WORKING PAPER NO. 55

The Production Function of Cognitive Skills: Nutrition, Parental Inputs and Caste Test Gaps in India

October 2009

Florencia Lopez Boo

Young Lives
An International Study of Childhood Poverty

WORKING PAPER NO. 55

The Production Function of Cognitive Skills:

Nutrition, Parental Inputs and Caste Test Gaps in India

October 2009

Florencia Lopez Boo

Young Lives, Department of International Development, University of Oxford, Oxford OX1 3TB, UK

The Production Function of Cognitive Skills:
Nutrition, Parental Inputs and Caste Test Gaps in India

Florencia Lopez Boo

First published by Young Lives in October 2009

ISBN: 978-1-904427-61-2

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

Available from:
Young Lives
Department of International Development
University of Oxford
3 Mansfield Road
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Fax: +44 (0)1865 281801
E-mail: younglives@younglives.org.uk
Web: www.younglives.org.uk

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

Contents

Abstract

This paper explores the determinants of children's cognitive outcomes using novel panel data from two cohorts of children in India. As in Todd and Wolpin's study (2007), I do not find evidence supporting restrictive models that assume test scores depend only on contemporaneous inputs. Rather, the results show that lagged inputs affect the production of current skills. In models where past inputs are not observed or imperfectly measured, past nutritional status and/or past test scores turn out to be a good proxy-indicator of this variable, which is evidence for the 'sufficiency' assumption. I allow for the endogeneity of nutrition using an instrumental variable approach and find that a 1 standard deviation increase in height-for-age z-scores at the age of 1 leads to cognitive test scores that are about a quarter of a standard deviation higher at age 5. Results for the Older Cohort have a similar magnitude. I also study the behaviour of inputs and find that parents seem to 'reinforce' children for early favourable outcomes rather than 'compensate' them for adverse scores; and they do so more in lower caste families and particularly with boys.

JEL Classification Codes: J13, J15, J24

Keywords: cognitive outcomes, health, nutrition, children and test score gap

Acknowledgements

The author wishes to thank M. Niaz Asadullah for very helpful discussions. The paper also benefited from excellent comments received from Geeta Kingdon and also from Paul Glewwe, Stefan Dercon, Patrice Engle, Jeremie Gignoux, Juan Pablo Rud and the Research Department of the IDB seminar participants. This research was funded by Young Lives (www.younglives.org.uk), a long-term 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 the Bernard van Leer Foundation, Inter-American Development Bank (in Peru), International Development Research Centre (in Ethiopia), and the Oak Foundation. 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, or those of the Inter-American Development Bank or their board of directors.

The Author

Florencia Lopez Boo is an Economist at the Inter-American Development Bank (IDB) and a Research Affiliate at IZA. She received a doctorate in Economics from the University of Oxford in 2008. She also holds an MSc in Development Economics from the Catholic University of Louvain and Namur University (Belgium). Before joining the IDB as a Young Professional (YP), she was a Research Assistant at STICERD/LSE and College-Lecturer at Oxford.

Her research interests include labour economics (children's and young people's work–education decisions, wage inequality), development economics (inequality and poverty) and early childhood development (cognitive skills, health and parental investment). At the IDB, she co-coordinates the early childhood development policy and research agenda and supports the impact evaluation design and implementation of conditional cash transfers and other social protection programmes in Latin America.

Young Lives is core-funded by the UK Department for International Development (DFID) for the benefit of developing countries, with sub-studies funded by International Development Research Centre (IDRC) in Ethiopia, UNICEF in India, Irish Aid in Vietnam and the Bernard van Leer Foundation.

The views expressed here are those of the author(s). They are not necessarily those of the Young Lives project, the University of Oxford, DFID or other funders.

1 Introduction and literature review

Much importance has been given lately to the examination of the relevance of cognitive skills,[1] mainly because of its power to explain school and professional success (Heckman 1995).[2] It is also known that cognitive abilities are formed relatively early in life and become less malleable as children age. At least two factors contribute to shaping children's ability: genetic endowment and the home environment, along with inputs received early in life. Many studies highlight the importance of genetic inheritance (Teasdale and Owen 1983; Wilson 1983). Nevertheless, these and other studies (from studying twins) also show that even if about 50 per cent of the variance in child development is due to genetic factors, a child's genetic expression is very much influenced by his or her environmental inputs.[3]

These findings have motivated a large amount of research on the disparities between socio-economic groups that can arise from a lack of early childhood investments or interventions (Heckman 2000; Carneiro and Heckman 2003; Carneiro et al. 2003; Heckman 2005; Cunha and Heckman 2007). Using secondary survey data for developed countries, researchers have addressed many of the issues related to early parental invesments or interventions like ‘Head Start’ and the cumulative nature of cognitive abilities in a variety of ways (Lally et al. 1988; Murnane et al. 1995; Garces and Currie 2002; DeCicca 2007; Temple and Reynolds 2007). Despite these recent advances in the literature, rarely does research on early childhood use developing country data.

Nonetheless, the motivation for research on developing countries comes from the fact that limited or inappropiate nutrition and stimulation early in life and sometimes discrimination in the intra-household allocation of resources (Grantham-McGregor et al. 2007; Kingdon 2002, 2005), problems so intrinsic to the developing world, might prevent the development and catch up of disadvantaged children. The existing evidence in less developed countries is mainly based on data from Latin America and the Lancet series (Engle et al. 2007; Grantham-McGregor et al. 2007; Walker et al. 2007). Researchers have shown the positive effects of pre-school, cash tranfers and nutritional status on school performance or cognitive outcomes of mostly school-aged children in this region.[4,5]

1 I use the words ‘skills’ and ‘abilities’ interchangeably, following Griliches (1977) and Cunha and Heckman (2008).

2 Heckman and Rubinstein (2001) and Cunha and Heckman (2008) have shown that the development of non-cognitive abilities also starts early in life, is equally affected by a child's environment, and plays an important role for a successful path in life. However, these skills remain malleable (Carneiro et al. 2003). That is why this paper concentrates on cognitive outcomes, whose early deficiencies have been proven much harder to remedy (O'Connor et al. 2000).

3 For instance, Scarr and Weinberg (1983) found that young siblings (black/interracial adoptees) were intellectually similar to their natural brothers/sisters. They hence conclude that younger children are more influenced by differences in their family environment.

4 For instance, Berlinski et al. (2008) used panel data and found that one year of pre-school increases average third grade scores by 8 per cent of a mean of the distribution of test scores in Argentina. Some exceptions that do study pre-school aged children are: for rural Nicaragua, Macours et al. (2008), who found significant effects on cognitive outcomes, especially language, from a randomised cash transfer programme. Moreover, Gertler and Fernald (2004) as well as Paxson and Schady (2007) found similar evidence with the Mexican Programme *Oportunidades* and the Ecuadorian Programme *Bono de Desarrollo Humano* for children 36 to 72 months.

Particularly, Schady (2006) surveys the few Latin American studies available and finds that differences in test performance by socio-economic status (SES) persist as children age. Divisions by SES in India are defined according to whether a household belongs to a certain caste. Indeed, most of the existing studies on determinants of school participation and attainment in India today acknowledge socio-religious differences in the population (Dreze and Kingdon 2001; Kingdon 2002; Dostie and Jayaraman 2006; Gang et al. 2008). In these studies, low participation 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 show the effects of early childhood conditions on cognitive skills or school participation.

In the recent past, the government of India has introduced a range of policy interventions (i.e., affirmative action in educational progress) targeting social groups like Scheduled Castes (SC) and Scheduled Tribes (ST).[6] Some evidence shows that these interventions have been successful (Jenkins and Barr 2006).[7] However, apart from Balhotra and Zamora (2007) and Asadullah et al. (2009), none of the existent studies emphasises the importance of one's caste membership as a determinant of educational attainment. More importantly, no research contributes to the understanding of how disparities in parental inputs and nutrition contribute to the process of skill formation by caste over time.

The objective of this analysis is therefore to fill an important gap in the literature by investigating determinants of a child's development of cognitive skills over two phases of childhood: pre-school and school ages. To the best of my knowledge, this is the first study using panel data to assess, firstly, the causal relation between nutrition and cognition at pre-school age in India; and secondly, the *direct* impact of parenting practices (i.e., home input score index) and the *indirect* impact of better parenting that improves a child's overall nutrition status (i.e., height-for-age z scores) on current and later receptive vocabulary test scores (Peabody Picture Vocabulary Test) and general intelligence test scores (Ravens Test); as well as the behaviour of parental investments with respect to past performance and nutrition; and thirdly, whether there are socio-economic gradients/gaps that emerge already in childhood and how they are mediated by early family investments (Heckman 1995, 2000; Fryer and Levitt 2004) in a developing country. I therefore complement and expand the literature on the nutrition–cognitive development nexus in developing countries (Engle et al. 2007) by taking into account parental practices (i.e., compensating/reinforcing early outcomes) and caste gradients.[8] For that, I exploit the novel and rich measures of parental inputs, anthropometrics and cognitive outcomes provided in the Young Lives

---

5 Investments in nutrition have been shown to be one of the most important predictors of later cognitive development and success in schooling and the labour market in developing countries. For a review, see Grantham-McGregor et al. (2007). For experimental studies from a longitudinal study from Guatemala, see Hoddinot et al. (2008), Behrman et al. (2008) and Maluccio et al. (2009). More related to this paper is the experimental work of Grantham-McGregor and her co-authors, who provides compelling evidence that parenting interventions may be more effective than health interventions in improving cognitive outcomes in Jamaica (1991, 1997). The important interventions are nutritional while there is not much evidence on morbidity effects on cognitive development other than malaria (through iron status) and possibly diarrhoea.

6 For further information on these groups, see Appendix 2.

7 As a matter of fact, the relative disadvantages experienced by SC children today are fewer than those faced by Muslims, while STs have made enormous progress.

8 Parental reinforcing practices are defined as those practices that use nutrition or other inputs to reinforce differences in children's learning ability or 'endowment' (i.e., more gifted children will receive more inputs as a reward), while compensating practices are those that attempt to equalise learning outcomes.

longitudinal data. This data consists of two cohorts of children (the 'younger' and the 'older') surveyed over two rounds four years apart (Round 1 in 2002 and Round 2 in 2006). Existing research with the Young Lives data so far has only explored various determinants of schooling and nutrition outcomes for the four sample countries.[9] However, none of these earlier studies look at the effect of direct and indirect early childhood investments on cognitive outcomes taking advantage of the panel nature of the data.

Furthermore, the methodology used here allows me to go beyond previous empirical studies. Todd and Wolpin (2007) 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. I therefore examine determinants of cognitive outcomes in different specifications for all ages in both cohorts. This structural production function analysis makes considerable progress in sorting out the causal relationship between nutrition, home inputs and cognitive outcomes. The challenge in estimating this relationship is that of other inputs being missing, a problem solved by the rich Young Lives data. However, yet another problem is that unobserved child, parents or household-specific factors may affect both nutrition (home inputs) and cognitive outcomes, which may lead to a correlation even though no causation exists. Particularly, since parental taste for child quality and a child's genetic ability are unobserved, Ordinary Least Squares (OLS) estimations of the nutrition–learning nexus, as well as the parental investment–learning nexus, are likely to be biased. This research will then take into account the endogeneity of height-for-age (HAZ hereafter) explicitly, while also exploring the determinants of parental inputs. I argue that since one important indicator of child malnutrition (low birth-weight) 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.[10]

I find strong evidence that better cognitive outcomes are related to a better nutritional status in early childhood, evidence which survives numerous robustness checks, including the history of past parental inputs in the cumulative specification. Past test scores are an important determinant of current test scores, which supports the *self-productivity* effects present in Cunha and Heckman (2008). Using a specification that incorporates these features, I analyse test score gaps between lower castes (LC) and upper castes (UC) and find that equalising nutrition at the average levels of UC children would close the caste test score gap by about 18.7 per cent, while by equalising home inputs at the average levels of UC children would close the caste test score gap by 9.5 per cent. Furthermore, when analysing the behaviour of parental inputs over time, estimates show a 'compensating' attitude for adverse early nutritional endowments and a 'reinforcing' attitude for favourable past test scores, particularly in LC households and with boys. This paper proceeds as follows. The second section describes the methodology and conceptual framework for modeling the cognitive skills production function and considers also its empirical application and statistical problems. The third section gives details on the data and variables used to

9 See Young Lives Working papers numbers 8, 16, 18, 20, 27 and 32 at www.younglives.org.uk. For instance, Galab et al. (2006) found a clear interplay between school and home for 8-year-olds. The authors assert that home environment remains crucial and argue that children perform better in school if parents place high value on schooling and actively support their children at home.

10 These IVs are interpreted as an aspect of the material conditions at birth which are free of influences by the home environment and the investment behaviour of family members since the birth of the child. If the results are robust, it would imply that there is a pathway from birth-weight or mothers height, through height-for-age, to cognitive achievement that is separate from family-specific influences.

represent cognitive outcomes, nutrition, home inputs and other controls, while the fourth section briefly gives some background on caste schooling gaps in India and the fifth section presents some descriptive statistics. The sixth section shows estimates of the cognitive skills production function. Parental investment demand functions are presented and I use the estimated cognitive skills production function to evaluate caste disparities in test scores. The last section concludes.

2 Methodology

In this section I outline the basic theoretical model that provides the basis for the empirical estimation. The empirical approach and its associated challenges are also discussed below.

2.1 Basic theoretical framework

The main interest of this paper lies in investigating the various *direct* and *indirect* determinants of current cognitive skill formation, with a particular focus on the effects of a child's past cognitive skill levels, past (and current) nutritional status and parental investment. For that reason, I discuss both a cognitive skills and nutrition production function, as well as parental inputs demand functions.

2.1.1 Cognitive achievement production function

I follow Todd and Wolpin (2003, 2007) and Cunha and Heckman (2008) in writing the skill level of child i at age t [11] as a function of the child's level of skills, current parental investment and other contemporaneous variables, including child, caregiver and household current characteristics.

$$\theta_{it} = f(I_{it}, H_{it}, X_{it}, \mu_{it}, \varepsilon_{it}) \quad (1)$$

where i denotes individuals, θ_{it} denotes a child's cognitive skill level for age *t* with *t* between 0 and T, I_{it} denotes parental investment at age t, H_{it} denotes nutrition[12] at age t and X_{it} denotes a vector of characteristics of child i, caregiver and home of child i, μ_{it} is the *expression* of the child's endowed mental capacity ('ability') and ε_{it} is an error term that includes the effect of the history of unobserved inputs and measurement error, hence ε_{it} is expected to be serially correlated.

