inclusion of income (whose coefficient is positive and significative albeit small). Column (4) is the OLS specification with the full set of controls, including past level of HAZ. We see that boys outperform girls; older children perform significantly better (a characteristic of the PPVT test); higher order children perform worse and children in households with more integrants perform worse in the PPVT score. Leaving in urban areas promotes the production of cognition. More educated parents and caregivers, as well as richer (or wealthier, see footnote of Table 3 families foster cognition. Caste dummies do not change much and sign and magnitude are as expected, at the exception of the UC dummy that switches sign and magnitude when we add child and household characteristics. The latter is probably due to the fact that, in a contemporaneous model setting, most caste differentials are captured by observables such as mothers education, location and income or wealth. This is consistent with the literature showing that once the children are placed in “more favourable” circumstances (i.e., when parents, especially mothers are literate and when infrastructure facilities are better), inter-caste differences in enrolment rates in school become insignificant (Sachar 2006).

In terms of the nutrition coefficient, the effect doubles from 11 per cent of 1SD to 22 per cent of 1SD when nutrition status at age 5 is instrumented with nutrition status at age 1 and mother’s height in column (5), and this difference is statistically significant. This result is very close (although in a different age) to Outes-Leon et al.(2010) who exploit within-siblings variation in height-for-age measures in Peru to explore its impact over cognitive skills of 5-year old children. [16] In the first stage, nutrition status at age 1 has a positive and significant effect on nutrition status at age 5. The R-squared of the first stage is 0.35 and the F 35.22.

5.2 Value Added

Tables 4 and 5 summarize our value added results. Table 4 assumes perfect

[16] The authors estimates results using OLS models, controlling for a range of important covariates, as well as instrumental variables techniques exploiting changes in food prices during 2006 to 2008 as well as household shocks prior to the outcome measurement. The IV results indicate that a one standard deviation increase in the height-for-age standardized measure is associated with an increase in about 0.17 to 0.21 standard deviations of the PPVT measure. These results are close to those obtained from OLS estimates, controlling for covariates. Though the robustness of the results is encouraging, it is important to recognize that both estimation exercises require stringent assumptions to identify causal effects

Table 4: Production function of cognition, restricted value-added equation

VARIABLES	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
Lag HAZ				0.0101	0.0042
				(0.011)	(0.009)
Change in HAZ					0.0271***
					(0.008)
Scheduled caste	0.1516***	0.2827***	0.2858***	0.1771***	0.1812***
	(0.033)	(0.048)	(0.047)	(0.050)	(0.052)
Scheduled tribe	-0.7004***	-0.6296***	-0.6246***	-0.4620***	-0.4446***
	(0.026)	(0.036)	(0.044)	(0.072)	(0.074)
Upper castes	0.0387	-0.0425	-0.0718	-0.1096*	-0.1109**
	(0.053)	(0.066)	(0.062)	(0.053)	(0.053)
Scheduled caste* Male dummy		-0.2554***	-0.2560***	-0.2540***	-0.2617***
		(0.044)	(0.043)	(0.039)	(0.041)
Scheduled tribe* Male dummy		-0.1509***	-0.1268***	-0.1749***	-0.2075***
		(0.028)	(0.034)	(0.030)	(0.031)
Upper castes* Male dummy		0.1138	0.1235	0.0991	0.0922
		(0.120)	(0.118)	(0.101)	(0.098)
Male dummy		0.2990***	0.2937***	0.3069***	0.3113***
		(0.036)	(0.036)	(0.037)	(0.036)
income			0.0015***	0.0010***	0.0010***
			(0.000)	(0.000)	(0.000)
Age of child (months)				0.0012	-0.0012
				(0.004)	(0.004)
First born dummy				-0.4370***	-0.4163***
				(0.066)	(0.076)
Last born dummy				-0.0801***	-0.0737***
				(0.020)	(0.019)
Urban dummy				-0.0748	-0.0725
				(0.093)	(0.094)
Household size				-0.0092	-0.0090
				(0.009)	(0.009)
Caregiver's education level				0.0102***	0.0109***
				(0.003)	(0.004)
Constant	0.1144***	-0.0377	-0.1141***	0.1721	0.3566
	(0.022)	(0.023)	(0.026)	(0.286)	(0.254)
Observations	1,457	1,457	1,457	1,437	1,423
R-squared	0.052	0.067	0.072	0.122	0.121

