YOUNG LIVES STUDENT PAPER

Disadvantaged at Birth? The Impact of Caste on Cognitive Development of Young Children in Andhra Pradesh, India

Arvind Nair

June 2009

Paper submitted in part fulfilment of the requirements for the degree of MSc in Economics for Development at the University of Oxford.

The data used in this paper comes from Young Lives, a longitudinal study investigating the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh), Peru and Vietnam over 15 years. For further details, visit: www.younglives.org.uk.

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

Disadvantaged at Birth?
The Impact of Caste on the Cognitive Development of Young Children in Andhra Pradesh, India

June 2009

Arvind Nair

Contents

Abstract

Caste-based discrimination has been a feature of Indian society for many centuries, and despite sustained efforts by the Indian government, caste-based inequity still persists in socio-economic status, educational attainment and in health outcomes. A potential source of caste inequity, which is as yet unexplored, is in the differential development of cognitive skills during early childhood. These skills have been associated with performance in school and even with longer term employment and wage outcomes. Current research highlights the fact that cognitive skills are malleable and can be influenced by external environmental factors such as parental education, household socio-economic status and child's health, especially during the initial years of a child's life. This paper examines the link between caste and these external environmental factors in order to highlight the channels through which caste might affect the cognitive development of young children. We use recent longitudinal data for a cohort of approximately 2000 young children from the Young Lives survey in the state of Andhra Pradesh in southern India. We find significant evidence of a differential in cognitive skill development by caste, and also that caste acts largely through lower parental education and lower household socio-economic status, suggesting an intergenerational mechanism of transmission of outcomes.

Acknowledgements

I would like to acknowledge my supervisor, Professor Stefan Dercon, whose support and guidance through the writing process made this essay possible. Also, a big thanks to Vinayak Uppal and Andrew Sniderman for reading and commenting on intermediate drafts, and to Alan Sanchez for providing guidance on the use of the indicators of psychosocial well-being. Finally, I owe a big debt to my family for their belief in me and their support through my studies.

Section 1: Introduction

Caste plays a complex stratifying role in Indian society and as a hereditary social code has, over centuries, governed marriage, occupation and conduct towards members of other castes. Caste-based discrimination has also been prevalent over centuries, including segregation by occupation with the lower castes subjected to less-skilled, lower paying jobs; a lack of access to education; and in the extreme, a form of geographic segregation and isolation (Srinivas, 1953). Special provisions were made in the Indian Constitution in 1950 to safeguard the economic and social interests of the erstwhile 'lower' castes- Scheduled Castes (SC) and Other Backward Castes (OBC)- and the marginalized tribal communities- Scheduled Tribes (ST).[1] These provisions include the reservation of jobs in the public sector, the state and central level civil service and reservation of seats in both private and public higher educational institutions for the SCs, STs and OBCs.

Despite efforts by the Indian government and the special provisions in the constitution, caste-based inequity still persists 60 years after independence. Using National Health Survey data from 92/93, Deshpande (2001) presents some suggestive evidence that SCs and STs are on average more deprived than other castes in terms of both human and physical capital outcomes. More rigorous studies, with controls for potential confounding variables, have also confirmed the correlation of caste with lower income, poor health and lower education outcomes (Borooah, 2005; Kijima, 2006).

A key concern is if the processes that bring about these caste-based inequities begin early in a child's life. In this context, a potentially important unexplored source of inequity is in the development of cognitive skills in early childhood. These cognitive skills have been shown to be highly positively correlated with school attainment and wage outcomes in adulthood (Grantham-Mcgregor et.al, 2007). Recent empirical and theoretical work, surveyed in Cunha, Heckman et al. (2006), suggests that these skills are malleable early in a child's life and can be affected by key environmental factors such as household socio-economic status, early child health and parental education.

[1] The castes and tribes for whom special provisions are provided are listed in a special schedule of the constitution, updated periodically through amendments enacted by the legislative assemblies. In practice, castes and tribes can be removed from the schedule but this has never happened in practice.

We investigate if caste is significantly correlated with each of these key environmental factors in order to understand the channels through which caste might affect the cognitive development of children. We begin by asking the question: do parents of SC/ST children have lower endowments in health and education? Then, controlling for the parental endowments, is there still a caste-differential in household socio-economic status? Do these differences subsequently translate into differences in children's health outcomes? And finally, how do these differences translate into differences in the development of cognitive skills by caste? We also investigate if caste mediates the impact of these environmental factors on the development of cognitive skills; for instance, does having more educated parents matter more for a non SC child than for an SC child?

The next section provides a brief literature review of the development of cognitive skills; Section 3 outlines the theoretical framework for the analysis of the effects of caste; Section 4 provides a more detailed description of the data and the various measures used in this study; Section 5 outlines the estimation strategy and presents results of various specifications and Section 6 concludes.

Section 2: Literature Review

Recent empirical and theoretical literature is increasingly focused on identifying environmental influences that might affect the cognitive development of young children. Empirical work in a developed country context has associated higher socio-economic status with higher cognitive achievement of the child (Smith et. al, 1997; Guo & Harris, 2000). This is a reduced form result; several plausible pathways between higher socio-economic status and better cognitive development have been proposed, including better pre-natal care, more material and nutritional resources and better parenting skills (Paxson et al., 2007). In a developing country context, empirical work has focused on a stylized association between the health of the child, especially if the child is malnourished and/or anaemic, and the development of cognitive skills (Pollitt et al., 1997; Powell et. al, 1995).

Recent empirical work by Paxson and Schady (2007), using data from young children in Ecuador, attempts to tie these strands together and finds a significant impact of the parent's socio-economic status, the health of the child as well as a key role for parenting inputs on cognitive development. In this study, we adapt Paxson and Schady's framework and identify parental education, household socio-economic status and health of the child as the key environmental influences.

Section 3: Conceptual Framework

Caste stratifies Indian society and creates social groups with a strong sense of identity. Thus, we can use the basic framework for identifying and analyzing endogenous social effects, as outlined in Manski (1993) and Durlauf (2006), in order to analyze the effects of belonging to a particular caste on cognitive development. Manski disaggregates the impact of belonging to a group on the outcomes of a particular member of the group into three channels: correlated individual effects whereby members of a group have similar outcomes on account of having similar individual characteristics; contextual effects whereby all members of the group are subject to some common external environment (for instance, discrimination based on caste); and endogenous effects whereby the behavior of an individual changes with the behavior of the group ('role model' or 'aspirations' effects).

We study the correlated individual effects of caste on cognitive skill development, by focusing on how caste affects the interaction between the key environmental influences that affect cognitive development. We begin by taking, as exogenous, parental endowment in education, and investigate if there is a difference in these endowments by caste:

$$\mathbf{P_i = F(Caste, U_i)} \qquad \textbf{-- (1)}$$

where P_i is parental endowment in education and U_i are long term location and community characteristics that may affect this endowment. Subsequently, we investigate if there remains a caste impact on household socio-economic status, after controlling for the endowments in education:

$$\mathbf{H_i = F(Caste, P_i, U_i)} \qquad \textbf{-- (2)}$$

where H_i is household socio-economic status; P_i is parental endowment in education and U_i are community level characteristics that may influence household's socio-economic status. Next, we ascertain if there remains a residual caste impact on children's health outcomes, controlling for parental and household characteristics:

$$\mathbf{CH_i = F(Caste, P_i, H_i, U_i) \quad \text{-- (3)}}$$

where CH_i are child health outcomes. Finally, children's health, household characteristics and parental characteristics all influence children's cognitive skill development:

$$\mathbf{Y_i = F(Caste, P_i, H_i, CH_i, PS_i, U_i) \quad \text{-- (4)}}$$

It must noted that, at this stage, in a departure from previous empirical studies, we also examine if parental psycho-social competencies (PS_i) seem to play a role in determining children's cognitive skill development.

In addition, we also ascertain if caste mediates the environmental factors at each step of the process. For instance, an interesting question is whether having an educated parent matters less for an SC than it does for an OC in terms of determining socio-economic status. Or whether, increased socio-economic status can enable an SC child to catch up in terms of improving his/her cognitive outcomes. We re-estimate Equations 2 through 5 including linear interactions between caste and the explanatory variables. For instance, Equation 5 will be re-estimated as follows:

$$\mathbf{Y = F(Caste, P_i, H_i, PS_i, CH_i, P_i \times Caste, H_i \times Caste, PS_i \times Caste, CH_i \times Caste, U_i) \quad \text{-- (5)}}$$

The residual caste effect on cognitive skills outcomes, after controlling for parental, household, child and community characteristics, is a combination of other correlated effects (for instance, through parental occupation, child schooling outcomes, etc.), potential contextual effects due to caste-based discrimination and endogenous/'role model' effects.