In order to capture the (unobserved) genetic endowment with which a child is born, Todd and Wolpin propose to take first differences of a linear specification of equation (1); that is, substracting θ_{it-1} from both sides where first differences of skill levels are interpreted as value-added, which gives:

$$\theta_{it} = f(\theta_{it-1}, I_{it}, H_{it}, X_{it}, \mu_{i0}, \varepsilon_{it}) \quad (2)$$

11 I will omit the fact that child *i* is in household *j* for notation convenience.

12 I use H as this discussion is based on health (H) production functions.

In brief, this value-added specification relates a cognitive skills outcome measure to contemporaneous family input measures (inputs applied between the baseline measure and a current measure) and a lagged (baseline) skills measure, nutrition and other co-variates.[13] In the next sub-section, I show two versions of equation (2): the contemporaneous one (as specified here above for generality), and the cumulative one that includes H_{it-1} and I_{it-1}. The key assumption of this model is that the baseline skills and nutrition measures are taken to be a sufficient statistic for input histories, and in the versions of the model that do not incorporate endowments, the lagged test score and nutritional status are also taken as sufficient statistics for endowed mental capacity. When a baseline skills measure is not available (i.e., for the Younger Cohort), lagged nutritional status alone will be taken as a sufficient statistic. Evidence based on the value-added specification is generally regarded as more convincing than that based on a contemporaneous specification (Summers and Wolfe 1977; Hanushek 1996).

2.1.2 *Nutrition production function*

Nutritional status (H_{it}) is the product of genetics and a household production process, where time and market goods are inputs used to produce child nutrition.[14] Given wages and prices, parents maximise their utility subject to the full income constraint that includes both time and income. The production of child nutrition is then consistent with theoretical notions that relate H_{it} to a process of interaction of child, family and non-family factors (including health-inputs and the child's genetic endowment). Therefore, the nutrition production function is:

$$H_{it} = f(IH_{it-1}, X_{it}, E_i) \quad (3)$$

where H_{it} is the measure of child nutrition, IH_{it-1} - represents lagged health and nutrition-related inputs used to produce H_{it}, such as parents' time spent with child and other inputs, including the efficiency of parents (or parental quality). X_{it} are family (e.g., environment and home) and non-family (e.g., public health services, access to water and sanitation in the community, etc.) influences and E_i is the child genetic endowment.

2.1.3 *Parental inputs demand function*

The input demands are specified as functions of exogenous child and family characteristics, home characteristics, lagged test scores (or lagged nutrition outcomes, depending on specification), a children heterogeneity term (μ_i), and an independently distributed shock (ε_{it}), as follows:

$$I_{it} = f(\theta_{it-1}, H_{it}, X_{it}, \mu_i, \varepsilon_{it}) \quad (4)$$

$$IH_{it} = f(H_{it-1}, X_{it}, \mu_i, \varepsilon_{it}) \quad (5)$$

13 In order to perform a first difference, as in Todd and Wolpin (2007), I would have needed: i) longer panel data and ii) each child to have had two measurements on the same dependent variable, which unfortunately is not the case (PPVT is available at age 12 and Ravens at age 8).

14 The household production function approach assumes that parents increase their utility if children are in good health, which implies the existence of a demand function for child health or child nutritional status.

where I_{it} and IH_{it} represent inputs used to produce skills, θ_{it}, and nutrition, H_{it}, respectively. These input demand functions do have standard interpretations as demand functions, even when there are unmeasured inputs in the above equation, provided the prices of these inputs and other factors affecting their demand are controlled for in the vector X_{it}.[15] The coefficient on either θ_{it-1} or H_{it-1} reflects 'self-productivity' effects (Cunha and Heckman 2008).

2.2 Empirical strategy for estimation of the production function of cognitive skills

In order to facilitate the modeling of various outcomes and inputs, I have explicitly made assumptions about the timing of corresponding measurements. For instance, nutrition and cognitive outcomes are revealed at the end of period $t-1$ (or the very start of t), while decisions about nutrition and other inputs are made at the beginning of t. In addition, i.i.d. random shocks μ_{it} (for the child) and ε_{it} (for the household/family) are assumed to be realised at the start of time t. Now, assuming linearity in the production function (2), the estimation equations will take the following forms:[16,17]

The contemporaneous specification: This specification relates test scores in the *t*th. period to data only on contemporaneous inputs:

$$\theta_{it} = \alpha H_{it} + \gamma I_{it} + \delta X_{it} + \beta' \mu_{i0} + \varepsilon_{it} \qquad (6)$$

Here, test scores are being produced with current nutrition, H_{it}, both observed and unobserved contemporaneous inputs, I_{it}, and endowments, μ_{i0},[18] plus other factors, X_{it}, that include child, family, community and school characteristics, and ε_{it} is a random error term.

The cumulative specification: This specification expands the contemporaneous one by including observable lagged inputs and nutrition (H_{it-1} and I_{it-1}) and test scores, θ_{it-1}, as in the equation below:[19]

$$\theta_{it} = \alpha_1 H_{it} + \alpha_2 H_{it-1} + \gamma_1 I_{it} + \gamma_2 I_{it-1} + \delta X_{it} + \zeta\theta_{it-1} + \beta_t \mu_{i0} + \varepsilon_{it} \qquad (7)$$

Ideally, data will be available on *complete* historical inputs. However, measuring *cumulative* nutrition inputs is extremely difficult, so a simple alternative is to use the child's nutritional status as a summary statistic for the nutritional history up to that age. Thus, within a multi-period framework, H_{it-1} refers to the end of the previous period and is a sufficient statistic for all previous inputs and prices. It also has to be assumed that any omitted inputs and measurement error in test scores are uncorrelated with included inputs. Therefore, equations (6) and (7) are the ones to be finally estimated.

15 It is then assumed that parents are attempting to maximise a well-defined expected utility function subject to a per-period budget constraint with the optimal choice of inputs being obtained from conditional demand functions.

16 Pursuing a linear specification of equation (2) relies on the fact that the *expression* of unobserved genetic endowment is constant over time, which seems a rather strong assumption but is needed in order to facilitate the estimation.

17 The same specifications (contemporaneous and cumulative) are applied to the health production function, but are not presented here for the sake of brevity.

18 The fact that the coefficient on unobserved genetic endowments is a constant independent of age, β', yields equation (6).

19 Because ε_{it} is expected to be serially correlated, Todd and Wolpin (2007) propose first differencing this equation to take away the fixed effect.

2.2.1 Endogeneity of child nutritional status and home inputs

A common problem in the production function approach to studying child outcomes relates to endogeneity of particular regressors, such as nutritional status or home inputs. The main challenge of estimating either equation (6) or (7) arises from the possibility that at least one of the folllowing conditions needed for the OLS estimation to be consistent does not hold:[20]

1. $E(H_{it}, \varepsilon_{it}) = 0$ (c1)
2. $E(H_{it}, \mu_i) = 0$ (c2)

A violation of condition (c1) could arise if there are unobserved parental preferences in ε_{it} that could thus lead to an *omitted variable bias* and upwardly bias the estimates of the coefficient on nutrition or that of home inputs. It is conceivable that parents with a strong preference for child investments provide their children with other inputs that improve both child nutritional status and cognitive skills.

In turn, a violation of condition (c2) implies that nutrition and/or home inputs are functions of, among other things, child learning efficiency or 'endowment'', μ_i, which represents factors, such as ability and motivation, that are out of parents' control but are influenced by home environment as well as by genetics (Rosenzweig and Wolpin 1988). This leads to *simultaneity bias* in a production function, which could go in either direction. Two mechanisms for this has been suggested in the literature (a) contemporaneous inputs (and/or nutrition) and unobserved mental capacity might not be orthogonal if parents use nutrition or other inputs to reinforce differences in children's learning ability or 'endowment' (i.e., more gifted children will receive more inputs as a reward), and in this case, the estimated impact of height-for-age is biased upward when parents 'reinforce', but if they instead attempt to equalise learning outcomes ('compensate'), the bias is downward. However, even if that orthogonality condition were satisfied, OLS estimation would still be biased due to a second problem and (b) baseline skills, θ_{it-1}, and/or nutritional status, H_{it-1}, must be correlated with endowed mental capacity, μ_{i0}.

Finding a valid, strong instrument to deal with endogeneity due to (c1) and (c2) is challenging. Instead, the standard approach has been to follow a two-prong strategy, whereby as many household as child controls, together with IV, are jointly implemented (Glewwe et al. 2001; Alderman et al. 2006). The strategy through which the above issues are accounted for is described below in paragraphs (a) to (c).

(a) Controlling for family/community/school characteristics, lagged scores and lagged inputs

In order to solve the problem of endogeneity due to violation of (c1), a class of estimators used to 'control' for permanent unobservable factors makes use of variation across observations within which the unobservable factor is assumed to be fixed. One such 'fixed effect' estimator prominent in the literature uses variation that occurs within families/children. For that fixed effect to be performed, I would have required each child to have had two measurements on the same dependent variable to perform differencing of test scores, which unfortunately is not the case (PPVT is available at age 12 and Ravens at age 8). The inclusion of as many family and child controls as possible, added to past test scores and inputs, allows for a solution to this problem. I therefore standardise the test scores available for the Older Cohort children at two different ages, allowing this specification to be performed

20 For the sake of succinctness, I have not re-written conditions 1. and 2. for *I* or *IH* to reflect the endogeneity of parental inputs or health-related parental inputs.

in a modified version of the cumulative or valued-added (VA) specification, as in equation (7), where 'modified' stands for a VA specification that includes past test scores and past inputs and nutrition in the right-hand-side (RHS).[21,22]

Probably most importantly, quality of parental investment is unobserved (Waldfogel 2006), but some of the observed variables will hopefully capture such unobservables (e.g., breastfeeding, antenatal care, level of immunisation, etc.).

(b) Instrumental variables

To solve endogeneity due to condition (c2), I attempt firstly to use a lagged measure of nutritional status (HAZ) as a regressor;[23] and secondly, to instrument past nutritional status with birth-weight for the Younger Cohort. Birth-weight is a measure of innate endowments and, as Glewwe et al. (2001) argue, it should not be contaminated by parental investments on the basis of revealed innate ability, since ability is arguably only revealed some time after the birth itself. In that sense, this instrument deals with endogeneity due to condition (a), although it leaves (b) unresolved. Unfortunately birth-weight is unavailable for the Older Cohort, therefore, and on a similar basis, I use mothers' height, mothers' age and the interaction of the two for the Older Cohort (Thomas et al. 1990). See also Section 6.2 for alternative, but weaker, IVs used.

(c) Parental investment driven by observable differences between siblings

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 into account two of the most recognised possibilities, I will include birth order and gender (i.e., a child can be favoured because he is the first born or because of his gender).

Taking all elements together, the proposed strategy takes care of possible departures from conditions (c1) and (c2), particularly in the preferred specification where nutritional status and/or past scores are assumed a sufficient statistic for the history of observed and unobserved inputs. In brief, the regression technique employed for both cohorts will be IV-OLS and will have the following steps depending on the cohort being analysed:

- Younger: OLS estimation of equation (3)
- Older: OLS estimation of parental investment functions (4) and (5)
- Older: OLS-IV estimation of VA function (7)
- All: OLS-IV estimation of equations (6) and (7)

21 Ideally, I would like to be able to take first differences in tests and verify whether a child has advanced in terms of her skill levels over time (assuming a linear evolution of cognitive skills), as in Todd and Wolpin (2007). However, this is not possible in *absolute* terms with the available data, because I have two different test scores with different ranges. Therefore, I can only measure whether a child has moved up in the *relative* ranking.

22 In addition, a potential concern is that children may be at different points of the distribution of ability in these two dimensions, even at a given age, which might make it difficult to distinguish between a true increase in relative position vis-à-vis other children versus merely having measured a different aspect of ability. However, it is unlikely that this is a big problem given the existent evidence (Butler and Hakuta 2006) of high correlation between the two instruments. One paper shows a gender differential in this respect: Garrity and Donoghue (1976) show that girls obtained a significant correlation between the PPVT raw (or non-age-corrected) scores and the Ravens, whereas boys did not.

23 This could be itself endogenous in that it proxies for unobserved innate endowments. However, the IV estimation plus other robustness tests using data on birth-weight will solve this problem.

3 Data

3.1 General description

I use data from the Indian survey of the Young Lives project. Young Lives is an innovative longitudinal research project investigating the changing nature of childhood poverty. Young Lives is tracking two cohorts of 12,000 children in Ethiopia, India (Andhra Pradesh), Peru and Vietnam over 15 years through a quantitative survey and participatory qualitative research, linked to policy analysis.[24] At present I am able to use information from two rounds of data collection from Andhra Pradesh, India.[25]

In Round 1, 2,000 children aged around 1 (the Younger Cohort) and 1,000 children aged around 8 (the Older Cohort) were surveyed in 2002. Following up, Round 2 tracked the same children and surveyed them in 2006 at age 5 and 12, respectively. Between the two rounds, the attrition rate across the whole sample was only 0.9 per cent, which is very low for a study of this size (it has to be noted that the tracking effort was remarkable). In terms of the representativeness, despite a few biases (see Kumra 2008 in a technical note comparing the Young Lives survey to DHS), it is shown that the Young Lives sample in Andhra Pradesh covers the diversity of children in the country.[26] It is also caste representative.

The sampling process consists of the following four steps. First, six districts were selected based on the classification of poor/non-poor given by their relative levels of development. In the second stage, twenty sentinel sites within these districts were identified based on the same classification. Subsequently, one village was randomly selected from the five villages that comprised a sentinel site. The sample of children is representative of the three regions of Andhra Pradesh: Rayalaseema, Coastal Andhra and Telangana. Finally, the questionnaires were administered to a random selection of 100 households with a child born in 2001-02 (1 year old) and 50 households with a child born in 1994-95 (8 years old) per sentinel site in these villages. Data was collected through household questionnaires, child questionnaires and a community questionnaire. The estimation used in this study incorporates the Young Lives survey design by using regions as the stratification variable and the sentinel sites as the clustering variable.

24 Young Lives seeks to: (i) improve understanding of the causes and consequences of childhood poverty and to examine how policies affect children's well-being; and (ii) inform the development and implementation of future policies and practices that will reduce childhood poverty. Young Lives is a collaborative partnership between research and government institutions in the four study countries, the University of Oxford, the Open University, other UK universities and Save the Children UK.

25 Andhra Pradesh state is divided into 23 administrative districts, which are each subdivided into a number of mandals (in India, a sentinel site was defined as a mandal), dependent upon the size of the district. There are 1,125 mandals and around 27,000 villages in Andhra Pradesh. 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. Andhra Pradesh has three distinct agro-climatic regions: Coastal Andhra, Rayalaseema and Telangana. The sampling scheme adopted for Young Lives 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, with district poverty classification based on development ranking; 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 Andhra Pradesh District Poverty Initiative Programme (APDPIP).

26 However, Andhra Pradesh has more educated women and lower rates of malnutrition than the worst off states in the north.

Therefore, while not suitable for simple monitoring of child outcome indicators, the Young Lives sample will be an appropriate and valuable instrument for analysing causal relations, modelling child welfare and assessing longitudinal dynamics in Andhra Pradesh.