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: When taking income out of the regression, results remain unaltered. Also, see Appendix 6 for results that include wealth instead of income and show no change in the main coefficients

persistence. This restricted value added model is equivalent to estimating a first difference in PPVT. Under the assumption of perfect persistence the outcome variables changes and so does the interpretation of the coefficients in the regression. The outcome variable is now how much further away, or closer, a child is from the sample mean of test scores in a certain round relative to the previous one. We can think of it as of how a child changes their rank in the sample. Therefore the coefficients of the regression measure the contribution of a certain variable in changing the student's rank. Because PPVT scores are bounded by zero, the standardized test scores are also bounded. Hence we should expect that students at the top end of the score distribution do not change their rank much. The power of the regressors reduces significantly as they only explain half of the variation in the change in PPVT compared to explaining the level of PPVT at age 8 (R-sq is 0.12 now as opposed to 0.25 in the contemporaneous specification). Including the change in nutritional status (last column) do not improve the power of the regression. While past HAZ turns not significant, the effect of a change in nutrition from age 5 to age 8 is positive but reduces to 0.03. In this specification, the coefficients on the caste dummies tells whether belonging to a certain caste or tribe make children more likely to move up (or down if negative) the rank of test scores relative to the base category (the BC). The fact that students from the UC do not show an advantage moving up the rank is hence not surprising. UC perform significantly better than BC students and so have less room for improvement. As Andrabi et al (2011) we refer to ς as the parameter that links PPVT across periods, as persistence. The canonical restricted value-added or gain-score model assumes that $\varsigma = 1$ (for examples, see Eric A. Hanushek 2003). When $\varsigma < 1$, the PPVT score exhibits conditional mean reversion. Estimates of the 'treatment' effect, in our case γ (the coefficient on HAZ), that assume $\varsigma = 1$ will be biased if the baseline achievement of the 'treatment' and control groups differs and persistence is imperfect. This has led many researchers to advocate leaving a lagged measure of cognitive skills on the right-hand side. Both the estimate of persistence ς and the treatment effect γ may remain biased when estimated by standard methods. To address these concerns, we jointly estimate ς and γ in an unrestricted value-added specification in Table 5.

There are several findings. First, the estimates reject perfect persistence, as the coefficient on lag PPVT is statistically different from one across all specifications. Also, the learning persistence is low; only one-fifth (IV specification) to one-third (OLS in first column) of PPVT persists between age 5 and 9. That is, ς is between 0.21 and 0.35 rather than closer to 1. This means that gains from nutrition may be much smaller than those obtained by assuming that ς is close to 1.

Secondly, the OLS valued-added coefficient on past nutrition status is statistically different and smaller from the OLS in column (4) from Table 3 (the contemporaneous specification). We expected this result as in a contemporaneous specification past nutrition might well have been capturing an unobservable such as ability. However, when nutrition status at age 5 is instrumented with nutrition status at age 1, the effect increases by around 4 standard deviations (from 0.09 in column 4 to 0.16 in column 5); this shows that there there is still some heterogeneity not completely captured by the lagged test score, and that instrumenting is important and necessary. Even though the change in this

Table 5: Production function of cognition, value-added equation

VARIABLES	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)
Lag HAZ				0.0924***	0.1608***
				(0.006)	(0.018)
Lag PPVT	0.3461***	0.3477***	0.3237***	0.2216***	0.2141***
	(0.009)	(0.008)	(0.008)	(0.011)	(0.013)
Scheduled caste	0.0912***	0.1767***	0.1792***	0.1275***	0.1288***
	(0.024)	(0.034)	(0.033)	(0.036)	(0.038)
Scheduled tribe	-0.3164***	-0.2862***	-0.2630***	-0.1713***	-0.1622***
	(0.015)	(0.016)	(0.024)	(0.043)	(0.041)
Upper castes	0.2875***	0.1954***	0.1431**	-0.1075**	-0.1189***
	(0.034)	(0.063)	(0.060)	(0.043)	(0.041)
Scheduled caste* Male dummy		-0.1624***	-0.1601***	-0.1793***	-0.1849***
		(0.040)	(0.036)	(0.039)	(0.041)
Scheduled tribe* Male dummy		-0.0761**	-0.0234	-0.0571	-0.0758**
		(0.032)	(0.026)	(0.037)	(0.037)
Upper castes* Male dummy		0.1355	0.1564*	0.1561***	0.1447***
		(0.084)	(0.085)	(0.050)	(0.043)
Male dummy		0.2645***	0.2523***	0.2831***	0.2952***
		(0.018)	(0.019)	(0.013)	(0.011)
income			0.0032***	0.0018***	0.0017***
			(0.000)	(0.000)	(0.000)
Age of child (months)				0.0212***	0.0220***
				(0.005)	(0.004)
First born dummy				-0.2277	-0.2130
				(0.393)	(0.355)
Last born dummy				-0.1088***	-0.1020***
				(0.016)	(0.013)
Urban dummy				0.1634***	0.1688***
				(0.055)	(0.051)
Household size				-0.0240***	-0.0227***
				(0.005)	(0.005)
Caregiver's education level				0.0347***	0.0346***
				(0.003)	(0.003)
Constant	-0.0162	-0.1504***	-0.3135***	-1.7340***	-1.3328***
	(0.019)	(0.017)	(0.019)	(0.372)	(0.495)
Observations	1,457	1,457	1,457	1,409	1,396
R-squared	0.127	0.145	0.170	0.288	0.283

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: When taking income out of the regression, results remain unaltered. Also, see Appendix 6 for results that include wealth instead of income and show no change in the main coefficients

coefficient is in the same direction as in the cross-sectional specification, as expected, the value added coefficient is significantly lower. Therefore, whereas the contemporaneous equation in Table 3 under-estimates the coefficient on HAZ with the IV, and the restricted value-added model in Table 4 suggests that the change in nutrition has only an effect of 0.03, our panel estimates suggest a large and significant effect of 0.16 (IV) per cent of one standard deviation increase in HAZ.