To tease out the contextual effect, we need a credible measure of caste-based discrimination and subsequently, need to test if this measure is significant in its effect on the development of cognitive skills. In the absence of a direct measure in the sample, we are forced to employ an indirect measure: the exogenous variability in the size of the SC/ST population across the various survey sites, which implicitly assumes that the presence of more SC/ST in the population reduces the level of discrimination they face (the limitations of this exercise are detailed in the empirical section).

In order to estimate endogenous effects, we need an exogenous measure of the cognitive achievement of the 'average', or 'normal', SC/ST child in several subpopulations (say sentinel sites) and then need to test if the achievement of SC/ST children in these subpopulations is correlated with the achievement of the 'average' child. The rather stringent data condition of exogenous variability in the cognitive skills of the 'average' child is absent in our data sample and thus endogenous effects cannot be estimated.

In modeling these effects, our aim is to provide a reduced form result and flag potential channels through which caste might act on cognitive skill development. We are not isolating specific pathways through which these effects might be transmitted. For instance, parental education affects household socio-economic status through numerous interlinked channels (such as through occupation, labor market outcomes, etc.) and we do not try to discover the exact channel through which this effect takes place, but instead ascertain that such a link seems to exist and that caste plays a role in this process.

Section 4: Sample, Measures and Descriptive Statistics

4.1: Sample

The data used in this paper comes from a longitudinal study of children collected by the Young Lives Project in the state of Andhra Pradesh (A.P.) in southern India. The survey is uniquely suited for our purpose as it has extensive data on the health, learning, community and household level outcomes of approximately 2000 children in two cohorts, born in 1994/95 and 2001/02, in 2002 and 2006/07. We use data for the younger cohort of children, as the focus is to understand the early age development of cognitive skills.

A 'sentinel site surveillance methodology' was used in the survey in order to select the younger and older cohorts of children. The methodology randomized households within a selected survey 'sentinel' site, which were chosen earlier on the basis of predetermined criteria meant to ensure a distinct pro-poor bias (as the Young Lives is a longitudinal study of childhood poverty), and the representation of the regional, urban versus rural, and ethnic diversity of the state. Kumra (2008), in a comparison of the Young Lives sample with Indian health survey data from 1998/99, finds that the sample is representative of the diversity of Andhra Pradesh, but that the average Young Lives household is wealthier than the average A.P. household. Kumra notes that this might show a failure of the sample to generate a pro-poor bias, but can also be plausibly attributed to the fact that poverty in Andhra Pradesh fell between 1998/99 and 2001/02.

As Young Lives is a longitudinal study, we need to be concerned with issues of attrition bias, which might result from non-random attrition either due to deaths or non-participation across the two rounds of the study. The attrition rate in the sample used in this study is 3.03%, which is on the lower end as compared to attrition rates in most developing country longitudinal studies (Outes-Leon & Dercon, 2008). Dercon et. al also show that attrition in the Young Lives sample is, to a limited extent, non-random and when tested on empirical models is unlikely to generate significant attrition bias.

There are 4 caste categories in the data: the Scheduled Castes (SC), Scheduled Tribes (ST), the Backward Castes (BC) and Other Castes (OC). The SC and ST categorization has been discussed earlier, and the OC group here corresponds to the erstwhile upper castes, and thus, the key variables of interest will be the difference in outcomes between the SC/ST and the OC subpopulations. As Table 1 indicates, the caste composition in the Young Lives sample overrepresents the SC and ST population as compared to their presence in Andhra Pradesh and in India. This is plausibly because SC/ST are more likely to be poor and thus, the pro-poor bias in the sampling decision will ensure that they are overrepresented. This may allow us to generate more precise estimates of the effects of caste but, of course, if the oversampling is correlated with unobservable factors that may influence cognitive skill outcomes, then we may be generating a bias. This is important to keep in mind while interpreting the results.

Table 1: Caste composition in the sample population

	Young Lives Sample	Andhra Pradesh (2001 Census)	India (2001 Census)
% of SC in total population	18.2	16.2	16.2
% of ST in total population	12.8	6.6	8.1

4.2: Measures

Cognitive Achievement: We proxy for the the development of cognitive skills using results from tests designed to measure cognitive achievement. Specifically, we use results from the Peabody Picture Vocabulary Test (PPVT) and the Cognitive Development Assessment (CDA) test, which were administered to the younger cohort of children in the second round in 2006/07. The PPVT was initially developed in 1959 by Dunn and Dunn as a test of receptive vocabulary, and is used in this study as a proxy for verbal ability. The CDA test was developed by the International Association for the Evaluation of Academic Achievement (IEA) as a test of cognitive achievement, and the mathematical section is used in this study as a proxy for children's quantitative ability. Further information on the validity and the possible cultural bias of these tests is discussed in the Young Lives survey justification documents (Young Lives, 2007) and in Cueto et. al (2009).

Child health outcomes: Child health outcomes are proxied for by the height for age z-index. This compares the height of a child to the median height in the WHO reference population. A low height-for-age measure is an indicator of the child's inability to reach his/her genetic potential, and as such, can be seen as a more long term indicator of deprivation (Gopalan, 2002). As per WHO guidelines, a height for age score less than -2 corresponds to moderate stunting. For a robustness check, we use weight-for-age z-index, but this is more succeptible to short term fluctuations than height.

Socioeconomic Status: Socioeconomic status is proxied for by the 'wealth index', which takes on a value between 0 and 1, and is constructed using data on consumer durables owned by the household, type of dwelling, access to drinking water, toilet facilities and number of people per room (see Young Lives justification documents (2007) for more information). As a robustness check, we also use the consumption expenditure of the household per capita, which is often used as a measure of welfare (Kijima, 2006).

Parental endowments in health and education: Parental education may be highly linked to children's cognitive development particularly with regard to the intellectual stimulus provided in the home environment (Paxson et al., 2007). In this study, we use data on the years of education completed by the mother and the father. We proxy for parental long-term health outcomes with maternal height.

Parental Psycho Social Competencies: Parental notions of 'self-efficacy', understood as a parent's sense of agency over their lives may be linked plausibly to the child's cognitive development through their parenting inputs (or lack thereof). A constructed index of 'self-efficacy' is used in this study, based on questions such as "I have no choice which school to send my child to". This index is standardized, with negative values reflecting a greater notion of self-efficacy. A similarly constructed index is used to measure parental notions of 'self-esteem' (based on questions such as "I am ashamed to wear my clothes"). The robustness and validity of these measures is described in greater detail in the Young Lives justification documents (Young Lives, 2007).

4.3: Descriptive Statistics

A preliminary glance at the descriptive statistics (Table 2) indicates the extent of deprivation among the population, with only 34% of the mothers having completed primary school and close to 27% of the children suffering from stunting prior to completing their first year. Using a poverty line of monthly consumption expenditure less than Rs. 617.8 in urban areas and Rs. 332.1 in rural area based on National Sample Survey Organisation (NSSO) estimates, we see an overall incidence of poverty of 12.67%.

Even more striking in the results is the caste disparity, with SC and ST children having overwhelmingly greater negative average outcomes than the OC children. More than 40% of ST children are stunted prior to the age of 1 as compared to 16% of OC children; OC households have a 70% higher average consumption per capita than SC households and the average maternal education is 3 times greater for OC than for ST. Some of the indicators show a relative convergence in outcomes across castes, such as the enrollment in preschool, but these are the exception rather than the norm.

It is interesting to note that while there is a wide caste-based discrepancy in the mean outcomes of most of the explanatory variables, there is less marked variation in the outcome cognitive skills variables across caste. Furthermore, the raw data picture seems to indicate that ST children have marginally higher average scores in the cognitive tests than OC children. This, of course, could mask significant disparities in distribution and could potentially be driven by biases in sampling (where children in a particularly well-off majority ST area could be driving the results). This will be investigated more thoroughly in the empirical section.