3.2 Description of key variables

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

Table 1. *Structure of key variables in Young Lives data*

	Cognitive skills test*	Nutritional status	Inputs: Parental (•) / School (◦)
Age 1			
(Younger Cohort, Round 1)	-	HAZ	• Freq. child sees father (if daily/weekly=1) • Whether birth in hospital or attended by trained person • Level of antenatal care (= 1 if medium/high,= 0 if none/low) • Whether child had a BCG, measles and polio immunisation • Whether mother was given iron folate tablets • Whether child had recommended 6 or + months breastfeeding
Age 5			
(Younger Cohort, Round 2)	Peabody Picture Vocabulary Cognitive Development Assessment	HAZ	• Freq. child sees father (if daily/weekly=1) • Whether father is dead • Whether child had complete immunisation ◦ Whether child attended pre-school ◦ Whether pre-school public/ private/religious or NGO
Age 8			
(Older Cohort, Round 1)	Raven's Coloured Progressive Matrices Reading level Writing level	HAZ	• Freq. child sees father (if daily/weekly=1) • Whether father is dead • Whether child works • Whether child performs household chores • Whether child reads as a hobby • Whether child does not have a hobby at all ◦ Whether child attended pre-school ◦ Whether pre-school public/ private/religious or NGO
Age 12			
(Older Cohort, Round 2)	Peabody Picture Vocabulary Reading level Writing level Numeracy level*	HAZ	• Freq. child sees father (if daily/weekly=1) • Whether father is dead • Whether child works • Hours child plays in a typical day (=1 if $\geq$2 hrs/day) • Whether child has complete immunisation ◦ Years of school ◦ Whether school is public or private

(*) I focus on the Ravens scores and the PPVT, while the writing, reading and CDA tests serve as robustness tests. Numeracy tests were not used due to the lack of variation and change in formulation of that question in Round 2. Note 1: HAZ stands for height-for-age z-score and is based on 2006 WHO standards. Note 2: Other measures used as robustness test of HAZ as proxy of *health* status for both cohorts in Round 1 are: (i) has child been sick in last two weeks (yes/no); (ii) how is health of child compared to others (same, better, worse); (iii) has the child had any long term health problem (yes/no); (iv) in the last three years, has the child had a serious illness or injury where you really thought they might die (yes/no). For the 1 year olds, I also built a child health index based on the sum of positive answers to the following episodes: loose or watery stools, blood, fever, cough, rapid breathing, vomit, serious loss of appetite, convulsions, unconsciousness and lethargy. Results using any of these are consistent with those that use HAZ. Note 3: complete immunisation means child had BCG, measles and polio immunisation.

3.2.1 Cognitive skills

I focus on two tests that measure different aspects of cognitive abilities: the Ravens test measures non-verbal reasoning while the Peabody Picture Vocabulary Test (PPVT hereafter) measures vocabulary knowledge.[27] See Appendix 1 for more details on these tests. A concern might be related to the effect of the language in which the tests were provided on the test scores. Actually, the questionnaires and the manuals for the field supervisors were translated into Telugu.[28] The translation of the questionnaire was an iterative process including translation, verification of relevance of each item, back-translation by an independent translator, and revisiting of details and specific wordings. In addition, an assistant who could speak Telugu, Urdu and English fluently was engaged in Hyderabad to assist with data collection from those households who could speak only Urdu. Similarly, local people helped out in some areas where some of the respondents were not fluent in Telugu (in the state border and tribal areas). Still, it is hard to assume that people responding to a vocabulary 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 (95 per cent of the 12 year olds in the Older Cohort and 90 per cent of the 5 year olds in the Younger Cohort).[29] Finally, the tests used in Young Lives to measure cognitive development are norm-referenced and therefore relative score interpretations are of primary interest (Cueto et al. 2009). For the rest of the analysis, I turn every outcome into a z-score by subtracting the mean and dividing by the standard deviation (SD). Standard scores allow for comparisons to be made across age groups and cohorts.

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, and by extension height-for-age, is also said to have a strong relationship with mental function and mortality.[30] Indeed, stunting, or low height-for-age, is a measure of chronic malnutrition, widely used in developing countries (Alderman 2000). HAZ is used as an input in the production function, however, for children aged 1, and given the lack of other outcome measures (i.e. cognitive outcomes), I only estimate a nutritional status production function as in equation (3) to see the relative importance of different factors.

---

27 I also use the Cognitive Development Assessment (CDA), a measure of quantitative ability, as a robustness test.

28 About 85 per cent of the Andhra Pradesh population identifies Telugu 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 Young Lives data, only 4.2 per cent of children speak a minority language.

29 Throughout the paper I 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. Moreover, it is interesting to note that the percentage of children speaking minority languages decreased significantly from Round 1 to Round 2, showing that there has been some assimilation.

30 Under the new standards, weight-for-height z-scores (indicating wasting) can only be computed up to the age of 60 months. Since a large proportion of children in the Younger Cohort, and all children in the Older Cohort, are above this age threshold, I do not report analysis of wasting here.

For the 8 year olds, I also use the Total Difficulties Score, a measure of mental health, based on responses to the 20 questions that comprise the Emotional Symptoms Score, Conduct Problems Score, Hyperactivity Score and the Peer Problems Score. Finally, for both the 8 and the 12 year olds, I also use as a robustness test the absolute value and z-score of the Body Mass Index (BMI).

3.2.3 Home input measures

Grantham-McGregor et al. (1991) find that both food suplementation and home stimulation were important, while Powell (2004) demonstrates how to intervene on stimulation for improved development. The findings show that parenting interventions may sometimes be more effective than health interventions in improving cognitive outcomes, underscoring the importance of assessing an index of parental home inputs. The data on the Younger Cohort provides information on mostly pre-school children,[31] while the data on the Older Cohort gives information on school-age children. And while peer interaction becomes essential for a school-age child, parental inputs remain vital. The changes in what matters for child development at different ages – whether this is breastfeeding, nutrition, child work, father's presence or school – is what I aim to capture with the home input measures.

I calculate the total home input score (or index) in two ways: firstly, the 'raw' index is an equal-weight summation of responses (modified so each has a [0,1] domain) of the individual items shown in Table 1 for a given cohort, age and round.[32] Secondly, the index is calculated as the output of factor analysis of the correlation matrix of all parental investment indicators, also shown in Table 1. Given the strong assumption associated with an equal-weight summation, the score presented in the rest of the paper is the factor score, whose calculation and tabulation is briefly described in Appendix 3.[33] Some of the items can be directly linked to cognitive skills in the sense that they are related to learning-specific skills. For example, primary caregivers are asked whether the child does household chores (which can be taken as ``time not-spent on learning''), or whether the child dad is alive and how often the child sees him (which can be taken as a proxy of how much fathers do care about their children). Other items are not so easily tied to cognitive development, but may be thought of as investments in child well-being and in creating an environment conducive to learning. For example, caregivers are asked about child's hobbies (such as reading, writing, playing with others, playing alone, etc), how many hours the child plays and other health-related questions (i.e. immunisation, breastfeeding, etc.).[34] Also, in the Older Cohort, paid work emerges as an important issue in Round 2, despite the legislation banning it. *A priori*, work could have an ambigous effect on cognitive skills, as it can serve as training but it could also pull away children from school. For this reason, its inclusion in the home input score is important.

---

31 However, it is worth noting that 87 per cent of the children were in pre-school by the age of 5. This will be taken into account in the analysis.

32 This index will clearly suffer from measurement error as this is not exactly a HOME (Home Observation Measurement of the Environment Score)-type measure like the one presented in the NLS-CS questionnaire.

33 Eigen values of the correlation matrix and the factor loadings are available upon request.

34 Cunha and Heckman (2008) argue that these types of index can be arbitrary and instead of creating an index of parental inputs, they estimate an index that best predicts latent skill dynamics. See Helmers and Patnam (2009) for an application to the case of India.

Lastly, for those 5, 8 and 12 year olds, I also use school indicators, which could be interpreted as home and/or school input. For instance, if children go to (better) private schools, this could be taken as a higher 'home input', as parents made the decision of the type of school they wished their child to attend.

3.2.4 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 characteritics, such as the wealth index (Filmer and Pritchett 1998) and household size, are also included. The wealth index has three components: housing quality, consumer durables and services. In the calculation of these variables, if any of the component variables are missing, then the resulting variable is treated as missing.[35] Geographical dummies included are: Coastal Andhra and Rayalaseema (with Telangana being the base category); and whether the household is located in an urban or rural area. I turn now to some background on caste disparities in India.

4 Background: Caste disparities in India

In developed countries like the USA, early disadvantages of black minorities in test scores is a well-documented phenomenon (Todd and Wolpin 2007). Similarly, in India, and despite the lack of data on test scores, it is widely believed that people belonging to LCs and non-Hindu (typically Muslim) faith groups are economically deprived, which clearly contributes to poorer educational performance.[36] Still, none of the existing studies has investigated the effect of the conditions affecting these social groups early in their lives on tests scores, which is important from a policy point view.[37]

Most of the existing studies on determinants of school participation and attainment in India today acknowledge socio-religious differences in the population and document the profile of educational achievement by caste, religion and gender (Dreze and Kingdon 2001; Kingdon 2002).[38] More recent evidence of social disparity in educational outcomes uses large

35 The Housing Quality Index is based on the number of rooms per person in the household and the main materials used for the walls, roof and floor. The Consumer Durable Index is based on the number of assets owned by the household. A typical set of 11 assets is considered: radio, refrigerator, bicycle, television, motorbike/scooter, car, mobile phone, landline telephone, fan, almairah (wardrobe) and clock. Productive assets (e.g., sewing machines) are not included in this calculation. For each asset owned by the household, a 1 is added to the index; the result is then divided by 11 to give a value between 0 and 1. The Services Index is based on whether or not the dwelling has electricity, the source of drinking water, type of toilet facility and the main type of fuel used for cooking. If the dwelling has electricity then 1 is added to the index. If drinking water is piped into the dwelling or the yard then 1 is added. If the household have their own toilet facility (not shared with other households) then 1 is added and if paraffin, kerosene, gas or electricity is used for cooking another 1 is added. The result is then divided by 4 to give a value between 0 and 1.

36 See Appendix 2 for a brief explanation of the caste system in India.

37 Helmers and Patnam (2009) do investigate early childhood development in India, but they do not focus on caste differentials.

38 Using the Public Report on Basic Education (PROBE) survey, Dreze and Kingdon (2001) find that educational disadvantages are much the same among Backward Classes (BC), SC and ST children, though somewhat larger for the latter category in the case of initial enrolment. Meanwhile, Kingdon (2002) and Dostie and Jayaraman (2006) report some evidence of caste disadvantage in schooling, even after controlling for family background and personal attributes.

nationally representative datasets and finds that children born into SC/ST households achieve much less than those from upper caste Hindu families (Borooah and Iyer 2005; Desai and Darden 2006; Dostie and Jayaraman 2006; Bhalotra and Zamora 2007; Rajaram and Jayachandran 2007; Asadullah et al. 2009). On the other hand, one study that uses the 1992/3 and 1998/9 rounds of the National Family Health Survey (NFHS) indicate that ST children are less likely to complete school but, in contrast to the attendance results, children from SC and from Muslim families are not less likely to complete (Balhotra and Zamora 2007).

These educational gaps are also mirrored by economic disparity. Recently, researchers have also looked at wage earnings disparities by religion and caste (Das 2003; Reilly and Dutta 2005; Dutta 2006; Bhaumik and Chakrabarty 2007). Almost all of these studies use various rounds of the National Sample Survey (NSS) data and report lower wages for Muslims and SC/STs. Such wage gaps are not surprising if there are differences in educational endowment.[39] Moreover, given that returns to education in India rise with levels of education (Dutta 2006; Kingdon and Unni 2001), any caste educational gap will translate into caste gaps in labour market earnings. For example, Gang, Sen and Yun (2008) find that differences in educational attainment explain about 25 per cent of the poverty gap between both the Scheduled Caste and Schedule Tribe and non-Scheduled-Hindu households in India. Moreover, in post-liberalisation years, returns to schooling has increased significantly, mainly in urban areas (Kijima 2006), thereby increasing the risk of rising economic disparity between castes. It is interesting to note that, according to the PROBE survey, SCs/STs have been able to catch up with Muslims and, partly, with non-SC/ST Hindus. This may be due to the targeting of SC/ST households by special programmes that establish schools or improve infrastructure and provide incentives for enrolment. Job reservation for LCs, too, may have had an indirect effect by providing the economic means to educate children and simultaneously increase the economic returns to education, even if there is some evidence of discrimination against LCs (Sidiqqui 2009).

In terms of potential explanations, one may observe inequality in educational outcomes for at least two reasons: firstly, because of differences in supply and demand-side related factors across caste groups; or secondly, because of differences in norms and practices.

On the demand-side, LC children may be more disadvantaged in terms of family factors, such as poverty, labour demands on children, or lack of parental motivation or inputs (the focus of this paper). LC parents have on average lower levels of education than upper caste Hindus. LCs also have limited ownership of land in rural areas and hence are likely to be poor. They are also primarily engaged in manual work or work regarded as demeaning (Das 2003), where demand for child labour may be high. For this reason, their (perceived) returns from education may be low, which in turn may lead to a reduced demand for education. Indeed, research on the intra-household allocation of education expenditure in India yields evidence of lower budgetary allocation of household resources to education amongst the Scheduled population (Scheduled Castes and Scheduled Tribes together) than the non-Scheduled population (Tilak 2002; Kingdon 2005).[40] A study using 1993-94 data showed that higher levels of village development and parental education resulted in higher enrolment rates for all communities. Interestingly, once the children are placed in ‘more

39 For instance, disparities between men and women in skills (i.e., girls lag behind boys at school) and labour market outcomes are well-documented (Kingdon 2002).

40 Equally, SC/STs are highly unlikely to be in non-farm self-employment and are the least likely builders of minority enclaves. Overall, SC/STs are overwhelmingly casual labourers (Das 2003).

favourable' circumstances (i.e., when parents, especially mothers, are literate and infrastructural facilities are better), inter-community (Hindus/SC–ST/Muslims) differences in enrolment rates become insignificant (Sachar 2006).[41]

Also, when ethnic-religious differences manifest in terms of linguistic differences, demand and supply-side factors might interact to leave children of certain castes disadvantaged. Many ST children and households prefer their mother tongue as a medium of instruction for education,[42] but few (pre-)schools accommodate this. Consequently, they often select for their children to remain out of school, which means they subsequently become disadvantaged in terms of learning outcomes.

There is still a general lack of descriptive research on the extent and nature of caste gaps in test scores in India, and no studies on pre-school aged children, mainly due to the lack of data on cognitive skills at different points in time. Apart from Balhotra and Zamora (2007) and Assadullah et al. (2009), none of the existent studies emphasise the importance of one's caste membership as a determinant of educational attainment and cognitive development, despite well-documented differences in outcomes by caste and the policy relevance of such research. The discussion presented in this paper intends to give insights on the mechanisms behind caste test score gaps, mainly focusing on differences in nutrition status and on how parents from different castes 'invest' in their children. Therefore, this new knowledge will enable us to pinpoint policy implications for tackling not only caste disparities in cognitive skills but also childhood poverty in India.