Third, and related to the above point, the nutrition effect is highly sensitive to the persistence parameter. Since nutrition is an input that could be thought as being continually applied, leading to a large baseline gap in cognitive skills, this is expected. We find that incorrectly assuming perfect persistence significantly understates and occasionally yields the wrong sign for nutritions impact on the PPVT.

It is worth commenting caste dummies in relation to the contemporanous specification. Column (1) shows that the introduction of lag PPVT decreases the positive coefficient of the UC from 0.42 to 0.29, and triples the (negative) coefficient of ST from -0.11 to -0.32, therefore increasing the gap between castes (the SC coefficient stays quite similar). The upward bias of the UC coefficient in the contemporaneous specification makes sense as probably this coefficient was capturing unobservables in endowments and inputs (positively correlated with the error term) that are now in our ς coefficient. On the other hand, now that we have a control for past history up to age 5, the ST children show larger disadvantages. This disadvantage decreases substantially (by almost half) when we include the full set of controls in column (4) and in the IV model in column (5). The latter is a worrying finding which means that even when controlling for a wide set of controls, STs are lagging behind. And indeed, recent analysis of data on the distribution of public facilities indicates that areas with SC concentrations gain in access to several facilities (e.g., schools, health centers), while those with STs and Muslims remain disadvantaged (Banerjee and Somanathan 2007). The result from Table 3 in which the UC dummy switches sign and magnitude when we add child and household characteristics remains in the value added specification (column 4), as well as the . Moreover, if we compare column (2) and column (3), we see these are caste-specific effects regardless of the inclusion of income at the exception of the interaction of UC with male which turns significant at the 10 per cent level. In our prefered specification in column 5 we see that UC boys have 15 per cent of 1SD advantage over UC girls (as shown in figure 1 , while SC and STs boys show a disadvantage that was not apparent in the unconditional analysis of that figure).

We follow the education literature and include family income as one of the measures of parental investments. We find that children with family income one standard deviation above the mean (XX vs mean) perform YY% of 1 SD in their test score at age 8. It is interesting to see that caste differentials are not driven by differentials in parental investments. By comparing columns 2 and 3, we find that adding family income does not change significantly castes differentials.

5.3 Robustness Check

As we mentioned in Section 2, a problem of our estimates are endogeneity that may arise from parents choosing investments levels for nutritional status and inputs for the production of cognition after observing the child quality and as a response of parental preferences. Fixed effect estimators allow us to control for permanent unobserved factors. Unfortunately only three countries in the younger cohort of the YL data have siblings scores: Ethiopia, Peru and Vietnam. Even though we do not observe all information about the children's siblings, we do observe for a sub-sample their siblings birthweight. As a robustness exercise we run the same set of regressions including the sibling's average birthweight as a regressor. If unobserved family characteristics associated with children's health are sibling invariant, variation within family can be used to identify children's health influence in children's performance on test score.

Table 6 presents the results for the fixed effects for our preferred IV specification, but now for the current sub-sample of children that do have siblings being sampled (around one quarter of the full sample). Column (1) replicates Column 5 of Table 5, while column (2) adds sibling's average birthweight. The main result is that our coefficient of interest of lagged HAZ is not significantly different and it is, at most, 1SD higher in in the fixed effects specification which is an encouraging result as this means that our main results are not being driven by unobserved family fixed effects, or at least those that are fixed across siblings.

Given that one of our controls (maternal height) might be very correlated with the sibling's average birthweight, we run a specification taking this variable out in the subsequent columns 3 and 4, but results remain unaltered.

Table 6: Production function of cognition, value-added equation with family fixed effects.