Simple statistical tests are used to ascertain whether there is a jointly statistically significant difference in average outcomes across castes. The ANOVA F-statistic and the Chi-Statistic are used as parametric test of differences across continuous and categorical variables respectively, and the Kruskal-Wallis statistic is a non-parametric alternative. At the 1% significance level, these tests show that the average outcomes for all castes are not all equal. Two-sample t-tests are then conducted to test for pairwise differences in means of each of these variables across castes (such as comparing the wealth index for SC with wealth index for BC) and over 97% of the pairwise differences are found to be significant at the 1% level (t-tests for pairwise differences not reported due to space constraints).

Table 2: Descriptive Statistics

	Complete Sample	SC	ST	BC	OC	ANOVA F or Chi-Statistic	Kruskal-Wallis Chi-Squared
Wealth Index (mean)	0.327 (0.004)	0.27	0.218	0.327	0.443	104.55***	248.31***
Consumption per capita (mean) / Rs.	765.1 (11.46)	702.61	560.65	759.08	956.49	36.89***	169.7***
Absolute poverty	12.67	9.88	29.2	9.32	12.68	73.57***	-
Father's education (mean) / years	5.43 (0.11)	4.63	3.69	5.08	7.97	56.87***	150.89***
Father completed primary school	46.21	42.37	27.6	42.44	69.27	129.9***	-
Mother's education (mean) / years	3.53 (0.1)	2.19	1.93	3.16	6.48	96.22***	206.55***
Mother completed primary school	34.12	25.42	16	30.54	60.49	180.71***	-
Stunting (year 1)	27.04	31.36	42.4	26.26	15.86	59.49***	-
Stunting (year 5)	34.22	37.29	40	36.87	22.19	34.42***	-
Parental Self-efficacy (mean)	0 (1)	0.09	0.46	-0.09	-0.15	25.32***	75.76***
Parental Self-esteem (mean)	0 (1)	0.158	0.02	0.01	-0.16	6.62***	21.78***
Log PPVT score (mean)	5.69 (0.003)	5.64	5.75	5.67	5.75	38.38***	109.09***
Log CDA score (mean)	5.69 (0.004)	5.65	5.7	5.68	5.74	15.67***	50.75***
Male	53.31	50.29	56.8	51.12	58.78	9.25***	-
Urban	25.6	16.1	5.6	23.25	50.97	210.9***	-
Preschool	86.91	84.46	79.6	89.38	87.8	18.92***	-
Affected by Drought	27.8	32.48	18.8	32.26	19.27	38.05***	-
Number of observations	1947	354	250	933	410		

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. All results are in percentages unless otherwise stated. The ANOVA F- test and the Kruskal-Wallis Chi-Squared test are used to test for joint difference in means for continuous variables and the Chi-squared test is used for categorical variables.

Section 5: Empirical Exploration and Results

5.1: 'Exogeneity' of Caste

A major assumption underpinning this study is the exogeneity of the assignment of an individual to a particular caste. If there is significant choice in caste assignment, and significant mobility of individuals across caste, then any analysis of the effects of caste that does not take into account this endogenous group formation will have a significant bias. There has been debate about the degree of fluidity in the caste system, dating back to Srinivas' theory (M.N.Srinivas, 1957) on the 'sanskritization' of lower castes and the movement of entire castes up the caste spectrum through the adoption of the norms of upper castes. If the evidence is confined to a movement of entire castes up or down the spectrum, then it does not affect our analysis as it is conducted at a point in time. However, if there is evidence of individuals changing their castes or marrying into an upper caste seeking social mobility, then the exogeneity assumption is no longer valid (the upper caste variable will contain an unobserved effect of the desire for social mobility which may be correlated with our outcomes of interest). However, there is no evidence of this in our data, as caste is inherited and inter-caste marriage is still not the norm: 99.5% of the children have the same caste as their mother and over 99% of the parent's share their caste. Therefore, assumptions regarding the exogeneity of caste can be made plausibly.

5.2: Caste and Parental Endowments

We initially examine the correlation between caste and parental education and long term parental health. We take these parental endowments in education and health as given exogenous outcomes. This may be problemmatic as parental endowments might themselves be influenced by their household socio-economic status and other processes from their childhood, but for the purposes of this study, it is safe to assume that these endowments are unlikely to be determined by the household's current socio-economic status and present living conditions. This is reasonable considering that adult education is not widespread in India and adult height is widely seen as a long-term indicator of health outcomes unlikely to be affected by short-run fluctuations (Gopalan, 2002).

We estimate a simple OLS regression of the endowments on children's caste, while controlling for certain long-run community variables such as residing in an urban area and the specific region of Andhra Pradesh (which have had different growth trajectories over the last generation). We also control for sentinel site by including dummies in order to ensure that we only estimate the coefficients using heterogeneity arising from within sentinel sites. OLS estimates are inefficient if we ignore the sentinel site clustering design of the data (as errors may not be i.i.d).[2] We thus present results with standard errors that are robust to clustering. The following equation is estimated:

$$P_i = \alpha + \beta_1 * SC_i + \beta_2 * ST_i + \beta_3 * BC_i + Controls + \mu_i \quad \text{-- (6)}$$

where P is a specific parental endowment; SC, ST and BC are dummy variables for the caste of the child; and the controls include urban/rural and regional dummies.

The results (Table 3) indicate a substantial and significant caste differential in average parental endowments. The father of an SC child has on average nearly two years less education than an OC father (42% of a sd) and an ST father is even more disadvantaged at nearly four years less years of education (74% of a sd). The differences are also stark for the mother of an ST child, as she has, on average, 4 years less education (87% of a sd) and is approximately 2 cm (34% of a sd) shorter than an OC mother.

Table 3: Caste and Parental Endowments

	(1)	(2)	(3)
	Father's education	**Mother's education**	**Mother's height**
SC	-2.075*** (-3.421)	-2.663*** (-4.977)	2.247*** (-3.426)
ST	-3.628*** (-8.441)	-3.899*** (-7.810)	-1.933** (-2.717)
BC	-1.659*** (-3.827)	-2.004*** (-4.303)	-0.414 (-0.823)
Constant	5.695** (2.636)	4.697*** (9.448)	149.0*** (99.72)
Observations	1941	1940	1908
R-squared	0.194	0.328	0.051

*** p<0.01, ** p<0.05, * p<0.1. Robust t statistics in parentheses. Omitted category is Other Castes, Rural, Rayalseema. Sentinel site dummies included and errors robust to clustering.

5.3: Caste, Parental Endowments and Socio-Economic Status

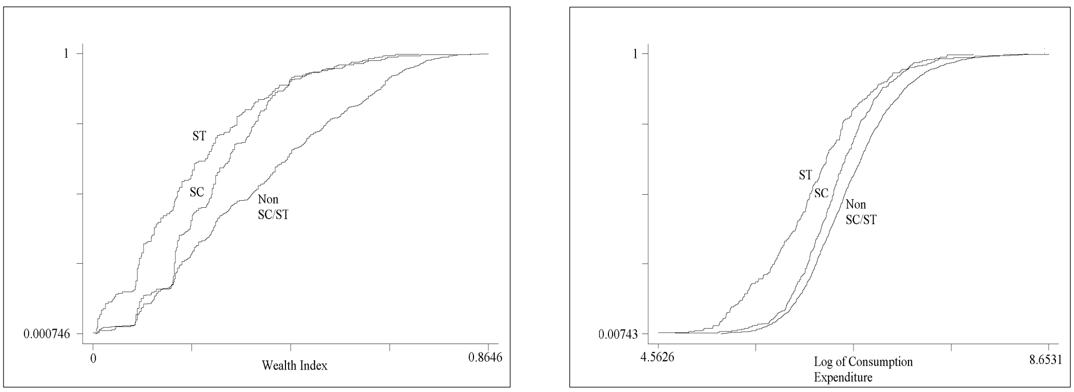
As per the theoretical framework, we now examine the link between caste and socio-economic status, while controlling for parental education. The lower socio-economic status of SC and ST households is a stylized fact that has been explored in detail in several empirical studies (Kijima, 2006; Deshpande, 2001). This stylized fact is borne

[2] We are correcting here for potential correlation between unobservables of households in the same sentinel site.

out in our sample, as can be seen in the cumulative density functions of the wealth index and the consumption expenditure per capita of households by caste. The cumulative density functions for ST households lie above that of the non SC/ST households below a maximum plausible poverty line. We take the maximum poverty line threshold for consumption per capita to be Rs. 617.8 (the urban poverty line according to the NSSO estimates) and for the wealth index to be its median value 0.291. Using the method devised by Davidson & Duclos (2000), we find that the vertical distances between the ST and non SC/ST curves are significantly different from zero for all values lower than the maximum plausible poverty line. This implies that non SC/ST is first order stochastically dominant over the ST function, and thus ST households face a higher poverty incidence (under any measure of poverty) than non SC/ST households.