5 Descriptive statistics

The means and SDs of all relevant variables are presented in Tables 2 and 3 below. 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)[43] and the fourth column for UCs, or UC (defined as the 'other castes' category in the data). The last two columns show p-values for the difference of means between LCs and BCs, and UCs and LCs, respectively. The latter is the most relevant comparison for this paper.

41 Less relevant for this paper are the supply side factors. For instance, minority groups may live in remote locations where access to schools is poor. Alternatively, areas with a lower caste population may differ in terms of ethnic mix, which has implications for the supply of schools, either because social or ethnic heterogeneity may adversely affect access to local public goods by inhibiting social groups from working collectively in order to extract public goods from the state (Alesina et al. 1999);.or else because the allocation of state funds for the provision of public goods may reflect the relative power of specific ethnic groups. Indeed, recent analysis of data on the distribution of schools in India indicates that areas with SC concentrations gain in access to several facilities (e.g., high schools, health centres, piped water), while those with STs and Muslims remain disadvantaged (Banerjee et al. 2005; Banerjee and Somanathan 2006).

42 Sixty-one per cent of the Younger Cohort's ST children and 37 per cent of the Older Cohort's ST children report a local dialect as being their first language.

43 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. I have therefore separated out this group and explicitly controlled for BC membership in the results section. I have also further split the lower caste group between SC and ST.

Table 2. *Younger Cohort: Means and t-tests, main variables by caste*

	All	LC	BC	UC	t test (p-val) LC-BC	t test (p-val) UC-LC
Cognitive scores						
PPVT	42.03	41.11	38.52	51.48	0.11	0.00
CDA Q	9.52	9.23	9.43	10.13	0.15	0.00
PPVT (z-sc)	0.00	-0.02	-0.10	0.30	0.11	0.00
CDA Q (z-sc)	0.00	-0.11	-0.03	0.23	0.15	0.00
Child						
% PPVT in Telugu	0.90	0.89	0.96	0.77	0.00	0.00
Coastal Andhra	0.35	0.35	0.36	0.32	0.67	0.37
Rayalaseema	0.30	0.28	0.28	0.39	0.87	0.00
Telangana	0.35	0.37	0.36	0.29	0.78	0.01
ST	0.18	0.59				
SC	0.13	0.41				
BC	0.48					
UC	0.21					
R2 pre-school	0.87	0.82	0.89	0.88	0.00	0.02
R2 private	0.28	0.13	0.27	0.51	0.00	0.00
Child nutrition						
R1 HAZ	-1.29	-1.52	-1.30	-0.93	0.01	0.00
R1 stunted	0.27	0.36	0.26	0.16	0.00	0.00
Birth-weight	2,766	2,691	2,750	2,847	0.22	0.00
R2 HAZ	-1.61	-1.75	-1.66	-1.29	0.11	0.00
R2 stunted	0.34	0.39	0.36	0.22	0.36	0.00
Home						
R2 urban	0.35	0.37	0.28	0.45	0.00	0.01
R2 hh size	5.52	5.29	5.60	5.67	0.01	0.00
R2 wealth index	0.36	0.28	0.36	0.48	0.00	0.00
CG and DAD						
R1 CG age	23.76	23.57	23.56	24.52	0.98	0.00
R1 CG edu	3.25	1.90	2.89	6.06	0.00	0.00
R1 DAD edu	5.34	4.95	5.24	6.12	0.00	0.00
R1 CG depress	0.30	0.41	0.26	0.23	0.00	0.00
R2 CG edu	3.50	2.07	3.14	6.41	0.00	0.00
R2 DAD edu	5.37	4.21	5.00	7.91	0.00	0.00
Parental investment						
R1 antecare index	1.94	1.70	1.99	2.18	0.00	0.00
R1 birth hospital	0.59	0.43	0.62	0.77	0.00	0.00
R1 leftnobreast	0.30	0.15	0.31	0.50	0.01	0.00
R1 seedad	0.98	0.97	0.98	0.98	0.25	0.73
R1 deaddad	0.00	0.00	0.00	0.00	0.66	0.81
R1 complete imm	0.72	0.73	0.69	0.78	0.15	0.06
R1 irontabs	0.95	0.92	0.95	0.97	0.01	0.00
R2 seedad	0.94	0.94	0.95	0.93	0.62	0.38
R2 deaddad	0.02	0.02	0.02	0.02	0.58	0.74
R2 complete imm	0.93	0.91	0.95	0.94	0.01	0.04

Source: Young Lives India data, Younger Cohort. Prefix R1 (R2) means the value comes from Round 1 (Round 2). Means are taken on children observed on both rounds (N=1950). LCs are: Scheduled Caste (SC) and Scheduled Tribes (ST). Other Backward Classes (BC) include Muslims, while UCs are those classified in the Young Lives data as Other Castes. HAZ=height-for-age z-score and CG=caregiver. Mothers' height is not available for the Younger Cohort and therefore not reported in this table.

Table 3. *Older Cohort: Means and t-tests, main variables by caste*

	All	LC	BC	UC	t test (p-val)	
					LC-BC	UC-LC
Cognitive scores						
PPVT	135.42	130.83	134.06	145.37	0.28	0.00
Ravens	22.98	22.20	22.77	24.59	0.14	0.00
PPVT (z-sc)	0.00	-0.11	-0.03	0.25	0.28	0.00
Ravens (z-sc)	0.00	-0.15	-0.04	0.31	0.14	0.00
R1 Writing	2.34	2.28	2.28	2.55	0.99	0.00
R1 Reading	3.08	2.87	3.06	3.43	0.01	0.00
Child						
% PPVT in Telugu	0.95	0.99	0.97	0.86	0.08	0.00
Coastal Andhra	0.34	0.36	0.37	0.26	0.75	0.03
Rayalaseema	0.31	0.27	0.27	0.43	0.99	0.00
Telangana	0.35	0.37	0.36	0.31	0.76	0.16
ST	0.20	0.67				
SC	0.10	0.33				
BC	0.49					
UC	0.21					
R1 years in school	2.46	2.40	2.46	2.56	0.34	0.05
R2 years in school	6.52	6.32	6.57	6.69	0.01	0.00
R2 public	0.63	0.71	0.64	0.47	0.04	0.00
Child nutrition						
R1 HAZ	-1.52	-1.61	-1.55	-1.34	0.36	0.00
R1 TDS	0.57	0.58	0.63	0.40	0.38	0.01
Home						
R2 urban	0.34	0.36	0.30	0.43	0.07	0.12
R2 hhsize	5.20	5.21	5.16	5.27	0.67	0.74
R2 wealth index	0.37	0.29	0.37	0.47	0.00	0.00
CG and DAD						
R1 CG age	31.04	31.35	30.73	31.33	0.19	0.96
R1 CG edu	2.34	1.19	2.07	4.65	0.00	0.00
R1 DAD edu	5.08	4.44	5.01	6.19	0.00	0.00
R2 CG edu	2.61	1.31	2.38	5.06	0.00	0.00
R2 DAD edu	4.50	2.88	4.19	7.60	0.00	0.00
Mother's height	150.3	149.7	150.2	151.5	0.55	0.04
Parental investment						
R1 chldwork	0.06	0.10	0.05	0.03	0.02	0.00
R1 seedad	0.94	0.91	0.96	0.93	0.01	0.33
R1 deaddad	0.03	0.04	0.02	0.04	0.07	0.87
R1 child HH chores	0.33	0.45	0.29	0.26	0.00	0.00
R1 reads as hobby	0.15	0.11	0.16	0.18	0.03	0.03
R1 no hobby	0.03	0.07	0.02	0.00	0.00	0.00
R2 deaddad	0.08	0.10	0.07	0.09	0.09	0.65
R2 chldwork	0.21	0.25	0.22	0.09	0.33	0.00
R2 immunisation	0.39	0.38	0.35	0.51	0.33	0.00
R2 hours play	3.78	4.04	3.74	3.47	0.06	0.01

Source: Young Lives India data, Older Cohort. See notes for Table 2. Means are taken on children observed in both rounds (N=994).

Overall, and based on the analysis of these two tables of descriptive statistics and the existing literature, I find that Andhra Pradesh has achieved progress on children development indicators since the mid-1990s (Galab et al. 2008). However, even though LCs and BCs have become wealthier and increasingly urban in both the Younger and the Older Cohort, significant differences remain based on sector (rural versus urban), caste and region.

Disparities between castes are particularly important in cognitive tests, nutrition outcomes, wealth, caregivers' and fathers' level of education and parental inputs. In general, I find a significant advantage amongst UCs. For the sake of succinctness, I only report results for the Younger Cohort from Table 2, unless otherwise stated (that is, if trends are dissimilar between cohorts, I mention this explicitly). Also, results for Round 2, are mostly reported between parenthesis.

In both the PPVT and CDA-Q tests, there is no significant difference betwen BCs and SC/STs, while UCs do significantly better (25 per cent in PPVT and 7.4 per cent in CDA-Q) than LCs and BCs. The level of stuntedness increased over time for both the UCs and the LCs, but remained about 20 percentage points lower for UCs (at 16 per cent when they were 1, and at at 22 per cent four years later).

In terms of residence, more UCs than LCs live in Rayalaseema and less UC than LC live in Telangana. About 50 per cent of UCs are urban as opposed to about 25 per cent of backward classes. In both cohorts, there is a remarkable increase in urbanisation rates among LCs, who went from about 10 per cent to around 40 per cent. Households are larger in UCs and in BCs; while LCs are poorer than BCs, who themselves are poorer than UCs (as per the wealth index). Over time, all castes and cohorts are becoming richer, but as inequality increases, UCs end up benefiting more from growth.

UC mothers are significantly older than LC (on average, by one year) and have two and three times more years of schooling than BCs and LCs, respectively, averaging a total of 6.41 years of education completed. For fathers, the differences, though significant, are not so ample: UC fathers have about one to three years more education than BC fathers, and two to four years more than LC.

Caregiver's depression (not available for the Older Cohort) is a particularly strong psycho-social well-being risk factor, and again differences between castes are large and significant: 23 per cent of UC, 26 per cent of BC and 41 per cent of LC caregivers report being depressed as per the Self-responding Questionnaire.[44]

In terms of the inputs variables, from the LCs, only 43 per cent of children were born in a hospital or with a medically trained person, as oposed to 77 per cent of children from UCs and 62 per cent of BCs. Besides, 97 per cent of mothers of UCs and 95 per cent of those of BCs had been given iron folate tablets/syrup during the antenatal visits, against 92 per cent of LC mothers. The level of antenatal care is also better for UCs and BCs, while LCs are more in a low-medium level of care according to the index of antenatal care (ranging from 0=none to 3=high). There is almost no difference between castes in the percentage of children who see their fathers daily or weekly, while the difference in months left without

44 Young Lives uses the Self-responding Questionnaire (SRQ20) to determine and measure the psycho-social well-being of the caregiver. They count the number of 'Yes' responses to the 20 relevant questions (i.e., headache, poor appetite, bad sleep, etc).

breasfeeding is significant.[45] Lastly, differences in timely immunisation are significant only between LCs (73 per cent) and UCs (78 per cent). By age 5, these rates go up to 95 and 91 per cent, respectively.

The difference in the type of pre-school children attend is striking: 51 per cent of UCs are going to private and 10 per cent to NGO-run schools, while these figures fall to 13 and 3 per cent, respectively, for LCs.[46]

In Table 3 one can observe that the UC 8 year olds have 11 per cent higher Ravens scores than LC, while 12-year-olds belonging to an UC score also 11 per cent higher in the PPVT.[47] Disparities are also significant in nutrition outcomes at age 8: 32 per cent of LCs and 35 per cent of BCs are stunted, as opposed to 25 per cent of UCs (there is actually an increasing gap in stunting between LC and UC children in Round 2).[48] Another important predictor of children's success is parental health and nutrition, and a good proxy at hand in Young Lives data is mother's/caregiver's height: as expected, LC mothers are 2 cm shorter than UC mothers.

About a fifth of the 12-year-old children in the sample reportedly engage in paid work and most of them are actually in rural areas. Breaking down these figures also reveals some clear patterns: child labour is significantly higher among the LCs and lower among the UCs. Differences within the LCs are important: STs present a rate of 30 per cent of working children in Round 2, which is 9 percentage points higher than the SCs' rate. It is clear that as time passes, figures go up and differences widen. The household chores figures also show important caste gaps.

As with the Younger Cohort, the largest difference found is in the type of school different castes attend: 71 per cent of LCs go to public schools, compared to 64 per cent of BCs and 47 per cent of UCs.

Now turning to the more cognitive-enhancing activities, it is shown that 18 and 16 per cent of UC and BC children, respectively, report reading as a hobby, while only 11 per cent of LC children report such a hobby. Meanwhile, at age 12, UCs play less hours/day than LCs.

Lastly, in terms of health care usage, 51 per cent of UC children had received complete immunisation by age 8, while that figure is only 38 per cent for LCs and 35 per cent for BCs.[49]

Overall, I observe important differences in caregivers, home and parental investments by caste, suggesting that these could be behind the disparities found in cognitive and nutrition outcomes. I now turn to the exploration of these hypotheses by means of differents types of specifications, as discussed in the Methodological Section.

45 When measured from the benchmark of six mandatory months of exclusive breastfeeding recommended, I find that LC children are left only ten days without breastfeeding, while for the BC and UC, these figures are 20 and 30 days, respectively.

46 Eighty-eight per cent of UCs and 82 per cent of LCs go to pre-school.

47 Bhalotra and Zamora (2007) also find a 10 percentage points advantage of upper castes in relation to lower castes for school attendance and completion in a similar age group.

48 Just to mention other health-related measures: the mental health index ('total difficulties scores') at age 8 presents 58 per cent of LC children as 'normal', and 69 per cent of BC and 40 per cent of UC as such. Also, while 49 per cent of caregivers from the UCs report their children's health status as being better than that of others, only 36 per cent of LCs report this to be the case.

49 In these descriptive tables, I do not report differences in transportation time to health facilities by caste. However, it is worth noting that in the Younger Cohort, on average, an LC family will take about 35 to 100 minutes to reach the health facility, while a BC family will take 25 and an UC 21 minutes.

6 Results

6.1 Younger Cohort: The cumulative nutrition production function at age 1 and caste effects

A better environment and child health *per se* are known to be associated with skill accumulation (Grantham-McGregor et al. 2007). However, nutritional status is also, at least partially, the outcome of received home inputs and household characteristics. Therefore, prior to studying the nutritional status–cognitive skills nexus, I estimate the cumulative version of equation (3). Another relevant question will be to assess whether the inclusion of the different castes dummies still influences early child nutrition after controlling for a rich set of co-variates.