VARIABLES	IV (1)	IV (2)	IV (3)	IV (4)
Lag HAZ	0.07415**	0.03261	0.05348**	0.04645*
	(0.035)	(0.042)	(0.022)	(0.024)
Lag PPVT	0.16552***	0.17197***	0.16724***	0.16882***
	(0.042)	(0.041)	(0.044)	(0.044)
sibling's average birthweight		0.00020***		0.00016**
		(0.000)		(0.000)
Maternal height	0.00121	0.00154		
	(0.001)	(0.001)		
Scheduled caste	0.20849**	0.18741*	0.17953*	0.15881
	(0.088)	(0.098)	(0.085)	(0.099)
Scheduled tribe	-0.16000***	-0.08783	-0.18192**	-0.13465
	(0.055)	(0.066)	(0.077)	(0.084)
Other castes	-0.20947	-0.20723	-0.20814	-0.20941
	(0.173)	(0.187)	(0.193)	(0.207)
Scheduled caste*Male dummy	-0.24778	-0.20154	-0.22848	-0.18657
	(0.170)	(0.187)	(0.176)	(0.200)
Scheduled tribe*Male dummy	-0.07552	-0.12118	-0.04562	-0.07880
	(0.275)	(0.281)	(0.292)	(0.295)
Upper castes*Male dummy	0.18594	0.21430	0.18192	0.20833
	(0.174)	(0.186)	(0.194)	(0.210)
Male dummy	0.56500***	0.52528***	0.56523***	0.53754***
	(0.040)	(0.048)	(0.040)	(0.050)
Family income	0.00108**	0.00113***	0.00113**	0.00113**
	(0.000)	(0.000)	(0.000)	(0.000)
Age of child (months)	0.03239***	0.03110***	0.03089***	0.03043***
	(0.006)	(0.006)	(0.006)	(0.006)
First born dummy	2.21024***	2.14473***	2.25213***	2.16724***
	(0.112)	(0.104)	(0.122)	(0.117)
Last born dummy	-0.00135	0.00229	0.00659	0.01337
	(0.059)	(0.060)	(0.066)	(0.066)
Urban dummy	0.02394	0.00719	0.03550	0.02357
	(0.056)	(0.053)	(0.057)	(0.056)
Household size	0.03148***	0.02908***	0.03213***	0.03130***
	(0.006)	(0.006)	(0.008)	(0.007)
Caregiver's education level	-0.00386	-0.00797**	-0.00337	-0.00540
	(0.004)	(0.004)	(0.005)	(0.005)
Constant	-8.84862	-7.71325	-9.36895	-8.66949
	(9.096)	(8.808)	(10.133)	(9.864)
Observations	302	302	308	308
R-squared	0.325	0.330	0.327	0.330

Robust standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Notes: When taking income out of the regression, results remain unaltered. Also, see Appendix 6 for results that include wealth instead of income and show no change in the main coefficients

5.4 Caste test score gaps

Using the production function estimates from column (5) in Table 5 (preferred IV specification), we examine the extent to which differences in HAZ and family income (our proxy for parental inputs) can account for caste disparities in test scores. We examine the fit of the model by comparing the actual values of test score gaps by age to the gap predicted under the model by caste group. The estimated model captures key features of the data, such as the magnitude of the gap for each of the groups (see columns 1 and 2 in Table 7). On average our estimates slightly underestimate the true performance across children of both castes.

The estimated production function coefficients do vary by caste and gender, therefore the gap in the predicted test scores will arise from both "pure" caste-effects (by gender) and by caste differences in various inputs. Given this, in column 3 we examine how the predicted test score gaps vary if we compensate low–HAZ-LC children with the average level of HAZ observed for UC children (i.e., we will assign the average HAZ of an UC boy to each LC boy with HAZ score below the average HAZ of an UC boy; and do the same with girls). It is shown in Table 7 that if HAZ is equalised in this way, then the PPVT test score gap would be reduced by 21 percent (0.09 points, o 9 per cent of a standard deviation). It is interesting to note that the absolute gap between UC and LC for boys is closed by more when equalising HAZ, 0.09. For girls, 0.086 of the gap would be closed by leveling nutritional status. However, the relative gap is closed by more for girls (32.2 per cent vs. 17.3) because their caste-gap is much smaller. Another interesting point is to note the amount of "units of nutrition" we are assigning individually; we are actually giving more to girls in our simulation because many LC girls are doing much, much worse than the average UC girl.

The simulations in column 4 show that once again a similar absolute gap would be close by leveling family income (0.04), but this represents an average reduction of 9.4 and a 7.2 per cent for boys and 15.4 per cent for girls.

These simulations show that while the caste-gap for boys is larger, the amount needed to compensate girls is larger. The latter seems an indication of pro-male discrimination. This is consistent with the findings of some recent papers on India on differential treatment of girls and boys: Rose (2000) and Barcellos et al. (2012), for example, demonstrate differences in time allocation of mothers in Indian households with and without sons. Jayachandran and Kuziemko (2012) identify gender differences in the duration of breastfeeding of young children. In general, however, literature on gender bias in intra-household allocation often does not find evidence of differential treatment of children. The closest result to ours would be that of Zimmerman who shows that girls' school enrollment is more vulnerable to rainfall shocks than that of boys, with 6-10 year old children driving these effects (Zimmerman, 2011)

However, the fact that UC families seem to be discriminating more against their girls (gender gap was shown by gap in the PPVT at age 8 which is 41 percent of one standard deviation for UC, while is only 13 percent for the LC) is a new finding in this extensive literature.