The SC and the non SC/ST wealth index density functions intersect at a value lower than the maximum poverty threshold, and the vertical distances at all values below the threshold between SC and non SC/ST households are not all significantly different from zero. We are not able to establish stochastic dominance, but, from the moments of the distribution, we can say that SC households seem to have lower socio-economic status than non SC/ST households. The average consumption expenditure of an SC household is Rs. 702, with a variance of Rs. 22, as compared to a mean of Rs. 830 with a variance of Rs. 23 for a non SC/ST household.

Cumulative Density of Wealth Index and Consumption Expenditure by Caste



We now estimate a simple cross sectional OLS regression of socio-economic status of household on caste, while controlling for parental education and for community level characteristics such as residing in an urban area, specific regions of Andhra Pradesh and also if the household has been affected by drought (which is the most significant adverse event affecting the state in the early 2000s). As before, sentinel site dummies are included and standard errors are robust to clustering. Specifically, the equations estimated are of the form:-

$$\mathbf{H_i = \alpha + \beta_1 SC_i + \beta_2 ST_i + \beta_3 BC_i + \beta_4 P_i + Controls + \mu_i} \qquad \text{-- (7)}$$

where H is household socio-economic status; SC, ST, BC are caste dummies and P is the parental endowment in education.

The results (Table 4) indicate that the link between caste and socio-economic status is robust even after controlling for educational endowments. In our initial specifications [1] and [4] for the wealth index and the consumption per capita respectively, we see that being an SC or ST is associated with substantially lower socio-economic status (an ST household has a 52% of a sd lower consumption expenditure than an OC household). Once we control for education, which itself shows a significant correlation with socio-economic status, we still have substantial and significant negative coefficients for SC and ST, but the magnitude is reduced (an ST household now has a 23% of sd lower consumption expenditure than OC household). This suggests that caste acts on socio-economic status both through lower endowments in education and also independently of the education channel.

The interaction effects, estimated in specifications [3] and [6] show some interesting trends although most of the interactions seem to be insignificant. The SC dummy interacted with father's education shows a small negative impact (3% of a sd) which may suggest that an extra year of education does not pay off as much in terms of increasing socio-economic status for an SC father as compared to an OC father.

It must be noted here that we do not attempt to explicitly model a pathway through which caste might affect socio-economic status. In one of the specifications (not reported due to space constraints), we do control for occupation of the primary earner which may be a major channel through which caste might affect socio-economic status independent of education. The results from this specification show, unsurprisingly, that certain occupations such as farming and casual labor are associated with significantly lower socio-economic status. We do not see a change in the sign and significance of the key results: the residual caste effects seen in specifications [2] and [5], and the interaction effects for SC father's education and ST mother's education remain significant. However, the magnitude of these effects is reduced (for instance an ST household has a 19% of a sd instead of a 23% of a sd lower consumption expenditure than an OC household, after controlling for education) indicating that some of the caste effect is channeled through occupation.

Table 4: Caste, Parental Endowments and Socio-Economic Status

	(1)	(2)	(3)	(4)	(5)	(6)
	Wealth Index	**Wealth Index**	**Wealth Index**	**Consumption per capita**	**Consumption per capita**	**Consumption per capita**
SC	-0.0777*** (-5.936)	-0.0355** (-2.770)	-0.00247 (-0.153)	-185.6*** (-3.539)	-94.76* (-1.975)	64.07 (1.563)
ST	-0.126*** (-5.368)	-0.0598*** (-3.013)	-0.0745** (-2.615)	-264.5*** (-4.467)	-119.6* (-1.972)	-44.59 (-0.487)
BC	-0.0445*** (-3.932)	-0.0113 (-1.389)	-0.0104 (-0.889)	-82.88** (-2.407)	-10.78 (-0.347)	121.7** (-2.149)
Father's education		0.0090*** (8.369)	0.0107*** (6.001)		18.79*** (6.468)	35.35*** (5.143)
SC*father's education			-0.00542** (-2.652)			-24.43** (-2.642)
ST*father's education			0.00215 (0.576)			-7.414 (0.849)
BC*father's education			-0.00183 (-0.820)			-21.84*** (-2.925)
Mother's education		0.0090*** (7.178)	0.0079*** (4.183)		19.75*** (6.055)	16.33** (2.537)
SC*mother's education			-0.0034 (-1.698)			-1.166 (-0.130)
ST*mother's education			0.00337 (0.937)			2.581 (0.255)
BC*mother's education			0.00258 (1.076)			5.238 (0.611)
Constant	0.283*** (26.14)	0.222*** (5.799)	0.222*** (5.369)	758.0*** (18.3)	714.5*** (12.27)	607.4*** (9.976)
Observations	1940	1935	1935	1944	1939	1939
R-squared	0.49	0.602	0.608	0.15	0.22	0.226

*** p<0.01, ** p<0.05, * p<0.1. Robust t statistics in parantheses. Omitted category is rural, both parents zero years of education, rayalseema, other castes and non drought affected. Sentinel site dummies are included and errors presented robust to clustering.

5.4: Caste and Child Health Outcomes

As per the theoretical framework, we now examine the link between caste and children's health outcomes, and ascertain if the caste effect (if present) is largely mediated through parental education and health and household's socio-economic status.

Children's health outcomes prior to the age of one may have a direct and significant link to his/her early cognitive development, as Duc (2009) has highlighted using Young Lives survey data from Vietnam. We, therefore, begin by estimating a simple OLS regression of height-for-age z scores at age one (CH_{i1}) on caste, other observable child characteristics such as age and gender (X_{i1}), household characteristics such as size, number of children and socio-economic status (H_{i1}), community variables such as location and region (M_{i1}), and parental education and health (P_i). Sentinel site dummies are included and errors are presented robust to clustering.

$$\mathbf{CH_{i1} = \alpha+\beta_1SC_i+\beta_2ST_i+ \beta_3BC_i + \beta_4P_i + \beta_5X_{i1}+\beta_6H_{i1}+ \beta_7M_{i1} + \varepsilon_i +\mu_i} \quad \textbf{--(8)}$$

This specification suffers from the exclusion of unobserved time invariant attributes such as the child's genetic endowment and parenting inputs, which may be correlated with some of our variables of interest and independently with children's health outcomes. This may bias our results and also yield inconsistent estimates. We denote the time invariant unobservables as ε in the specification above.[3]

In order to account for these time invariant attributes, we estimate a differenced specification (in a two round panel setting, a differenced specification is exactly equivalent to a fixed effects specification), but with the time-invariant observables (such as gender, caste and parental endowments) interacted with a time dummy so that their effect is not differenced out as well. This also implies that time-invariant observables are allowed to have a varying effect on the child health in the two time periods. Errors are presented robust to clustering.[4] The differenced specification is reported below:

$$\mathbf{\Delta CH_{it} = \alpha+\beta_1SC_i+\beta_2ST_i+ \beta_3BC_i +\beta_4P_i +\beta_5X_i + \beta_6\Delta X_{it} +\beta_7\Delta H_{it} + \beta_8\Delta M_{it} + \Delta\mu_{it}} \quad \textbf{--(9)}$$

ΔX_{it}, ΔH_{it} and ΔM_{it} are time variant child, household and community characteristics respectively; P and X are time invariant household and parental characteristics and the constant term should be interepreted as a coefficient on the time dummy.

---

[3] The specification also suffers from the exlcusion of time variant attributes such as the provision of nutritional inputs, despite the presence of extensive controls.

[4] This accounts for inconsistency created by potential correlation between residuals of households in the same sentinel site.