Table 4. *The production function of nutrition, cumulative specification for 1 year olds*

	OLS, dependent variable: HAZ	
	(1)	**(2)**
Age of child (months)	-0.10***	-0.10***
	(0.01)	(0.01)
male dummy	-0.13**	-0.13**
	(0.06)	(0.06)
firstborn	0.74	0.74
	(0.76)	(0.76)
lastborn	-0.04	-0.03
	(0.07)	(0.07)
Coastal Andhra	0.04	0.05
	(0.08)	(0.08)
Rayalaseema	0.79***	0.78***
	(0.08)	(0.08)
Urban dummy	-0.07	-0.08
	(0.10)	(0.10)
Household size	-0.02*	-0.03*
	(0.01)	(0.01)
Wealth index	0.85***	0.75***
	(0.23)	(0.23)
Age of caregiver	0.02**	0.02**
	(0.01)	(0.01)
Caregiver's education level	0.02**	0.02**
	(0.01)	(0.01)
Father's education level	0.02	0.02
	(0.02)	(0.02)
Home Input 1 yr old (factors-scored)	0.09**	0.08*
	(0.04)	(0.04)
Scheduled Caste		-0.09
		(0.09)
Scheduled Tribe		-0.11
		(0.11)
Upper Caste		0.09
		(0.09)
Constant	-1.05***	-0.97***
	(0.25)	(0.25)
Observations	1769	1769
R-squared	0.15	0.15
F-Test: caregiver characteristics & parental inputs	4.60	3.99
Prob > F	0.00	0.00
F-Test: caregiver characteristics & parental inputs & caste		3.14
Prob > F		0.00
Chi-square	24.40	20.05
Prob > F	0.00	0.00

Note 1: Standard errors in parentheses. Note 2: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Note 3: Home Input 1 yr old is an index (based on factor analysis) of parental investment as described in the text. Note 4: Telangana Region is the base category.

Table 4 shows that age of child in months is negatively correlated with height-for-age z scores, probably capturing some unobserved factor/s that worsen children's health as they age.[50] Girls appear healthier (i.e., better nourished) than boys, a finding along the lines of a well-known fact in the demographic literature on male-to-female ratios in child mortality over the first year of life.[51] Children living in Rayalaseema seem to be healthier than children living in Telangana, which is not a surpise as this is one of the state's most impoverished areas after suffering from agrarian crisis for several consecutive years. Not surprisingly, child nutrition is negatively correlated with household size, while it is positively correlated with wealth. However, birth order and urban/rural status do not have any significant effects on HAZ.

Caregiver characteristics are significant at the 95 per cent level, with education and age of the caregiver having the same effects. For each additional year of caregiver's education, HAZ scores are on average 2 per cent of 1 SD higher. On the other hand, the father's level of education does not have any influence on the child's production function of nutrition. Lastly, for each 'point' increase in the home input score, HAZ scores are on average 9 per cent of 1 SD higher, a larger effect than the education of the caregiver. In the second column, after the inclusion of the lower (negative and non-significant) and upper (positive and non-significant) castes dummies, all coefficients stay significant and similar in magnitude, except for the wealth index. This result is consistent with the fact that the covariates included capture most of the differences between castes.[52]

6.2 Younger Cohort: The cumulative production function of skills at age 5, nutrition and home inputs effects

Different versions of equation (7) are estimated below: the OLS and the IV version are shown in Tables 5 and 6, respectively. In Table 4, column (1) shows the baseline specifications at which age, region and the urban and wealth index dummies are significant. The level of education of the caregiver and father; and whether the child is ST, also have significant effects. Of most interest is the coefficient on past height-for-age on the PPVT, which implies that raising HAZ by 1 SD will increase child test score by 11 per cent of 1 SD, a bigger effect than any of the parental background effects. In column (2), after the inclusion of the (non-significant) pre-school dummy, all variables remain unaltered, suggesting that pre-school attendance is not a particularly important determinant of language abilities. (It has to be noted that 87 percent of children attended pre-school at age 5 and 44 per cent also claim to be already enrolled in primary school, despite being under the official starting age. It is likely that the free midday meal scheme explains this). In column (3), I also include the current home input index, which, even if positive, turns out to be insignificant. Home inputs at age 1 (in the factors scored version but also each individual variable of the index) were included in an older version of the paper, but are not reported after finding that past nutrition is a better

50 Children moving from the six months exclusive breastfeeding to other types of food could be one such a factor, given the lack of pure water and sanitation services in some communities.

51 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. 2007). 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. However, this is not inconsistent with evidence showing that male-to-female ratio of *born children* is low in India (i.e., fewer girls than boys are born). The literature speculates that this is due to the pre-natal pro-male discrimination followed by selective abortion of female foetuses (Jha et al. 2006).

52 The observations that are lost on the way from the original 1,950 to 1,769 (181 observations) are made up of missing observations from PPVT scores on non-Telugu speakers, father's education, HAZ, birth order and the home index at age 1, in which the antenatal care indicator has most of the missing observations. As will be shown later, selection is not an issue as results remain the same when non-Telugu speakers are included, as well as when imputing missing values (not reported).

indicator of past inputs than the lagged home input index, or any of its individual components.[53]

In column (4), the significant coefficient on current nutrition suggests that PPVT scores depend not only on lagged but also contemporaneous nutrition, providing evidence against the contemporaneous specification. Lastly, column (5) replicates column (4), but now the dependent variable is the CDA. Results are qualitatively similar, except for change of sign (from positive to negative) in the coefficient on Coastal Andhra and the importance of pre-school attendance.[54]

In terms of the performance of different castes, the effect of belonging to an ST household on the PPVT remains consistently positive and significant throughout, except for the CDA for 5 year olds and (as will be shown in the next sub-section) the Ravens, suggesting that there is some feature of the PPVT at which STs fare significantly better than BCs, even after netting out differences in child, caregiver, home and parental characteristics. On the other hand, in the CDA specification, SCs are performing significantly worse, as expected.[55]

Table 5. *The production function of skills, cumulative specification for 5 year olds (OLS)*

	Peabody Picture Vocabulary Test z scores				**CDA test z scores**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Age of child (months)	0.04***	0.04***	0.04***	0.04***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
male dummy	0.05	0.05	0.05	0.05	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
firstborn	0.19	0.19	0.20	0.23	0.44
	(0.51)	(0.51)	(0.51)	(0.51)	(0.53)
lastborn	0.00	0.00	0.00	0.01	-0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Coastal Andhra	0.14**	0.14**	0.14**	0.13**	-0.24***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
Rayalaseema	0.08	0.08	0.08	0.09	-0.03
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Urban	0.38***	0.38***	0.38***	0.38***	0.26***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Household size	-0.01	-0.01	-0.01	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Wealth index	0.36**	0.36**	0.35**	0.29*	-0.09
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Caregiver's education	0.04***	0.04***	0.04***	0.04***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Father's education	0.02***	0.02***	0.02***	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

53 Moreover, including past inputs in the specification in column (3) – with or without past health – shows that past home inputs matter in the production function of cognitive skills, which allows for the rejection of the contemporaneous specification. However, when past health is included along with the past home inputs index, these are not jointly significant and the coefficient on past health is larger.

54 The 338 observations that are lost from the original 1,950 to 1,612 are made up of the following: first, I restrict the sample to those whose first language is Telugu, while the rest are missing observations from PPVT, fathers education, HAZ and birth order. As will be shown later, selection is not an issue as results remain the same when including non-Telugu speakers as well as when imputing missing values (not reported, but available upon request).

55 To test the relative importance of overall family background with respect to health and home inputs, I carried out an F-test for joint significance. The size of the F-statistics in column (4) are smaller than those in columns (1) to (3). This is suggestive of the possibility that family background and health (or inputs) decrease in importance in relation to cognitive skills, once current health is included.

	Peabody Picture Vocabulary Test z scores				CDA test z scores
	(1)	(2)	(3)	(4)	(5)
past HAZ	0.11***	0.11***	0.11***	0.09***	0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Scheduled Caste	-0.04	-0.04	-0.04	-0.05	-0.11*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Scheduled Tribe	0.51***	0.51***	0.50***	0.49***	0.06
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)
Upper Castes	0.02	0.02	0.02	0.01	0.00
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Pre-school dummy		-0.03	-0.03	-0.03	0.16**
		(0.07)	(0.07)	(0.07)	(0.06)
Home inputs score – 5yr olds			0.09	0.09	0.08
			(0.09)	(0.09)	(0.09)
Current HAZ				0.07***	0.10***
				(0.03)	(0.03)
Constant	-3.11***	-3.08***	-3.00***	-2.81***	-2.94***
	(0.38)	(0.38)	(0.39)	(0.40)	(0.39)
Observations	1612	1612	1612	1612	1841
R-squared	0.28	0.28	0.28	0.28	0.16
Test join signif:parental background & nutrition	34.35	34.37	34.47	30.69	19.04
Prob > F	0.00	0.00	0.00	0.00	0.00
Test join signif:parental background & inputs	-	-	36.10	34.63	18.88
Prob > F	-	-	0.00	0.00	0.00

Note 1: Standard errors in parentheses. Note 2: *** p < 0.01, ** p < 0.05, * p < 0.1 Note 3: The Home Input score index is composed by the variables of care as described in the text. Note 4: Columns (3) and (4) are the Cumulative specification.

Table 6 replicates the specifications in columns (3) and (4) from Table 5 for a truncated sample for the PPVT and the CDA, respectively. Here past (or current) nutrition is instrumented with birth-weight in columns (3), (4), (7) and (8).[56]

In sum, results show that OLS estimates in columns (1) and (5) understimate the IV estimate effect of past HAZ on PPVT and CDA z-scores shown in columns (3) and (7): for each SD increase in past HAZ, PPVT scores are on average 24 per cent of 1 SD higher, while CDA scores are 46 per cent of 1 SD higher. OLS in columns (2) and (6) also understimates the effects of current nutrition, but only in the case of CDA scores as shown in columns (4) and (8). The1st stage (Cragg-Donald) F-statistics are over 35 for both cognitive tests in the case of specifications with past nutrition, and over 10 in the specifications that include current nutrition instead. These F-statistics seem to show that birth-weight works via past nutrition better than via current nutrition, as one would have expected. Overall, this downward bias of the OLS estimate of HAZ (i.e., that the OLS estimate is smaller than the IV) seems to indicate that nutrition is more important than what one might think *a priori*. Actually parents, after observing low birth-weight, might be attempting to compensate by adjusting their inputs upwards and as a result equalising learning outcomes with those of healthier children. This

56 There is information on birth-weight for about 50 per cent of the sample. Given this loss of observations, I have tried other IVs, namely, the parental inputs variables used to calculate the home input score at age 1. Results showed that OLS has a downward bias at estimating the coefficient on HAZ which is qualitatively (and quantitatively) close to the results that use birth-weight as an IV. Moreover, first-stage F-tests also showed that these IVs were strong and valid instruments.

will push down the coefficient on HAZ. This parental behaviour is consistent with evidence from the Philipines by Liu et al. (2009).[57]

The implication of these results is important: the effect of past HAZ on PPVT and CDA scores is independent of the source of variation of HAZ, whether it is due to the variation in birth-weight, home, household, region, home inputs or caregivers' characteristics, including caste. Given the consistency of the effect of past HAZ in Table 5 and the OLS estimations in Table 6 (0.10 on PPVT and 0.06 on CDA), and the IV results in Table 6 (0.24 on PPVT and 0.46 on CDA), one can conclude that the nutrition-cognition nexus is important in Andhra Pradesh, while home inputs do less to mediate caste disadvantages, as will be shown with simple simulations in the last sub-section.

Table 6. *The production function of skills, cumulative specification for 5 year olds (instrumenting HAZ with birth weight)*

	Peabody Picture Vocabulary Test z scores				**CDA test z score**			
	OLS		**IV**		**OLS**		**IV**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Caregiver's education level	0.05***	0.05***	0.05***	0.05***	0.03***	0.03***	0.03***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Father's education level	0.02*	0.02*	0.02	0.01	0.02**	0.02**	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Past HAZ	0.15***		0.24*		0.11***		0.46***	
	(0.03)		(0.14)		(0.03)		(0.14)	
Current HAZ		0.19***		0.54		0.14***		0.34***
		(0.04)		(0.34)		(0.04)		(0.14)
Home inputs 5yr (factors-sc)	0.18	0.16	0.17	0.14	0.16	0.15	0.15	0.10
	(0.15)	(0.15)	(0.15)	(0.16)	(0.14)	(0.14)	(0.15)	(0.18)
Scheduled Caste	-0.03	-0.09	0.00	-0.10	-0.06	-0.10	0.07	-0.11
	(0.11)	(0.11)	(0.12)	(0.12)	(0.11)	(0.11)	(0.13)	(0.14)
Scheduled Tribe	0.58**	0.53**	0.60*	0.53*	0.11	0.08	0.21	0.11
	(0.15)	(0.15)	(0.16)	(0.16)	(0.14)	(0.14)	(0.16)	(0.18)
Upper Castes	0.11	0.10	0.11	0.08	-0.03	-0.04	-0.02	-0.11
	(0.09)	(0.09)	(0.09)	(0.10)	(0.08)	(0.08)	(0.09)	(0.11)
Constant	-4.34***	-3.96***	-4.44***	-3.55***	-3.64***	-3.34***	-3.97***	-2.23**
	(0.64)	(0.64)	(0.67)	(0.79)	(0.58)	(0.58)	(0.65)	(0.89)
Observations	713	713	713	713	795	795	795	795
R-squared	0.31	0.31	0.30	0.23	0.18	0.18	0.02	0.02
CD Wald F-stat	-	-	33.37	10.64	-	-	37.61	14.12

Note 1: Standard errors in parentheses. Note 2: *** p < 0.01, ** p < 0.05, * p < 0.1 Note 3: The Home Input score index is composed of the variables of care described in the text. Note 4: The C-D Wald F stat is the Cragg-Donald Wald F-statistic of weak identification. Note 5: Child, region, pre-school and home controls are always included.

57 Another important fact is that father's education becomes non-significant in all IV specifications; and that home inputs stay non-significant in all specifications.

6.3 Older Cohort: The contemporanous specification of the production function of skills at age 8, nutrition and home inputs effects

For the 8 year olds in the Older Cohort, I am only able to estimate equation (6).[58] Results in Table 7 show that current health (both physical and mental) influence the contemporaneous scores of the Ravens test in the expected direction and in a similar fashion.[59] For each SD increase in the total difficulties score (TDS), a mental health indicator, Ravens scores are on average 7.5 per cent of 1 SD lower, and for an SD increase in current HAZ, Ravens scores range from an average of 6 to 8 per cent of 1 SD higher, which indicates that physical and mental health are equally important at age 8.[60]

It is interesting to note that even the effect of caregiver's education is less important than nutrition (coefficient= .05), and that neither father's education nor any of the caste dummies are a significant determinant of the Ravens scores, even if they have the expected sign (positive for UCs and negative for SC/STs). Most importantly, after controlling for the number of years spent in school in column (2) and successively adding a control for the current home input score index in column (3), the coefficients on height-for-age and mental health remain unaltered.

Nevertheless, schooling does matter and its coefficient doubles that of either HAZ or TDS. Moreover, given the non-significant coefficient on the home inputs score, one could think that HAZ is already capturing the effects of (past and current) parental investment on test scores, supporting again the sufficiency assumption.[61] Lastly, the low R-squared value seems to indicate that there is still an important portion of the variance that cannot be captured with the data at hand, probably the most important being unobserved child endowments.