Moreover, it is interesting to note that our results are also partly consistent with those of Sachar, who finds that once the children are placed in "more favourable" circumstances (i.e., when parents, especially mothers are literate

Table 7: PPVT caste gap closed by nutrition: IV-Value Added specification

	Actual caste gap (1)	Predicted caste gap (2)	Closed by nutrition (3)	Closed by FI (4)
All	0.442	0.435	0.091	0.041
			(21.0%)	(9.4%)
Boys	0.567	0.552	0.095	0.040
			(17.3%)	(7.2%)
Girls	0.266	0.270	0.087	0.042
			(32.2%)	(15.4%)

Note 1: The percentage of the gap closed is in parentheses.

and when infrastructure facilities are better), inter-community (Hindus/SC-ST/Muslims) differences in enrolment rates become insignificant (Sachar 2006). In our case, the gaps still exist, although to a lesser extent.

6 Conclusions and further research

This paper explores children's cognitive outcomes using novel panel data from India for children 55 through 102 months of age. With a value added specification, we found that a 1 standard deviation increase in height-for-age z-scores at the age of 5 leads to cognitive test scores that are about a 16 per cent of a standard deviation higher at age 8. Our analysis suggests that the differences in income levels between castes found in adulthood arise early in childhood. After controlling for a wide range of controls; upper caste children show a substantial advantage in vocabulary tests, but most importantly, they show a much more pronounced inequality between boys and girls than their lower caste counterparts. Compensating low caste children with the average nutritional status of their upper caste counterparts would close around one fifth of the caste cognitive differentials. That gap would be closed by one third in the case of girls, because their caste gap is smaller, showing that UC families seem to discriminate more against girls than LC families. Using a sub-sample of the data with information about siblings' birthweight in a unique way, we find that family fixed effects explain 1 SD of the overall nutrition-cognition effect and are not driving our results at all.

In terms of further research, it might be worth analysing with the Older Cohort data that spans children from 8 to 16 year-olds how our results extrapolate to older ages and to other cognitive tests; and also see how vocabulary is related to later adulthood outcomes such as employment, earnings, marriage market, early pregnancy, etc.

Appendix Tables

Table 8: Production function of cognition, contemporanous equation

VARIABLES	OLS (1)	OLS (2)	OLS (3)	IV (4)
Lag HAZ			0.10***	0.20***
			(0.006)	(0.017)
Scheduled caste	0.06**	0.12***	0.12***	0.12***
	(0.021)	(0.032)	(0.039)	(0.042)
Scheduled tribe	-0.11***	-0.10***	-0.01	-0.01
	(0.017)	(0.029)	(0.040)	(0.040)
Upper castes	0.42***	0.32***	-0.14**	-0.15***
	(0.032)	(0.069)	(0.051)	(0.048)
Scheduled caste* Male dummy		-0.11**	-0.14***	-0.15***
		(0.045)	(0.047)	(0.050)
Scheduled tribe* Male dummy		-0.04	0.02	-0.01
		(0.051)	(0.044)	(0.045)
Upper castes* Male dummy		0.15*	0.20***	0.18***
		(0.073)	(0.042)	(0.040)
Male dummy		0.25***	0.27***	0.29***
		(0.019)	(0.015)	(0.016)
Age of child (months)			0.02***	0.03***
			(0.005)	(0.005)
First born dummy			-0.14	-0.12
			(0.475)	(0.418)
Last born dummy			-0.12***	-0.11***
			(0.013)	(0.010)
Urban dummy			0.17***	0.17***
			(0.038)	(0.036)
Household size			-0.01***	-0.01***
			(0.003)	(0.003)
Wealth index			0.58***	0.52***
			(0.062)	(0.063)
Caregiver's education level			0.03***	0.03***
			(0.002)	(0.002)
Father's education level			0.01***	0.01***
			(0.002)	(0.002)
Constant	-0.09***	-0.21***	5.78***	5.49***
	(0.020)	(0.022)	(1.719)	(1.660)
Observations	1,457	1,457	1,408	1,395
R-squared	0.027	0.043	0.274	0.265