We also estimate a dynamic specification which allows for the fact that past health outcomes might substantially impact current health outcomes (Strauss & Thomas, 2008). However, this kind of specification introduces endogeneity as lagged health status is correlated with the residuals by construction (as CH_{i1} is correlated with μ_1 which is by construction correlated with $\Delta\ \mu$). A popular instrument for lagged health status is birthweight (Federov & Sahn, 2005), but since, in our data birthweight is only available for 43% of the sample, we use caregiver's perception of birth size (very small, small, average, large, very large) which is available for the entire sample. This makes it superior to birthweight even given the fact that this variable has only five categories. We estimate the following dynamic specification:

$$\Delta CH_{it} = \beta_1 SC_i + \beta_2 ST_i + \beta_3 BC_i + \beta_4 P_i + \beta_5 X_i + \beta_6 CH_{i1} + \beta_7 \Delta X_{it} + \beta_8 \Delta H_{it} + \beta_9 \Delta M_{it} + \Delta\ \mu_{it} \quad \text{-- (10)}$$

The results from all of the above health specifications, for the child height-for-age z-scores, are presented in Table 5. The first specification estimates equation 8 with only caste and other time-invariant community, household and child controls. Lower castes seem to have strongly more negative initial health outcomes: as much as a 37% drop in sd of the child's height for age z-score for an ST child. This correlation is reduced with the introduction of parental endowments in education and health and socio-economic status in [2]. The maternal height channel is especially significant with a 1 standard deviation increase in maternal height associated with a 18% of a sd increase in the child's height-for-age z score. This, along with the strong association of caste with maternal height seen earlier, suggests that caste acts on child health strongly through the parental health channel. We also estimate a specification with interaction effects (not reported here due to space constraints), in order to ascertain if caste mediates some of these channels, and largely these effects seem to be insignificant.[5]

Specification [3] is the differenced specification, not controlling for parental endowments and socio-economic status, which are introduced subsequently in specification [4]. These differenced specifications show insignificant caste coefficients (even in specification [3]) and insignificant effects of socio-economic status and parental endowments. This contrasts with our initial OLS results (in [1] and [2]), as well

[5] Excepting the impact of child's health for an extra year of father's education is significantly lower for an SC and an ST child as compared to an OC child.

as with the results from a pooled OLS regression of height-for-age z-scores (in both rounds) which is reported in [5] and [6]. This could be because our OLS and POLS results are driven largely by time-invariant unobservables and the true caste, parental endowments and socio-economic status impact is minimal. Alternatively, due to the presence of measurement error (which could increase in differences than in levels), we could be getting more attenuation and generating more biased estimates using the panel dimension of the data as compared to limiting ourselves to cross sectional estimates.

We estimate the dynamic specification in [7] and [8] and since we instrument lagged health status with birth size, we report the Kleibergen-Papp LM and Wald F-Statistics testing the identification and weak identification of the equation.[6] We can reject the null of underidentification and weak identification for each of the specifications. The lagged health status shows a substantial and significant negative coefficient, reflecting that past health status is an important determinant of current health status (but since the coefficient is not close to -1, we can rule out full 'catch-up', i.e., a period of growth that compensates completely for a period of deprivation). The caste effects return in this dynamic specification, and are slightly negative.

These results suggest a rather grim story. The SC/ST children have lower initial height-for-age z-scores, thus indicating lower initial health outcomes, which seem to be largely driven through the lower parental health endowment channel (if we believe the initial specifications are not driven completely by time-invariant attributes). The dynamic specification indicates that, after controlling for the initial lower health outcomes, SC/ST children seem to be drifting further away from their OC counterparts between the ages of 1 and 5. The same picture is borne out with the specifications for weight-for-age z-scores which are not reported.

[6] In the absence of i.i.d errors , due to the clustering design of the data, the canonical Anderson LM and the Cragg-Donald F Statistic cannot be used and the alternative Kleibergen-Papp statistics must be employed.

Table 5: Caste and Child Health Outcomes

	(1) Height for age Round 1	(2) Height for age Round 1	(3) Change in height for age	(4) Change in height for age	(5) Height for age (POLS)	(6) Height for age (POLS)	(7) Change in height for age	(8) Change in height for age
SC	-0.276** (-2.581)	-0.074 (-0.615)	-0.018 (-0.322)	-0.0387 (-0.598)	-0.275*** (-3.33)	-0.104 (-1.06)	-0.103* (-1.915)	-0.0735 (-1.308)
ST	-0.524*** (-3.689)	-0.242* (-1.738)	0.0909 (0.697)	0.071 (0.528)	-0.468*** (-4.88)	-0.229** (-2.35)	-0.0559 (-0.509)	-0.0185 (-0.179)
BC	-0.214** (-2.557)	-0.11 (-1.495)	-0.0298 (-0.733)	-0.0512 (-0.998)	-0.212*** (-3.10)	-0.133** (-2.13)	-0.0766** (-2.134)	-0.0743* (-1.873)
Father's educ		0.01 (1.2)		0.00952 (1.053)		0.007** (2.43)		0.0119* (1.656)
Mother's educ		0.0147 (1.69)		-0.00857 (-1.119)		0.007* (1.76)		-0.00001 (-0.0028)
Mother's height		0.040*** (6.931)		-0.00398 (-1.096)		0.039*** (7.70)		0.00849 (1.16)
Wealth Index Rd 1		0.710*** (3.37)						
Wealth Index						0.794*** (5.05)		
ΔWealth Index				0.135 (0.585)				0.06 (0.382)
Lagged health status							-0.293*** (-2.674)	-0.303** (-2.345)
Age of Child	-0.0993*** (-7.522)	-0.10*** (-7.939)			-0.047*** (6.25)	-0.049*** (-6.96)		
ΔAge			-0.141*** (-9.669)	-0.137*** (-8.693)			-0.111*** (-6.267)	-0.111*** (-5.906)
Time					2.162*** (5.47)	2.312*** (5.97)		
Male	-0.0896* (-1.746)	-0.0865* (-1.798)	-0.0116 (-0.249)	-0.00774 (-0.153)	0.100*** (2.63)	0.088*** (2.46)	-0.0429 (-1.047)	-0.0438 (-1.001)
Constant	0.19 (0.85)	-6.35*** (-6.605)	6.926*** (8.772)	7.662*** (7.034)	-0.204 (-0.73)	-7.45*** (-9.23)	5.568*** (5.426)	4.187* (1.958)
Kleibergen-Papp LM							8.559***	7.597***
Kleibergen-Papp Wald F							15.198	12.187
Observations	1927	1885	1915	1871	3859	375	1889	1849
R-squared	0.172	0.221	0.15	0.151	0.1167	0.173	0.465	0.478

*** p<0.01, ** p<0.05, * p<0.1. Robust t statistics in parentheses. Lagged health status is instrumented with birth size. Controls for drought, household size, region and sentinel sites are included. Errors presented robust to clustering.

5.5: Caste and Cognitive Skill Development

We now turn to a composite picture of the impact of caste on the development of cognitive skills of the child. As per Paxson & Schady (2007), we see cognitive skill development as subject to environmental influences, including parental endowments in health and education and household socio-economic status. As per Duc (2009), we envision the child's early age health outcomes significantly impacting the development of cognitive skills. We thus estimate the following specification:

$$\mathbf{Y_{i2} = \alpha+\beta_1SC_i+\beta_2ST_i+ \beta_3BC_i +\beta_4P_i +\beta_5X_i+\beta_6M_i+\beta_7CH_{i1} +\beta_8H_{i2} + \varepsilon_m+\mu_i} \quad \textbf{-- (11)}$$

where Y_{i2} is a measure of the child's cognitive achievement, as proxied by the scores in the PPVT and CDA tests in Round 2; SC, ST and BC are caste dummies; P is a vector of parental endowments in education, health and parental psycho-social competencies; X is a vector of child attributes such as age, sex, number of siblings; and M is a vector of community variables such as region and rural/urban; CH_{i1} is a measure of the child's initial health status; and H_{i2} is a measure of current household socio-economic status. As in all earlier regressions, we include dummies for sentinel sites and present standard errors robust to clustering.