Table 7. *The production function of skills, cumulative specification for 8 year olds*

	Ravens test score (OLS)		
	bsline	**bsline + school**	**bsline + school + HI**
	(1)	**(2)**	**(3)**
Caregiver's education level	0.05***	0.05***	0.05***
	(0.01)	(0.01)	(0.01)
Father's education level	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)
HAZ	0.08**	0.07**	0.06**
	(0.03)	(0.03)	(0.03)
TDS	-0.07*	-0.08*	-0.08*
	(0.04)	(0.04)	(0.04)
Scheduled Tribe	-0.02	-0.03	-0.04
	(0.08)	(0.08)	(0.08)
Scheduled Caste	-0.05	-0.02	-0.03
	(0.12)	(0.12)	(0.12)
Upper Caste	0.13	0.12	0.12
	(0.09)	(0.09)	(0.09)
years school		0.15***	0.14***
		(0.04)	(0.04)

58 Past health status and past home inputs are not available.

59 However, it is only the mental health index (and not HAZ) that has an influence on writing and reading levels (these results are available upon request).

60 Given the lack of good instruments, it is not possible to instrument health in this cross section specification.

61 F-tests show the joint significance of caregiver characteristics with (respectively): health, school and home inputs variables.

	Ravens test score (OLS)		
	bsline	**bsline + school**	**bsline + school + HI**
	(1)	**(2)**	**(3)**
Home Input 8 yr old (factors-scored)			0.03
			(0.04)
Constant	-2.68***	-2.42***	-2.34***
	(0.84)	(0.83)	(0.84)
Observations	912	912	908
R-squared	0.13	0.14	0.14
F-Test: parental background & nutrition	6.86	6.30	6.25
Prob > F	0.00	0.00	0.00
Test join signif: parental background & school	-	8.16	8.26
Prob > F	-	0.00	0.00
Test join signif: parental background & inputs	-	-	7.76
Prob > F	-	-	0.00

Note 1: Standard errors in parentheses. Note 2: *** p < 0.01, ** p < 0.05, * p < 0.1. Note 3: The Home Input score index is composed by the variables as described in the text. Note 4: The highest the Total Diff Score, the worst the mental health. Note 5: Child, region, pre-school and home controls are always included.

6.4 Older Cohort: The value added specification of the production function of skills at age 12

As described in Section 2, the data requirements for implementing the cumulative specification of the cognitive skills production function are demanding, as one needs a complete history of inputs, beginning at the child's conception. The Older Cohort from the Young Lives data allows me to: (i) use part of that inputs history, and (ii) solve the endogeneity problems due to condition (c1) as I can account for unobserved endowments with cognitive skills measures available at two points in time for the same child.[62] Thus, beyond estimating the cumulative and contemporaneous specification and dealing with endogeneity, one major contribution of this paper is that I am able to exploit the panel structure of the data.

Table 8 shows the cumulative specification, estimated by OLS in columns (1) to (4) and by IV methods in columns (5) to (7). The model that allows for child-specific unobserved endowments nests the cumulative model with endowments that are orthogonal to included inputs. Under the null that endowments are uncorrelated with inputs, the OLS estimator applied to equation (6) from Section 2 is consistent, but under the alternative it is inconsistent.

The baseline equation (6) is estimated in column (1) and column (2) adds castes dummies and home inputs at age 12. Equation (7) is estimated in column (3) and (4) where home inputs and Ravens score, both at age 8, are included to test the contemporanous versus the cumulative 'value added' specification. In column (5), I replicate column (4) but now instrument past nutrition with mother's height and the interaction of mother's height and age. While variation in mothers' height is found in any healthy population, it also reflects heterogeneity in these women's early childhood disease susceptibility. To the extent that disease susceptibility is genetically inherited, mother's height should be correlated with the nutrition determined component of child height.[63,64] Past HAZ is significant at the 86 per cent

62 Endogeneity due to condition (c2) is solved with the IV approach (even if problem (b) is somehow left unresolved). See Section 2.

63 Subsequent columns are robustness tests of the specification in column (5). Reading and writing scores are added in column (6) and the full sample is used in column (7) (i.e., it includes children whose first language is not Telugu, that is, they speak Hindi, Urdu, Oria, Kannada, Marati, Tamil or local dialects).

level in columns (5) and (7), while the IV version in column (6) shows that increasing 1 SD past HAZ will increase average PPVT scores by around 32 per cent of 1 SD at the 90 per cent level of significance, which shows a downward bias of OLS, and then an indication of compensating behaviour when poor health outcomes are observed, consistent with results from the Younger Cohort.[65]

Table 8. *The production function of skills, cumulative specification for 12 year olds (instrumenting past HAZ with mothers' height and age)*

	Peabody Picture Vocabulary Test z scores						
			OLS			IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Caregiver's education level	0.03***	0.03**	0.03**	0.02*	0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Father's education level	0.02*	0.01	0.01	0.01	0.02*	0.02	0.01*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
past HAZ	0.00	-0.01	-0.01	-0.01	0.26	0.32*	0.36
	(0.03)	(0.03)	(0.03)	(0.03)	(0.21)	(0.16)	(0.22)
Scheduled Caste		-0.04	-0.04	-0.04	-0.03	0.00	-0.03
		(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Scheduled Tribe		0.23**	0.23**	0.25**	0.30*	0.31*	0.27*
		(0.11)	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)
Upper Caste		-0.02	-0.02	-0.03	-0.07	-0.06	0.00
		(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	(0.09)
Home inputs score, age 12		0.21***	0.21***	0.20***	0.19***	0.16***	0.18***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Home inputs score, age 8			0.01	0.01			
			(0.03)	(0.03)			
Ravens (z-score), age 8				0.13***	0.11***	0.08**	0.10***
				(0.03)	(0.03)	(0.04)	(0.04)
Writing level child, age 8						0.11**	
						(0.05)	
Reading level child, age 8						0.10**	
						(0.04)	
Constant	-0.71	-0.97	-0.94	-0.43	-0.24	-0.55	-0.48
	(1.10)	(1.05)	(1.06)	(1.05)	(1.11)	(1.20)	(1.10)
Observations	863	863	863	863	863	832	961
R-squared	0.19	0.27	0.27	0.28	0.21	0.18	0.13
F-Test: parental background & nutrition	4.22	3.15	3.15	3.15	-	-	-

64 The following results are not reported but worth describing: the age of the child does not seem to exert any influence on cognitive skills, while gender appears for the first time to have a significant impact on cognitive skills (being male increases average PPVT by 12 per cent of 1 SD consistently across specifications). First-borns have higher scores, but this effect disappears as soon as caste and home inputs controls are added. The urban, wealth index and regional dummies are all important as in previous regressions.

65 I have also included past BMI scores and z-scores (not reported) as a measure of health for 8 year olds and the results are remarkable similar.

	Peabody Picture Vocabulary Test z scores						
	OLS				IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F-Test: parental background & inputs	-	19.25	16.49	16.49	-	-	-
F-Test: Ravens- age 8 & nutrition	-	-	-	8.22	-	-	-
F-Test: Ravens- age 8 & inputs	-	-	-	52.81	-	-	-
Sargan stat	-	-	-	-	0.00	0.063	0.01
C-D Wald F stat	-	-	-	-	9.22	7.29	8.67

Note 1: Standard errors in parentheses. Note 2: *** p < 0.01, ** p < 0.05, * p < 0.1. Note 3: The Home Input score index is composed of variables as described in the text. Note 4: C-D Wald F stat is the Cragg-Donald Wald F-statistic of weak identification. Note 5: Child, region, school and home controls are always included.

Both parents' education variables affect cognitive skills in the expected direction. However, caregiver's education is more important than father's education in the OLS specifications, a trend that reverses in the IV specification (columns 5 and 7). This makes sense as the IVs are related to mothers' endowments. When comparing these sets of results with those of the previously analysed age groups, the most salient feature is that caregiver's educational level matters much more at age 5 and 8 (coefficient=0.04 to 0.05) than at age 12 (coefficient=0.02), which is understandable since outside factors start to influence cognitive skills as the child grows older and goes to school. Still, the coefficient on father's education is very stable with a coefficient of around 0.02 for all ages. Meanwhile, the Scheduled Tribes positive coefficient shows an advantage with respect to the base category (BCs), which is also true for the IV version. Furthermore, there is a positive, consistent and significant association of home inputs at age 12 and PPVT scores, even if that relationship is not clear for past home inputs (at age 8) in column (3).

A similar specification test can be used to examine the support for the VA model, which augments the contemporaneous specification with a lagged test score measure. The key assumption of the value-added model is that the lagged test score is a sufficient statistic for historical inputs and, in the versions of the model that do not incorporate endowments – like the OLS specifications in column (4) to (7) – it is also taken to be a sufficient statistic for endowments. To test the first assumption, I have included lagged input measures in the value-added specification, which should have no additional explanatory power under the sufficiency assumption. The estimate in column (4) shows that for test scores presented here, the lagged home input measure at age 8 is not statistically significantly different from zero. I interpret these results as evidence for the ‘sufficiency’ assumption.

Moreover, given the concern that the Ravens scores can be affected by school attendance, I have added reading and writing levels in column (6). It can be seen that the Ravens at age 8 has a consistent positive effect on the PPVT at age 12: increasing 1 SD in the Ravens test will significantly increase average PPVT scores by somewhere between 8 and 13 per cent of 1 SD. This is clear evidence against a contemporaneous specification. As there are no lagged input measures (say, home inputs at age 7 and less or lagged scores/nutrition at age 7 or less), I do not know whether their omission could be engendering an overstatement of the impact of a unit increase in Ravens score. Writing and reading levels have a quantitatively similar (positive and significant) impact on PPVT scores. Using the full sample in column (7) does not change the results, confirming that home inputs, lagged test scores (at a lower level of significance) and father’s education are the main determinants of performance in the PPVT test, with other factors like caste and nutrition having, unexpectedly, little effect.[66]

A logical extension of this analysis is to consider effects by the sign (positive if z-score above mean or negative if z-score below mean) and the magnitude of the changes in Ravens scores. I do so by showing a non-parametric regression of the z score of PPVT (in standard units) as a function of the z score of the Ravens test.[67] Results in Appendix Figure 1 are based on column (6) in Table 8 and show the analyisis for:(i) full sample, (ii) by gender and (iii) by caste. The figure shows an upward sloping relationship between PPVT and Ravens scores for the whole sample. It is shown that a one log-unit increase in the Ravens would increase the PPVT by 20 per cent.

I next extend this analysis to consider differences by gender, again focusing on the sign and magnitude of the effect of the Ravens. The figure suggests that boys and girls benefit from positive changes in the Ravens in a similar way. On the other hand, negative changes in the Ravens at the very bottom of the distribution (below -2) are more harmful to girls than to boys (i.e., they have a larger effect), a result that is consistent with the fact that families protect boys more than girls at the bottom of the distribution of the Ravens. Put differently, this suggests that household behavioural responses to very bad test scores play an important role in determining later cognitive skills.

More striking is the analysis by caste, which suggests that UCs benefit more from positive changes in the Ravens than LCs. However, overall, negative changes in the Ravens are much more harmful to LCs than to UCs, again suggesting important differential dynamic effects by socio-economic status.

6.5 Older Cohort: Parental investment demand functions

The estimations above allow input choices to be correlated with a child's fixed endowment but assumes that, conditional on endowment, input choices do not respond to earlier test score realisations. It is plausible, however, that parents might adjust their input choices in response to their child's earlier test score outcomes, as suggested by results in Appendix Figure 1. To analyse this in more detail, a demand function for parental investment is estimated below, as in equations (4) and (5). Results in Table 9 seem to indicate that, conditional on endowments, exogeneity of the input choices (one of the identifying assumptions) is rejected for the Ravens and the reading test score measures at a 1 per cent level. On the whole, these results provide evidence that input choices are correlated with endowments and with the unobserved components of achievement realisations (conditional on endowments), as expressed by past cognitive skills.

66 The relative importance of overall family background is confirmed by the F-test, which always returns a highly significant F-statistic. The joint significance of family background and health is not as clear as it is with respect to home inputs (at age 8 and/or 12). The importance of the Ravens test is also confirmed by an F-test which is also highly significant, particularly with respect to home inputs.

67 Specifically, I regress the PPVT on a set of co-variates and predict the residual from this regression. I also regress Ravens on the standard covariates and predict the residuals from this regression. Finally, locally weighted least squares are used to depict the relation between the residual from the PPVT z-score and the residual from the Ravens z-score regression. This approach is closely related to the two-stage procedure. For presentational purposes, the figure is trimmed at the 1st and 99th percentiles of Ravens deviation.

The positive estimated coefficients of Ravens, writing and reading scores on the home input measures columns (1) and (2), the school decisions column (3), the immunisation column (4) and the play time column (5) show that parents seem to be 'reinforcing' good earlier outcomes such as test scores rather than 'compensating' for poor earlier outcomes like nutrition. This is the opposite of the result found for nutrition outcomes in both the Younger and the Older Cohort with an IV approach.

Lastly, Table 10 replicates Table 9 by gender and by caste. It is interesting that the 'reinforcing' behaviour found earlier seems to be driven by boys and LC children for the Ravens and the writing tests, but not for the reading test. However, the coefficient on reading for males is larger than the coefficient for females. The finding on boys is consistent with Appendix Figure 1 and with the idea of pro-male discrimination in India in the intra-household allocation of resources. However, this result also gives additional information, as discrimination also happens *a posteriori* (i.e., after observing some objective measure of achievement). The reading of the results for the LCs is more complex. Probably, parents under extreme conditions of poverty will have to spread their resources more stringently, and therefore end up allocating more resources to the more able children, rather than the neediest in the household.

Table 9. *Parental investment functions, 12 year olds*

	Home inputs		Schooling	Immunisation	Play time
	(1)	(2)	(3)	(4)	(5)
Ravens (z score), age 8	0.17***	0.12**	0.11***	0.02	0.14**
	(0.05)	(0.05)	(0.04)	(0.01)	(0.07)
Writing level, age 8		0.11	0.11**	0.00	-0.07
		(0.07)	(0.06)	(0.02)	(0.10)
Reading level, age 8		0.26***	0.28***	-0.04***	0.04
		(0.05)	(0.04)	(0.01)	(0.07)
Age of child (months)	0.03***	0.03***	0.04***	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
male dummy	-0.09	-0.14*	-0.11	0.04	0.63***
	(0.09)	(0.09)	(0.07)	(0.02)	(0.12)
firstborn	0.38***	0.37***	0.25**	0.05	-0.16
	(0.12)	(0.12)	(0.10)	(0.03)	(0.17)
lastborn	0.06	0.11	0.08	0.01	0.20
	(0.11)	(0.11)	(0.09)	(0.03)	(0.15)
Constant	0.53	0.22	-0.51	0.41	4.14*
	(1.61)	(1.57)	(1.32)	(0.45)	(2.24)
Observations	926	926	926	935	935
Core controls	yes	yes	yes	yes	yes
R-squared	0.31	0.35	0.25	0.42	0.24

Note 1: Standard errors in parentheses. Note 2: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Note 3: Core controls refer to child, caregiver, home and region.