Robust standard errors in parentheses
** p<0.01, ** p<0.05, * p<0.1

Table 9: Production function of cognition, restricted value-added equation

VARIABLES	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Lag HAZ			0.0185	0.0150
			(0.0108)	(0.00942)
Scheduled caste	0.152***	0.283***	0.179***	0.193***
	(0.0333)	(0.0476)	(0.0578)	(0.0609)
Scheduled tribe	-0.700***	-0.630***	-0.330***	-0.320***
	(0.0265)	(0.0356)	(0.0795)	(0.0857)
Upper castes	0.0387	-0.0425	-0.123**	-0.117*
	(0.529)	(0.0655)	(0.0581)	(0.0606)
Scheduled caste* Male dummy		-0.255***	-0.259***	-0.277***
		(0.0444)	(0.0438)	(0.0453)
Scheduled tribe* Male dummy		-0.151***	-0.152***	-0.185***
		(0.0283)	(0.0392)	(0.0427)
Upper castes* Male dummy		0.114	0.0689	0.0505
		(0.120)	(0.102)	(0.101)
Male dummy		0.299***	0.285***	0.289***
		(0.0358)	(0.0313)	(0.0320)
Age of child (months)			0.00606*	0.00345
			(0.00340)	(0.00316)
First born dummy			-0.412**	-0.372*
			(0.191)	(0.190)
Last born dummy			-0.0665***	-0.0572***
			(0.0187)	(0.0182)
Urban dummy			-0.0913	-0.0768
			(0.0842)	(0.0860)
Household size			-0.00399	-0.00409
			(0.00789)	(0.00799)
Wealth index			0.510***	0.400***
			(0.0748)	(0.0903)
Change in Wealth Index				0.338***
				(0.085)
Change in HAZ				0.0187**
				(0.008)
Caregiver's education level			0.00953*	0.0116**
			(0.00457)	(0.00476)
Father's education level			-0.0149***	-0.0129***
			(0.00268)	(0.00288)
Constant	0.114***	-0.0377	-4.064	-3.588
	(0.0215)	(0.0230)	(3.092)	(2.895)
Observations	1,457	1,457	1,436	1,419
R-squared	0.052	0.067	0.151	0.152

Robust standard errors in parentheses
** p<0.01, ** p<0.05, * p<0.1

Table 10: Production function of cognition, value-added equation

VARIABLES	OLS (1)	OLS (2)	OLS (3)	IV (4)
Lag HAZ			0.09***	0.16***
			(0.005)	(0.019)
Lag PPVT	0.35***	0.35***	0.22***	0.21***
	(0.009)	(0.008)	(0.013)	(0.014)
Scheduled caste	0.09***	0.18***	0.14***	0.14***
	(0.024)	(0.034)	(0.041)	(0.043)
Scheduled tribe	-0.32	-0.29***	-0.07	-0.07
	(0.015)	(0.016)	(0.045)	(0.044)
Upper castes	0.29***	0.20***	-0.13***	-0.14***
	(0.034)	(0.063)	(0.046)	(0.043)
Scheduled caste* Male dummy		-0.16***	-0.18***	-0.18***
		(0.040)	(0.044)	(0.046)
Scheduled tribe* Male dummy		-0.08**	-0.03	-0.05
		(0.032)	(0.038)	(0.038)
Upper castes* Male dummy		0.14	0.17***	0.15***
		(0.084)	(0.045)	(0.040)
Male dummy		0.26***	0.28***	0.29***
		(0.018)	(0.012)	(0.012)
Age of child (months)			0.02***	0.02***
			(0.005)	(0.004)
First born dummy			-0.20	-0.19
			(0.411)	(0.370)
Last born dummy			-0.11***	-0.11***
			(0.013)	(0.011)
Urban dummy			0.11**	0.12***
			(0.047)	(0.044)
Household size			-0.01**	-0.01***
			(0.004)	(0.004)
Wealth index			0.57***	0.54***
			(0.062)	(0.062)
Caregiver's education level			0.03***	0.03***
			(0.003)	(0.003)
Father's education level			0.01***	0.01***
			(0.002)	(0.002)
Constant	-0.02	-0.15***	3.91**	3.70**
	(0.019)	(0.017)	(1.733)	(1.666)
Observations	1,457	1,457	1,408	1,395
R-squared	0.127	0.145	0.305	0.301

Robust standard errors in parentheses
** p<0.01, ** p<0.05, * p<0.1

Appendix 1: The caste system in Andhra Pradesh

The caste system is still extremely important in India in various spheres, not least politically. The 'Other Castes' (also called 'Upper Castes', as I have defined them here) category comprises mostly of 'forward castes' who traditionally enjoy a more privileged socio-economic status; at the other end of the spectrum, Scheduled Castes (SCs) and Scheduled Tribes (STs) are traditionally disadvantaged communities. SCs are the lowest in the traditional caste structure. They were formerly known as the 'untouchables' and now call themselves Dalit. In rural Andhra Pradesh, SC colonies are located separately, and in most cases away from the main villages. These colonies are named after the caste and even in the official records are often called harijana wada (or Dalit colonies). They have been subjected to discrimination for centuries and therefore had no access to basic services, including education. National legislation aims to prohibit 'untouchability' and discrimination. STs are the indigenous people, living in and dependent on forests. Different groups of tribes live in different parts of Andhra Pradesh and vary in their culture, language and lifestyles. Though a good number of them are mainstreamed and live in plain areas, a considerable proportion continues to live in isolated hilltops and has little access to services. Backward Classes (BCs) are people belonging to a group of castes who are considered to be backward in view of the low level of the caste in the structure. In Andhra Pradesh, the BCs are further divided into four groups (ABCD) and some caste groups are placed in each of these sub-groups. Recently, the High Court has ordered the inclusion of a fifth sub-group, E, and Muslims have been placed into this category.