The results from these basic cross-sectional regressions suffer from several sources of endogeneity. A potential source of endogeneity is simultaneity, especially with cognitive skills and household socio-economic status. As Paxson et al. note, parents can adjust their investment in the child according to their perceived sense of the child's ability- this adjustment could be complementary in that more resources could be diverted to a 'brighter' child or compensatory whereby more resources are diverted to a child percieved to be weaker. As these investments are plausibly made diverting resources away from invesments in housing, etc. (which feeds into the wealth index), we get a link between cognitive achievement and household socio-economic status in the opposite direction than what we are estimating. We try to alleviate this problem by instrumenting wealth index in Round 2 with the wealth index in Round 1 (which are likely to be correlated), as it is less likely that parents get a sense of the child's cognitive ability prior to the age of 2 and therefore cannot make the necessary adjustment in their investment in the child (Glewwe, Jacoby, & King, 2001).[7]

Another major source of endogeneity is the impact of unobserved community, parental/household and child attributes which may be correlated with some of our explanatory variables and with the development of cognitive skills. Since we do not

---

[7] Instrumenting the wealth index with a lag may also account for measurement errors, as the lag is plausibly uncorrelated with errors introduced in the current round measurement of wealth

have a measure of the child's cognitive skills from the first round, we cannot estimate a differenced specification (as with child health) in order to account for the time-invariant unobserved attributes.[8] We instead attempt to address this issue through extensive control variables and through the use of instruments for some of the explanatory variables.

In order to ameliorate the impact of any community level unobservables, we control extensively for region, urban versus rural and in addition, we estimate community fixed effects (ε_m) through dummies for sentinel sites in order to ensure that we are only picking up within sentinel site heterogeneity. We attempt to control for many observable parental/household attributes such as education, health, parental psycho-social competencies, whether affected by drought and household socio-economic status. We do not have controls for certain key variables, such as parenting skills, which may plausibly influence child's health outcomes prior to the age of one and cognitive achievement. We attempt to alleviate this problem by instrumenting child height for age with birthsize, which is unlikely to be correlated independently with cognitive achievement.[9]

However, instrumenting child health at age one with birthsize, at best, accounts for the impact of the parental/household level unobservables on a particular explanatory variable. We cannot control for the effect of the unobservables on all of our explanatory variables, and moreover, we are not able to control for biases arising from some child-level unobservable attributes (potentially 'innate' ability). However, extensive controls for child, household, parental education and community attributes may prevent these unobservables from having a major independent impact on the child's cognitive ability. Also, it is unlikely that especially the child-level unobservables will be correlated with caste. Therefore, even though our results are not causal, we argue that they still reflect a reduced form effect of caste and the channels through which it acts on the development of cognitive skills.

Keeping this caveat in mind, we now turn to the results from the basic specification for the log of PPVT scores, instrumenting for wealth index in Round 2 and child's initial height-for-age with wealth index in Round 1 and birthsize respectively, as

[8] Even the differenced specification cannot account for unobservables that a re time varying.
[9] Birthsize is not a perfect instrument as unobservable parental attributes, such as the desire for a child, may still be correlated with birthsize and with cognitive skill development.

reported in [1] through [3] in Table 6. Specifications [4] through [6] report the equivalent specifications for the log of the CDA scores.

The PPVT results indicate a robust effect of caste in the initial specification; being an SC child is associated with a 31% of a sd lower log PPVT score. This effect disappears almost entirely in the subsequent specifications and seems to be almost entirely channeled through parental education and household socio-economic status. Parental education and socio-economic status appear to be highly significant- a 1 year increase in father's education is associated with a 2% of a sd increase in log PPVT scores, and a standard deviation increase in the wealth index associated with a 12% of a sd increase. Child health prior to the age of one is also significant, but has a smaller impact and the parental psycho-social competencies do not seem to be significant. The interaction effects [3] and [6] remain largely insignificant, except for those relating to the parental psycho-social competencies which will be discussed subsequently.

The results for the CDA scores tell a stylistically similar story to the PPVT results, with some important differences. The substantial and significant negative impact of caste in the initial specification is reduced substantially, and made largely insignificant, in subsequent specifications. As with the PPVT scores, the role of parent's education and household socio-economic status seems especially significant, and the role of the child's initial health outcomes also significant but to a lesser degree. The differences lie in the fact that the residual effect for an SC child remain significant for the CDA scores even after controlling for the other channels, although the effect is reduced in magnitude as compared to the initial specification. Also, the role of parental psycho-social competencies seem to be significant, but small in magnitude. The interaction effects [6] are largely insignificant, but they do suggest maternal education has a greater payoff for the ST child than the OC child.

The interaction effects also reveal some interesting trends for indicators related to parental psycho-social competencies. Separate regressions estimated (but not included due to space constraints) show a degree of association between caste and these psycho-social competencies; with SC/ST generally having lower sense of self-esteem and self-efficacy than the OC even after controlling for education and socio-economic status. It is interesting to note that the SC interaction with self-esteem (for both PPVT and CDA) is highly significant and positive, suggesting that an decrease in the sense of self-esteem of the parent (note that in our measures, a higher self-esteem index is associated with lower self-esteem) has a lower deleterious impact on the cognitive outcomes of the SC child than that of the OC child. This trend seems to be reversed when we look at self-efficacy, which shows significant and negative coefficients for both SC (PPVT) and ST (CDA), suggesting a grim scenario where a lower sense of self-efficacy for the parent has a greater negative impact on the SC/ST child.

Overall, these results indicate that most of the caste effect is mediated through parental education and to a lesser degree through household socio-economic status and children's initial health outcomes. It must be noted that each of the explanatory variables could be masking a whole set of interrelated mechanisms: for instance, lower socio-economic status could be acting on cognitive skill development through impacting the schooling decisions made by the parents (public versus private). We are agnostic about identifying the exact mechanism through which caste, or any of the other explanatory variables, might be acting. However, we take a closer look at the schooling outcomes mechanism as it might be of particular interest in the context of Andhra Pradesh.

Table 6: Caste and Cognitive Skills

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of PPVT score	**Log of PPVT score**	**Log of PPVT score**	**Log of CDA score**	**Log of CDA score**	**Log of CDA score**
SC	-0.0508*** (-3.433)	-0.00653 (-0.483)	0.362 (0.883)	-0.0639*** (-6.749)	-0.0253** (-2.485)	0.881 (1.251)
ST	-0.0436* (-1.778)	0.0138 (0.662)	0.0598 (0.11)	-0.0467** (-2.105)	0.0132 (0.686)	-0.267 (-0.480)
BC	-0.0288** (-2.422)	0.00395 (0.373)	0.266 (0.813)	-0.0403*** (-3.741)	-0.011 (-1.231)	0.496 (1.157)
Wealth Index Round 2		0.151** (1.995)	0.122 (1.544)		0.0671 (0.62)	0.00505 (0.042)
SC*wealth index			0.0805 (0.419)			0.0846 (0.552)
ST*wealth index			0.18 (1.229)			0.223 (1.453)
BC*wealth index			-0.033 (-0.416)			0.0741 (0.809)
Height for age Round 1		0.0330* (1.723)	0.0333 (0.966)		0.0535* (1.717)	0.0635 (1.121)
SC*height-for-age			-0.00264 (-0.0573)			-0.0159 (-0.263)
ST*height-for-age			0.0189 (0.265)			-0.0888 (1.078)
BC*height-for-age			-0.00517 (-0.114)			-0.0126 (-0.201)
Father's education		0.00318*** (3.012)	0.00543 (1.459)		0.00403*** (3.198)	0.00627 (1.263)
SC*father's education			-0.00398 (-0.962)			-0.00157 (-0.256)
ST*father's education			-0.000791 (-0.164)			-0.00432 (0.639)

BC*father's education			-0.00229 (-0.581)			-0.00408 (-0.790)
Mother's education		0.00480*** (3.782)	0.00701*** (2.584)		0.00359** (2.381)	0.00102 (0.512)
SC*mother's education			-0.00588 (-1.452)			0.00182 (0.587)
ST*mother's education			-0.0054 (-0.760)			0.0117** (2.233)
BC*mother's educ			-0.00124 (-0.387)			0.00302 (1.023)
Mother's height		-0.00083 (-0.861)	0.000189 (0.194)		-0.000801 (-0.483)	0.0017 (1.12)
SC*mother's height			-0.00229 (-0.976)			-0.00631 (1.498)
ST*mother's height			-0.000108 (-0.0371)			0.000473 (-0.154)
BC*mother's height			-0.00148 (-0.832)			-0.00352 (-1.539)
Self efficacy		-0.00578 (-1.347)	-0.00426 (-0.640)		-0.0141*** (-2.631)	0.00182 (0.206)
SC*self-efficacy			-0.0115 (-1.039)			-0.0291*** (-2.591)
ST*self-efficacy			-0.0230* (-1.779)			-0.0218 (-1.275)
BC*self-efficacy			0.00977 (1.292)			-0.0141 (-1.508)
Self esteem		0.0039 (1.115)	-0.00768 (-1.315)		0.00794** (1.967)	-0.0142 (-1.145)
SC*self-esteem			0.0213*** (2.645)			0.0298** (2.467)
ST*self-esteem			-0.0029 (-0.301)			0.0304 (1.547)
BC*self-esteem			0.00895 (1.415)			0.0241* (1.713)
Kleibergen-Paap LM Statistic		8.299***	2.171		7.067***	1.925
Kleibergen-Paap Wald F-Statistic		7.846	0.645		5.098	0.515
Constant	5.302*** (71)	5.277*** (31.52)	5.122*** (22.65)	5.281*** (47.71)	5.186*** (21.38)	4.906*** (17.4)
Observations	1656	1590	1590	1841	1763	1763
R-squared	0.275	0.353	0.366	0.12	0.121	0.121

Robust t statistics in parentheses. Wealth Index in Round 2 and Height-for-age z index in Round 1 are instrumented by wealth index in Round 1 and birthsize respectively. Controls for region, sentinel site, gender, age of child, urban/rural, drought and household size included but not reported due to space constraints. Errors presented robust to clustering.