Table 10. *Parental investment functions: Home inputs, 12 year olds, by gender and caste*

	Female	Male	SC/ST	Upper Caste
	(1)	(2)	(3)	(4)
Ravens (z score), age 8	0.05	**0.18*** **	**0.12** **	0.13
	(0.07)	(0.07)	(0.05)	(0.08)
Writing level, age 8	0.04	**0.21** **	**0.20** **	-0.22
	(0.09)	(0.10)	(0.08)	(0.13)
Reading level, age 8	**0.19*** **	**0.32*** **	**0.24*** **	**0.24** **
	(0.07)	(0.07)	(0.06)	(0.10)
Age of child (months)	0.04***	0.01	0.03**	0.02
	(0.02)	(0.01)	(0.01)	(0.02)
male dummy	0.00	0.00	-0.02	-0.55***
	(0.00)	(0.00)	(0.10)	(0.16)
firstborn	0.36**	0.41**	0.27*	0.71***
	(0.17)	(0.17)	(0.14)	(0.21)
lastborn	0.11	0.11	0.07	0.22
	(0.15)	(0.15)	(0.13)	(0.19)
Constant	-1.57	1.82	0.03	2.18
	(2.30)	(2.15)	(1.86)	(2.71)
Observations	474	452	726	200
Core controls	yes	yes	yes	yes
R-squared	0.36	0.37	0.34	0.43

Note 1: Standard errors in parentheses. Note 2: *** p < 0.01, ** p < 0.05, * p < 0.1. Note 3: Core controls refer to child, caregiver, home and region.

6.6 Caste test score gaps

Using the production function estimates from column (6) in Table 8 (preferred IV specification), I examine the extent to which differences in HAZ and inputs can account for caste disparities in test scores. The specification test results of the previous section rejected all of the more restrictive specifications and gave support only for the VA model, which allows for child specific endowment effects (via the inclusion of lagged scores). I 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 and the slight decrease in the gap over time (see Table 11 and Appendix Figure 2). Because the estimated production function coefficients do not vary by caste,[68] the decreasing gap in the predicted test scores arises from caste differences in various inputs. Lastly, I examine how the predicted test score gaps vary if I set the levels of HAZ and home inputs for LC at the average levels observed for UC children (i.e., what the gap would have been if LC children received the UC average levels of home inputs). It is shown in Table 11 that if HAZ is equalised at the average UC level, then the PPVT test score gap would be reduced by 18.7, while if home inputs are equalised at the average UC level, then the PPVT test score gap would be reduced by 9.5 per cent. Again, it

68 Except for the STs coefficient in some specifications, but not in the preferred one.

is interesting to note that the gap between UC and LC for boys is closed by more than one fifth when equalising HAZ and only 5 per cent when equalising the home inputs index. For girls, one fourth of the gap would be closed by home inputs and 17.5 per cent by leveling nutritional status . The larger percentage of the gap closed by home inputs for girls seems an indication of pro-male discrimination in terms of home inputs.

Table 11. *PPVT caste gap closed by home inputs and by nutrition: Value added specification with lags for Older Cohort*

	Actual caste gap	Predicted caste gap	Closed by nutrition	Closed by home
All	11.1	10.5	1.96	1.00
			(18.7%)	(9.5%)
Boys	11.8	9.2	1.98	0.50
			(21.5%)	(5.4%)
Girls	10.3	11.9	2.08	3.18
			(17.5%)	(26.7%)

Note 1: The percentage of the gap closed is in parentheses. Note 2: The predicted caste gap in the Ravens for 8 year olds is 12.1 per cent.

7 Conclusions and further research

This paper has explored the determinants of children's cognitive skills using data from two cohorts of children for Andhra Pradesh, India. As in Todd and Wolpin's (2007) study for the USA, I do not find evidence that supports restrictive models that assume test scores depend only on contemporaneous inputs. Alternatively, the results for both cohorts show that both contemporaneous and lagged test scores (or lagged inputs) affect the production of current skills. Results also show that it is important to allow for unobserved child-specific endowment effects and endogeneity of inputs. In models where past inputs are not observed, past nutritional status turns out to be a very good proxy-indicator of this variable. Allowing for the endogeneity of past nutrition in the OLS model using an instrumental variable approach, I find that a 1 SD increase in HAZ at the age of 1 leads to PPVT scores that are on average 24 per cent of 1 SD higher at age 5.

I also find that parents are compensating (i.e., by ‘purchasing’ more health inputs) children for their adverse early health outcomes but reinforcing children for early favourable cognitive outcomes. The latter behaviour seems to apply mainly to boys and LC children.

Using a specification that incorporates the features above, I analyse test score gaps between LCs and UCs: the estimates show that policies that would aim to equalise HAZ and/or home inputs of UCs and LCs would close a significant proportion of the test score gap (up to a fourth of the gap for girls and a fifth for boys). However, a comparison of the efficiency of such policies would require information both about their ability to modify behaviour and about the costs of implementation.

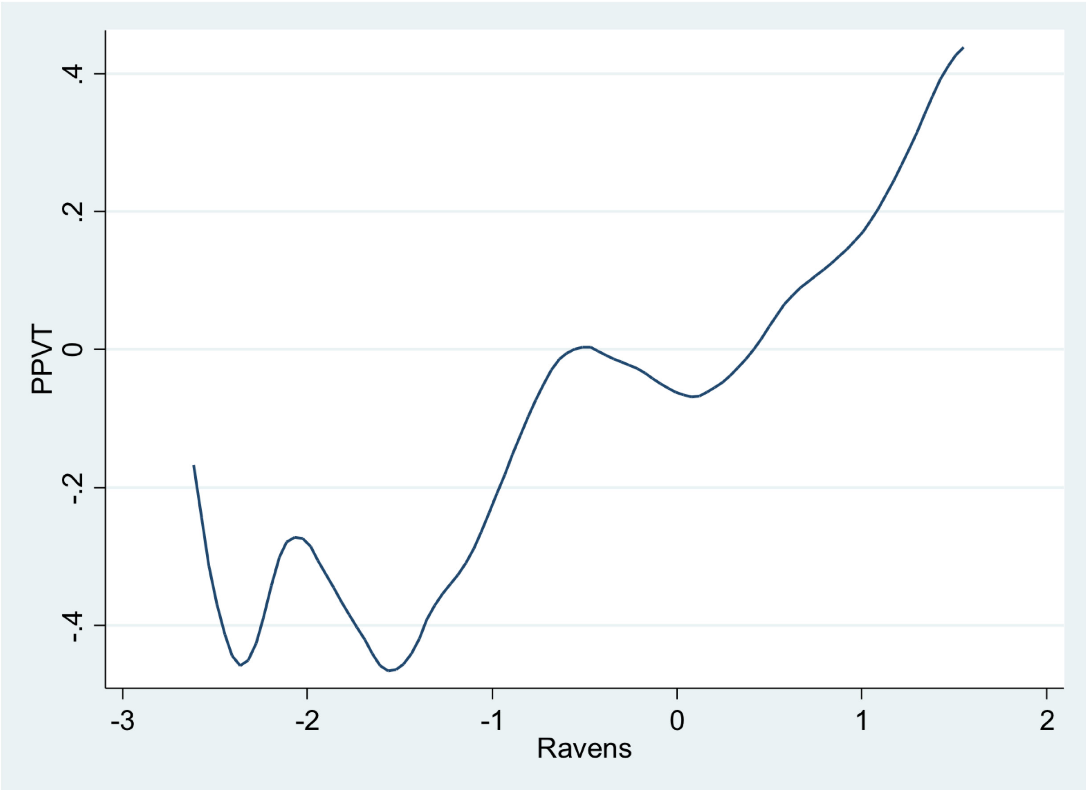
In terms of further research, it might be worth analysing with the availability of Round 3 in 2010 (specifically for the Older Cohort) whether the early childhood conditions are related to later adulthood outcomes (e.g., employment, earnings, etc.). Given the fact that some of these children will become parents themselves, more information can be gathered about the inter-generational tranfer of ‘care-behaviour’.

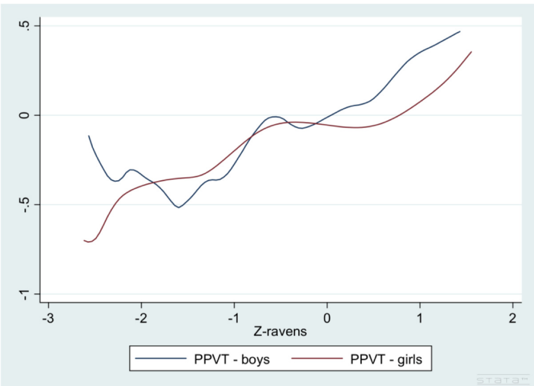
Appendix

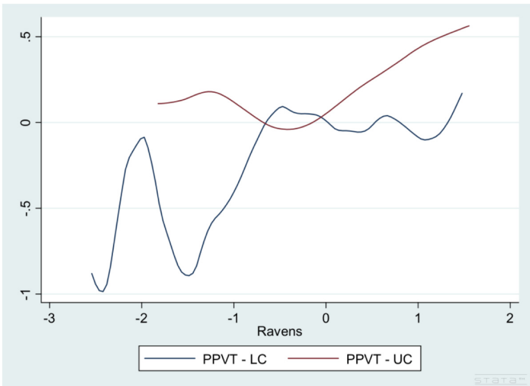
Figures

Appendix Figure 1.

Non parametric estimation of Ravens effects on PPVT scores

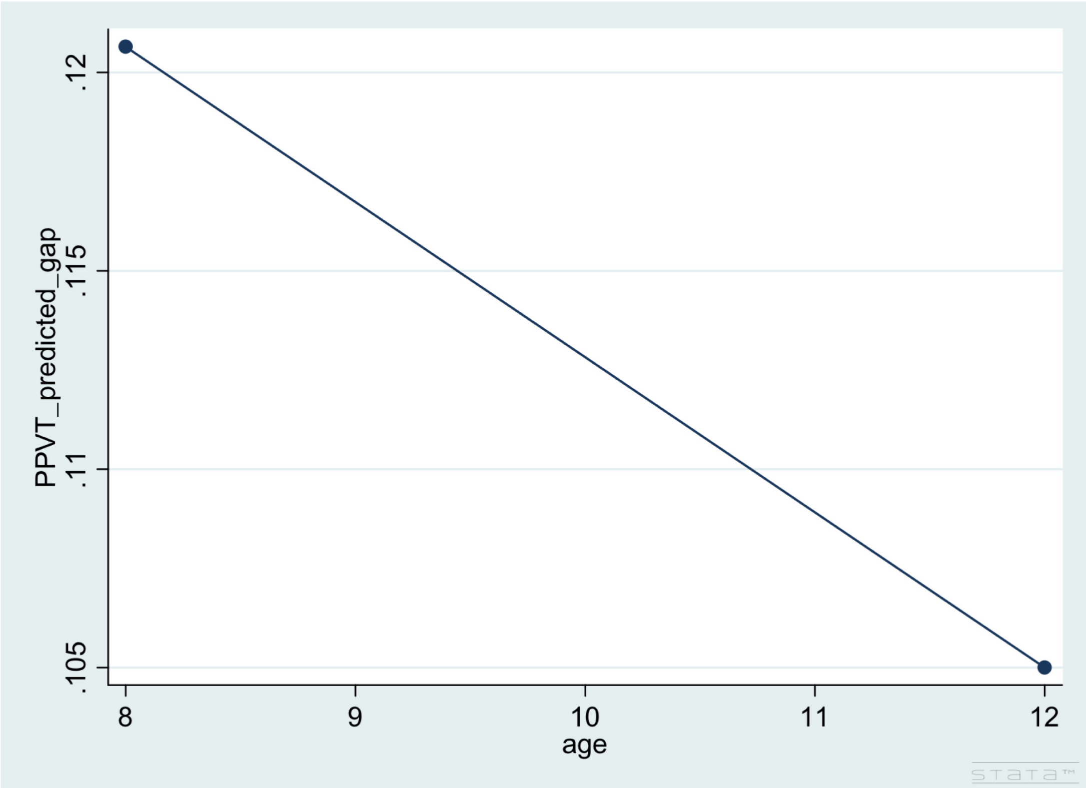






Source: Author's calculations based on Young Lives data. Locally-weighted regression, with bandwidth=0.75 and 100 intervals to perform regressions between p1 and p99th.

Appendix Figure 2. *Ravens and PPVT scores gap*



Source: Author's calculations based on Young Lives data.

Appendix 1

Cognitive tests description

The Ravens is a measure of non-verbal reasoning ability. According to its creator, the test measures the two main components of general intelligence: the ability to think clearly and make sense of complexity, which is known as eductive ability; and the ability to store and reproduce information, known as reproductive ability. However, others have interpreted this test as showing just how good visual decodification skills in children are. This test has been used to assess the cognitive abilities of children in several international studies in developing countries, such as Guatemala, Kenya, Egypt and Mexico. Factor analytical studies show that the test is a good indicator for Spearman's g-factor. The standard version of the Ravens test consists of 5 scales (A–E), with 12 items in each scale. Each item contains a figure with a missing piece, below which alternative pieces are placed to complete the figure. Each set involves a different principle for obtaining the missing piece; within a set, the items are arranged in increasing order of difficulty. According to the instructions given by the trainer, Ravens Colored Matrices version was administered. The Colored Matrices version consists of three scales: Test-A (12 items), Test-B (12 items) and Test-AB (12 items). Subscales A and B measure aspects related to cognitive processes, while subscale AB measures the intellectual capacity of the children. The test is supposed to be relatively free of cultural bias. The score is the number of correct responses to the items.

On the other hand, the PPVT is a a test of vocabulary recognition that has been widely used as a general measure of cognitive development. Earlier studies that have used the PPVT include Rosenzweig and Wolpin (1994), Blau (1991) and McCulloch and Joshi (2002). The PPVT consists of 17 sets of 12 words each. The child looks at pictures on an easel and identifies the picture which matches the word the interviewer reads out. Children start the test at a particular set depending on their age. They then move up or down depending on their responses. The response is always between 01 and 04. Training Items C and D (designed for children 8 years and older) are used for children aged 12 and Training Items A and B (designed for children under 8 years) are used for children aged 5 in the Young Lives sample.

Appendix 2

The caste system in India

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.

Appendix 3

Parental investment as a latent variable and factor analysis

Parental investment is a latent variable and therefore has to be estimated given available indicators through confirmatory factor analysis. More specifically, I am interested in a single measure for the respective latent variable and therefore employ a one-factor model. In general terms, the one-factor model assumes the following form:

$$x_{it} = b_{i0t} + b_{i1t}\alpha_t + n_{it} \qquad (8)$$

where x represents observed measures of the latent variable with $i = 1,...m_t$ denoting the different available indicators for the specific latent variable (as listed in Table 1); α_t is the factor for the latent variable (in this case, parental investment) and v_{it} is an error term where α_t and v_{it} are unobserved; b_{i1t} represents factor loadings and b_{i0t} is a measure-specific intercept. In order to estimate the model, we have to make several distributional assumptions. First, the factor and the error term are uncorrelated and have an expected value of zero. Second, the errors are independent over time and across children. Thirdly, we assume that the relationship between the factor and the observed variables is linear. Finally, the scale of the common factor is fixed by setting the first factor loading equal to one. In brief, I perform a factor analysis of a correlation matrix by means of a maximum-likelihood estimation of equation 8. This analysis specifies the maximum-likelihood factor method assuming multivariate normal observations. This estimation method is equivalent to Rao's canonical-factor method, and maximises the determinant of the partial correlation matrix. Hence this solution is also meaningful as a descriptive method for non-normal data. The factor score is then predicted as the conditional mean of the latent variable given the observed variables.