References

Aaronson, D. (1998): "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," Journal of Human Resources, 33(4), 915–946.

Almond, D., and J. Currie (????): "Human capital development before age five," in Handbook of Labor Economics, ed. by O. Ashenfelter, and D. Card, vol. 4, Part B. New York, Columbia University.

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.

Arnold, F., S. Parasuraman, A. P., and M. Kothari (2009): "Nutrition in India. National Family Health Survey (NFHS-3), India, 2005-06," Mumbai: International Institute for Population Sciences; Calverton, Maryland, USA: ICF Macro.

Asadullah, M., U. Kambhampati, and F. Lopez Boo (2009): "Social divisions in school participation and attainment in India," IDB-Research Department Working Papers number 692.

Baird, S., J. Friedman, and N. Schady (2011): "Aggregate Income Shocks and Infant Mortality in the Developing World," The Review of Economics and Statistics, 93(3), 847–856.

Banerjee, A., and R. Somanathan (2007): "The Political Economy of Public Goods: Some Evidence from India," Journal of Development Economics, 82(2), 287–314.

Barcellos, S., L. Carvalho, and A. Lleras-Muney (2012): "Child Gender and Parental Investments in India - Are Boys and Girls Treated Differently?," NBER Working Paper.

Becker, G. S., and N. Tomes (1986): "Human Capital and the Rise and Fall of Families," Journal of Labor Economics, 4(3), S1–39.

Behrman, J., M. Calderon, S. Preston, J. Hoddinott, R. Martorell, and A. Stein (2009): "Nutritional Supplementation of Girls Influences the Growth of their Children: Prospective Study in Guatemala," American Journal of Clinical Nutrition, 90, 1372–1379.

Bhalotra, S., and B. Zamora (2010): "Social divisions in education in India," A. Sharif and R. Basant (eds.), Handbook of Muslims in India, Delhi: Oxford University Press.

Carneiro, P., F. Cunha, and J. Heckman (2003): "Interpreting the evidence of family influcence on child development," Paper presented at The Federal Reserve Bank of Minneapolis and The McKnight Foundation conference (in cooperation with the University of Minnesota) on The Economics of Early Childhood Development: Lessons for Economic Policy. October 17 2003.

Cunha, F., and J. Heckman (2007): "The technology of skill formation," American Economic Review, 97(2), 31–47.

——— (2008): "Formulating, identifying and estimating the technology of cognitive and non-cognitive skill formation," Journal of Human Resources, 43(4), 739–782.

Dostie, B., and R. Jayaraman (2006): "Determinants of School Enrollments in Indian Villages," Economic Development and Cultural Change, 54(2), 405–21.

Dreze, J., and G. G. Kingdon (2001): "School Participation in Rural India," Review of Development Economics, 5(1), 1–24.

Engle, P., M. Black, J. Behrman, M. Cabral de Mello, P. Gertler, L. Kapiriri, R. Martorell, M. Young, and the International Child Development Steering Group (2007): "Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world," Lancet, 369, 22942.

Engle, P., L. Fernald, H. Alderman, J. Behrman, C. OGara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, S. Iltus, and the Global Child Development Steering Group (2011): "Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low and Middle Income Countries," Lancet, 378(9799), 1339–53.

Fernald, L., A. Weber, E. Galasso, and L. Ratsifandrihamanana (2011): "Socioeconomic gradients and child development in a very low income population: Evidence from Madagascar," Dev Sci, 14(4), 832–47.

Filmer, D., and L. Pritchett (1999): "The Effect of Household Wealth on Educational Attainment: Demographic and Health Survey Evidence," World Bank Policy Research Working Paper number 1980.

Gang, I. N., K. Sen, and Myeong-Su Yun (2008): "Poverty In Rural India: Caste And Tribe," Review of Income and Wealth, 54(1), 50–70.

Heckman, J. (1995): "Lessons from the Bell curve," Journal of Political Economy, 103(5), 1091–1120.

——— (2005): "Lessons from the Technology of Skill Formation," Working Papers 11142, National Bureau of Economic Research, Inc.Cambridge, MA.

Heckman, J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," Journal of Labor Economics, 24(3), 411–482.

Helmers, C., and M. Patnam (2011): "The formation and evolution of childhood skill acquisition: Evidence from India," Journal of Development Economics, 95(2), 252–266.

Hoddinott, J., J. A. Maluccio, J. R. Behrman, R. Flores, and R. Martorell (2008): "The Impact of Nutrition During Early Childhood on Income, Hours Worked, and Wages of Guatemalan Adults," The Lancet, 371(1), 411–416.