5.6: Caste, Schooling Outcomes and Cognitive Skill Development

Almost all of the sample are attending either preschool or formal school, and over 37% of the sample are attending a private formal school. Schooling decisions, both in terms of attendance and the type of school to send the child, are plausibly correlated with parental education, household socio-economic status, health of the child as well as caste. We investigate this in OLS regressions [1] and [2] reported in Table 7. Specification [1] indicates that the total time enrolled in preschool and formal school (in months) is only correlated with the child's age and is not significantly correlated with caste, socio-economic status or parental education.[10] However, as per the LPM regression estimated in [2], higher socio-economic status, and having more educated parents, clearly increases the likelihood of a child attending a private formal school, while SC children seem less likely to go to a private formal school even after controlling for other factors.

In order to analyze the impact of schooling on cognitive skill development, we must correct for selection as it may be the case that 'smarter' children or children from a family where learning is already prized are sent to private school. We attempt to control for household level unobservables that might drive both schooling and cognitive skill development, by including the percentage of siblings in a private school[11] but we cannot control for child-level unobservables ('innate ability'). With this caveat, the results for the CDA and PPVT scores (specifications [3] through [6] in Table 7) indicate that time spent in school and private schooling seems to have little independent impact on test scores, and can thus be primarily seen as a channel through which socio-economic status and parental education acts on cognitive development.

Table 7: Caste, Schooling Outcomes and Cognitive Skill Development

	(1)	(2)	(3)	(4)	(5)	(6)
	Time in School	**Private School (LPM)**	**Ln CDA**	**Ln CDA**	**Ln PPVT**	**Ln PPVT**
SC	0.15 (0.194)	-0.12*** (-2.82)	-0.0473*** (-4.591)	-0.0244** (-2.260)	-0.0311** (-2.253)	-0.00512 (-0.405)
ST	-1.38 (-0.79)	-0.075 (-1.42)	-0.0295 (-1.651)	0.0105 (0.578)	-0.0224 (-1.032)	0.0122 (0.609)
BC	0.65 (1.08)	-0.027 (-0.72)	-0.0332** (-2.814)	-0.0124 (-1.357)	-0.0193* (-1.867)	0.00338 (0.33)
Wealth Index Rd 1	2.808 (1.43)					
Wealth Index Rd 2		1.176*** (4.75)		0.0823 (0.726)		0.134 (1.589)
Father's education	0.072 (1.42)	0.003 (1.27)		0.00388*** (2.966)		0.00309*** (2.989)
Mother's education	0.087 (1.37)	0.011*** (3.21)		0.00317** (2.278)		0.00453*** (3.653)
Mother's height	0.064 (1.53)	-0.001 (-0.63)		-0.000647 (-0.381)		-0.000738 (-0.752)
Self Efficacy	0.372	-0.007		-0.0132**		-0.00436

[10] This is consistent with the evidence that Andhra Pradesh has achieved widespread enrollment in primary and pre-primary school, and students drop out only from upper-primary onwards.
[11] Since private school is more expensive, once we control for household's socio-economic status, this variable could plausibly account for the household's inherent desire to get a child educated.

	(1.52)	(-0.80)		(-2.348)		(-0.989)
Self Esteem	-0.414 (-1.43)	-0.004 (-0.57)		0.00864* (1.92)		0.00318 (0.865)
Height for age Round 1	-0.386 (0.36)	0.0111 (1.585)		0.0516* (1.663)		0.0316* (1.716)
Male	0.452 (1.07)	0.058*** (3.46)	0.0014 (0.161)	0.00921 (1.091)	0.00224 (0.338)	0.0069 (1.348)
Age of child	0.856*** (9.66)	-0.0007 (0.01)	0.00553*** (4.121)	0.00956*** (3.616)	0.00508*** (4.91)	0.00718*** (5.166)
Time in School			0.000509 (1.583)	0.0000255 (0.0576)	0.000861 (1.68)	0.000341 (0.747)
Private School			0.0591*** (4.445)	0.0104 (0.672)	0.0739*** (7.852)	0.0190* (1.693)
% siblings in private school			-0.005 (-0.36)	-0.018 (-1.17)	0.015 (1.24)	-0.004 (-0.46)
Kleibergen Papp LM	7.502***	7.62***		4.726**		6.632***
Kleibergen Papp Wald F	11.595	5.842		2.747		5.044
Constant	-38.98*** (-6.321)	0.366 (1.085)	5.369*** (62.14)	5.167*** (22.22)	5.287*** (83.88)	5.239*** (31.52)
Observations	1841	1884	1799	1723	1619	1555
R-squared	0.21	0.442	0.142	0.132	0.313	0.364

*** p<0.01, ** p<0.05, * p<0.1. Robust t statistics in parentheses. Wealth index in Round 2 and height for age in Round 1 instrumented with wealth index in Round 1 and birthsize. In specification [1], we use wealth index in Round 1 as children attend preschool from age 3, which is prior to the collection of data relevant for calculation of Wealth Index in Round 2.

5.7: Contextual Effects of Caste

As per the theoretical framework, we now examine potential contextual effects of caste on cognitive development, i.e., effects that caste-based discrimination may have on cognitive development of SC/ST children. Results presented in Table 6 indicate that the residual effects of caste on cognitive achievement are limited, once the other channels such as parental education and socio-economic status have been controlled for (excepting SC children for the CDA scores). Of course, this residual effect is not merely a combination of endogenous and contextual effects, as we may be capturing other correlated individual effects not controlled for (such as occupation, schooling, etc.)

We therefore make a separate attempt to capture potential contextual (discrimination) effects of caste in our sample.[12] We use an indirect measure of discrimination by assuming that an increase in the percentage of the SC/ST population in a sentinel site will reduce the 'severity' of the discrimination they face. This is contentious as it is conflates numerical superiority with other forms of social and

[12] The endogenous effect cannot be isolated due to data limitations as stated earlier.

political influence, but in the absence of other plausible measures, this may be a valid starting point.[13]

We reestimate the specifications from Table 6, but include an interaction of the SC (ST) dummy with the percentage of the SC (ST) population within the cluster site. We report the results from the re-estimation in Table 8. The interaction effects are insignificant, which indicates that, for an SC (ST) child, an increase in the proportion of the SC (ST) population within a cluster site does not have any significant impact on his/her scores in the cognitive skills tests. Or, given our assumptions, that a decrease in level of discrmination does not impact the cognitive skill development of an SC/ST child.

However, this does not indicate that discrimination itself is absent, but instead that once we account for lower parental endowments, lower socio-economic status and child's health (all of which may separately be impacted by caste-based discrimination), we are unlikely to see any further discrimination effects.[14]

Table 8: SC/ST outcomes in 'dominant' cluster sites

	(1)	(2)
	Ln CDA	**Ln PPVT**
SC	-0.029 (-1.642)	-0.0234 (-0.990)
ST	0.0448 (1.592)	0.0176 (0.558)
BC	-0.0109 (-1.208)	0.00313 (0.295)
Proportion SC in cluster site	-1.238*** (-9.429)	0.0944 (1.17)
Proportion ST in cluster site	-13.68*** (-4.047)	-2.264 (-0.954)
SC * Proportion SC in cluster	0.0116 (-0.244)	0.0677 (-0.837)
ST * Proportion ST in cluster	-0.0832 (-1.242)	-0.0115 (-0.192)
Wealth Index Round 2	0.0543 (0.502)	0.150* (1.902)
Height for age Rd 1	0.0539* (1.73)	0.0330* (1.713)
Father's	0.00411***	0.00323***

[13] This measure of dominance is also flawed, because: (a) in our 20 cluster sites, we do not get a sufficient exogenous variability in the SC/ST population; (b) we make the assumption that the proportion of the SC/ST population in the sentinel site is a valid measure of the SC/ST population in the true population, which, as earlier noted, may not be tenable as the sample overrepresents SC/ST in general. However, this is an initial attempt in the absence of other credible measures.