References

Alderman, H., J. Hoddinott and B. Kinsey (2006) 'Long-term Consequences of Early Childhood Malnutrition', *Oxford Economic Papers*, 58.3: 450-74.

Alderman, Harold (2000) 'Anthropometrics' in Margaret Grosh and Paul Glewe (eds) *Designig HH Survey Questionaniares for Developing Coutries: Lessons from 10 Years of LSMS Experience*, Oxford: Oxford University Press: 251-272.

Alesina, Alberto Reza Baqir and William Easterly (1999) 'Public Goods And Ethnic Divisions', *Quarterly Journal of Economics* 114.4:1243-84.

Asadullah, M. Niaz and Uma Kambhampati and Florencia Lopez Boo (2009) *Social Divisions in School Participation and Attainment in India,* Inter-American Development Bank/Research Department Working Paper 692, Washington DC: Inter-American Development Bank.

Baird, S., J. Friedman and N. Schady (2007) *Aggregate Income Shocks and Infant Mortality in the Developing World*, World Bank Policy Research Working Paper 4346, Washington DC: World Bank.

Banerjee, A.V., L. Iyer and R. Somanathan (2005) 'History, Social Divisions and Public Goods in Rural India', *Journal of the European Economic Association* 3.2-3: 639-47.

Banerjee, A.V. and R. Somanathan (2007) 'The Political Economy of Public Goods: Some Evidence from India' *Journal of Development Economics* Pages 82.2: 287-314.

Behrman, Jere R., John Hoddinott, John A. Maluccio, Erica Soler-Hampejsek, Emily L. Behrman, Reynaldo Martorell, Manuel Ramirez and Aryeh D. Stein (2008) *What Determines Adult Cognitive Skills? Impacts of Pre-School, School-Years and Post-School Experiences in Guatemala*, IFPRI Discussion Paper No. 826, Washington DC: IFPRI.

Bhalotra, S. and Bernarda Zamora (2007) 'Social Divisions in Education in India' in A. Sharif and R. Basant (eds) *Handbook of Muslims in India*, Delhi: Oxford University Press.

Bhaumik, S.K. and M. Chakrabarty (2007) *Is Education the Panacea for Economic Deprivation of Muslims? Evidence from Wage Earners in India, 1987-2004*, William Davidson Institute Working Paper 858, University of Michigan, Ann Arbor MI: William Davidson Institute.

Borooah, V.K., and S. Iyer (2005) '*Vidya*, *Veda*, and *Varna*: The Influence of Religion and Caste on Education in Rural India', *Journal of Development Studies* 41.8: 1369-1404.

Butler, Y.G. and K. Hakuta (2006) 'Cognitive Factors in Children's L1 and L2 Reading', *Academic Exchange* 10.1: 23-27.

Carneiro, Pedro, Flavio Cunha and James Heckman (2003) 'Interpreting the Evidence of Family Influence 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, 17 October 2003.

Carneiro, Pedro and James Heckman (2003) 'Human Capital Policy' in James Heckman, Alan B. Krueger and Benjamin Friedman (eds) *Inequality in America: What Role for Human Capital Policies?*, Cambridge MA: MIT Press.

Cueto, Santiago, Juan Leon, Gabriela Guerrero and Ismael Muñoz (2009) 'Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 2 of Young Lives', Young Lives Technical Note 15, Oxford: Young Lives.

Cunha, Flavio and James Heckman (2008) 'Formulating, Identifying and estimating the Technology of Cognitive and Non-Cognitive Skill Formation', *Journal of Human Resources* 43.4: 739-82.

Cunha, Flavio and James Heckman (2007) 'The Technology of Skill Formation', *American Economic Review* 97.2: 31-47.

Das, M. (2003) 'Ethnicity and Social Exclusion in Job Outcomes in India: Summary of Research Findings', unpublished paper, Washington DC: World Bank Institute.

DeCicca, Philip (2007) 'Does Full-Day Kindergarten Matter? Evidence From the First Two Years of Schooling', *Economics of Education Review* 26.1: 67-82.

Desai, Sonalde and Cecily Darden (2006) 'In the Margins: Social Inequality in Children's Educational Outcomes in India', paper presented at the Population Association of America annual conference, Los Angeles, 30 March–1 April 2006.

Dostie, Benoit and Rajshree Jayaraman (2006) 'Determinants of School Enrollments in Indian Villages', *Economic Development and Cultural Change* 54.2: 405-21.

Dreze, Jean and Geeta Gandhi Kingdon (2001) 'School Participation in Rural India', *Review of Development Economics* 5.1: 1-24.

Dutta, P. V. (2006) 'Returns to Education: New Evidence for India, 1983--1999', *Education Economics* 14.4: 431-51.

Engle, Patrice L., Maureen M. Black, Jere R. Behrman, Meena Cabral de Mello, Paul J. Gertler, Lydia Kapiriri, Reynaldo Martorell, Mary Eming 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: 229-42.

Filmer, Deon and Lant Pritchett (1998) *The Effect of Household Wealth on Educational Attainment: Demographic and Health Survey Evidence*, World Bank Policy Research Working Paper 1980, Washington DC: World Bank.

Fryer, R. and S. Levitt (2004) 'Understanding the Black-White Test Score Gap in the First Two Years of School', *Review of Economics and Statistics* 86.2: 447-64.

Galab, S., P. P. Reddy and Rozana Himaz (2008) *Young Lives Round 2 Survey Report Initial Findings: Andhra Pradesh, India*, Oxford: Young Lives.

Gang, I.N., Kunal Sen and Myeong-Su Yun (2008) 'Poverty In Rural India: Caste And Tribe', *Review of Income and Wealth* 54.1: 50-70.

Garces, E., Duncan Thomas and Janet Currie (2002) 'Longer-Term Effects of Head Start; *American Economic Review* 92.4: 999-1012.

Garrity, L. I. and J.T. Donoghue (1976) 'Pre-school Children's Performance on the Raven's Coloured Progressive Matrices and the Peabody Picture Vocabulary Test', *Educational and Psychological Measurement* 36.4: 1043-7.

Gertler, P.J. and L. C. Fernald (2004) 'The Medium Term Impact of Oportunidades on Child Development in Rural Areas', mimeo, Berkeley CA: University of California.

Glewwe, P., H.G. Jacoby and E.M. King (2001) 'Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis', *Journal of Public Economics* 81.3: 345-68.

Grantham-McGregor, S.M., Y. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp and T.I.C.D.S. Group (2007) 'Developmental Potential in the First 5 years for Children in Developing Countries',*The Lancet* 369.9555: 60-70.

Grantham-McGregor, S.M., C.A. Powell, S.P. Walker and J.H. Himes (1991) 'Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study', *The Lancet* 338.8758: 1-5.

Grantham-McGregor, S.M., S.P. Walker, S.M. Chang and C.A. Powell (1997) 'Effects of Early Childhood Supplementation with and Without Stimulation on Later Development in Stunted Jamaican Children', *American Journal of Clinical Nutrition* 66. 2: 247-53.

Griliches, Z. (1977) 'Estimating the Returns to Schooling: Some Econometric Problems', *Econometrica* 45.1: 1-22.

Hanushek, E.A. (1996) 'School Resources and Student Performance' in G. Burtless (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington DC: Brookings Institution.

Heckman, J. (1995) 'Lessons from the Bell Curve', *Journal of Political Economy* 103.5: 1091-120.

Heckman, J. (2000) 'Policies to foster Human Capital', *Research in Economics* 54.1: 3-56.

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

Heckman, J. and Y. Rubinstein (2001) 'The Importance of Non-Cognitive Skills: Lessons from the GED Testing Program', *American Economic Review* 91.2: 145-9.

Helmers, C. and M. Patnam (2009) *The Formation and Evolution of Childhood Skill Acquisition: Evidence From India*, Young Lives Working Papers no. 38., Oxford: Young Lives.

Hoddinott, J., J. 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: 411-16.

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

Jha, P., R. Kumar, P. Vasa, N. Dhingra, D. Thiruchelvam and R. Moineddin (2006) 'Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households', *The Lancet* 367.9506: 211-18.

Kijima, Y. (2006) 'Why did Wage Inequality Increase? Evidence from Urban India 1983-99', *Journal of Development Economics* 81.1: 97-117.

Kingdon, G. (2002) 'The Gender Gap in Educational Attainment in India: How Much can be Explained?', *Journal of Development Studies* 39.2: 25-53.

Kingdon, G. (2005) 'Where Has All the Bias Gone? Detecting Gender Bias in the Household Allocation of Educational Expenditure in India', *Economic Development and Cultural Change* 53.1: 409-51.

Kingdon, G. and J. Unni (2001) 'Education and Women's Labour Market Outcomes in India', *Education Economics* 9.2: 173-95.

Kumra, N. (2008) 'An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India', Young Lives Technical Note 2., Oxford: Young Lives.

Lally, J. Ronald, Peter L. Mangione and Alice S. Honig (1988) 'The Syracuse University Family Development Research Program: Long-Range Impact on an Early Intervention with Low-Income Children and Their Families' in D.R. Powell and I.E. Sigel (eds) *Parent Education as Early Childhood Intervention: Emerging Direction in Theory, Research, and Practice. Annual Advances in Applied Developmental Psychology,* Vol 3, Norwood NJ: Ablex Publishing Corporation.

Liu, H., T. Mroz and L. Adair (2009) 'Parental Compensatory Behaviours and Early Child Health Outcomes in Cebu, Philippines', *Journal of Development Economics* 90.2: 209-30.

Macours, K., N. Schady and R. Vakis (2008) *Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood Evidence from a Randomized Experiment*, Impact Evaluation Series No. 25, World Bank Policy Research Working Paper 4759, Washington DC: World Bank.

Maluccio, J.A., J. 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.

Murnane, R. J., J.B. Willett and F. Levy (1995) 'The Growing Importance of Cognitive Skills in Wage Determination', *The Review of Economics and Statistics* 77.2: 251-66.

O'Connor, T., M. Rutter, C. Beckett, L. Keaveney, J. Kreppner, T. English and R.A.S. Team (2000) 'The Effects of Global Severe Privation on Cognitive Competence: Extension and Longitudinal Follow-Up', *Child Development* 71.2: 376-90.

Paxson, C. and N. Schady (2007) *Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador*, World Bank Policy Research Working Paper 4226, Washington DC: World Bank.

Powell, C., H. Baker-Henningham, S. Walker, J. Gernay, S. Grantham-McGregor (2004) 'Feasibility of Integrating Early Stimulation into Primary Care for Undernourished Jamaican Children: Cluster Randomised Controlled Trial', *British Medical Journal* 329.89: 1-4.

Rajaram, S. and V. Jayachandran (2007) 'Child Labour and Schooling Outcomes among Children in India: An Analysis of Levels, Trends and Differentials', Paper accepted for the population association of America 2007 annual meeting Marriott Marquis, New York, 29 March 2007.

Reilly, B. and P.V. Dutta (2005) *The Gender Pay Gap and Trade Liberalisation: Evidence for India*, PRUS Working Papers 32, Brighton: Poverty Research Unit at Sussex, University of Sussex.

Rosenzweig, M. and K. Wolpin (1988) 'Heterogeneity, Intrafamily Distribution and Child Health', *Journal of Human Resources* 23.4: 437-61.

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

Scarr, S. and R. Weinberg (1983) 'The Minnesota Adoption Studies: Genetic Differences and Malleability', *Child Development* 54.2: 260-7.

Schady, N. (2006) 'Early Childhood Development in Latin America and the Caribbean', *Economica* 6.2: 185-213.

Summers, A. and B. Wolfe (1977) 'Do Schools Make a Difference?', *American Economic Review* 67.4: 639-52.

Teasdale, T. and K. Owen (1983) 'Heredity and Familial Environment in Intelligence and Educational Level – A Sibling Study', *Nature* 309.2: 620-2.

Temple, J.A. and A.J. Reynolds (2007) 'Benefits and Costs of Investments in Preschool Education: Evidence from the Child-Parent Centres and Related Programs', *Economics of Education Review* 26.1: 126-44.

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.

Tilak, J.B.G. (2002) *Determinants of Household Expenditure on Education in Rural India*, NCAER Working Paper Series 88, New Delhi: National Council of Applied Economics Research.

Todd, P.E. and K.I. Wolpin (2003) 'On the Specification and Estimation of the Production Function for Cognitive Achievement', *Economic Journal* 113.485: F3-F33.

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.

Waldfogel, J. (2006) *What Children Need*, Boston: Harvard University Press.

Walker, Susan P., Theodore D. Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A. Wasserman, Ernesto Pollitt, Julie A. Carter and the International Child Development Steering Group (2007) 'Child Development: Risk Factors for Adverse Outcomes in Developing Countries', *Lancet* 369: 145–57.

Wilson, R. (1983) 'The Louisville Twin Study: Development Synchronies in Behaviour', *Child Development* 54.2: 298-316.

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

Young Lives is an innovative long-term international research project investigating the changing nature of childhood poverty.

The project seeks to:

- improve understanding of the causes and consequences of childhood poverty and to examine how policies affect children's well-being
- inform the development and implementation of future policies and practices that will reduce childhood poverty.

Young Lives is tracking the development of 12,000 children in Ethiopia, India (Andhra Pradesh), Peru and Vietnam through quantitative and qualitative research over a 15-year period.

Young Lives Partners

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

Ethiopian Development Research Institute, Ethiopia

Centre for Economic and Social Sciences, Andhra Pradesh, India

Save the Children – Bal Raksha Bharat, India

Sri Padmavathi Mahila Visvavidyalayam (Women's University), Andhra Pradesh, India

Grupo de Análisis para el Desarollo (Group for the Analysis of Development), Peru

Instituto de Investigación Nutricional (Institute for Nutritional Research), Peru

Centre for Analysis and Forecast, Vietnamese Academy of Social Sciences, Vietnam

General Statistics Office, Vietnam

The Institute of Education, University of London, UK

Child and Youth Studies Group (CREET), The Open University, UK

Department of International Development University of Oxford, UK

Statistical Services Centre, University of Reading, UK

Save the Children UK (staff from the Rights and Economic Justice team in London as well as staff in India, Ethiopia and Vietnam).

Young Lives
An International Study of Childhood Poverty

Department of International Development
University of Oxford
3 Mansfield Road, Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Fax: +44 (0)1865 281801
Email: younglives@younglives.org.uk

www.younglives.org.uk