Jayachandran, S., and I. Kuziemko (2012): "Why Do Mothers Breastfeed Girls Less Than Boys? Evidence and Implications for Child Health in India," The Quarterly Journal of Economics, 126(3), 1485–1538.

Jenkins, R., and E. Barr (2006): "Social Exclusion of Scheduled Caste Children from Primary Education in India," mimeo UNICEF, India, New Delhi.

Kingdon, G. (2002): "The Gender Gap in Educational Attainment in India: How Much Can be Explained?," Journal of Development Studies, 39(2), 25–53.

——— (2005): "Where Has All the Bias Gone? Detecting Gender Bias in the Household Allocation of Educational Expenditure in India," Economic Development and Cultural Change, 53(2), 409–51.

Kumra, N. (2008): "An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India," Young Live technical note no. 2.

Maluccio, J., J. Hoddinott, J. R. Behrman, A. Quisumbing, R. Martorell, and A. Stein (2009): "The Impact of Nutrition During Early Childhood on Education among Guatemalan Adults," Economic Journal, 119(537), 251–66.

Martorell, R., M. P, J. Maluccio, A. Stein, and R. J.A. (2010): "The nutrition intervention improved adult human capital and economic productivity.," Journal of Nutrition, 140(2), 411–414.

Naudeau, S., S. Martinez, P. Premand, and D. Filmer (2011): "Cognitive development among young children in low-income countries. In No Small Matter: The Impact of Poverty, Shocks, and Human Capital Investments in Early Childhood Development," Alderman H ed. The World Bank.

Nores, M., and S. Barnett (2010): "Benefits of Early Childhood Interventions across the World: (Under) Investing in the Very Young," Economics of Education Review, 29, 271–282.

Organization, W. H. (2006): Neo-natal and Peri-natal Mortality: Country, Regional, and Global Estimates. WHO Press, Geneva.

Outes-Leon, L., C. Porter, and A. Sanchez (2011): "Early Nutrition and Cognition in Peru: A Within-Sibling Investigation," IDB WORKING PAPER SERIES No. IDB-WP-241.

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.

Rose, E. (2000): "Gender Bias, Credit Constraints and Time Allocation in Rural India.," The Economic Journal, 110(465), 738–758.

ROSENZWEIG, M., AND K. WOLPIN (1988): "Heterogeneity, intrafamily distribution and child health," Journal of Human Resources, 23(4), 437461.

SACHAR, R. (2006): Social, Economic and Educational Status of the Muslim Community of India: A Report. Prime Minister's High Level Committee Cabinet Secretariat Government of India.

SAMEROFF, A., R. SEIFER, AND C. BALDWIN (1993): "Stability of Intelligence from Preschool to Adolescence, The influence of Social and Family Risk Factors," Child Development, 64, 80–97.

SANCHEZ, A. (2009): "Early Nutrition and Cognitive Achievement in Preschool Children in Peru," Working Paper NO. 57. Department of International Development, University of Oxford, Oxford OX1 3TB, UK.

SCHADY, N. (2011): "Parental education, vocabulary, and cognitive development in early childhood: Longitudinal evidence from Ecuador," Am J Public Health, 101(12), 2299–307.

SHONKOFF, J., AND D. PHILLIPS (2000): "From Neurons to Neighborhoods: The Science of Early Childhood Development," National Academy Press, Washington, DC.

THOMAS, D., J. STRAUSS, AND M. HENRIQUES (1990): "Child Survival, Height for Age and Household Characteristics in Brazil," Journal of Development Economics, 33(2), 197–234.

TODD, P., AND K. WOLPIN (2003): "On the specification and estimation of the production function for cognitive achievement," Economic Journal, 113(485), F3–33.

——— (2007): "The production of cognitive achievement in children: Home, school and racial test score gaps," Journal of Human Capital, 1(1), 91–136.

(UNESCO), UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (2008): Education for All Global Monitoring Report 2008, no. 9780199532636 in OUP Catalogue. Oxford University Press.

VICTORA C., L. ADAIR, C. F. P. H. R. M. L. R., H. S. ON BEHALF OF THE MATERNAL, AND C. U. S. GROUP (2008): "Undernutrition 2: Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital," The Lancet, 371(9609), 340–357.

WALKER, S., T. WACHS, S. GRANTHAM-MCGREGOR, M. M. BLACK, C. A. NELSON, S. L. HUFFMAN, H. BAKER-HENNINGHAM, S. M. CHANG, J. D. HAMADANI, B. LOZOFF, J. M. MEEKS GARDNER, C. A. POWELL, A. RAHMAN, AND L. RICHTER (2011): "Inequality in early childhood: risk and protective factors for early child development," Lancet, 378(9799), 1325 – 1338.