[14]Subject to the validity of numerical superiority as a plausible sign of lower discrmination; and proportion of SC/ST within the cluster site as a measure of numerical superiority.

education	(3.246)	(3.029)
Mother's education	0.00369** (2.406)	0.00480*** (3.783)
Mother's height	-0.00084 (-0.505)	-0.000833 (-0.855)
Self Efficacy	-0.0143*** (-2.637)	-0.00575 (-1.341)
Self Esteem	0.00773* (1.935)	0.00386 (1.091)
Constant	5.454*** (21.3)	5.293*** (31.46)
Kleibergen Papp LM	7.086***	8.325***
Kleibergen Papp Wald F	5.096	7.853
Observations	1763	1590
R-squared	0.121	0.354

*** p<0.01, ** p<0.05, * p<0.1.Robust t statistics in parentheses. Controls for sentinel site, region, rural, household size and drought included but not reported due to space constraints. Errors presented robust to clustering. We instrument Wealth Index in Round 2 and Height for Age Round 1 with Wealth Index in Round 1 and birthsize respectively.

Section 6: Concluding Remarks

The results indicate that caste plays a significant role in the early age development of cognitive skills, with SC/ST children showing lower cognitive achievement than children from the Other Castes. SC/ST children seem to, on average, have less educated parents, come from households with lower socio-economic status and have lower initial health outcomes. The caste effect on the development of cognitive skills is chanelled largely through lower parental education and lower household socio-economic status (with a limited role for the child's initial health outcomes). The results thus seem to suggest intergenerational transmission of outcomes: SC/ST children have, on average, less educated and less well-off parents, which negatively impacts the development of cognitive skills in childhood, and this, in turn, harms the SC/ST child's prospects in school and puts him/her on a lower educational trajectory.

As we have a longitudinal data set, we can subsequently trace the extent to which this initial cognitive skills gap translates into a school achievement gap by caste. Does this gap remain relatively stable, reduce or increase over time? Several mechanisms could potentially make the gap wider. The lower socio-economic status of the SC/ST parents may not enable them to send their children to private schools (as already seen in the sample); or the lack of educated parents could make SC/ST children have an inferior learning environment at home. As learning is cumulative, a child who has fallen behind in the initial years may have to face increasing levels of difficulty catching up in later years of education. Caste-based discrimination in school may excarcebate this gap: the PROBE report (1999) extensively documents discrimination of SC/ST children by teachers and the treatment of SC/ST children as less 'smart' than other students.

On the other hand, if efforts taken by the government have had an impact, then this gap may reduce over time. Government efforts, through the Universal Education Program (*Sarva Shiksha Abhiyan*) which targets the education of SC/ST children through mid day meals, establishment of schools and provision of scholarships to SC/ST children (Department of Education, 2006) have helped reduce the primary school enrollment rate gap by caste. However, the closing of the enrollment gap has, as yet, not been accompanied with a reduction in the significant gap in drop-out rates, with SC/ST students far more likely to drop out in middle and high school than non SC/ST students (AP Districts at a Glance, 2007). Perhaps the impact of the government's efforts will be seen only as the current cohort enters middle/high school.

The key questions are: is there a significant correlation in the initial cognitive skills of the child and their performance in school? And further, is there evidence that government policies targeting children in lower primary school can redress some of the inequities caused by the differential in cognitive skills? If subsequent data for this sample indicates that the answer to these questions is in the affirmative, then there is a strong policy case for aggressively targeting resources for SC/ST children at the primary or lower primary stage.

References

Cueto, S., Leon, J., Gabriela, G., & Munoz, I. (2009). Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives. *Young Lives Technical Note No. 15* .

Cunha, F., Heckman, J., Lochner, L., & Masterov, D. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek, & F. Welch, *Handbook of the Economics of Education.* Amsterdam: North Holland.

Davidson, R., & Duclos, J. (2000). Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica , 68* (6), 1435-1464.

Department of Education, Government of India. (2006). *Sarva Shiksha Abhiyan: Fourth Joint Review Mission.* New Delhi: Aide Memoire.

Deshpande, A. (2001). Caste at Birth? Redefining Disparity in India. *Review of Development Economics , 5* (1), 130-144.

Directorate of Economics and Statistics, Government of Andhra Pradesh. (2007). *Districts at a Glance: Andhra Pradesh 2007.* Hyderabad: Directorate of Economics and Statistics.

Duc, L. T. (2009). The effect of early age stunting on cognitive achievement among children in Vietnam. *Young Lives Working Paper No. 45* .

Durlauf, S. (2006). Groups, Social Influences and Inequality. In S. Durlauf, S. Bowles, & K. Hoff, *Poverty Traps* (pp. 141-175). New York: Russell Sage Foundation.

Federov, L., & Sahn, D. (2005). Socioeconomic Determinants of Children's Health in Russia: A Longitudanal Study. *Economic Development and Cultural Change , 53* (2), 479-500.

Glewwe, P., Jacoby, H., & King, E. (2001). Early Childhood Nutrition and Academic Achievement: A Longitudanal Analysis. *Journal of Public Economics* , 345-368.

Gopalan. (2002). Undernutrition: Measurement and Implications. In S. Osmani, *Nutrition and Poverty.* Oxford: Clarendon Press.

Grantham-Mcgregor, S., Cheung, Y., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. a. (2007). Development Potential in the First 5 Years for Children in Developing Countries. *Lancet , 369*, 60-70.

Guo, G., & Harris, K. (2000). The Mechanisms Mediating the Effects of Poverty on Children's Intellectual Development. *Demography , 37* (4), 431-447.

Kijima, Y. (2006). Caste and Tribe Inequality: Evidence from India, 1983-1999.

Economic Development and Cultural Change , 54 (2), 369-404.

Kumra, N. (2008). An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India. *Young Lives Technical Note No. 2* .

M.N.Srinivas. (1957). Caste in Modern India. *Journal of Asian Studies , 16* (4), 529-548.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies , 60*, 531-542.

Outes-Leon, I., & Dercon, S. (2008). Survey Attrition and Attrition Bias in Young Lives. *Young Lives Technical Note No. 5* .

Paxson, C., & Schady, N. (2007). Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health and Parenting. *Journal of Human Resources , 42* (1), 49-84.

Pollitt, E., Watkins, W., & Husaini, M. (1997). Three-Month Nutritional Supplementation in Indonesian Infants and Toddlers Benefits Memory Function 8 Years Later. *American Journal of Clinical Nutrition* , 1357-1363.

Powell, C., Walker, S., Himes, J., Fletcher, P., & Grantham-Mcgregor, S. (1995). Relationships between physical growth, mental development and nutritional supplementation in stunted children: the Jamaican study. *Acta Paedactica , 84* (1), 22-29.

PROBE team. (1999). *Public Report on Basic Education in India.* New Delhi: Oxford University Press.

Smith, J., Brooks-Gunn, J., & Klebanov, P (1997). Consequences of Living in Poverty for Young Children's Cognitive and Verbal Ability and Early School Achievement. In J. Duncan, & J. Brooks-Gunn, *Consequences of Growing up Poor* (pp. 132-189). New York: Russell Sage Foundation.

Strauss, J., & Thomas, D. (2008). Health over the life course. In T. Schultz, & J. Strauss, *Handbook of Development Economics Volume 4.* Amsterdam: North Holland Press.

V.K.K.Borooah. (2005). Caste, Inequality and Poverty in India. *Review of Development Economics , 9* (3), 399-414.

Young Lives. (2007). *Justification of the Content of the Household Questionnaires. Round 2: 2006/07.* Retrieved March 15, 2007, from Young Lives: www.younglives.org.uk/pdf/research/HH_Justification_Doc.pdf